PHENOTYPING; FROM PLANT, TO DATA, TO IMPACT AND HIGHLIGHTS OF THE THE INTERNATIONAL PLANT PHENOTYPING SYMPOSIUM - IPPS 2018

EDITED BY: Trevor Garnett, Carolyn J. Lawrence-Dill, Tony Pridmore,
Michelle Watt, Cyril Pommier, Roland Pieruschka and
Kioumars Ghamkhar

PUBLISHED IN: Frontiers in Plant Science

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-397-2
DOI 10.3389/978-2-88966-397-2

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

PHENOTYPING; FROM PLANT, TO DATA, TO IMPACT AND HIGHLIGHTS OF THE THE INTERNATIONAL PLANT PHENOTYPING SYMPOSIUM - IPPS 2018

Topic Editors:
Trevor Garnett, University of Adelaide, Australia
Carolyn J. Lawrence-Dill, Iowa State University, United States
Tony Pridmore, University of Nottingham, United Kingdom
Michelle Watt, Julich-Forschungszentrum, Germany
Cyril Pommier, INRA Centre Versailles-Grignon, France
Roland Pieruschka, Julich Research Centre, Germany
Kioumars Ghamkhar, AgResearch Ltd, New Zealand

Citation: Garnett, T., Lawrence-Dill, C. J., Pridmore, T., Watt, M., Pommier, C., Pieruschka, R., Ghamkhar, K., eds. (2021). Phenotyping; From Plant, to Data, to Impact and Highlights of the The International Plant Phenotyping Symposium - IPPS 2018. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-397-2

Table of Contents

frontiers in Plant Science

EDITORIAL
published: 04 December 2020
doi: 10.3389/fpls.2020.618342

Editorial: Phenotyping; From Plant, to Data, to Impact and Highlights of the International Plant Phenotyping Symposium - IPPS 2018

Cyril Pommier[1,2]*, *Trevor Garnett*[3]*, *Carolyn J. Lawrence-Dill*[4]*, *Tony Pridmore*[5]*, *Michelle Watt*[6]*, *Roland Pieruschka*[7]* and *Kioumars Ghamkhar*[8]*

[1] Université Paris-Saclay, INRAE, URGI, Versailles, France, [2] Université Paris-Saclay, INRAE, BioinfOmics, Plant Bioinformatics Facility, Versailles, France, [3] The Plant Accelerator, Australian Plant Phenomics Facility, School of Agriculture, Food and Wine, University of Adelaide, Urrbrae, SA, Australia, [4] Iowa State University, Ames, IA, United States, [5] University of Nottingham, Nottingham, United Kingdom, [6] Faculty of Science, School of BioSciences, University of Melbourne, Parkville, VIC, Australia, [7] Forschungszentrum Jülich, BG-2: Plant Sciences, Institute for Bio- and Geosciences, Jülich, Germany, [8] Grasslands Research Centre, AgResearch, Palmerston North, New Zealand

Edited and reviewed by:
Eetu Puttonen,
National Land Survey of Finland, Finland

***Correspondence:**
Cyril Pommier
cyril.pommier@inrae.fr
Trevor Garnett
Trevor.Garnett@grdc.com.au
Carolyn J. Lawrence-Dill
triffid@iastate.edu
Tony Pridmore
Tony.Pridmore@nottingham.ac.uk
Michelle Watt
watt.m@unimelb.edu.au
Roland Pieruschka
r.pieruschka@fz-juelich.de
Kioumars Ghamkhar
kioumars.ghamkhar@agresearch.co.nz

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 16 October 2020
Accepted: 10 November 2020
Published: 04 December 2020

Citation:
Pommier C, Garnett T, Lawrence-Dill CJ, Pridmore T, Watt M, Pieruschka R and Ghamkhar K (2020) Editorial: Phenotyping; From Plant, to Data, to Impact and Highlights of the International Plant Phenotyping Symposium - IPPS 2018. Front. Plant Sci. 11:618342. doi: 10.3389/fpls.2020.618342

Keywords: phenotype, plant science, high throughput phenomics, database (all types), phenomic integration, artificial intelligence, image analysis, data management

Editorial on the Research Topic

Phenotyping; From Plant, to Data, to Impact and Highlights of the International Plant Phenotyping Symposium - IPPS 2018

The aim of this Research Topic is to provide a series of research articles on a range of subjects in Plant phenomics (Tardieu et al., 2017) from the use of appropriate sensors for capturing morphological and physiological traits to smart ways of processing, extracting and managing "clean" data. Presentation of new approaches to data acquisition, processing and analysis as well as prerequisites for automation are also among the objectives of this Research Topic.

Plant phenomics is the use of sensors, cameras, and algorithms for trait quantification in plants including model species crops, forages, vegetables as well as forest and fruit trees. The relationships between plants and their environment including soil microbes can affect this quantification bringing in new challenging parameters into the equation. Data may be acquired in a range of experimental conditions including laboratories, greenhouse, and field or natural experimental site within forests. Data from the latter can be used for biodiversity studies as well where the scope, measurement means, and objectives can be shifted for that purpose.

This Research Topic presents a series of articles with an insight into recent advances in plant phenomics. There are 12 research articles and one opinion paper covering the heterogeneity and complexity (Watt et al., 2020) of this rapidly developing scientific domain. The majority of the articles address problems associated with data acquisition and analysis using innovative computational methods, and a few discuss artificial intelligence. This reflects the current trend in plant phenomics which leans toward recovering the maximum amount of knowledge and information from the data deluge triggered by high throughput phenotyping. The significance of data management is also highlighted.

In this Research Topic, phenomics data acquisition at a very large scale, is demonstrated by Hao et al.. Their work shows how the airborne LiDAR approach allows estimation of the response of forest ecosystems regarding climate change and carbon density. Likewise, Zhang et al. used aerial vectors to enable high throughput data acquisition for understanding leaf development in rapeseed. This approach was taken to obtain usable information for precision farming, including precision fertilization, irrigation, and yield prediction. High throughput phenotyping in automated greenhouse is discussed by Nguyen et al.. They show the production of long time series of images allowing measurement of top view area and shoot biomass to help the estimation of nitrogen use efficiency in wheat.

The imaging types of data produced can also be rather diverse, including color images (as in Hao et al.; Cho et al.; Bateman et al.) but also hyperspectral (Bruning et al.) or Fluorescence (Hupp et al.; Méline et al.) imaging. Méline et al.. suggest promise for estimation of plant response to biotic stress in a non-destructive way in a model plant. This opens up future perspectives for crop research through translational biology.

Major recent innovations in automation and integration of plant phenomics have mobilized the development of deep and machine learning methods and tools and the efforts to address issues with data processing. Dobrescu et al. discuss multitask learning (MTL) to infer two morphological and one classification trait at the same time. Bateman et al. present a new method and name it local context network (LC-Net), which is designed to measure biomass of individual species in a mixed sward using convolutional neural networks.

Atanbori et al. demonstrate the use of conditional Generative Adversarial Network (GAN) to improve the quality of root counting and measurement to train data for missing classes of root images/data.

Semantic approaches are important not only to correctly describe the data but also to extract and organize processed information. Two papers in this Research Topic explore how semantic annotation can be fully or partially automated. Braun and Lawrence-Dill show how *in silico* text mining and natural language processing approaches can be used to extract information from phenotype descriptions, enabling better exploitation of the data gold mine that lies in the literature. Adhikari et al. is more applied and field-based and uses both expert annotation and artificial intelligence approaches to train an automatic weeding method for Paddy fields. Over the past 10 years, scientific communities have proposed several approaches to help data reusability following the FAIR (Wilkinson et al., 2016) principles. This encompasses biologist-friendly data standards such as MIAPPE (Papoutsoglou et al., 2020), ontologies designed to help biologists describe their datasets (Shrestha et al., 2012), or to support data scientists engaged in data integration (Cooper et al., 2017). Technical standards such as the Breeding API (Selby et al., 2019) and databases implementing all of these components such as GnpIS (Pommier et al., 2019) and PHIS (Neveu et al., 2018) have been well-received by the community. In this Topic, MaizeDIG explores the deep integration of phenomic and genomic data using BioDIG (Biological Database of Images and Genomes) web-based software.

The only Opinion paper (Sadras) in this topic gives interesting thoughts on the outcome of the 5th International Plant Phenotyping Symposium (IPPS) 2018, and especially the relationship between theory and our ability to understand and make efficient use of the existing data.

Overall, we believe that the articles published in this Research Topic provide important new knowledge in the broad field of plant phenomics. Our application of sensors, AI, semantics, and data standards and sharing on plant biology is only just beginning. Combining all these tools and techniques to better understand plants and to improve crops are showing immense promise. It is crucial to continue investigating the morphology, physiology, and gene x environment interactions while simultaneously developing and deploying novel approaches to improve the quantification and repeatability of these studies. We hope that the work presented in this Research Topic will help to both consolidate the field of plant-phenotyping and shed light on the significance of methods in decoding plant traits in different environments.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Cooper, L., Meier, A., Laporte, M.-A., Elser, J. L., Mungall, C., Sinn, B. T., et al. (2017). The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res.* 46, D1168–D1180. doi: 10.1093/nar/gkx1152

Neveu, P., Tireau, A., Hilgert, N., Nègre, V., Mineau-Cesari, J., Brichet, N., et al. (2018). Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Phytol.* 221, 588–601. doi: 10.1111/nph.15385

Papoutsoglou, E. A., Faria, D., Arend, D., Arnaud, E., Athanasiadis, I. N., Chaves, I., et al. (2020). Enabling reusability of plant phenomic datasets with MIAPPE 1.1. *New Phytol.* 227, 260–273. doi: 10.1111/nph.16544

Pommier, C., Michotey, C., Cornut, G., Roumet, P., Duchêne, E., Flores, R., et al. (2019). Applying FAIR principles to plant phenotypic data management in GnpIS. *Plant Phenomics* 2019, 1–15. doi: 10.34133/2019/1671403

Selby, P., Abbeloos, R., Backlund, J. E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O. E., et al. (2019). BrAPI—an application programming interface for plant breeding applications. *Bioinformatics* 35, 4147–4155. doi: 10.1093/bioinformatics/btz190

Shrestha, R., Matteis, L., Skofic, M., Portugal, A., McLaren, G., Hyman, G., et al. (2012). Bridging the phenotypic and genetic data useful for integrated breeding through a data annotation using the Crop Ontology developed by the crop communities of practice. *Front. Physiol.* 3:326. doi: 10.3389/fphys.2012.00326

Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055

Watt, M., Fiorani, F., Usadel, B., Rascher, U., Muller, O., and Schurr, U. (2020). Phenotyping: new windows into the plant for breeders. *Annu. Rev. Plant Biol*. 71, 689–712. doi: 10.1146/annurev-arplant-042916-041124

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:160018. doi: 10.1038/sdata.2016.18

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 10 July 2019
doi: 10.3389/fpls.2019.00917

Estimating the Aboveground Carbon Density of Coniferous Forests by Combining Airborne LiDAR and Allometry Models at Plot Level

Hongke Hao[1,2], Weizhong Li[2], Xuan Zhao[3], Qingrui Chang[1]* and Pengxiang Zhao[2]*

[1] College of Natural Resources and Environment, Northwest A&F University, Yangling, China, [2] College of Forestry, Northwest A&F University, Yangling, China, [3] College of Landscape Architecture and Arts, Northwest A&F University, Yangling, China

Forest carbon density is an important indicator for evaluating forest carbon sink capacities. Accurate carbon density estimation is the basis for studying the response mechanisms of forest ecosystems to global climate change. Airborne light detection and ranging (LiDAR) technology can acquire the vertical structure parameters of forests with a higher precision and penetration ability than traditional optical remote sensing. Combining top of canopy height model (TCH) and allometry models, this paper constructed two prediction models of aboveground carbon density (ACD) with 94 square plots in northwestern China: one model is plot-averaged height-based power model and the other is plot-averaged daisy-chain model. The correlation coefficients (R^2) were 0.6725 and 0.6761, which are significantly higher than the correlation coefficients of the traditional percentile model ($R^2 = 0.5910$). In addition, the correlation between TCH and ACD was significantly better than that between plot-averaged height (AvgH) and ACD, and Lorey's height (LorH) had no significant correlation with ACD. We also found that plot-level basal area (BA) was a dominant factor in ACD prediction, with a correlation coefficient reaching 0.9182, but this subject requires field investigation. The two models proposed in this study provide a simple and easy approach for estimating ACD in coniferous forests, which can replace the traditional LiDAR percentile method completely.

Keywords: LiDAR, AGB, ACD, allometry model, NPC, CHM, TCH

Edited by:
Cyril Pommier,
INRA Centre Versailles-Grignon,
France

Reviewed by:
András Zlinszky,
Institute of Ecology Research Center
(MTA), Hungary
Eric R. Casella,
Forestry Commission England,
United Kingdom

***Correspondence:**
Qingrui Chang
changqr@nwsuaf.edu.cn
Pengxiang Zhao
zhaopengxiang@nwsuaf.edu.cn

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *07 March 2019*
Accepted: *28 June 2019*
Published: *10 July 2019*

Citation:
Hao H, Li W, Zhao X, Chang Q and Zhao P (2019) Estimating the Aboveground Carbon Density of Coniferous Forests by Combining Airborne LiDAR and Allometry Models at Plot Level. Front. Plant Sci. 10:917.
doi: 10.3389/fpls.2019.00917

INTRODUCTION

Forest carbon storage accounts for 82.5% of terrestrial vegetation carbon storage, which is the main component of the vegetation carbon sink (Cusack et al., 2014; Kauranne et al., 2017). Accurate calculations of forest carbon stocks are a hot topic in the field of forest carbon sink research. At present, large-scale estimations of forest carbon sinks are mainly realized by means of traditional optical remote sensing. Generally, the relationship between field survey data and remote sensing extraction parameters is established first and then extrapolated to the whole research scope; this technique is essentially remote sensing-assisted sampling surveying (Drake et al., 2003; Chi et al., 2017; Koju et al., 2018).

Traditional optical remote sensing can extract the spectral information and horizontal structure information of vegetation. However, with increasing biomass, saturation occurs easily, which affects the estimation accuracy of forest carbon storage (Zhao et al., 2016). Light detection and ranging (LiDAR) detects the distance between a sensor and target by emitting laser pulses and receiving reflections from the ground object. Thus, LiDAR can acquire high-precision three-dimensional information of the object. Furthermore, LiDAR has a certain penetrating ability and can obtain vertical structure information of forests, improving the estimation accuracy of forest height and structure and forest carbon storage (Dubayah and Drake, 2000; Naesset and Bjerknes, 2001; Hudak et al., 2002; Gwenzi and Lefsky, 2014).

With the big-data progress and increasing storage space in recent years, airborne LiDAR has become an important means of forest resource surveys and carbon storage research (Guo et al., 2017; Swetnam et al., 2017). Data processing methods are mainly divided into plot-based inversion and individual tree-based inversion. However, due to the large number of trees, complex spatial structure of forests and canopy shielding effect, single-tree segmentation algorithms are not yet mature. Therefore, developing plot-based inversion methods is indispensable (Ayrey et al., 2017; Dechesne et al., 2017).

There are two main approaches for the estimation of carbon density based on plots. One approach is the use of a variety of machine learning algorithms to establish the relationship between measured carbon density and LiDAR percentile metrics, which can make full use of the information contained in the point cloud to obtain increasing precision (Zhao et al., 2011; McRoberts et al., 2016). However, the modeling process is a black box operation, and the prediction results are difficult to explain. The other approach is the establishment of LiDAR inversion models directly based on allometry models (Mascaro et al., 2011; Asner and Mascaro, 2014). The premise of this method is that there is a similar allometric growth law for plot-level biomass and single-tree biomass. The key to this approach is finding the appropriate allometric growth model and the corresponding LiDAR extraction parameters.

A multiple linear regression model based on LiDAR percentiles is a popular method for estimating forest carbon density or biomass, which is widely used and has acceptable precision in different forest area (e.g., Boudreau et al., 2008; Zhao et al., 2009; Ferraz et al., 2016; Jimenez-Berni et al., 2018). Among them, Naesset and Gobakken (2008) explained 88% variation in aboveground biomass (AGB) and 85% belowground biomass using LiDAR derived variables in boreal forest (1395 circular sample plots with size 200–400 m^2); Levick et al. (2016) used similar methods to obtain the fitting accuracy of 92% in 1 ha plots and 68% in 0.05 ha plots in temperate forest; Cao et al. (2016) established two regression models for estimating the AGB using multi-temporal LiDAR data of subtropical forest with R^2 of 0.74 and 0.79 in 0.09 ha plots, respectively; Dubayah et al. (2010) estimated AGB in tropical forest with an R^2 of 0.90 in 0.5 ha plots. Previous studies have demonstrated that the accuracy and form of percentiles models are closely related to the LiDAR instruments (Naesset, 2009; Silva et al., 2017) and plot size (Maltamo et al., 2011; Mascaro et al., 2011) except for intrinsic characteristics of forest.

In this study, we attempt to find a simple plot-based LiDAR extraction parameter, establish allometry models of the aboveground carbon density (ACD) of the northern coniferous forest, and evaluate the accuracy of these models. The objectives of this study are (1) the selection of the best parameter for ACD prediction from the following three plot-based LiDAR extraction parameters: top of canopy height (TCH), AVG (plot-averaged height), and Lorey's height (LorH); (2) the proposal of direct and indirect fitting models of TCH and ACD and comparison of their accuracy and (3) calculation of the ACD of the study area with the proposed models and comparisons of the results and spatial distribution characteristics.

MATERIALS AND METHODS

Study Area

The study was conducted 50 km southwest of Zhangye City, Gansu Province, Northwest China (**Figure 1**). The study area is approximately 264 ha, and its centre is at 100°15′E, 38°32′N. The elevation ranges from 2700 to 3200 m, the annual rainfall is 200 to 500 mm, and the monthly average temperature is 5.4 to 19.6°C. The main vegetation is natural pure Qinghai spruce (*Picea crassifolia*) forest, which has both naturally renewed young trees and tall over-ripe old trees. The ecoregion classification is "cascade pure conifer forest."

LiDAR Data and Processing

The LiDAR data used in this study were acquired on June 2008 using a LiteMapper 5600 instrument that recorded up to five returns per pulse, along with their intensity. The average flight altitude was 3560 m, the relative height over ground was 760 m, and the flight speed was 227 km/h. The laser scanner adopted RIEGL LMS-Q560, and the wavelength was 1550 nm. The laser pulse width was 3.5 ns, and the laser pulse divergence angle was less than or equal to 0.5 mrad. The LiDAR point cloud used the WGS84 coordinate system and the UTM projection zone 47 in the northern hemisphere. To increase the point density, the flight was repeated seven times over the study area with a side overlap of approximately 90%. As a result, the average point cloud interval was decreased to 0.54 m, and the average point cloud density was 3.43/m^2.

Subsequently, a set of metrics (**Table 1**) was derived from the point cloud using the LAStools software package[1]. The main processing steps were as follows: (1) the point cloud was filtered and classified to ground, trees and noise; (2) the normalized point cloud (NPC, also referred to as height above ground) was calculated; (3) height percentiles, density percentiles and canopy cover (CC) were derived from the NPC corresponding to each plot; (4) the digital surface model (DSM) and digital elevation model (DEM)were interpolated from the first echo and the last echo of the point cloud, respectively. The canopy height model (CHM) was the difference of the first two. (5) The TCH was

---

[1] http://www.lastools.org

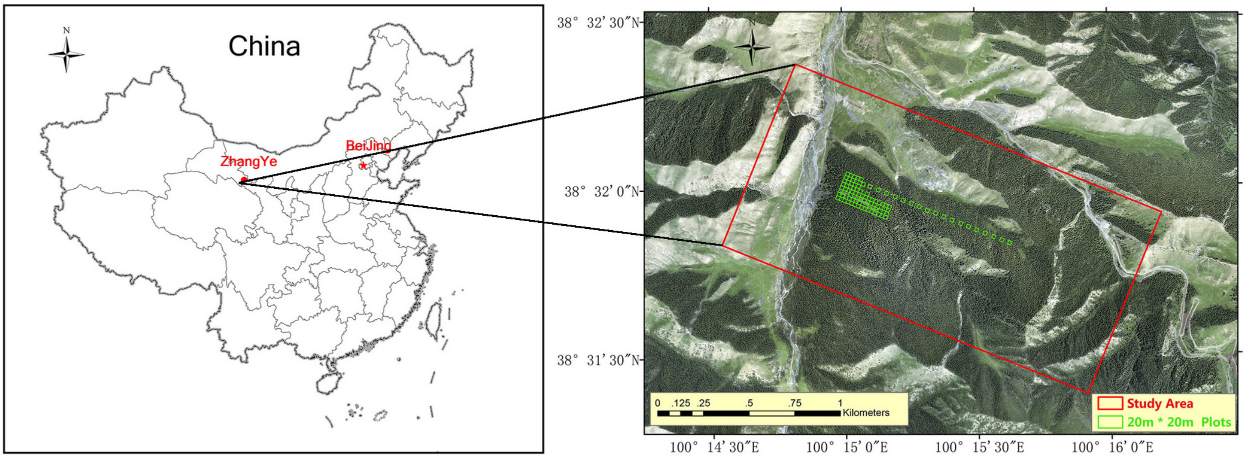


FIGURE 1 | Study area and 94 plots.

extracted from the CHM based on each plot (the mean value of 400 pixels per plot).

In addition, in order to explore the effect of CHM pixel size on ACD prediction, we generated 10 CHMs from NPC, with pixel sizes from 1 to 10 m. When the pixel location of CHM corresponds to a laser point, the point's height value is used as the pixel value. If the location corresponds to multiple laser points, the average value of height is used as the pixel value. For the pixels without corresponding laser point, inverse distance weighted (IDW) is adopted for interpolation, which can ensure smooth transition between the target pixel and surrounding pixels (Montealegre et al., 2015).

Field Data and Processing

To calibrate and validate the models, the plot data were acquired simultaneously with the LiDAR data. A total of 94 square plots (20 m × 20 m), which included 5734 trees, were used. The four corners and the centre of each plot were measured using differential GPS (DGPS), and the error was less than 10 cm. For each tree with a diameter at breast height (DBH) greater than 5 cm, the tree type, diameter, height to crown base, crown width in cardinal directions, crown class, and crown transparency were measured. DBH was measured on all trees using a diameter tape, and the heights of all trees were measured using a laser ranging hypsometer with theoretical accuracy up to the decimeter level. Considering the canopy occlusion and human error, the average accuracy of the measured tree height was better than 0.5 m.

TABLE 1 | Metrics derived from LiDAR and field investigation data.

LiDAR Metric	Description	Origin Source
TCH	Top of canopy height of plot	CHM (canopy height model)
h25…h95	Height percentiles	NPC (normalized point cloud)
d25…d95	Density percentiles	
CC	Canopy closure	
AvgH	Average height of plot	Field investigation
LorH	Lorey's height of plot	
BA	Base area of plot	

Using the species-specific allometry Eqs 1–4 in the study area (Wang et al., 1998; He et al., 2013), tree biomass components (stem, branch, foliage and fruit) were calculated from DBH and height. The AGB of each tree was equal to the sum of the AGB components and was then summed to obtain the AGB of each plot. These equations were constructed by destructive sampling and the fitting precision reached 0.9887, 0.9568, 0.8662, and 0.9340, respectively. Because the study area is a nature reserve and pure spruce forest with little human disturbance, it is believed that the field-estimated AGB of this study can also achieve such accuracy. Finally, the AGB was converted to the ACD using the conversion coefficient of 0.5034, which was attained using the potassium dichromate oxidation method on samples by Wang et al. (2000) in the same area. Additionally, Lorey's height (LorH; **Table 1**) was also calculated based on each plot using Eq. 5, and the LorH values were compared with the TCH values extracted from LiDAR. For the same purpose, the average height (AvgH) of the plots was also calculated.

$$\text{Biomass\_stem} = 0.0478 \times \left(D^2 \times H\right)^{0.8665} \quad (1)$$

$$\text{Biomass\_branch} = 0.0061 \times \left(D^2 \times H\right)^{0.8905} \quad (2)$$

$$\text{Biomass\_foliage} = 0.2650 \times \left(D^2 \times H\right)^{0.4701} \quad (3)$$

$$\text{Biomass\_fruit} = 0.0342 \times \left(D^2 \times H\right)^{0.5779} \quad (4)$$

$$\text{LorH} = \sum \text{BA}_i \text{H}_i / \sum \text{BA}_i \quad (5)$$

where D is DBH, H is tree height, BA_i and H_i are the basal area and the height of the ith tree, respectively, and a and b are regression fitting coefficients.

Plot Allometry Models

The use of an allometry model is the main means of forest biomass calculation. This type of model is obtained by the regression of the sample forest harvesting and tree-measuring metrics and is a single-tree model for specific tree species in a specific region. The present work imitates the form of single-tree models at the plot level to find a suitable plot-level LiDAR metric to replace traditional tree-measuring metrics.

An idealized and simple tree allometry equation for special species is:

$$\mathrm{AGB} = aD^{\mathrm{b}} \tag{6}$$

Since DBH is the most easily accessible and accurately measurable tree indicator, and there is an intrinsic relationship between the DBH and tree height, the model is widely used (Chave et al., 2005).

However, Eq. 6 cannot explain the variability of diameter and tree height growth caused by tree age, forest density, site conditions and management measures; the introduction of the tree height factor is necessary.

$$\mathrm{AGB} = aD^{b1}H^{b2} \tag{7}$$

where H represents the tree height (m), and a, $b1$, and $b2$ are fitted coefficients.

The essence of LiDAR is ranging, which can directly estimate tree height. Therefore, this paper applies Eqs 6 and 7 to Eqs 8 and 9, which are plot-averaged height-based allometry models, and fits the equations as follows:

$$\mathrm{ACD} = aH^{b} \tag{8}$$

$$\mathrm{ACD} = a\mathrm{BA}'^{b1}\mathrm{TCH}^{b2} \tag{9}$$

where ACD represents the aboveground carbon density (Mg C ha^{-1}), and H represents AvgH, LorH, or BA computed from field-measured data and TCH extracted from the CHM. BA' is the fitted result in Eq. 10.

$$\mathrm{BA}' = a + b\mathrm{TCH} \tag{10}$$

Percentile Model

Light detection and ranging is highly sensitive to the three-dimensional structure of forest, because laser pulses can penetrate the canopy and then record all echo signals from the ground to the canopy surface. Therefore, a series of LiDAR metrics, such as height percentile, density percentile, variation coefficient, etc., have been successively extracted to capture key information of forest canopy (Nilsson, 1996; Lefsky et al., 2002; Naesset, 2002; Nelson et al., 2004).

In this study, the height and density percentiles extracted from the NPC were used to regression fit the ACD calculated from the field investigation data. The model is as follows, and the independent variables are described in **Table 1**. The prediction results of this model (Eq. 11) were compared with the prediction results of the allometry models (Eqs 8, 9) proposed in this paper.

$$\begin{aligned}\mathrm{lnACD} = {} & \beta0 + \beta1\mathrm{lnh25} + \beta2\mathrm{lnh50} + \beta3\mathrm{lnh75} + \beta4\mathrm{lnh90} + \\ & \beta5\mathrm{lnh95} + \beta6\mathrm{lnd25} + \beta7\mathrm{lnd50} + \beta8\mathrm{lnd75} + \beta9\mathrm{lnd90} + \\ & \beta10\mathrm{lnd95} + \beta11\mathrm{lnCC} + \varepsilon\end{aligned} \tag{11}$$

Model Fitting and Evaluation

All models in this study were fitted by the least squares (OLS) method (Meng et al., 2018). This method is simple and reliable, avoiding the algorithm differences of different fitting methods and making the fitting results more contrasting. All power models were first converted to a linear model by natural log transformation for regression fitting to correct both non-normality and heteroscedasticity. Then, the correlation of coefficients (R^2) and back-log root-mean-squared errors (RMSE) were employed to compare the performance of the models, and 10-fold cross-validation analysis was used to evaluate the stability of the models.

Model Application

The study area was divided into a 20 m × 20 m grid using GIS software (**Figure 2**). The size and direction of the grid were the same as those of the field plots in order to reduce possible errors. Then, LiDAR parameters were extracted using each grid from the CHM and NPC and introduced into the allometry models (Eqs 8, 11) proposed in this paper for calculation. Thereby, ACD distribution maps of the study area were obtained, compared and evaluated. Since there was no point cloud at the boundary of the study area, grid incompleteness due to cropping did not affect the final prediction.

RESULTS

Field Investigation Data Analysis

Using the comparison between the simple power-law model (Eq. 8) of ACD and the three plot-averaged metrics (AvgH, LorH, and BA)calculated from the field inventory, we found that the BA explains 91.8% of the variation in ACD, which is much higher than the 39.5% explained by AvgH and 10.1% explained by LorH (**Table 2**), and the convergence of BA is much better than that of AvgH and LorH (**Figure 3**), which indicates that BA is the optimal plot-averaged indicator for the inversion of ACD. It is not surprising that BA is a stronger predictor of AGB than height is, because BA can be measured with a lot better accuracy than height, and because DBH is weighted higher than tree height in Eq. 7. We also found that the exponent b is close to 1 (b = 1.0300), which indicates that the ACD is nearly linearly related to BA, and the cross-validated R^2 value (0.9182) is reduced by 0.004 compared with the model-fitted R^2 (0.9143), which indicates that the model using BA tends to be applicable and stable. This result means that, when

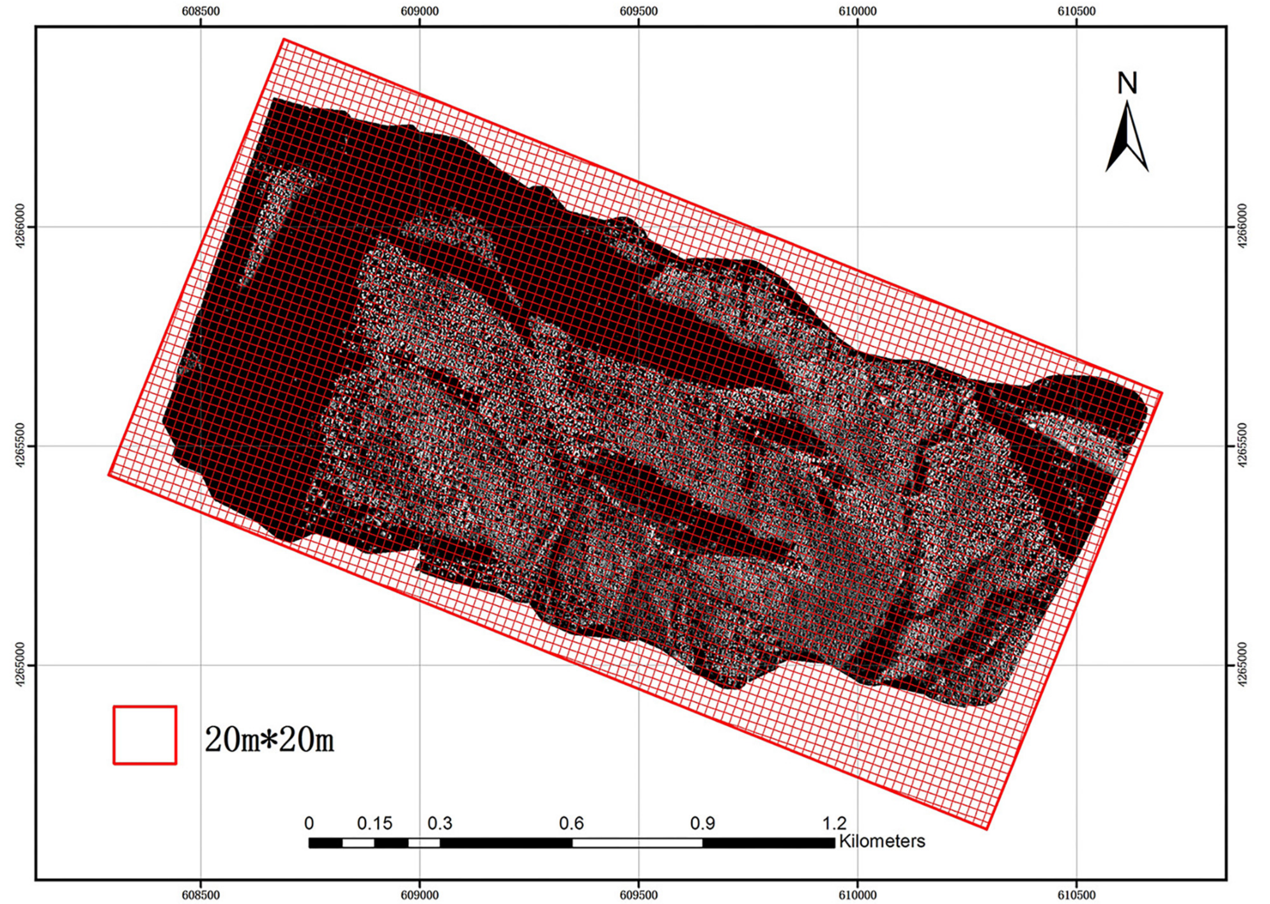


FIGURE 2 | Fishnet for the study area.

TABLE 2 | Summary for ACD estimation using AvgH, LorH, and BA.

Model	**Model parameters**		**R^2**	**Jackknifing R^2**	**RMSE (Mg C ha^{-1})**
	a	***b***			
ACD = aAvgH^b	8.7899	0.8026	0.3945	0.3587	13.5561
ACD = aLorH^b	7.8821	0.6691	0.1014	0.0614	15.9938
ACD = aBA^b	1.6528	1.0345	0.9182	0.9143	5.5440

we want to obtain the ACD of plots, we can discard the exhaustive field inventory data and only need to perform spatially explicit point-based measurements using the relascope or prism method.

Moreover, although AvgH and LorH are the most commonly used plot-averaged height indicators, when they were applied in Eq. 8 to predict ACD, the effect was poor, with R^2 values of 0.3954 and 0.1014, respectively, and RMSE values of 13.5561 (Mg C ha^{-1}) and 15.9938 (Mg C ha^{-1}), respectively; the results with AvgH are slightly better than those with LorH (**Table 2**). This suggests that the plot-averaged height alone does not account for the variation in the ACD. Therefore, the plot-level ACD estimation (EACD) based on LiDAR should exclude the AvgH and LorH steps and directly fit ACD with LiDAR-extracted metrics (**Figure 3D**).

TCH Models and the Comparisons

The plot-level LiDAR metric (TCH) was taken into the ideal simplest allometry (Eq. 8) and was subjected to log changes and linear fits. The result showed that TCH could explain 67.25% of the variation in ACD (**Figure 4A**), suggesting that TCH is a simple and effective predictor of ACD from LiDAR and that the classic allometry model (Eq. 6) can be extended from the tree-level to the plot-level scale.

Using regression by ordinary least squares, we modeled variation in BA to TCH for 94 plots, with resulting values of R^2 = 0.6066 and RMSE = 5.1749 m^2 ha^{-1} (Eq. 10, **Figure 4B**,

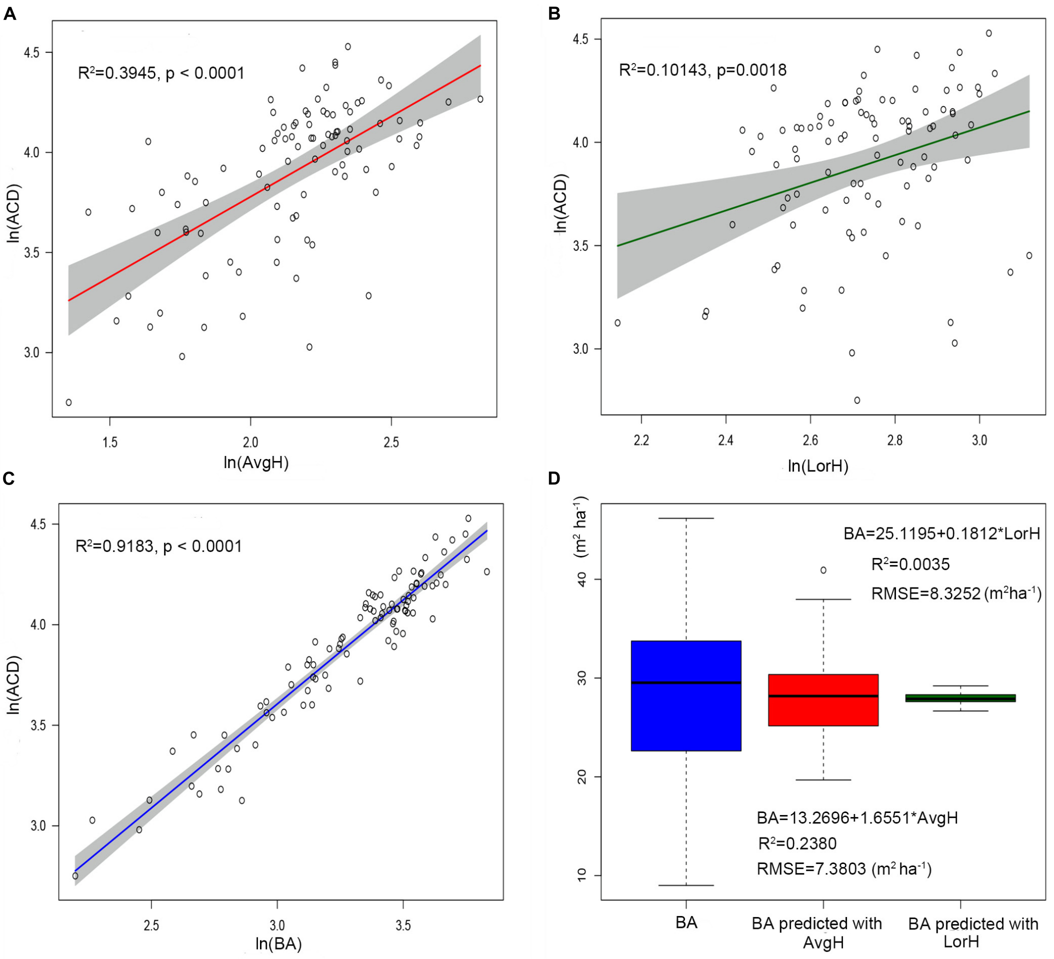


FIGURE 3 | The linear relationship between ln(AvgH) and ln(ACD) **(A)**, ln(LorH) and ln(ACD) **(B)**, ln(BA)and ln(ACD) **(C)**. The shadowed region shows the 95% confidence interval. **(D)** Boxplots of the field-surveyed BA and predicted BA using plot-averaged height (AvgH) and Lorey's height (LorH).

and **Table 3**). By substituting this regression result into Eq. 9, the EACD could be generated without field inventory data. However, in a comparison of the fitting results, we found that the scatter plots were almost the same (**Figures 4C,D**); R^2 only increased by 0.0036, and RMSE increased by 0.1163 Mg C ha^{-1} (**Table 3**). This strongly suggests that the daisy-chain method of TCH cannot achieve the same ACD prediction as the field-measured BA. Therefore, if we only use LiDAR-extracted TCH, the height-diameter model (Eq. 9) and height model (Eq. 8) have no essential difference in accuracy. Thus, this paper ultimately chose Eq. 8 as the final allometry model, and the parameter H used TCH.

Percentile Model and a Comparison of Results

The following result (Eq. 12) was obtained by the multiple regression fitting of the surveyed ACD of 94 plots and the LiDAR percentile metrics listed in **Table 1**.

$$\ln \mathrm{ACD} = 1.896 + 0.033 \ln h25 + 12.106 \ln d95 \qquad (12)$$

The ACD in the study area is closely related to h25 and d95. These two parameters can explain 59.1% of the ACD variation, with a RMSE of 11.6304 Mg C ha^{-1} (**Figure 5A**), and the prediction accuracy is lower than the 67.25% of Eq. 8 using TCH. It demonstrated that the height allometry model proposed in this paper can replace the traditional LiDAR percentile model with improved precision.

Figure 5B illustrates the difference between the predicted values and the survey values of ACD. The median values of the MLR model and the TCH model are near 52 Mg C ha^{-1}, which is slightly lower than the measured ACD. Furthermore, the range of predicted values of the TCH model is slightly smaller than the surveyed value range, which is larger than the range of the MLR model. Therefore, compared with the MLR model, the TCH model has a wider prediction range and can represent larger and smaller values of ACD.

Model Application

All grid values in the study area were calculated using our proposed TCH allometry model and percentile model, and then

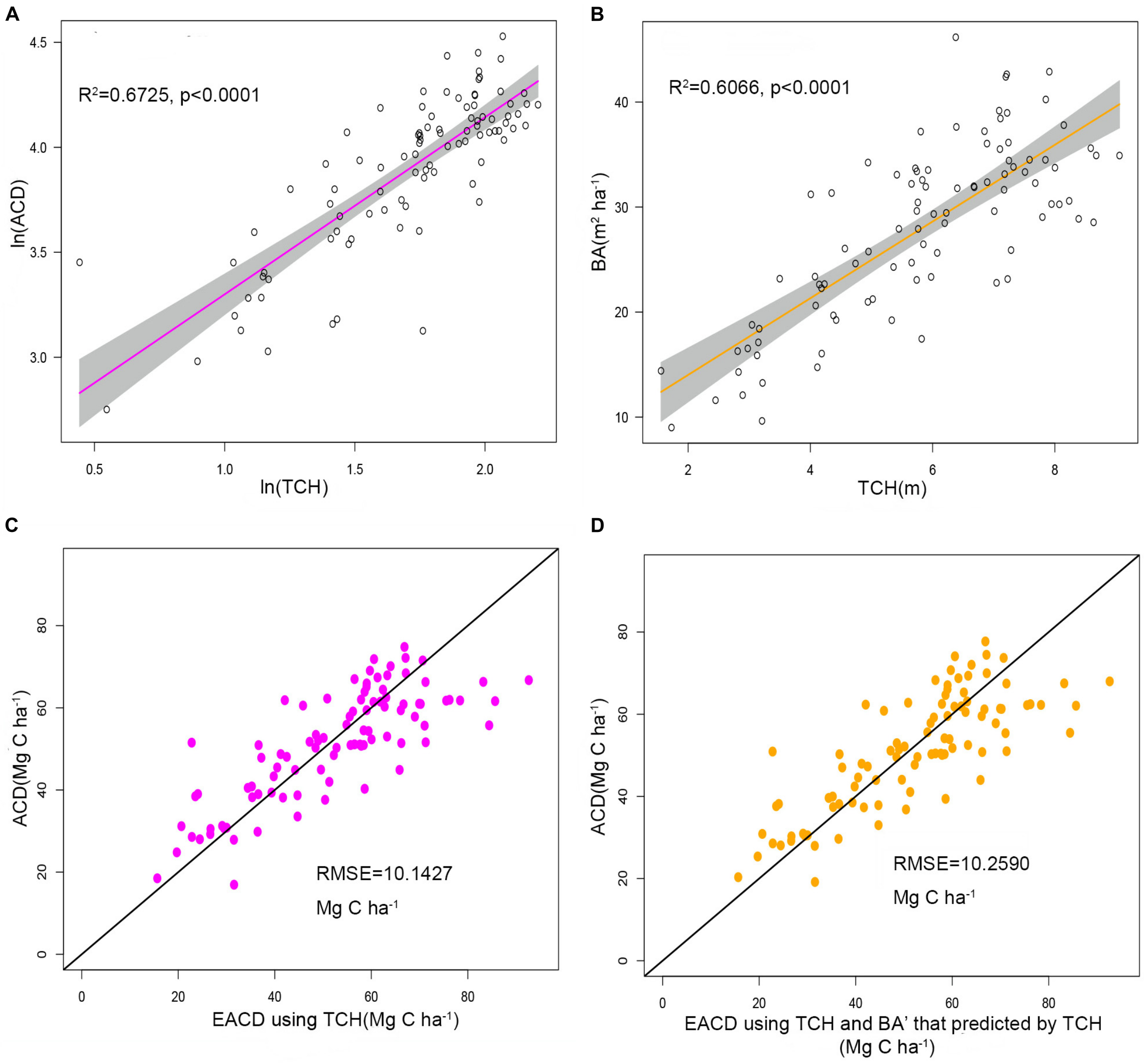


FIGURE 4 | (A) Linear relationship between ln(TCH) and ln(ACD). **(B)** Linear relationship between TCH and the basal area (BA). The shadowed region shows the 95% confidence interval. **(C)** ACD estimated using top of canopy height (TCH) from LiDAR compared to field-surveyed ACD. **(D)** ACD estimated using TCH and BA predicted from TCH compared to field-surveyed ACD. The black line is a 1:1 reference line.

TABLE 3 | Summary for ACD estimation using TCH only or TCH and BA' in pairs.

Model	**Model parameters**			**R^2**	**Jackknifing R^2**	**RMSE (Mg C ha^{-1})**
	a	***b*1**	***b*2**			
ACD = *a*TCH*b*1	11.6592	0.8436	–	0.6725	0.6585	10.1427
BA' = *a*+*b*1TCH	6.7117	3.6516	–	0.6066	0.5882	5.1749 (m^2 ha^{-1})
ACD = *a*BA'*b*1TCH*b*2	0.9393	1.1977	0.0018	0.6761	0.6022	10.2590

maps of ACD were produced. **Figure 6** shows that the spatial distribution of the two maps is very similar. The high-density area of the map from the percentile model is slightly larger than that of the map from the TCH model (blue circle), and the low-density area demonstrates the opposite trend (blue rectangle). In addition, the density distribution percentages of the two maps are basically the same as those shown in the two pie charts. According to the grid statistics, the average ACD from the TCH model is 41.49 Mg C ha^{-1}, and the maximum value is 104.70 Mg C ha^{-1}, which is slightly larger than the values of 40.13 and 95.46 Mg C ha^{-1} from the percentile model. This resulted in an overall aboveground carbon reserve of the study area of 5535.54 Mg for the TCH model and 5433.06 Mg for the percentile model; the difference between the two models is only 1.89%.

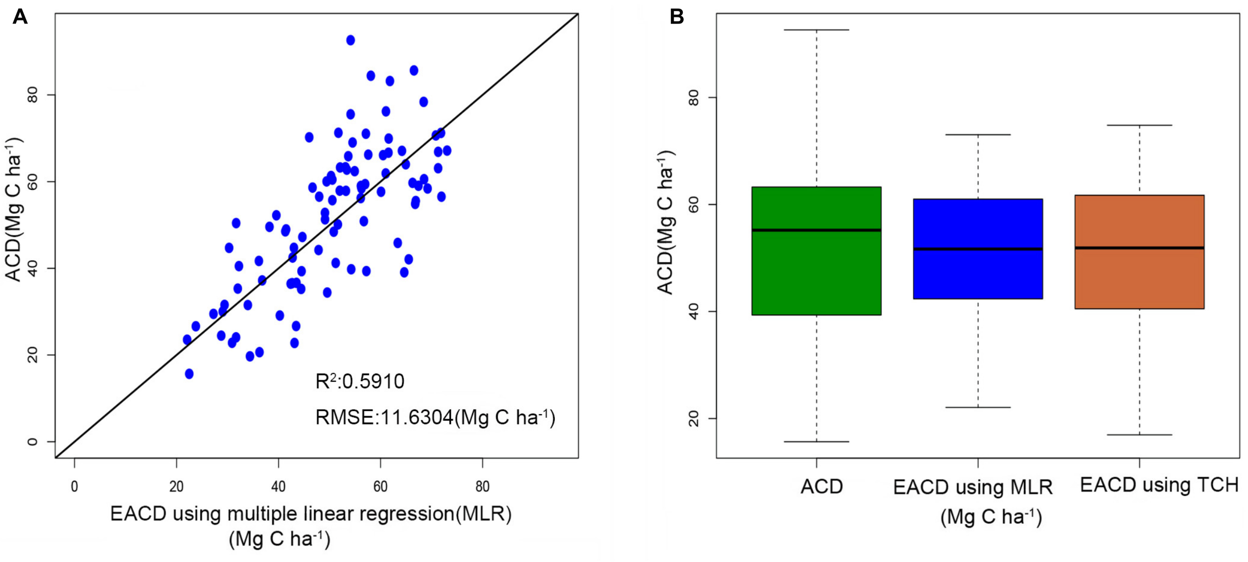


FIGURE 5 | **(A)** Estimated ACD values of the percentile model versus the field investigation. **(B)** Boxplots of field-surveyed ACD and ACD estimated using the percentile model and TCH model.

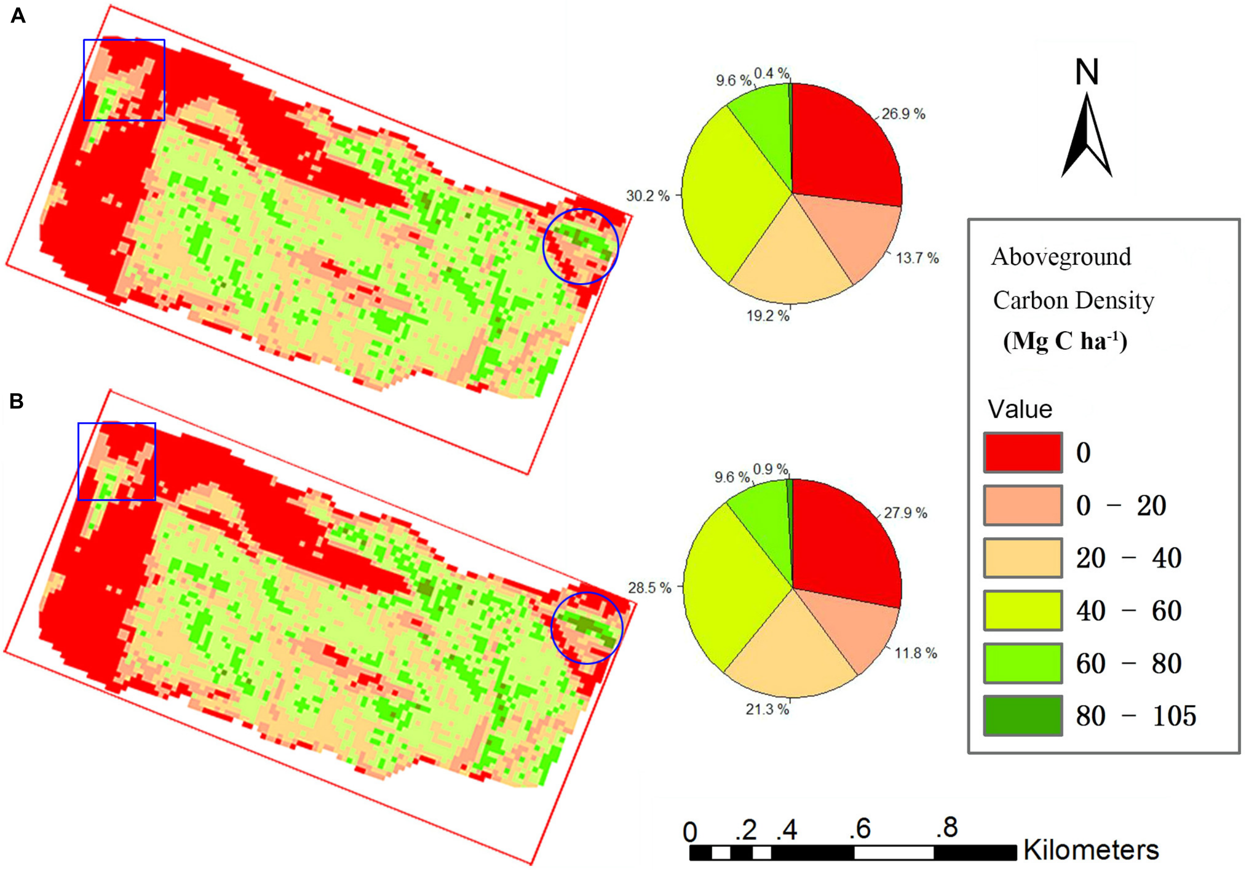


FIGURE 6 | Study area carbon density map predicted using the TCH model **(A)** and percentile model **(B)**. Square is low density area, Round is high density area.

Although the accuracies of the TCH and the MLR models are not much different, the TCH model is much simpler and easier than the MLR model.

DISCUSSION

Our original purpose was to find a suitable plot-averaged LiDAR parameter and use existing allometry models to quickly and accurately predict the forest carbon density. The exponential model of TCH captures 67.25% of ACD changes (**Table 3**) and has a higher accuracy than the traditional percentile model of this study. We also realized that the accuracy of our prediction is relatively low. The possible reasons are (1) the plot size of 20 m × 20 m is relatively small, and the edge effect is obvious; (2) point cloud density is not enough, and the conical crown of spruce is not captured accurately and (3) the penetration rate of point cloud is insufficient, and the detection of lower wood is limited. However, for the research objective, we did effectively improve the accuracy of LiDAR's prediction of ACD, simplify the prediction steps and solidify the form of the prediction model.

Moreover, since the TCH is derived from the mean of the CHM based on the plots, the TCH is also subject to the pixel size. We extracted 10 CHMs from the LiDAR point cloud, with pixel sizes from 1 to 10 m, and then extracted the corresponding TCHs to fit the ACD. As the pixel size increased, R^2 continually decreased, and the RMSE continually increased (**Figure 7**). This result indicated that the smaller the CHM pixel is, the better the fitting effect of TCH will be. This study was limited to a point cloud density of 3.43/m^2, so the minimum pixel size was 1 m. In addition, we found that when the pixel sizes were 5 and 7 m, the fitting effect fluctuated slightly, but this fluctuation did not affect the overall law. The reason for this finding requires further study.

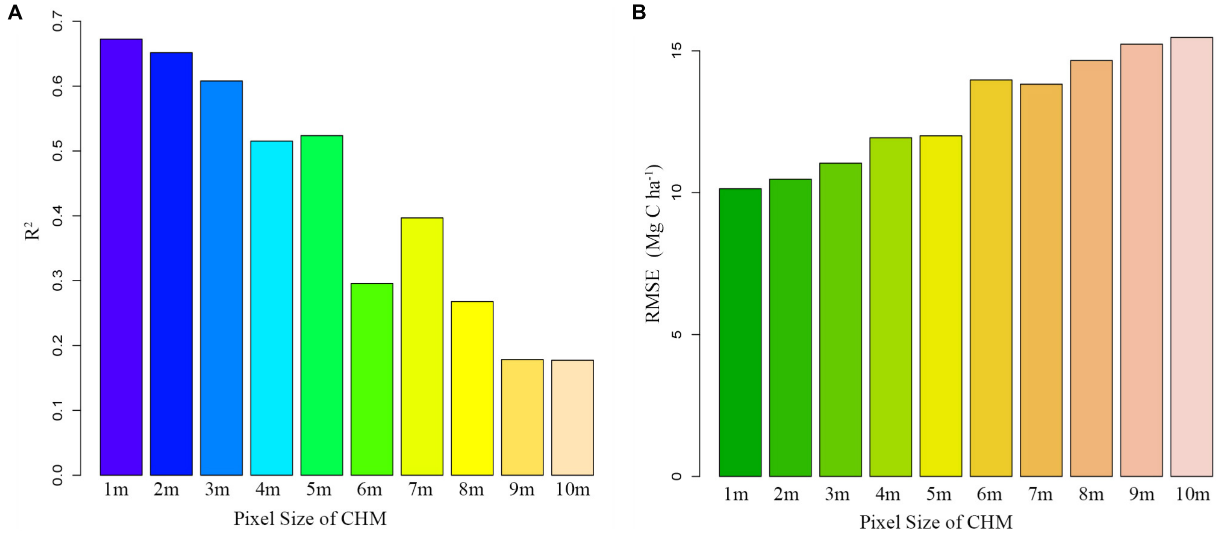


FIGURE 7 | Fitting trends of TCH and ACD under different CHM precisions. **(A)** Declining R^2; **(B)** Increasing RMSE.

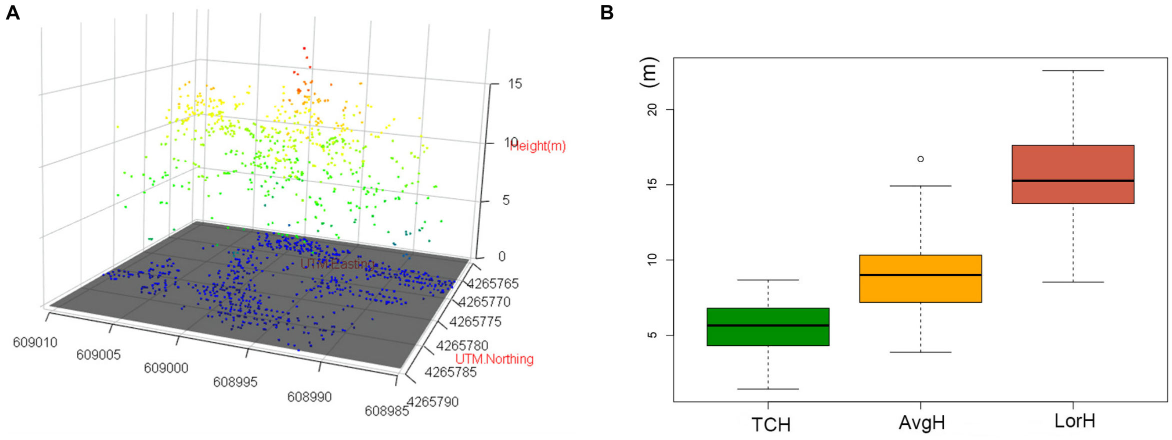


FIGURE 8 | (A) 3D scatter plot of the normalized point cloud (NPC) in plot 1. **(B)** Boxplots of TCH, AvgH and LorH. White bullet is mild outliers of boxplot.

Similarly, fitting accuracy is also limited by the size of the plot and the number of samples. A larger plot size means a smaller boundary effect, and a larger plot number means a smaller outlier influence (Ni et al., 2014; Gwenzi and Lefsky, 2017). However, larger plots are more expensive and time-consuming than smaller plots, so finding an optimal plot size and number in coniferous forest will be an important task for future study.

The three-dimensional visualization of the point cloud in plot no. 1 (**Figure 8A**) suggests that the forest point cloud includes the crown, some of the lower layer, some of the trunks and the woodland gap. Therefore, the TCH data derived from the point cloud also contain the above information. However, the average height of the plot (AvgH) ignores the forest gap and is therefore slightly higher than the TCH (**Figure 8B**). Although LorH is widely used for the estimation of forest biomass (Mitchard et al., 2012; Gwenzi and Lefsky, 2016), LorH is mainly used to evaluate site quality and mostly reflects the largest trees in the forest; therefore, its value is larger than that of TCH and AvgH (**Figure 8B**), and LorH is not applicable for fitting the ACD of irregular and mixed forests. This explains why TCH is the optimal ACD predictor.

We also recognized a flaw in the ACD prediction at the plot scale. Whether in the field measurement phase, the plot-based TCH extraction phase, or the final ACD prediction phase, our resolution is fixed at 20 m × 20 m. This inevitably leads to the conversion of the continuous ACD distribution in nature into a discontinuous distribution, which may cause a large jump phenomenon at the boundary. Therefore, selecting the appropriate interpolation algorithm to restore the continuity of the ACD will help improve the prediction accuracy of our proposed models (Loquin and Strauss, 2010; DeWitt et al., 2017). In addition, we only adopted linear regression fitting based on the least squares method in this paper, and although this method is simple and practical, it is not necessarily the best method. With the rise of machine learning in LiDAR research (Zhou et al., 2017; Jin et al., 2018; Lin et al., 2018), it will be necessary to compare various machine learning algorithms in future research to find the best way to fit the allometry models.

Finally, we must emphasize that although our proposed TCH-based allometric approach is an efficient LiDAR-assisted ACD prediction method, the allometry model used for plot calculation is generally targeted to a specific region and species (Picard et al., 2015; Duque et al., 2017), so it is necessary to re-select an appropriate allometry model for other tree species and ecological regions when our method is used. Moreover, developing a general ACD prediction model based on LiDAR for forests across ecological regions and species will be the focus of our future research.

CONCLUSION

Using the traditional allometry growth model theory, this paper proposed two models based on TCH extracted from LiDAR data. The first model was a simple power model (only using TCH) based on the diameter allometry, and the second model was a daisy-chain model ($\text{TCH} \rightarrow \text{BA}' \rightarrow \text{ACD}$) based on diameter-height allometry. A comparison of the results suggested there was little difference in the fitting accuracy and error distribution between models. In addition, this paper compared the traditional LiDAR percentile method with the proposed method and found that the latter method had a higher precision, fewer parameters, more concise steps and more stable forms than the former method. Furthermore, the implicit hypothesis in our study, the traditional allometry model of individual trees can be extrapolated to the plot scale, was confirmed. The LiDAR-assisted ACD estimation method proposed in this study will accelerate the application of airborne LiDAR technology in forest carbon density measurements and provide an accurate data basis for forest ecosystem research.

DATA AVAILABILITY

Publicly available datasets were analyzed in this study. This data can be found here: http://www.heihedata.org/data/.

AUTHOR CONTRIBUTIONS

HH and WL: conceptualization, methodology, and writing-original draft preparation. HH: software, resources, and data curation. XZ and PZ: validation. HH and XZ: formal analysis. HH, WL, and XZ: investigation. QC: writing, review, and editing. HH and PZ: visualization. QC and PZ: supervision, project administration, and funding acquisition.

FUNDING

This research was funded by the National Key Research and Development Program of China (Grant No. 2016YFD0600203), the Introducing the International Advanced Agricultural Science and Technology Program (948 Program) of China (Grant No. 2014-4-27), the National Natural Science Foundation of China (Grant No. 41701398), and Chinese Universities Scientific Fund (Grant No. 2452017108).

ACKNOWLEDGMENTS

The authors would like to thank Northwest Institute of Eco-Environment and Resources (CAS) for support in LiDAR data acquisition and also thank Gansu Province Qilian Water Resource Conservation Forest Research Institute for support in field investigation.

REFERENCES

Asner, G. P., and Mascaro, J. (2014). Mapping tropical forest carbon: calibrating plot estimates to a simple LiDAR metric. *Remote Sens. Environ.* 140, 614–624. doi: 10.1016/j.rse.2013.09.023

Ayrey, E., Fraver, S., Kershaw, J. A., Kenefic, L. S., Hayes, D., Weiskittel, A. R., et al. (2017). Layer stacking: a novel algorithm for individual forest tree segmentation from LiDAR point clouds. *Can. J. Remote Sens.* 43, 16–27. doi: 10.1080/07038992.2017.1252907

Boudreau, J., Nelson, R. F., Margolis, H. A., Beaudoin, A., Guindon, L., and Kimes, D. S. (2008). Regional aboveground forest biomass using airborne and spaceborne LiDAR in Quebec. *Remote Sens. Environ.* 112, 3876–3890. doi: 10.1016/j.rse.2008.06.003

Cao, L., Coops, N. C., Innes, J. L., Sheppard, S. R. J., Fu, L. Y., Ruan, H. H., et al. (2016). Estimation of forest biomass dynamics in subtropical forests using multi-temporal airborne LiDAR data. *Remote Sens. Environ.* 178, 158–171. doi: 10.1016/j.rse.2016.03.012

Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., et al. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia* 145, 87–99. doi: 10.1007/s00442-005-0100-x

Chi, H., Sun, G., Huang, J., Li, R., Ren, X., Ni, W., et al. (2017). Estimation of forest aboveground biomass in changbai mountain region using ICESat/GLAS and Landsat/TM Data. *Remote Sens.* 9:707. doi: 10.3390/rs9070707

Cusack, D. F., Axsen, J., Shwom, R., Hartzell-Nichols, L., White, S., and Mackey, K. R. (2014). An interdisciplinary assessment of climate engineering strategies. *Front. Ecol. Environ.* 12, 280–287. doi: 10.1890/130030

Dechesne, C., Mallet, C., Le Bris, A., and Gouet-Brunet, V. (2017). Semantic segmentation of forest stands of pure species combining airborne lidar data and very high resolution multispectral imagery. *ISPRS J. Photogramm. Remote Sens.* 126, 129–145. doi: 10.1016/j.isprsjprs.2017.02.011

DeWitt, J. D., Warner, T. A., Chirico, P. G., and Bergstresser, S. E. (2017). Creating high-resolution bare-earth digital elevation models (DEMs) from stereo imagery in an area of densely vegetated deciduous forest using combinations of procedures designed for lidar point cloud filtering. *Gisci. Remote Sens.* 54, 552–572. doi: 10.1080/15481603.2017.1295514

Drake, J. B., Knox, R. G., Dubayah, R. O., Clark, D. B., Condit, R., Blair, J. B., et al. (2003). Above-ground biomass estimation in closed canopy Neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Glob. Ecol. Biogeogr.* 12, 147–159. doi: 10.1046/j.1466-822X.2003.00010.x

Dubayah, R. O., and Drake, J. B. (2000). Lidar remote sensing for forestry. *J. For.* 98, 44–46.

Dubayah, R. O., Sheldon, S. L., Clark, D. B., Hofton, M. A., Blair, J. B., Hurtt, G. C., et al. (2010). Estimation of tropical forest height and biomass dynamics using lidar remote sensing at La Selva, Costa Rica. *J. Geophys. Res.* 115:G00E09. doi: 10.1029/2009jg000933

Duque, A., Saldarriaga, J., Meyer, V., and Saatchi, S. (2017). Structure and allometry in tropical forests of Choco. Colombia. *For. Ecol. Manage.* 405, 309–318. doi: 10.1016/j.foreco.2017.09.048

Ferraz, A., Saatchi, S., Mallet, C., Jacquemoud, S., Goncalves, G., Silva, C. A., et al. (2016). Airborne lidar estimation of aboveground forest biomass in the absence of field inventory. *Remote Sens.* 8:653. doi: 10.3390/rs8080653

Guo, X., Coops, N. C., Tompalski, P., Nielsen, S. E., Bater, C. W., and Stadt, J. J. (2017). Regional mapping of vegetation structure for biodiversity monitoring using airborne lidar data. *Ecol. Inform.* 38, 50–61. doi: 10.1016/j.ecoinf.2017.01.0051574-9541

Gwenzi, D., and Lefsky, M. A. (2014). Modeling canopy height in a savanna ecosystem using spacebome lidar waveforms. *Remote Sens. Environ.* 154, 338–344. doi: 10.1016/j.rse.2013.11.024

Gwenzi, D., and Lefsky, M. A. (2016). Plot-level aboveground woody biomass modeling using canopy height and auxiliary remote sensing data in a heterogeneous savanna. *J. Appl. Remote Sens.* 10:016001. doi: 10.1117/1.jrs.10.016001

Gwenzi, D., and Lefsky, M. A. (2017). Spatial modeling of lidar-derived woody biomass estimates collected along transects in a heterogeneous savanna landscape. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 10, 372–384. doi: 10.1109/Jstars.2016.2582148

He, Q. S., Chen, E. X., An, R., and Li, Y. (2013). Above-ground biomass and biomass components estimation using LiDAR data in a coniferous forest. *Forests* 4, 984–1002. doi: 10.3390/f4040984

Hudak, A. T., Lefsky, M. A., Cohen, W. B., and Berterretche, M. (2002). Integration of lidar and Landsat ETM plus data for estimating and mapping forest canopy height. *Remote Sens. Environ.* 82, 397–416. doi: 10.1016/S0034-4257(02)00056-1

Jimenez-Berni, J. A., Deery, D. M., Rozas-Larraondo, P., Condon, A. T. G., Rebetzke, G. J., James, R. A., et al. (2018). High throughput determination of plant height, ground cover, and above-ground biomass in wheat with LiDAR. *Front. Plant Sci.* 9:237. doi: 10.3389/fpls.2018.00237

Jin, S., Su, Y., Gao, S., Wu, F., Hu, T., Liu, J., et al. (2018). Deep learning: individual maize segmentation from terrestrial lidar data using faster R-CNN and regional growth algorithms. *Front. Plant Sci.* 9:866. doi: 10.3389/fpls.2018.00866

Kauranne, T., Joshi, A., Gautam, B., Manandhar, U., Nepal, S., Peuhkurinen, J., et al. (2017). LiDAR-assisted multi-source program (LAMP) for measuring above ground biomass and forest carbon. *Remote Sens.* 9:154. doi: 10.3390/Rs9020154

Koju, U. A., Zhang, J., Maharjan, S., Zhang, S., Bai, Y., Vijayakumar, D. B. I. P., et al. (2018). A two-scale approach for estimating forest aboveground biomass with optical remote sensing images in a subtropical forest of Nepal. *J. For. Res.* 31, 1–18.

Lefsky, M. A., Cohen, W. B., Parker, G. G., and Harding, D. J. (2002). Lidar remote sensing for ecosystem studies. *Bioscience* 52, 19–30.

Levick, S. R., Hessenmoller, D., and Schulze, E. D. (2016). Scaling wood volume estimates from inventory plots to landscapes with airborne LiDAR in temperate deciduous forest. *Carbon Balance Manag.* 11:7. doi: 10.1186/s13021-016-0048-7

Lin, Y., Jiang, M., Pellikka, P., and Heiskanen, J. (2018). Recruiting conventional tree architecture models into state-of-the-art LiDAR mapping for investigating tree growth habits in structure. *Front. Plant Sci.* 9:220. doi: 10.3389/fpls.2018.00220

Loquin, K., and Strauss, O. (2010). "Convolution filtering and mathematicalmorphology on an image: a unified view," in *Linear Filtering and Mathematical Morphology on an Image: A Bridge. ICIP: International Conference on Image Processing*, Cairo, 3965–3968.

Maltamo, M., Bollandsas, O. M., Naesset, E., Gobakken, T., and Packalen, P. (2011). Different plot selection strategies for field training data in ALS-assisted forest inventory. *Forestry* 84, 23–31. doi: 10.1093/forestry/cpq039

Mascaro, J., Detto, M., Asner, G. P., and Muller-Landau, H. C. (2011). Evaluating uncertainty in mapping forest carbon with airborne LiDAR. *Remote Sens. Environ.* 115, 3770–3774. doi: 10.1016/j.rse.2011.07.019

McRoberts, R. E., Domke, G. M., Chen, Q., Naesset, E., and Gobakken, T. (2016). Using genetic algorithms to optimize k-nearest neighbors configurations for use with airborne laser scanning data. *Remote Sens. Environ.* 184, 387–395. doi: 10.1016/j.rse.2016.07.007

Meng, R., Dennison, P. E., Zhao, F., Shendryk, I., Rickert, A., Hanavan, R. P., et al. (2018). Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements. *Remote Sens. Environ.* 215, 170–183. doi: 10.1016/j.rse.2018.06.008

Mitchard, E. T. A., Saatchi, S. S., White, L. J. T., Abernethy, K. A., Jeffery, K. J., Lewis, S. L., et al. (2012). Mapping tropical forest biomass with radar and spaceborne LiDAR in Lope National Park, Gabon: overcoming problems of high biomass and persistent cloud. *Biogeosciences* 9, 179–191. doi: 10.5194/bg-9-179-2012

Montealegre, A., Lamelas, M., and Riva, J. (2015). Interpolation routines assessment in ALS-derived digital elevation models for forestry applications. *Remote Sens.* 7, 8631–8654. doi: 10.3390/rs70708631

Naesset, E. (2002). Predicting forest stand characteristics with airborne scanning laser using a practical two-stage procedure and field data. *Remote Sens. Environ.* 80, 88–99. doi: 10.1016/s0034-4257(01)00290-5

Naesset, E. (2009). Effects of different sensors, flying altitudes, and pulse repetition frequencies on forest canopy metrics and biophysical stand properties derived from small-footprint airborne laser data. *Remote Sens. Environ.* 113, 148–159. doi: 10.1016/j.rse.2008.09.001

Naesset, E., and Bjerknes, K. O. (2001). Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sens. Environ.* 78, 328–340. doi: 10.1016/S0034-4257(01)00228-0

Naesset, E., and Gobakken, T. (2008). Estimation of above- and below-ground biomass across regions of the boreal forest zone using airborne laser. *Remote Sens. Environ.* 112, 3079–3090. doi: 10.1016/j.rse.2008.03.004

Nelson, R., Short, A., and Valenti, M. (2004). Measuring biomass and carbon in Delaware using an airborne profiling LIDAR. *Scand. J. For. Res.* 19, 500–511. doi: 10.1080/02827580410019508

Ni, X. L., Park, T., Choi, S. H., Shi, Y. L., Cao, C. X., Wang, X. J., et al. (2014). Allometric scaling and resource limitations model of tree heights: Part 3. Model optimization and testing over continental china. *Remote Sens.* 6, 3533–3553. doi: 10.3390/rs6053533

Nilsson, M. (1996). Estimation of tree heights and stand volume using an airborne lidar system. *Remote Sens. Environ.* 56, 1–7. doi: 10.1016/0034-4257(95)00224-3

Picard, N., Rutishauser, E., Ploton, P., Ngomanda, A., and Henry, M. (2015). Should tree biomass allometry be restricted to power models? *For. Ecol. Manage.* 353, 156–163. doi: 10.1016/j.foreco.2015.05.035

Silva, C. A., Hudak, A. T., Vierling, L. A., Klauberg, C., Garcia, M., Ferraz, A., et al. (2017). Impacts of airborne lidar pulse density on estimating biomass stocks and changes in a selectively logged tropical forest. *Remote Sens.* 9:1068. doi: 10.3390/Rs9101068

Swetnam, T. L., Gillan, J. K., Sankey, T. T., Mcclaran, M. P., Nichols, M. H., Heilman, P., et al. (2017). Considerations for achieving cross-platform point cloud data fusion across different dryland ecosystem structural states. *Front. Plant Sci.* 8:2144. doi: 10.3389/fpls.2017.02144

Wang, J., Che, K., Fu, H., Chang, X., Song, C., and Ke, H. (1998). Study on biomass of water conservation forest in qilian mountains. *J. Fujian Coll. For.* 18:5. doi: 10.13324/j.cnki.jfcf.1998.04.009

Wang, J., Che, K., and Jang, Z. (2000). Study on carbon balance of qinghai spruce forest in qilian mountains. *J. Northwest For. Univ.* 15, 9–14. doi: 10.3969/j.issn.1001-7461.2000.01.002

Zhao, K. G., Popescu, S., Meng, X. L., Pang, Y., and Agca, M. (2011). Characterizing forest canopy structure with lidar composite metrics and machine learning. *Remote Sens. Environ.* 115, 1978–1996. doi: 10.1016/j.rse.2011.04.001

Zhao, K. G., Popescu, S., and Nelson, R. (2009). Lidar remote sensing of forest biomass: a scale-invariant estimation approach using airborne lasers. *Remote Sens. Environ.* 113, 182–196. doi: 10.1016/j.rse.2008.09.009

Zhao, P., Lu, D., Wang, G., Wu, C., Huang, Y., and Yu, S. (2016). Examining spectral reflectance saturation in landsat imagery and corresponding solutions to improve forest aboveground biomass estimation. *Remote Sens.* 8:469. doi: 10.3390/rs8060469

Zhou, T., Popescu, S., Lawing, A., Eriksson, M., Strimbu, B., and Bürkner, P. (2017). Bayesian and classical machine learning methods: a comparison for tree species classification with LiDAR waveform signatures. *Remote Sens.* 10:39. doi: 10.3390/rs10010039

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 28 August 2019
doi: 10.3389/fpls.2019.01050

MaizeDIG: Maize Database of Images and Genomes

Kyoung Tak Cho[1], **John L. Portwood, II**[2], **Jack M. Gardiner**[3], **Lisa C. Harper**[2], **Carolyn J. Lawrence-Dill**[4,5,6], **Iddo Friedberg**[6,7] and **Carson M. Andorf**[2]*

[1] Department of Computer Science, Iowa State University, Ames, IA, United States, [2] USDA-ARS Corn Insects and Crop Genetics Research Unit, Iowa State University, Ames, IA, United States, [3] Division of Animal Sciences, University of Missouri, Columbia, MO, United States, [4] Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, United States, [5] Department of Agronomy, Iowa State University, Ames, IA, United States, [6] Bioinformatics and Computational Biology Graduate Program, Iowa State University, Ames, IA, United States, [7] Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, IA, United States

Edited by:
Jose Antonio Jimenez-Berni, Instituto de Agricultura Sostenible (CSIC), Spain

Reviewed by:
Zhen Su, China Agricultural University (CAU), China
Katsumi Sakata, Maebashi Institute of Technology, Japan
Daniel Arend, Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), Germany

***Correspondence:**
Carson M. Andorf
carson.andorf@ars.usda.gov

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *15 April 2019*
Accepted: *29 July 2019*
Published: *28 August 2019*

Citation:
Cho KT, Portwood JL, II, Gardiner JM, Harper LC, Lawrence-Dill CJ, Friedberg I and Andorf CM (2019) MaizeDIG: Maize Database of Images and Genomes. Front. Plant Sci. 10:1050. doi: 10.3389/fpls.2019.01050

Background: An organism can be described by its observable features (phenotypes) and the genes and genomic information (genotypes) that cause these phenotypes. For many decades, researchers have tried to find relationships between genotypes and phenotypes, and great strides have been made. However, improved methods and tools for discovering and visualizing these phenotypic relationships are still needed. The maize genetics and genomics database (MaizeGDB, www.maizegdb.org) provides an array of useful resources for diverse data types including thousands of images related to mutant phenotypes in *Zea mays* ssp. *mays* (maize). To integrate mutant phenotype images with genomics information, we implemented and enhanced the web-based software package BioDIG (Biological Database of Images and Genomes).

Findings: We developed a genotype-phenotype database for maize called MaizeDIG. MaizeDIG has several enhancements over the original BioDIG package. MaizeDIG, which supports multiple reference genome assemblies, is seamlessly integrated with genome browsers to accommodate custom tracks showing tagged mutant phenotypes images in their genomic context and allows for custom tagging of images to highlight the phenotype. This is accomplished through an updated interface allowing users to create image-to-gene links and is accessible *via* the image search tool.

Conclusions: We have created a user-friendly and extensible web-based resource called MaizeDIG. MaizeDIG is preloaded with 2,396 images that are available on genome browsers for 10 different maize reference genomes. Approximately 90 images of classically defined maize genes have been manually annotated. MaizeDIG is available at http://maizedig.maizegdb.org/. The code is free and open source and can be found at https://github.com/Maize-Genetics-and-Genomics-Database/maizedig.

Keywords: MaizeDIG, BioDIG, phenotype, genes, QTL, gene model, MaizeGDB

INTRODUCTION

One of the fundamental relationships in biology is how observable physical and biochemical characteristics (phenotypes) relate to the underlying combination of alleles of the genes within an organism (genotype). Maize is a good case study for exploring this relationship because maize is both agronomically important as the top production grain crop (http://faostat.fao.org/) and serves as a model species. Maize has numerous reference quality genome assemblies available for exploring the genotype to phenotype relationship (Schnable et al., 2009; Lu et al., 2015; Hirsch et al., 2016; Jiao et al., 2017; Sun et al., 2018; Springer and Anderson, 2018), and a great deal of research has been carried out in maize that links mutant phenotypes to genes or regions within the genome, for example, quantitative trait locus (QTL) (Schnable and Freeling, 2011). Other linkages from genetics and genomics data to phenotype include the identification of numerous QTLs and associations of single-nucleotide polymorphisms with phenotypes and traits through genome-wide association studies (Xiao et al., 2017). Many loci and traits are further annotated with accessible high-resolution images. MaizeGDB has more than 2,700 images mapped to genes and more than 1,000 mapped to a gene model with specific genomic coordinates available. In addition, there are an additional 200 images where the specific phenotype feature has been tagged within the image. Most of these images were donated by Dr. Gerald Neuffer, University of Missouri-Columbia. His research created thousands of mutations by treating maize pollen with the known mutagen ethyl methanesulfonate. Many of these were imaged and shared in the book *The Mutants of Maize* (Neuffer et al., 1968) and/or are available at the "Guide to Maize Mutant Phenotypes" website hosted at MaizeGDB (http://mutants.maizegdb.org/). This collection of digital images is under active curation.

Although many of these images are linked to a gene locus, prior to the efforts described here, there was no method to visualize these images within their genomic context *via* the MaizeGDB Genome Browser. This current situation hinders researchers working to prioritize candidate genes underlying traits of interest within a chromosomal region from being able to narrow their search for the causal gene in a quick and easy way.

The first maize genome sequence assembly was completed in 2009 (Schnable et al., 2009). The most recent version of the reference assembly, B73 RefGen_v4, is based on PacBio sequencing and high-resolution optical mapping with an N50 value of 1.18 Mb (Jiao et al., 2017). It also has nearly 40,000 structurally annotated protein-coding genes. In addition, the rapidly decreasing cost of genome sequencing and assembly has led to the availability of a number of new high-quality maize genome assemblies from different maize lines. This suggests that use of a single reference genome may no longer continue to best serve the needs of the maize research community. In addition to B73, reference quality maize genome assemblies and annotations released in the past few years include CML247 (Lu et al., 2015), PH207 (Hirsch et al., 2016), Mo17 (Yang et al., 2017; Sun et al., 2018), European Flints EP1 and F7 (Unterseer et al., 2017), and W22 (Springer et al., 2018). In the near future, we anticipate the release of genome assemblies for dozens, if not hundreds, of *Zea* lines. The high level of phenotypic and genomic diversity within maize—any two given maize lines can be as different as humans and chimpanzees (Buckler et al., 2006)—means these additional genome assemblies will have great value.

Currently, there is a pressing need to better integrate phenomic data within the context of multiple genome assemblies. The current methodology of associating images to genomic data at MaizeGDB is relatively complex: Each gene can have multiple alleles (denoted as variations at MaizeGDB); each image is linked to its appropriate allele; a gene can be linked to a gene model with genomic coordinates. A tool is needed to make simplified and more direct connections between the phenotype and genotype, which in turn will allow researchers to explore a genome or gene region simultaneously in both a genomic and phenomic context. There currently exist tools and resources to help manage, curate, and analyze images such as BisQue at CyVerse (Kvilekval et al., 2010) and BioDIG at GMOD (Oberlin et al., 2013). BisQue is a cloud-based platform that provides support to organize images, integrate metadata, and build complex analyses with the images/metadata. BioDIG is a stand-alone software to integrate images with genomic data. To address the need to curate phenotype images, link to genomic data, and visualize the relationships within the context of a model organism database, we chose to implement an enhanced version of BioDIG.

In this article, we present an open-source, web-based software called MaizeDIG (SciCrunch.org tool reference ID: SCR_016987), a multiple maize reference genome implementation of BioDIG, with enhancements that enable tagging and linkages between images and genes/gene models or QTLs, support for multiple reference assemblies, creation of custom autogenerated genome browser tracks, and expanded search capabilities.

MaizeGDB

The Maize Genetics and Genomics Database (MaizeGDB; http://www.maizegdb.org) is the model organism database for maize (Portwood et al., 2019). MaizeGDB's overall aim is to provide a single point of access to maize research data and tools for integration, visualization, and discovery. Stored at MaizeGDB is comprehensive information on genes, alleles, mutants, stocks, molecular markers, gene product information, phenotypic images and descriptions, metabolic pathway information, pedigree information, reference data, sequences including multiple genome assemblies and annotation sets, contact information for maize researchers, and more. Curation of high-quality and high-impact datasets has been the foundation of the MaizeGDB project since its inception over 25 years ago (Schaeffer et al., 2011; Andorf et al., 2016; Harper et al., 2016). Since the first maize genome sequence assembly was announced in 2009 (Schnable et al., 2009), MaizeGDB has provided access to genome assemblies and annotations through GBrowse-based genome browsers and other tools (Sen et al., 2009). Currently, MaizeGDB supports over 10 independent maize genome sequence assemblies of different maize lines, each with their own annotation set(s). A comparison of these genome assemblies reveals a high level of sequence diversity between maize lines, allowing maize researchers to investigate what genotypic combination of alleles leads to the great variety of phenotypes in maize, especially agriculturally important phenotypes such as grain yield and planting density. Genotype-to-phenotype hypotheses can be facilitated by genome visualization tools. One way to integrate high-quality phenotypic data with a reference genome assembly is to link images of phenotypes to genes that have known physical locations on the genome assembly. MaizeGDB currently hosts 4,367 unique mutant phenotype images, of which 2,396 are linked to a locus on the genome. Detailed descriptions of specific phenotypic features within an image can provide additional clarity. To address these needs, MaizeGDB has adapted the BioDIG system to create a maize-specific database of images and genomes.

BioDIG

BioDIG is a web-based biological database for linking images to genomic data (Oberlin et al., 2013). To link an image to a gene, an image tagging tool called MaizeDIG Workbench is provided. In the workbench, a user can outline and highlight regions within the image, called a "tag" and create links from the tag to genomic coordinates (gene links). BioDIG is designed to work generally for all types of images, but the tagging feature is particularly useful for phenotypes in complex images or where the exact phenotype may be difficult to identify. To handle genomic data, the Generic Model Organism Database (GMOD) Chado database schema is used because it is compatible with standard data formats (e.g., Generic Feature Format or GFF files), and it is easily extensible to add features to genes. The front-end (client-side) user interface is based on the Django framework (https://www.djangoproject.com/) and jQuery.

BioDIG was designed for general organisms, and it needs genome and image data of the target organism. BioDIG provides a tool to build and update genome data with a GFF. However, it requires extra database installations and configurations. We extended BioDIG to use genome and image data from MaizeGDB directly, and it can reduce potential errors or overheads related to the handling of genome and image data because MaizeGDB has well-curated genome data and images for maize. In addition, BioDIG's functionality was extended by adding novel features such as multiple genome support, enhanced searching, and images shown in dynamic genome browser tracks.

Connecting Genes to Phenotype Through Gene Links

A large amount of both genomics and phenomics data is available for maize with many methodologies, applications, and tools being utilized. Even with these valuable tools in place, it is still a challenge to visualize and discover relationships between phenotype and genotype. In MaizeDIG, we have defined the relationship between a gene and its phenotype as a gene link. Gene links can be used to give a quick visual representation of the phenotype in any context in which a gene is presented at MaizeGDB. The gene link comprised an image tag and a gene. A single image can have multiple phenotypic features, and each of those features may be represented by a single or multiple genes. Likewise, a mutant allele of a gene can display multiple phenotypes shown in different images. The exact phenotypic features within an image can be annotated with the tagging tool in which a curator can highlight the section of the image by drawing either a rectangular region or freestyle region, and these tagged regions can be linked to a gene(s). These tagged regions along with their gene links represent a map of the phenotype-genotype relationship that can be used in the framework of the MaizeGDB database to address a variety of situations that require an understanding of a gene and the corresponding phenotypes of its alleles. MaizeGDB gene pages also provide functional annotations that are linked to the image. These annotations include Plant Ontology, Gene Ontology, and other ontologies terms, their expression, and descriptions of the mutant phenotypes.

MaizeDIG allows a novel way to visualize these relationships as a track on the MaizeGDB Genome Browser. The MaizeGDB Genome Browser lets a user browse the genomic DNA sequence at varying levels of complexity to see aligned biological features within the context of that genomic region. The MaizeDIG track allows a user to browse a chromosomal region and see images of tagged mutant phenotypes for genes of interest. Some applications of using this track include determining gene function, chromosome walking to genes, identifying enriched biological processes or metabolic pathways, and identifying regulatory regions adjacent to genes.

MATERIALS AND METHODS

MaizeDIG is a genotype-phenotype linking database based on BioDIG that has been enhanced to show phenotype images on a genome browser and handle multiple maize genome assemblies. MaizeGDB has over 2,700 mutant phenotype images, and many of these digital images are linked to a gene model. MaizeDIG has a set of web-based tools that allows searching, tagging, annotating, and linking images with genes, gene models, and alleles. In addition, MaizeDIG has been integrated with several MaizeGDB Genome Browsers simultaneously. Once an image is tagged to a gene, it becomes available as a custom track for any maize genome assembly that has a gene model associated to that genomic region. There are four functional aspect of MaizeDIG: data handling, image curation, image search, and genome browser integration. Each of these aspects has been enhanced relative to its original implementation of BioDIG. **Figure 1** shows the four main modules, and we discuss details of these in the following subsections.

Data Process

The raw data used to build the MaizeDIG database are images and genomic mapped loci (e.g., genes). The general data process for MaizeDIG is to load an image, tag the location that best identifies a phenotype, link the tag to a gene or QTL, and provide additional metadata. To make a link between gene and image, both image and genomic data must exist that can be linked together.

Image Data

All phenotype images at MaizeGDB have been provided by individual researchers. To date, the photographs in MaizeDIG have been different than ones in publication to keep from copyright law

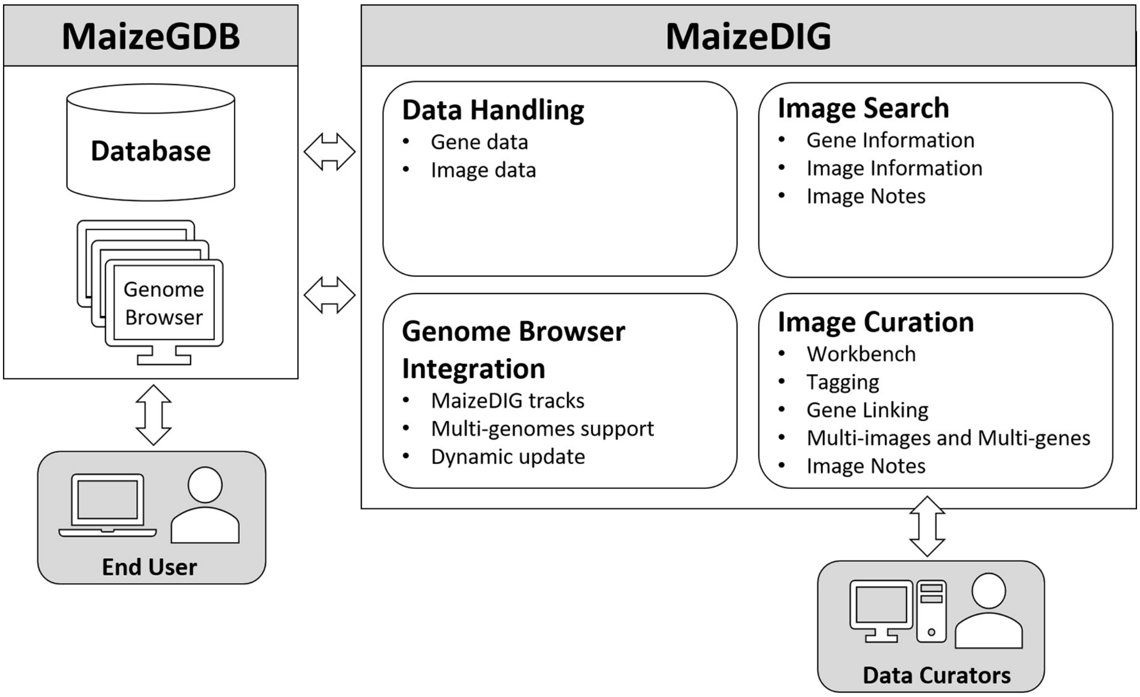


FIGURE 1 | Structure of MaizeDIG. MaizeDIG's structure is divided into four major components: data handling, image curation, image search, and genome browser integration.

violation. Given more recent Open Access licensing, we anticipate the use of images directly from publications with the appropriate Creative Commons license. MaizeGDB has 2,396 images that are each linked to the gene representing multiple alleles that display the imaged phenotype. Most images are of phenotypes caused by mutations in single genes. Images are loaded through the MaizeGDB curation tools, but in the absence of a model organism database, BioDIG can be used as the primary way of uploading images.

Genomic Data

MaizeGDB acts as the steward of maize genome assemblies. Currently, MaizeGDB hosts genome assemblies for over 10 different maize inbred lines. In the absence of software that can display multiple genomes at the same time, MaizeGDB provides a genome browser for each genome assembly. Each assembly has its own set(s) of structural and functional gene annotations, that is, gene model sets. Classically defined genes are associated to these gene model sets. Information linked to a gene is therefore inherited by the associated gene models sets.

The next sections describe the individual components that make up the process of building a maize genotypic-phenotypic database and pipeline to integrate gene, gene model, and phenotype image data from the MaizeGDB database.

Gene Data

MaizeDIG is an extension of Chado (Mungall and Emmert, 2007), one of the most popular relational database schema for biological research. **Figure 2** shows the table schemas related to genotypic and phenotypic data. The user table deals with all user information including administration accounts. The feature table has the genome data needed from maize. The genelink table deals with linking information between the tag and feature. A full description of the image, genome, and user management modules and details on the underlying BioDIG database are described at GMOD and the original BioDIG article (Oberlin et al., 2013). Although MaizeDIG provides a system to manually link genes to an image, the majority of these associates are preloaded based on information at MaizeGDB. In the MaizeGDB database, images and phenotype descriptors are (independently) linked to alleles, alleles are linked to genes, genes are linked to gene models, and gene models are linked to genomic coordinates. Scripts have been developed to load these associations into the MaizeDIG database. The details of linking between image and gene are discussed in Image Curation section.

Image Data

To build an initial phenotype image database, 2,721 images with gene information were imported from MaizeGDB to MaizeDIG. These images are based on interesting maize mutations that were photographed, and extensive efforts were taken to map to specific genomic locations using whatever chromosome mapping technologies were available at the time. Indeed, this collection of new maize mutations launched many careers in maize genetics which contributed to diverse areas of plant science such as developmental biology, chromosome mechanics, hormone metabolism, photosynthesis, and more.

The images are mapped to gene(s) in MaizeGDB; hence, each image imported from MaizeGDB has an associated gene called primary gene. An image can have only one primary gene in MaizeDIG. The primary gene does not automatically become a gene link directly, but it functions as the recommended gene. In general, curators use the recommended gene to make a gene link, but it is also possible to choose a different gene(s) as well. There are examples where the same image may exist multiple times in the database with different genes assigned to it. MaizeDIG handles these redundancies by creating image groups, which internally recognizes redundancies and consolidates any gene or gene link data. In addition to gene data, if a caption is available for an image, it is loaded as a description in MaizeDIG.

Image Curation

MaizeDIG provides an image curation tool suite enabling tagging of phenotypic features, image-gene linking, visualization of image-gene data, and image searching. In this section, we provide details on each of these features.

Workbench

Image curation is done within the MaizeDIG Workbench (see **Figure 3**). The workbench is accessible only when a user is logged into his/her account. This allows the curation process to be saved and associated to a user/curator. The workbench is divided into three subsections: toolbar menu (left), image viewing/editing section (center), and detail information section (right). The recently viewed images menu on the left-side toolbar menu is useful for keeping track of the 10 most recently viewed images. The image viewing/editing section has a menu bar on the top. The menu bar includes tags, tag groups, gene linking, and user note features. Additional information such as Image Description, Gene ID, Gene Symbol, Gene Name, Image ID, and User Notes is shown on the right side of the image. The gene ID/symbol/name are associated to a gene, but they do not function as a gene link until activating the gene link using the "Gene Link- > Add New Link to Tag" tool. The image viewing/editing box can show all tags belonging to the image, and even though each tag has an ownership with an associated "user_id" (**Figure 2**), all tags are displayed on the image.

Tagging

An image can represent a phenotype by the image itself without additional modification, but in many cases, the phenotypic image will benefit from additional visual annotation to provide clarity. The tagging process specifies a certain region within an image and labeling it with a tag ID. **Figure 3** shows a custom tagging on image. The tagging tool provides eight colors and two drawing styles (rectangular and free pen). Multiple tags can be created on a single image.

Gene Linking

Once an image has been tagged, a gene link between the tag and a gene can be created by adding a gene to the tag. The "Add New Link to Tag" tool provides a way to complete this task, which is called creating a "genelink." A genelink is required to have a one-to-one relationship between a tag and a gene. A gene symbol or gene model can be assigned to a tag, and

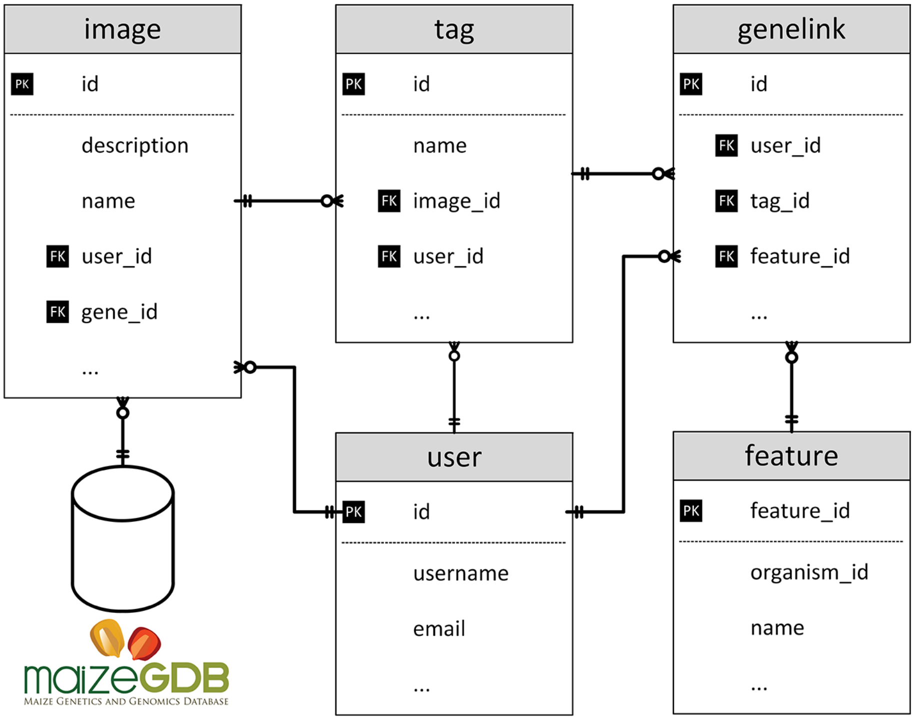
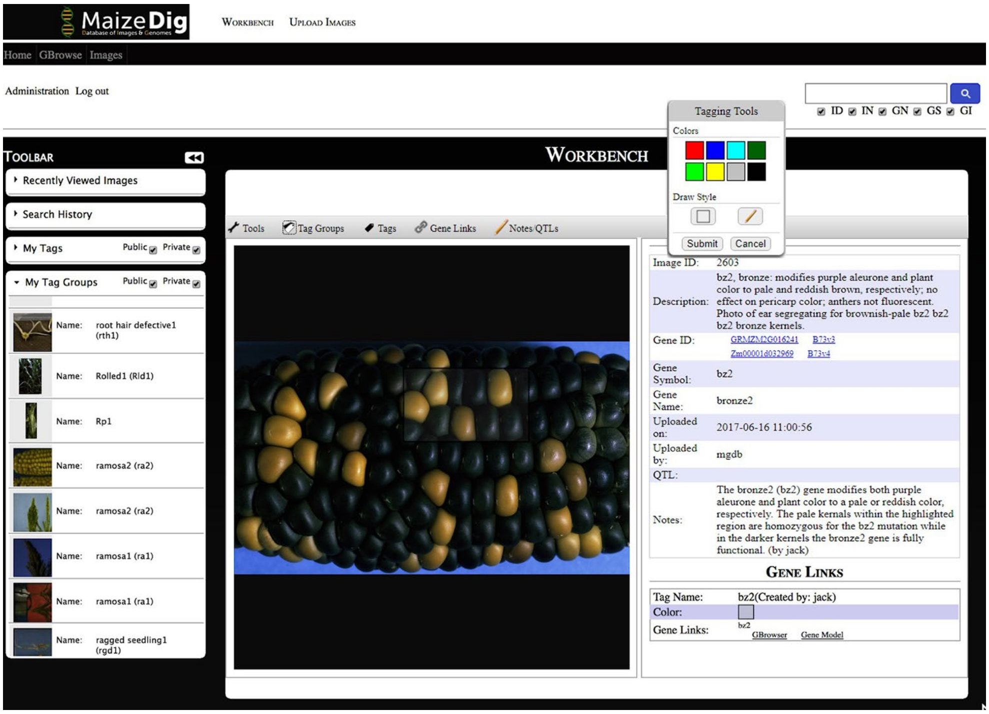


FIGURE 2 | Database structure of MaizeDIG. This image shows the database schema for the five major database tables used in MaizeDIG.



Name: root hair defective1 (rth1)
Name: Rolled1 (Rld1)
Name: Rp1
Name: ramosa2 (ra2)
Name: ramosa2 (ra2)
Name: ramosa1 (ra1)
Name: ramosa1 (ra1)
Name: ragged seedling1 (rgd1)

Image ID:	2603
Description:	bz2, bronze: modifies purple aleurone and plant color to pale and reddish brown, respectively; no effect on pericarp color; anthers not fluorescent. Photo of ear segregating for brownish-pale bz2 bz2 bz2 bronze kernels.
Gene ID:	GRMZM2G016241 B73v3 Zm00001d032969 B73v4
Gene Symbol:	bz2
Gene Name:	bronze2
Uploaded on:	2017-06-16 11:00:56
Uploaded by:	mgdb
QTL:	
Notes:	The bronze2 (bz2) gene modifies both purple aleurone and plant color to a pale or reddish color, respectively. The pale kernals within the highlighted region are homozygous for the bz2 mutation while in the darker kernels the bronze2 gene is fully functional. (by jack)

GENE LINKS

Tag Name:	bz2(Created by: jack)
Color:	▨
Gene Links:	bz2 GBrowser Gene Model

FIGURE 3 | Maize curators workbench. A screenshot of the MaizeDIG Curators Workbench is shown with and image displaying the bronze2 (bz2) kernel phenotype segregating on a maize ear. Tools for image tagging, curation, and linking to the MaizeGDB Genome Browser can be accessed on the toolbar directly above the image. Mousing over each of the five tools activates a pull-down menu with a variety of options. The "Tagging Tools" pop-up that allows a user to add a tag to an existing tag group is shown. The collapsible vertical tool bar to the left of the image allows MaizeGDB curators to review their search history and recently viewed images as well their tags and tag groups. In the upper right-hand corner is a search box that allows curators to search over 2,300 images by Image Description (ID), Image Notes (IN), Gene Name (GN), Gene Symbol (GS), and Gene ID (GI). To the right of the image is the preloaded image description, Gene ID, Symbol, and Name as well as the Image ID. The phenotypic description as it will appear on the genome browser pop-up (minus the curator attribution) appears in the Notes section. The Gene Links information appears at the bottom once a successful gene link has been made between an image tag on the workbench and the genome browser.

typically, a single image is linked to a single gene. It is also possible that a single image can show multiple phenotypes or have phenotypes associated to multiple genes and therefore need to be mapped to the multiple genes. The gene link allows connecting to MaizeGDB database such as Genome Browser and Gene model page.

Linking to QTL Data

Most phenotype images at MaizeGDB are for monogenic phenotypes. MaizeGDB does contain more complex phenotypes and traits that have been associated to multiple loci across the genotype. These QTLs are also supported in MaizeDIG. Within the MaizeDIG Workbench, a QTL ID can be added to any image. Similar to a gene link, the QTL ID connects the image to QTL information in the MaizeGDB database.

Many-to-Many Relationships Between Images and Genes

In the phenotypic data handling section, we discussed the relationship of images and genes within the context of two different scenarios: multiple images are associated to a single gene, and multiple genes are linked to a single image. MaizeDIG has addresses these two scenarios with the "Image Group" tool. In the "multi-image to single-gene" case, it is straightforward for the gene to be assigned to multiple images. However, additional treatment is needed for the "multi-genes to single-image" situation. In this case, the image data are duplicated with different "Image ID" for each primary gene, but is still considered a single image within the "Image Group" tool. This ensures that all tags and gene links in the same group appear in each image.

Image Notes

Image notes provide another mechanism for communicating additional information in the image curation process. It is important to be able to share opinions/comments among curators. Curators can create/edit image notes regardless of its ownership. In addition, images can be queried by the image notes using the image search tool.

ENHANCEMENTS

Image Search

MaizeDIG offers an image search tool that has been enhanced relative to BioDIG as follows. A curated image has detailed information such as image description, gene model IDs, gene symbol, gene name, and image notes. Any of these categories can be searched individually or together using the image search box. With thousands of images loaded into MaizeDIG, the flexible search tool is needed to locate the correct image to curate. In addition, an image search history has been implemented in MaizeDIG that saves 10 most recent searched images and the search settings in an expandable box on the left side of the workbench.

Downloading Images and Metadata

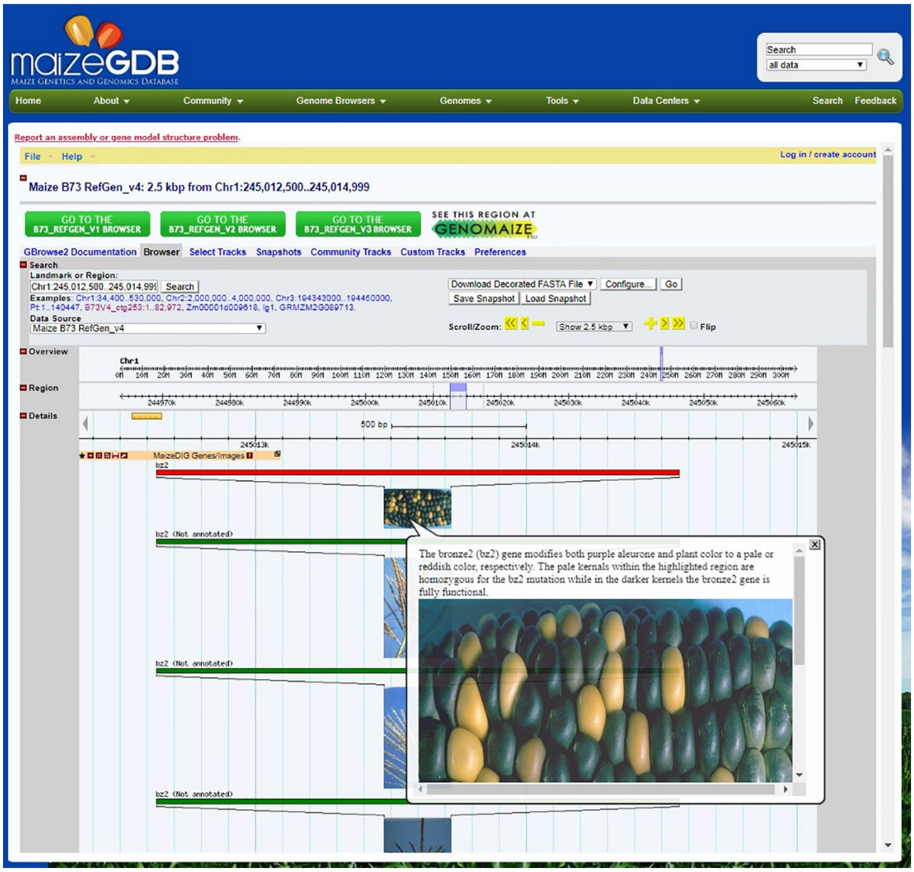
Images and associated metadata can be downloaded from either the workbench or an image record page. The "Tools|Download Image Data" menu item will open up a simple download form where a user can select the image attributes and data format. The attributes include URL of image, upload data/user, image tags, image file, tag groups, and gene links. The image tag data will include the coordinates and color options of the highlighted regions. The options for the output format are JSON and XML. After selecting the attributes and format, a zip file will be downloaded that includes the image and attribute file. Genome Browser integration MaizeGDB's GBrowse (Donlin, 2009) instance provides hundreds of tracks across multiple genome assemblies, but up until now did not include any phenotypic data. The genome browser is one of MaizeGDB's most accessed resources, receiving nearly 80,000 unique page views in 2018 according to usage statistics from Google Analytics, and so adding genotype-phenotype relations to this resource raises the accessibility and usability of these data. Therefore, we have created dynamically loaded tracks that contain all of the images, tags, gene links, and user notes in MaizeDIG. The tracks are available for all of our genome browsers that have gene model associations with the reference genome (B73) which as of this publication includes the following: B73v3, B73v4, W22v2, PH207, Mo17 (x2), B104, EP1, F7, and Zx-PI566673. As new genomes are incorporated into MaizeGDB, new MaizeDIG tracks will also be created for their respective genome browsers. All MaizeDIG tracks are dynamically updated when a user creates (or removes) a gene link to a tag. Upon creating a gene link, a new feature is inserted into the track at the location of that specified gene, and vice versa if the gene link is deleted. The feature shows a thumbnail of the image, and mousing over it reveals the full size of the image along with all of its tags, user notes, and other gene links. **Figure 4** shows an example of this track on the B73v4 genome browser.

CURATION CASE STUDY

After the initial bulk upload of over 2,300 images and metadata, it was both possible and advisable to manually annotate select images that present a clear representation of the relevant phenotype associated with a gene mutation. To accomplish this, MaizeGDB curators added manual annotations for every classically defined gene (Schnable and Freeling, 2011) in the database that were associated with a mutant phenotype image. In total, approximately 90 genes were curated, which enables the end user to unambiguously identify the exact phenotype in both manually curated and noncurated images. The manual curation approach taken by MaizeGDB curators varies somewhat depending on the particular phenotype and its presentation within the image. In ideal situations, the curator is able to deploy a "compare and contrast" approach if the mutant and nonmutant phenotype is present within the same image. However, this is not always possible, and a different curation strategy must be taken in these situations. In **Figure 3**, a screenshot taken from the curators MaizeDIG workbench displays an image of a maize ear, with a majority of kernels having a dark purple color, and interspersed between them are a few bronze-colored kernels. The *bronze2* gene (*bz2*) when fully functional causes the kernels to acquire a dark purple color while the bronze colored kernels represent the *bz2* gene in a nonfunctional or mutated state.

The manual curation process is initiated by the creation of a tag group from the "Tag Groups" pull-down menu directly above the image. From this menu, it is possible to create multiple tag groups and delete or modify them. In practice, it is rare to have multiple tag groups attached to an image, although certainly there are situations where this proves to be a useful feature. Once a tag group has been created and named, a tag can be created from the "Tags" pull down in the menu bar. Once the tag has been activated, a "Tagging Tools" pop-up appears (**Figure 3**) with a choice of both drawing style (freehand or box) and eight tag fill color options. After selecting both the drawing style and color fill option, the curator moves directly to the image and highlights the area of interest using the mouse. In this example, kernels with (bronze) and without (dark purple) the bz2 mutation are selected to be within the tagged area. Once the tagged area has been selected, and the submit button on the "Tagging Tools" pop-up is chosen, an additional "Submit Tag" pop-up appears (not shown in **Figure 3**), and a "Tag Group" can be chosen and a name attached to the new tag. Once an image tag has been created, it is necessary to create a "Gene Link" to make a link or connection between the gene/image on the curation workbench and the gene's location on the MaizeGDB Genome Browser, which is what the end user will see and interact with (**Figure 4**). This is accomplished by accessing the "Gene Links" pull-down menu directly above the image. If there are multiple tag options, a tag is selected, and the relevant gene symbol (*bz2* in this example) is entered in the "Gene or Locus" field.

The manual curation process is completed by the addition of a phenotypic description in the pop-up from the "Notes" section above the image (**Figure 3**) in the curation workbench. When completed, the phenotypic description appears both on the right-hand side of the image (**Figure 3**) in the workbench and in the image pop-up on the MaizeGDB genome browser (**Figure 4**). It is worth noting that the "Description" is seen only on the workbench view and not in the genome browser pop-up that is visible to the end user. Information in the "Description" section is pulled from MaizeGDB records and often contains information that is not relevant to the phenotype being presented. In general, MaizeGDB curators strive to keep the phenotypic description simple and avoid excessive use technical genetic jargon, while still maintaining a robust scientific description. At the outset of the MaizeDIG project, our goal was to provide simple explanations for the phenotype when presented in the context of the genome browser. For those interested, more detailed explanations can be found in the gene model records at MaizeGDB. MaizeGDB



Maize B73 RefGen_v4: 2.5 kbp from Chr1:245,012,500..245,014,999

Landmark or Region: Chr1:245,012,500..245,014,999

Examples: Chr1:34,400..530,000, Chr2:2,000,000..4,000,000, Chr3:194343000..194450000, Pt:1..140447, B73V4_ctg253:1..82,972, Zm00001d009618, lg1, GRMZM2G089713.

Data Source: Maize B73 RefGen_v4

The bronze2 (bz2) gene modifies both purple aleurone and plant color to a pale or reddish color, respectively. The pale kernals within the highlighted region are homozygous for the bz2 mutation while in the darker kernels the bronze2 gene is fully functional.

FIGURE 4 | MaizeDIG on the MaizeGDB Genome Browser. A screenshot of the MaizeDIG curated bronze2 (bz2) image and how it presents in the context of the MaizeGDB Genome Browser is shown. Mousing over any of the MaizeDIG images activates an image pop-up, while clicking on the image takes you to a webpage with a full screen image and phenotypic description. The majority of images in MaizeDIG are high-resolution and present well as a full screen image. Manually curated images have a red banner and are automatically sorted to the top of the image stock, while unannotated images have a green banner and are shown below. Note the multiple images shown for bz2. Some phenotypes may only have one image, while other phenotypes may have ten or more.

curates maize research papers and will use MaizeDIG to curate phenotype images found in these manuscripts.

CONCLUSION

MaizeDIG facilitates a major priority at MaizeGDB which is to make accessible high-quality data and create added value through curation, integration, or through providing tools that visualize and/or analyze the data. This tool enables the integration of phenotype and genotype data in a model species rich in both data types. This is done directly by making dynamic tracks on multiple MaizeGDB Genome Browsers, the most widely used tools at MaizeGDB, for genomic exploration. Having the phenotype images in their genomic context along with other mapped features (e.g., gene models, expression data, polymorphisms, etc.) enables users to prioritize gene candidates for phenotypes of interest. This data will be valuable in forward genetics by providing quick visual clues for prioritizing genes. For example, if a region of the genome has been associated with the phenotype "photosynthetic capacity," genes could be quickly identified by looking at mutant phenotypes related to leaves. In addition, these images can be used in a variety of settings to prioritize candidate genes identified through coexpression, differential expression, sequence similarity, domain sharing, functional enrichment, and subcellular localization. This represents a major step forward as experimental determination and/or validation are both time-consuming and expensive. MaizeDIG will be a valuable addition to the MaizeGDB manual annotation workflow.

These use cases are not unique to maize, and MaizeDIG's availability as a free and open-source resource will allow the application of integrating phenomic images and genomic data to other databases and projects.

DATA AVAILABILITY

The datasets analyzed for this study can be found in MaizeGDB at http://maizedig.maizegdb.org. The source code is free and open source and can be found at (https://github.com/Maize-Genetics-and-Genomics-Database/maizedig).

AUTHOR CONTRIBUTIONS

All authors were involved in conceptualization, writing, reviewing, and editing of the manuscript. CL-D, IF, and CA were responsible for Funding Acquisition. CA provided project administration. KC and JP developed the software and data visualization tools. JG and LH provided data curation.

FUNDING

This work was supported by the US Department of Agriculture-Agricultural Research Service (project no. 5030-21000-068-00-D) and an award from the Iowa State University Presidential Interdisciplinary Research Initiative to support the D3AI (Data-Driven Discovery for Agricultural Innovation) project (to CL-D, IF, and CA). For more information, see http://www.d3ai.iastate.edu/. IF was funded, in part, by NSF award ABI 1458359. Support to develop the “Guide to Maize Mutant Phenotypes” dataset was derived from the NSF’s ABI funding (DBI 0743804) to CL-D and Gerry Neuffer. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENT

We would like to acknowledge the MaizeGDB team for testing and providing feedback for MaizeDIG, Dr. Gerald Neuffer for providing his collection of mutant phenotype images to MaizeGDB, and Andrew Oberlin and personnel working in the Friedberg laboratory for support for BioDIG installation.

REFERENCES

Andorf, C. M., Cannon, E. K., Portwood, J. L., 2nd, Gardiner, J. M., Harper, L. C., Schaeffer, M. L., et al. (2016). MaizeGDB update: new tools, data and interface for the maize model organism database. *Nucleic Acids Res.* 44, D1195–D1201. doi: 10.1093/nar/gkv1007

Buckler, E. S., Gaut, B. S., and McMullen, M. D. (2006). Molecular and functional diversity of maize. *Curr. Opin. Plant Biol.* 9, 172–176. doi: 10.1016/j.pbi.2006.01.013

Donlin, M. J. (2009). Using the generic genome browser (GBrowse). *Curr. Protocols Bioinform.* 28(1), 9.9.1–9.9.25. doi: 10.1002/0471250953.bi0909s28

Harper, L., Gardiner, J., Andorf, C., and Lawrence, C. J. (2016). MaizeGDB: the maize genetics and genomics database. *Methods Mole. Biol.* 1374, 187–202. doi: 10.1007/978-1-4939-3167-5_9

Hirsch, C. N., Hirsch, C. D., Brohammer, A. B., Bowman, M. J., Soifer, I., Barad, O., et al. (2016). Draft assembly of elite inbred line PH207 provides insights into genomic and transcriptome diversity in maize. *Plant Cell* 28 (11), 2700–2714. doi: 10.1105/tpc.16.00353

Jiao, Y., Peluso, P., Shi, J., Liang, T., Stitzer, M. C., Wang, B., et al. (2017) Improved maize reference genome with single-molecule technologies. *Nature* 546, 524-527. doi: 10.1038/nature22971

Kvilekval, K., Fedorov, D., Obara, B., Singh, A., and Manjunath, B. S. (2010). Bisque: a platform for bioimage analysis and management. *Bioinformatics* 26, 544–552. doi: 10.1093/bioinformatics/btp699

Lu, F., Romay, M. C., Glaubitz, J. C., Bradbury, P. J., Elshire, R. J., Wang, T., et al. (2015). High-resolution genetic mapping of maize pan-genome sequence anchors. *Nat. Commun.* 6, 6914. doi: 10.1038/ncomms7914

Mungall, C. J., and Emmert, D. B. (2007). A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* 23, i337–i346. doi: 10.1093/bioinformatics/btm189

Neuffer, M. G., Jones, L., and Zuberer, M. S. (1968). *The Mutants of Maize*. Madison, WI, Crop Science Society of America.

Oberlin, A. T., Jurkovic, D. A., Balish, M. F., and Friedberg, I. (2013). Biological database of images and genomes: tools for community annotations linking image and genomic information. *Database (Oxford)* 2013, bat016. doi: 10.1093/database/bat016

Portwood, J. L., 2nd, Woodhouse, M. R., Cannon, E. K., Gardiner, J. M., Harper, L. C., Schaeffer, M. L., et al. (2019). MaizeGDB 2018: the maize multi-genome genetics and genomics database. *Nucleic Acids Res.* 47, D1146–D1154. doi: 10.1093/nar/gky1046

Schaeffer, M. L., Harper, L. C., Gardiner, J. M., Andorf, C. M., Campbell, D. A., Cannon, E. K., et al. (2011). MaizeGDB: curation and outreach go hand-in-hand. *Database (Oxford)* 2011, bar022. doi: 10.1093/database/bar022

Schnable, J. C., and Freeling, M. (2011). Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. *PloS One* 6, e17855. doi: 10.1371/journal.pone.0017855

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science* 326, 1112–1115. doi: 10.1126/science.1178534

Sen, T. Z., Andorf, C. M., Schaeffer, M. L., Harper, L. C., Sparks, M. E., Duvick, J., et al. (2009). MaizeGDB becomes ‘sequence-centric’. *Database (Oxford)* 2009, bap020. doi: 10.1093/database/bap020

Springer, N., Anderson, S. N., Andorf, C., Ahern, K., Bai, F., Barad, O., et al. (2018). The W22 genome: a foundation for maize functional genomics and transposon biology. *Nat. Genet.* 50, 1282–1288. doi: 10.1038/s41588-018-0158-0

Springer, N. M., and Anderson, S. N., (2018). The maize W22 genome provides a foundation for functional genomics and transposon biology. *Nat. Genet.* 50, 1282–1288. doi: 10.1038/s41588-018-0158-0

Sun, S., Zhou, Y., Chen, J., and Shi, J., (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat. Genet.* 50, 1289–1295. doi: 10.1038/s41588-018-0182-0

Unterseer, S., Seidel, M. A., Bauer, E., Haberer, G., Hochholdinger, F., Opitz, N., et al. (2017). European Flint reference sequences complement the maize pan-genome. *bioRxiv*. 103747. doi: 10.1101/103747

Xiao, Y., Liu, H., Wu, L., Warburton, M., and Yan, J. (2017). Genome-wide association studies in maize: praise and stargaze. *Mol. Plant* 10, 359–374. doi: 10.1016/j.molp.2016.12.008

Yang, N., Xu, X. W., Wang, R. R., Peng, W. L., Cai, L., Song, J. M., et al. (2017). Contributions of Zea *mays* subspecies mexicana haplotypes to modern maize. *Nat. Commun.* 8, 1874. doi: 10.1038/s41467-017-02063-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 15 October 2019
doi: 10.3389/fpls.2019.01239

Noninvasive Phenotyping of Plant–Pathogen Interaction: Consecutive *In Situ* Imaging of Fluorescing *Pseudomonas syringae*, Plant Phenolic Fluorescence, and Chlorophyll Fluorescence in *Arabidopsis* Leaves

Sabrina Hupp[1], Maaria Rosenkranz[1,2], Katharina Bonfig[1], Chandana Pandey[3]* and Thomas Roitsch[1,3,4]

[1] Department of Pharmaceutical Biology, University of Würzburg, Würzburg, Germany, [2] Department of Environmental Sciences, Institute of Biochemical Plant Pathology, Research Unit Environmental Simulation, Helmholtz Zentrum Muenchen, Neuherberg, Germany, [3] Department of Plant and Environmental Sciences, Section of Crop Science, University of Copenhagen, Copenhagen, Denmark, [4] Department of Adaptive Biotechnologies, Global Change Research Institute, CAS, Brno, Czechia

OPEN ACCESS

Edited by:
Roland Pieruschka,
Julich Research Centre,
Germany

Reviewed by:
Marek Zivcak,
Slovak University of Agriculture,
Slovakia
Tomasz Hura,
The Franciszek Górski Institute of
Plant Physiology (PAS), Poland

***Correspondence:**
Chandana Pandey
cp@plen.ku.dk

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 02 June 2019
Accepted: 05 September 2019
Published: 15 October 2019

Citation:
Hupp S, Rosenkranz M, Bonfig K,
Pandey C and Roitsch T (2019)
Noninvasive Phenotyping of
Plant–Pathogen Interaction:
Consecutive In Situ Imaging of
Fluorescing Pseudomonas syringae,
Plant Phenolic Fluorescence,
and Chlorophyll Fluorescence in
Arabidopsis Leaves.
Front. Plant Sci. 10:1239.
doi: 10.3389/fpls.2019.01239

Plant–pathogen interactions have been widely studied, but mostly from the site of the plant secondary defense. Less is known about the effects of pathogen infection on plant primary metabolism. The possibility to transform a fluorescing protein into prokaryotes is a promising phenotyping tool to follow a bacterial infection in plants in a noninvasive manner. In the present study, virulent and avirulent *Pseudomonas syringae* strains were transformed with green fluorescent protein (GFP) to follow the spread of bacteria *in vivo* by imaging Pulse-Amplitude-Modulation (PAM) fluorescence and conventional binocular microscopy. The combination of various wavelengths and filters allowed simultaneous detection of GFP-transformed bacteria, PAM chlorophyll fluorescence, and phenolic fluorescence from pathogen-infected plant leaves. The results show that fluorescence imaging allows spatiotemporal monitoring of pathogen spread as well as phenolic and chlorophyll fluorescence *in situ*, thus providing a novel means to study complex plant–pathogen interactions and relate the responses of primary and secondary metabolism to pathogen spread and multiplication. The study establishes a deeper understanding of imaging data and their implementation into disease screening.

Keywords: green fluorescence protein (GFP), plant–pathogen interaction, imaging PAM, chlorophyll fluorescence imaging, phenolic compounds

INTRODUCTION

Plants are regularly attacked by several pathogens, such as bacteria, fungi, viruses, oomycetes, nematodes, and others. Due to a lowered performance of an infected plant, pathogen invasion can lead into severe economical losses on economically important field and forest sites (Gutierrez-Arellano and Mulligan, 2018). To defend themselves, plants have developed various strategies in which not only secondary but also carbohydrate metabolism plays complex roles (Trouvelot et al., 2014; Jammer et al., 2015).

Several studies exist on the importance of secondary metabolites in plant defense against pathogen-induced biotic stresses (Großkinsky et al., 2016a; Großkinsky et al., 2016b; Jing et al., 2018; Rosa et al., 2018; Zaynab et al., 2018). Phenolic compounds, the most ubiquitous secondary metabolites in plants, are stress induced and serve in specific roles of plant defense, e.g., as deterrents to pathogens and herbivores and by protecting against UV radiation and oxidative stress. Most of them (flavonoids, tannins, hydroxycinnamate esters, and lignin) have common origin from shikimate acid pathway *via* phenylpropanoids (Ge et al., 2018; Thakur et al., 2018).

The roles of primary metabolites in plant defense against pathogens are less exploited than those of secondary metabolites, even if many biologists have shown a correlation between sugar quantities and plant defense responses (Bonfig et al., 2010; Carvalho et al., 2019; Dong and Beckles, 2019). Furthermore, several studies have proven decrease in photosynthesis simultaneously with pathogen spread on a plant leaf (Bonfig et al., 2006; Berger et al., 2007; Dong et al., 2016; Lu and Yao, 2018; Tischler et al., 2018; Mahlein et al., 2019). The observation may be caused by decreased performance of photosynthetic apparatus and change of the pathogen-attacked site from source to a sink. Alternatively, up-regulation of invertases (Kuska et al., 2018; Su et al., 2018) may lead to observed phenomenon. The physiological background of decreased photosynthetic performance of a pathogen-attacked leaf is, however, still under debate.

Not only plants but also pathogens have developed various strategies to ensure more beneficial outcome in plant–pathogen interactions. Plant pathogens can successfully use plant carbohydrates for their own energy use (Kuska et al., 2018). They are likely to redirect the carbohydrate metabolism in plant leaves to ensure enough nutrients for themselves (Ökmen and Doehlemann, 2014). Several pathogenic species possess sucrose-hydrolyzing enzymes (Chaliha et al., 2018; Kanwar and Jha, 2018), which can help them to secure the plant-provided nutrients directly at the infection site. Indeed, Berger et al. (2007) suggested that change of an attacked leaf from a source to a sink could be due to a pathogen manipulation of plant primary metabolism. Recently, it was shown that certain pathogen species are indeed able to modify the host's photosynthesis to stay active, thereby creating conditions favorable to its own survival (Xue et al., 2018). Whether the plant-provided nutrients could also enhance the survival of the pathogen in other manners, e.g., by playing a role in pathogenesis, has to be elucidated in the future. Multicolor fluorescence imaging (MCFI) has also been used as a promising tool for disease detection in plant phenotyping (Murchie and Lawson, 2013; Barón et al., 2016; Pérez-Bueno et al., 2016). However, for the detection of plant stress phenotyping, the most commonly applied sensor and imaging techniques are digital RGB (red–green–blue) imaging; spectroscopy; thermography; fluorescence; three-dimensional, by, for example, stereo cameras and LIDAR (light detection and ranging); real-time camera set-ups; RNA-seq analysis; and, to a lesser extent, tomography (Quemada et al., 2014; Großkinsky et al., 2017; Ghosal et al., 2018; Pineda et al., 2018; Dobos et al., 2019; Polonio et al., 2019; Sperschneider, 2019).

Accumulation of bacteria in plants was studied mainly by reisolation of the bacterial cells (Raacke et al., 2006; Aydi-Ben-Abdallah et al., 2019; Liu et al., 2019). Similarly, the analysis of phenolic compounds in plant tissues commonly proceeded through diverse extraction methods (Torti et al., 1995; Giorgetti et al., 2018; Proestos et al., 2018). The techniques provide an invasive method to achieve information on pathogen accumulation or plant defense response. However, the disadvantages of invasive techniques are clear; the determination has to be done from detached leaves; the changes cannot be followed over time; and moreover, the precise location of the pathogen is not possible.

To assess the impact of biotic stress on host plant, various imaging techniques are currently used in plant physiology (Barón et al., 2016). Some of these techniques include MCFI and chlorophyll fluorescence (Chl-F) imaging. Plant health status is monitored by MCFI, and it is based on recording the blue (F440), green (F520), red (F680), and far red (F740) fluorescence by leaves when they are excited with UV light (Polonio et al., 2019). Particularly, the blue–green fluorescence is a valuable technique to study secondary metabolism, because phenolic compounds from the phenylpropanoid pathway are the primary emitters of that fluorescence. The extent of absorbance of light by the epidermal polyphenols can be derived on the basis of the ratio of Chl-F emission intensities induced by a standard red beam and a Ultraviolet-visible spectroscopy (UV-VIS) beam. Similarly, red Chl-F emitted by photosystem II (PSII) provides information on the photosynthetic performance of plants in terms of activity and indirect information on the CO_2 assimilation rate (Murchie and Lawson, 2013).

Novel technologies such as hyperspectral imaging and Chl-F imaging offer an elegant, noninvasive means to explore indirectly the bacteria spread within the plant tissue (Rolfe and Scholes, 2010; Großkinsky et al., 2017; West et al., 2017; Bohnenkamp et al., 2019; Kuska et al., 2019). In our own study (Bonfig et al., 2006), the decrease in maximum PSII quantum yield after avirulent *Pseudomonas syringae* infection was detectable already 3 h after the inoculation of bacteria into the tissue. However, the method is not directly measuring the accumulation of bacteria, and the changes in photosynthesis could be due to other, undetermined, reasons. A direct noninvasive manner to locate and quantify bacterial pathogen in the plant leaf tissue was missing until Wang et al. (2007) decided to transform *P. syringae* cells with a plasmid containing bright fluorescing green fluorescence protein (GFP) uv-gene. The group was able to monitor the bacterial expansion in the whole plant level under long-wavelength UV light. Later, confocal laser scanning microscopy (Riedel et al., 2009) and fluorescence microscopy (Parente et al., 2008) were applied to detect GFP-expressing bacteria. Labeling prokaryotic cells by GFP has become a routinely applied technique to visualize cells in plant living tissue (Lozoya-Pérez et al., 2018; Yang et al., 2019).

Our interest is, in addition to detecting bacterial cells, to follow a pathogen invasion on a leaf by simultaneously monitoring direct changes in plant primary and secondary metabolisms. Here we present a technique to monitor different fluorescing sources on plant leaves *in situ* by using imaging PAM fluorescence (Heinz Walz GmBH) system as a tool. In the present study, we detect the GFP, Chl-F, and phenolic fluorescence within a short time from one intact leaf by utilizing different wavelengths and filters. The possibility to combine the three measurements in one instrument provides a clear advantage in characterization of the plant–pathogen interactions in the future. For initializing and valuing the technique, *Arabidopsis thaliana* and *P. syringae* were used as model systems. *Pseudomonas syringae* is a hemibiotrophic pathogen (Preston, 2000) that can invade several, also economically important plant species.

MATERIALS AND METHODS

Plant and Bacterial Materials

Arabidopsis thaliana, cv. Columbia 0, were cultivated at 22°C, L9:D15, and a photosynthetic photon flux density of 180 µmol photons $m^{-2} s^{-1}$ in climate chambers (Binder, Germany). We used 5- to 8-week-old *Arabidopsis* rosettes in the experiments. One of the first fully expanded leaves was chosen for the measurements. One *Arabidopsis* plant was treated as one biological replicate.

Pseudomonas syringae DC3000 and DC3000rpm were cultured in 28°C either on LB agar plates or by shaking in Kings medium B. The medium contained the appropriate antibiotics as follows: 50 µg mL^{-1} rifampicillin for both *P. syringae* strains, 5 µg ml^{-1} tetracyclin for the avirulent strain, and 100 µg ml^{-1} kanamycin for pPNptGreen-expressing bacteria. For plant infection, *P. syringae* were harvested by centrifuging and resuspended in 10 mM $MgCl_2$ until optical density $(OD)_{600} = 0.2$, which is equal to approximately 1×10^8 cell-forming units (cfu) ml^{-1}. Plants were infected by infiltrating the appropriate bacterial suspension by 1-ml plastic syringe (without needle) through the stomata into the leaf tissue. The individuals that were treated with the pathogens were selected randomly among the cultivated plants.

Electrocompetent *P. syringae* cells were obtained by growing bacteria in 500 ml of Kings medium B under optimal conditions until $OD_{600} = 0.6 \pm 0.1$. The bacteria were harvested by centrifuging and resuspended to 500 ml 10% glycerol (4°C). The harvest was repeated, and after each harvest, the pellet was resuspended first to 250 ml and then to 150 ml and finally to 3 ml cold 10% glycerol. The obtained bacteria cells were stored at −80°C until use.

Creating pPNptGreen Construct and Transforming it Into *P. Syringae*

The plasmid pPNptGreen (13,199 bp) carrying a GFP gene sequence and kanamycin resistance was obtained as a gift from G. Beattie (Department of Plant Pathology, Iowa State University, Iowa, USA). The competent *P. syringae* cells were transformed with pPNptGreen by electroporation (2.5 kV) and spread on LB agar plates. The colonies carrying the pPNtpGreen construct were selected by kanamycin resistance and additionally by detecting green fluorescence under UV light.

Measurement of the Growth of the Fluorescing and Wild-Type Bacteria in *Arabidopsis*

The reisolation of bacterial cells from plant leaves (Raacke et al., 2006) was applied to compare the growth of fluorescing bacteria to that of the wild type. Four leaves per plant were infected with 10^5 cfu ml^{-1} of *P. syringae*. The infection sites were harvested with a corkbore (*r* = 0.7 cm) 24, 48, and 72 h after the infiltration. All the four infection sites from one plant were pooled together and fine powdered in 500 µL 10 mM $MgCl_{2,}$ with a pestle in an Eppendorf tube. The obtained suspension was diluted with $MgCl_2$ until 1 ml and further until 1:100 or up to 1:100,000. A hundred microliters of each dilution was spread on agar plates; the plates were grown under 28°C for 48 h after which the colonies were calculated.

Quantifying the Fluorescence Signal From *P. Syringae*

Known concentrations (10^{10}, 10^9, 10^8, 10^7, 10^6m, and 10^5 cfu ml^{-1}) of fluorescing *P. syringae* in 10 mM $MgCl_2$ were used to valuate two different methods:

First, the fluorescence was quantified by a fluorometer (Fluoroscan, Ascent, Germany). The *P. syringae* suspensions (200 µL) in different concentrations were pipetted on a black microtiter plate. The GFP was excited at 485 nm, and fluorescence was measured at 538 nm. The background fluorescence of nonfluorescing control bacteria was subtracted from the obtained values. Second, to validate the signal detected by imaging PAM, the fluorescence of single *P. syringae* drops in different concentrations on black, nonfluorescing background was quantified. Photographs were taken with imaging PAM (for details of the technique, see below), and the fluorescence signal in the middle of the drop was measured. As a control, the nonfluorescing *P. syringae* strains were used.

Detecting GFP Under Fluorescence Binocular

The fluorescing bacteria were detected with fluorescence binocular, which was equipped with a special GFP3 filter (instruction manual imaging PAM, Heinz-Walz GmbH, Germany). The fluorescence were measured each 24 h until 96 h after the infection. Photographs of the fluorescing bacteria were taken with a camera (Spotlight Color) and analyzed by the photograph software SpotAdvanced.

Detection of Chl-F and GFP by Imaging PAM

Chlorophyll fluorescence measurements were performed as described in Bonfig et al. (2006). A maximum saturation pulse was applied on dark adapted (ca. 20 minutes) plants. Further fluorescence parameters were subsequently measured with actinic light intensity set at 76 µmol $m^{-2} s^{-1}$. Measurements were performed every 20 s, from 50 to 290 s, duration of the light pulses being set at 8 (instruction manual imaging PAM, Heinz-Walz GmbH). We report PSII quantum yield ($Y(II) = (Fm' - F')/Fm'$) and Fv/Fm ($Fv/Fm = (Fm - Fo)/Fm$) similar as has been previously described (for review e.g. Maxwell and Johnson, 2000; Murchie and Lawson, 2013). In addition to Chl-F, GFP was detected by imaging PAM (Walz, Germany) equipped with a special long-pass filter with an angle of 645 nm. The GFP was excitated at 450 nm. The measurements were performed with maximum intensity of measuring light and gain (the amplitude of fluorescence signal) = 8 (instruction manual imaging PAM, Heinz-Walz GmbH). Green fluorescent protein fluorescence was detected from the below sites of the leaves.

Detection of Phenolic Fluorescence by Reactive Oxygen Species Head

Phenolic compounds were detected by a special application developed for imaging PAM. The standard blue power LED lights were substituted with special UV-A power LEDs having an emission peak at 365 nm. All the wavelengths above 400 nm were filtered from the excitation light and a short-pass interference filter blocked the transmission above 650 nm, which is essential for excluding chlorophyll fluorescence from the detected signal. The reactive oxygen species (ROS) head was operated analog to imaging PAM, as described previously by Hideg and Schreiber (2007). Measuring light is applied as short (10–200 ls) pulses at low frequency (1–8 Hz). Two images are measured: one during the pulse and one directly afterward, from which a difference image was derived. This eliminates eventually disturbing ambient background light (Hideg and Schreiber, 2007). Phenolic fluorescence was detected from the below sites of the leaves.

Statistical Analysis

Statistical and correlation analyses were performed with SPSS for Windows (release 15.0) and Sigma Plot for Windows (version 10.0). Linear regression analysis was performed to describe the dependency of fluorescence intensity on bacterial density.

RESULTS AND DISCUSSION

Selection of Transgenics

The possibility to transform a fluorescing protein into a prokaryote is a promising tool to follow bacterial infection in a nondestructive manner in plants (Wang et al., 2007; Parente et al., 2008). In the present work, GFP was successfully transformed into two strains of bacterial pathogen *P. syringae*; in a virulent strain pv. *tomato* DC3000 and in an avirulent strain pv. *tomato* DC3000rpm. The positive transformants were selected by kanamycin resistance and additionally by picking the transgenic, green-fluorescing colonies from agar plates under UV light. The highest fluorescing colonies were chosen for further studies.

Detection of GFP-Fluorescing Bacteria in Plant Leaves by Binocular

The GFP-transformed *P. syringae* DC3000 were detected in *Arabidopsis* leaves 24 h after an infection with 1×10^7 cfu ml^{-1} pathogen (**Figure 1A**) under a binocular. When lower concentration (≤1×10^6 cfu ml^{-1}) of pathogens was initially applied or dip-inoculation-technique used, the fluorescence was detected earliest at 48 h after the infection. An application of 1×10^5 cfu ml^{-1} was detectable only 96 h after the infection (data not shown).

The progeny of both *P. syringae* DC3000 strains (each applied in concentration of 1×10^7 cfu ml^{-1} on individual leaves) was followed by binocular over 96 h. At the first time point (24 h after the infection), the virulent strain was hardly detectable; however, the fluorescence intensity increased by each 24-h period and was at its strongest 96 h after the infection (**Figures 1A–D**). Unlike the virulent strain, the fluorescent signal of avirulent *P. syringae* DC3000rpm strain was not well detectable. First, the detection was possible only when at least 1×10^8 cfu ml^{-1} bacterial concentration was initially infiltrated into the leaf tissue. Twenty-four hours after the infection, only a very low signal was detected (**Figure 1E**). The signal increased slightly over the time, and the signal was stronger in later time points, when the time points 24 h and 72 h or 48 h and 96 h were compared to each other (**Figures 1E–H**). The fact that avirulent strain of *P. syringae* showed lower increase in fluorescence signal over time than the virulent strain was likely due to an incompatible interaction between avirulent strain and the plant. The fast reaction to avirulent strain by programmed cell death testifies for incompatible interaction (Bonfig et al., 2006) and for faster defense response against avirulent than virulent strain in *Arabidopsis*. The fast defense

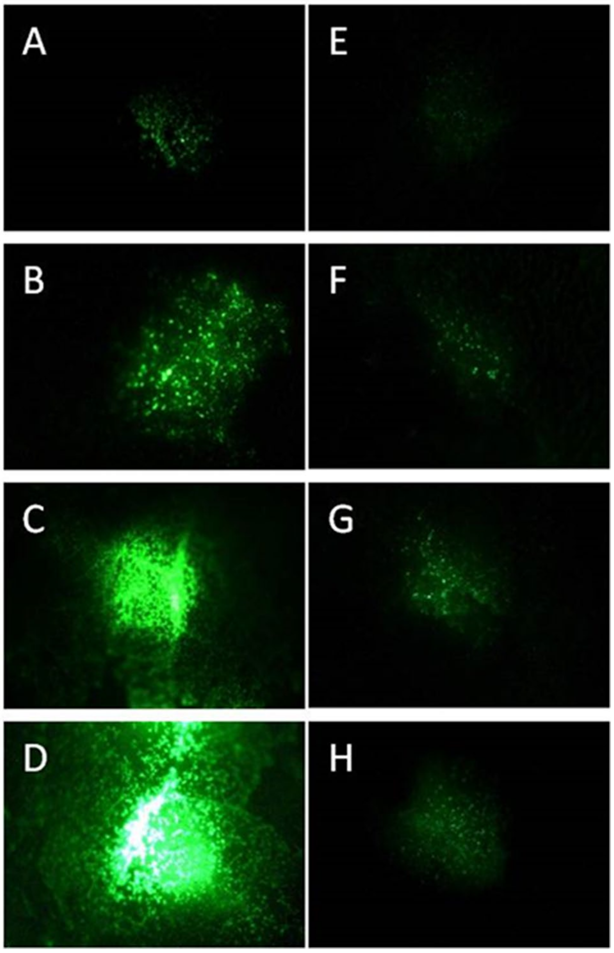


FIGURE 1 | Progeny of GFP synthesizing *P. syringae* DC3000 **(A–D)** and DC3000rpm **(E–H)** detected as fluorescence signal 24 h **(A, E)**, 48 h **(B, F)**, 72 h **(C, G)**, and 96 h **(D, H)** after infiltration with 1×10^7 cfu ml^{-1} for DC3000gfp and 1×10^8 cfu ml^{-1} for DC3000rpm_gfp.

response against the microbial pathogen was followed by low pathogen progeny and, logically, low GFP fluorescence that partly remained under the detection limit. Due to these complications, the avirulent strain was excluded from part of the further studies.

As expected, an infected leaf showed further developed necrosis in the middle of the infiltration site than on the edges of it (**Figure 2A**). Interestingly, 2 weeks after the infection with the virulent *P. syringae* strain, the highest fluorescence signal was measured on the edges of the infection site, whereas the signal was lower directly in the middle of the infiltration site (**Figure 2B**). Thus, the strongest fluorescence signal was detected on the areas in which the bacteria invaded so far untouched plant cells. Furthermore, our data show that the bacterial density was at its highest on the edges of the infection site 2 weeks after the infection, whereas a lower bacterial density at the necrotic sites was found. The result suggests that the bacteria spread on the plant leaf from necrotic spot further into not yet infected areas.

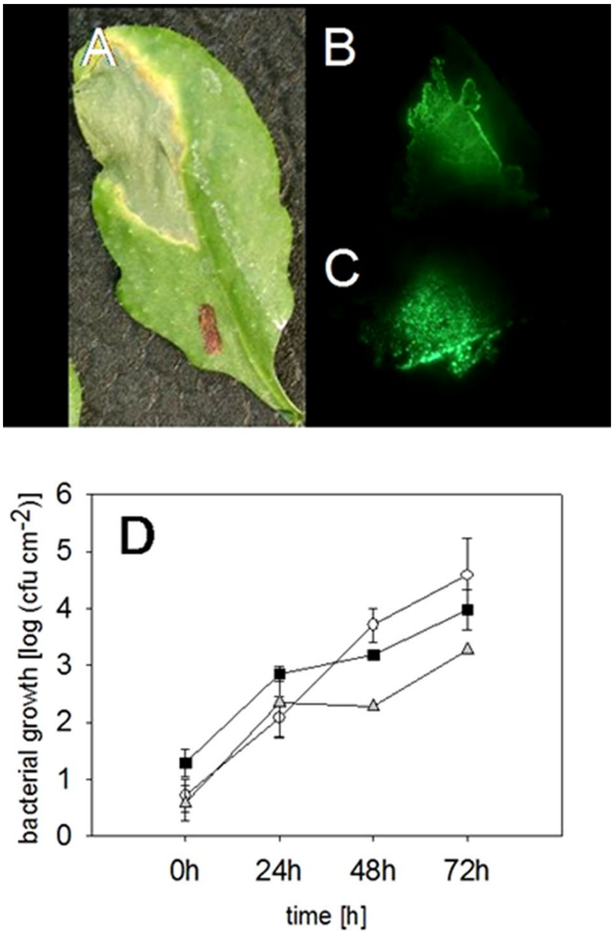


FIGURE 2 | Locating and quantifying the *P. syringae* DC3000_gfp cells in differentially infected *Arabidopsis* leaves. Typical symptoms in the *Arabidopsis* leaves due to an infection with *P. syringae* applied by a needless syringe **(A)** and the location of the bacterial cells in the same leaf by a fluorescence signal 2 weeks after the infection **(B)**. The highest fluorescing signal was detected in the leaf veins compared to other tissue 72 h after an infection with *P. syringae* **(C)**. The accumulation of fluorescing and nonfluorescing *P. syringae* DC3000 and fluorescing DC3000rpm cells in *Arabidopsis* leaves. DC3000 (■); DC3000gfp (○); DC3000rpm_gfp (▲); n = 3 **(D)**. No significant differences (Kruskal-Wallis) were found between the different strains.

Furthermore, in the infected *Arabidopsis* leaves, it was obvious that the virulent *P. syringae* cells accumulated stronger in the vascular tissue of the plant leaves than on the other sites. Higher fluorescence signal was often, even if not always, found in the middle vein or in the smaller veins of the plant leaves (**Figure 2C**). Such an accumulation was detected at different time points by binocular, but only when at least 1×10^{8} cfu ml^{-1} bacterial concentration was initially applied into the leaves.

Whether virulent *P. syringae* use the veins to move from one site to another or prefer them due to a higher nutrient quantity is not known at present.

The Progeny of Wild-Type and GFP-Transformed *P. Syringae*

The fluorescing *P. syringae* developed symptoms in a similar manner with the nonmarked wild-type strain. Infiltration of the virulent, fluorescing strain into leaf tissue developed necrosis surrounded by so-called "chlorotic halos" at the infection site (**Figure 2A**). We also found no differences in the accumulation of the fluorescing and wild-type pathogens (for the time point 72 h: Kruskal-Wallis, $P = 0.191$, n = 3) (**Figure 3A**). The bacteria reached a concentration of approximately 615 ± 230 cfu cm^{-2} 24 h after the infiltration and the amount of cfu was approximately doubled within each further 24-h periods. Seventy-two hours post infection, a concentration of approximately 4.5×10^{4} cfu ml^{-1} was reached (**Figure 2D**).

As the fluorescing bacterial cells grew similar to the nonmarked cells, the plasmid pPNtpGreen likely did not interfere the plant–pathogen interaction. Wang and colleagues, who studied several different *P. syringae* strains (Wang et al., 2007), previously showed similar results. In general, our data prove that the fluorescing bacteria can be used instead of nonfluorescing wild-type bacteria in wide range of studies that aim to explore *P. syringae* interaction with its host.

Quantification of GFP-Transformed *P. syringae*

Fluorometric assay, compared to the traditional reisolation technique of bacteria, provides an easy and accurate tool to rapidly quantify the bacterial density on a plant leaf. The bacterial densities were detectable down to 1×10^{6} cfu ml^{-1} by the fluorometric assay. The fluorescence signal detected by fluorometer showed linear increase with increasing concentration of cell units ($R = 0.982$ for virulent and $R = 0.975$ for avirulent strain, **Figure 3A**). No differences in the signal between virulent and avirulent bacteria were found (Mann-Whitney U, $P > 0.1$ for all the time points, n = 3 ± 1), testifying for a comparable expression of the GFP in both of the strains.

We moreover quantified the fluorescence signal of different *P. syringae* concentrations by imaging PAM. Measuring the fluorescence of *P. syringae* drops in different concentrations

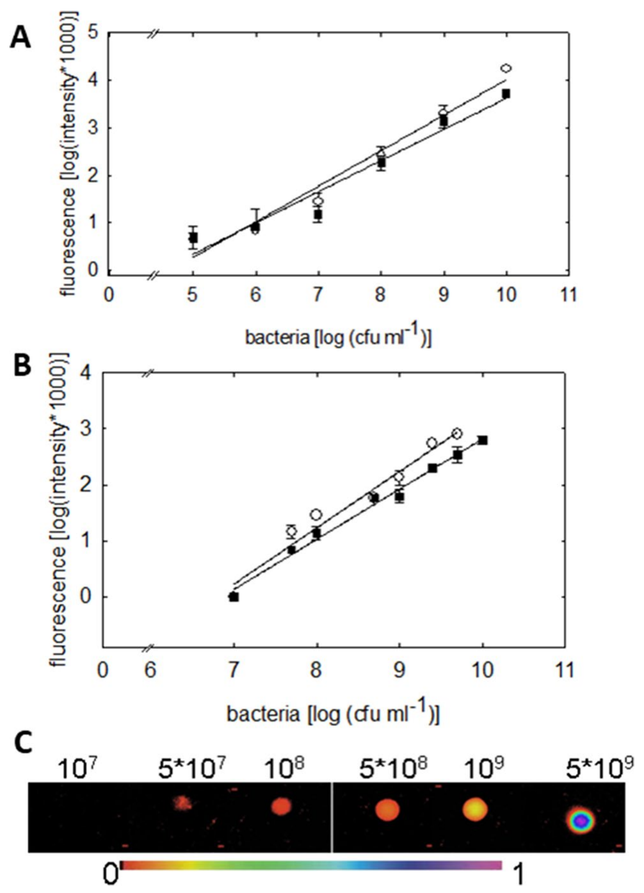


FIGURE 3 | Quantification of the green fluorescence intensity emitted by *P. syringae* DC3000_gfp (○) or DC3000rpm_gfp (■) cells. **(A)** Fluorometric quantification of cells in six different concentrations, from 1×10^5 up to 1×10^{10} cfu ml^{-1}, is shown. The fluorescence signal shows linear dependency on the concentration of the bacteria ($R = 0.982$; $y = 0.748x - 3.468$ (DC3000) and $R = 0.9750$; $y = 0.656x - 2.937$ (DC3000rpm); n = 4 ± 2). **(B)** Quantification of green fluorescence intensity by imaging PAM. The fluorescence was measured by imaging PAM from 20 µL of *P. syringae* cells in 10 mM $MgCl_2$ in increasing concentrations. The fluorescence signal shows linear dependency on bacterial concentration; $R = 0.98$; $y = 1x - 6.8$ (DC3000) and $R = 0.99 = 0.99$; $y = 0.89x - 6.12$ (DC3000rpm); n = 4 ± 1. In **(C)**, an example of fluorescence signals in various *P. syringae* concentrations is shown (from 1×10^7 up to 5×10^9 cfu ml^{-1}) detected by imaging PAM. The false scale color is given in the bottom.

(**Figure 3C**) made it possible to validate the GFP detection by imaging PAM (**Figure 3B**). The lower detection limit of imaging PAM was identified at 5×10^7 cfu ml^{-1}, and the saturation of the signal was detected at 1×10^{10} cfu ml^{-1}. The fluorescence signal depended linearly on bacterial concentration, $R = 0.98$ for virulent and $R = 0.99$ for avirulent strain. In support with the fluorometric quantification, also here no differences between virulent and avirulent strain were found (Mann-Whitney *U*, $P > 0.1$ for all the time points, n = 4 ± 1), further verifying a comparable expression of the GFP in both of the strains. Detection of pathogen in plants on multiple scales facilitates the advancement and current development.

Simultaneous Detection of GFP and Chl-F by Imaging PAM

Next we followed simultaneously the progeny of fluorescing virulent *P. syringae* in *Arabidopsis* leaves and the plant primary metabolism performance in a noninvasive manner by imaging PAM. Infection with GFP-marked *P. syringae* was detectable by imaging PAM (excitation with 450-nm wavelength) 20 and 24 h post infection at which time points fluorescence from wild-type bacteria site was not yet detectable (**Figures 4A–F**).

Imaging PAM made it possible to visually prove that the bacterial cells were indeed located on the area in which also plant's photosynthetic performance was affected. On the same leaf site on which bacteria were inoculated, the photosynthetic performance of the plant leaves decreased in synchrony with the spreading bacteria. A decrease in quantum yield of PSII (**Figures 4G–I**) and in maximum efficiency of PSII (**Figures 4J–L**) was detected from the initial values. The result supports the previous studies showing that reducing photosynthesis is an effective method to defend against biotrophic pathogens (Berger et al., 2007; Garavaglia et al., 2010). No differences were found in photosynthetic performance between the GFP-labeled and wild-type *P. syringae*–infected leaves. The lowest maximum quantum yields of PSII (Fv/Fm) were detected at the very sites where the bacteria were initially inoculated (**Figure 4**).

The detection of the fluorescing, avirulent *P. syringae* strain was possible only ≥24 h post infection and only when initially a high amount of *P. syringae* (1×10^8 cfu ml^{-1}) was applied (data not shown). At these later time points (≥24 h), it was not possible to distinguish GFP from plant phenolic fluorescence by imaging PAM.

The visualization of the bacterial cells by imaging PAM can be especially useful in investigating the role of plant primary metabolites in plant defense responses. The location of the bacteria can also be

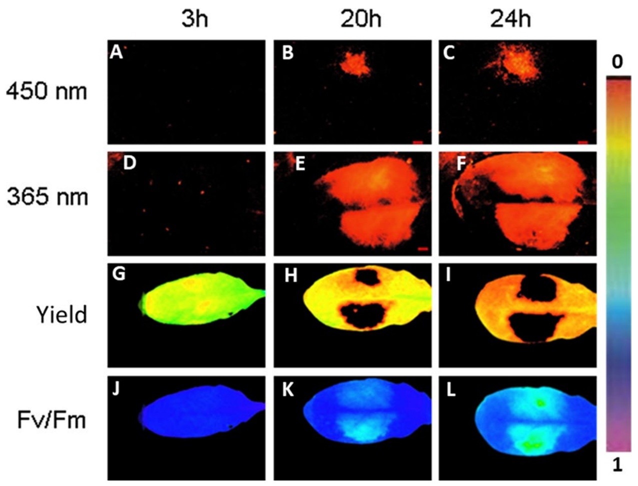


FIGURE 4 | An example of *Arabidopsis* leaf infected with GFP-labeled *P. syringae* DC3000 on one side of the leaf (upper side) and with wild-type *P. syringae* on the other site of the leaf (lower side), both in concentration 1×10^7 cfu ml^{-1}. The fluorescence signal is detected after 3, 20, and 24 h by exciting with either 450 nm (imaging PAM) **(A–C)** or 365 nm (ROS head) **(D–F)**. Quantum efficiency of PSII **(G–I)** and maximum efficiency of PSII (Fv/Fm) **(J–L)** are shown for the same time points. The false color scale is given on the right.

assigned more accurately than by, e.g., reisolation of bacteria. One of the advantages of GFP detection by imaging PAM over binocular is the possibility to take all the fluorescence pictures in similar position with exactly the same distance between the leaf and camera.

Combining GFP, Phenolic, and Chl-F Detections *In Situ*

The fluorescing nature of the phenolic compounds allows these secondary metabolites to be detected under UV or blue light. As GFP can be detected with the excitation peak at 450 nm, at which wavelength also phenolic fluorescence is excited, we applied furthermore a special application of imaging PAM, so-called ROS head (Hideg and Schreiber, 2007), which sends wavelength of 365 nm. Using this special application, we detected phenolic fluorescence, in *Arabidopsis* leaves infected with GFP synthesizing *P. syringae* DC3000 on one leaf side and with wild-type *P. syringae* on the other side. Our results reveal that, contrary to imaging PAM, by ROS head a strong plant fluorescence signal was detected 20 and 24 h post infection with any *P. syringae* strain (**Figures 4E, F**). Both of the wavelengths (UV in ROS head and blue light in imaging PAM) were thus used in the studies to be able to distinguish plant autofluorescence from GFP.

To determine in which time frames GFP and phenolic fluorescence can be distinguished from each other, the fluorescence signal from several infected plant leaves (one side of the leaf infected with GFP labeled, the other side with nonmarked *P. syringae*) was recorded each half an hour during at least 24 h. The measurements were done by applying either 365 nm (ROS head), so that only plant phenolic fluorescence was excited, or 450 nm (imaging PAM), which excites both GFP and phenolic fluorescence. The results show that with imaging PAM GFP is detectable 15.3 ± 2.5 h after infection, whereas phenolic fluorescence can be seen only 23.7 ± 4.4 h after the infection (**Figures 5A, C**; $P < 0.01$ (Student *t* test).

The recordings with ROS head (excitation at 365 nm) show that phenolic fluorescence appears approximately at the same time point (16.6 ± 1.7 h) (**Figures 5A, B**) as GFP can be detected with imaging PAM. Thus, imaging PAM was "blind" to phenolic fluorescence until a certain level of fluorescence signal was achieved. With ROS head, the GFP-labeled bacteria were detected probably due to combination of phenolic fluorescence and GFP signal already 13 ± 0.7 h post infection. The earlier detection point is probably due to GFP that can be excited by UV light in certain extent. Our results show that with certain limits GFP can be distinguished from phenolic fluorescence and vice versa.

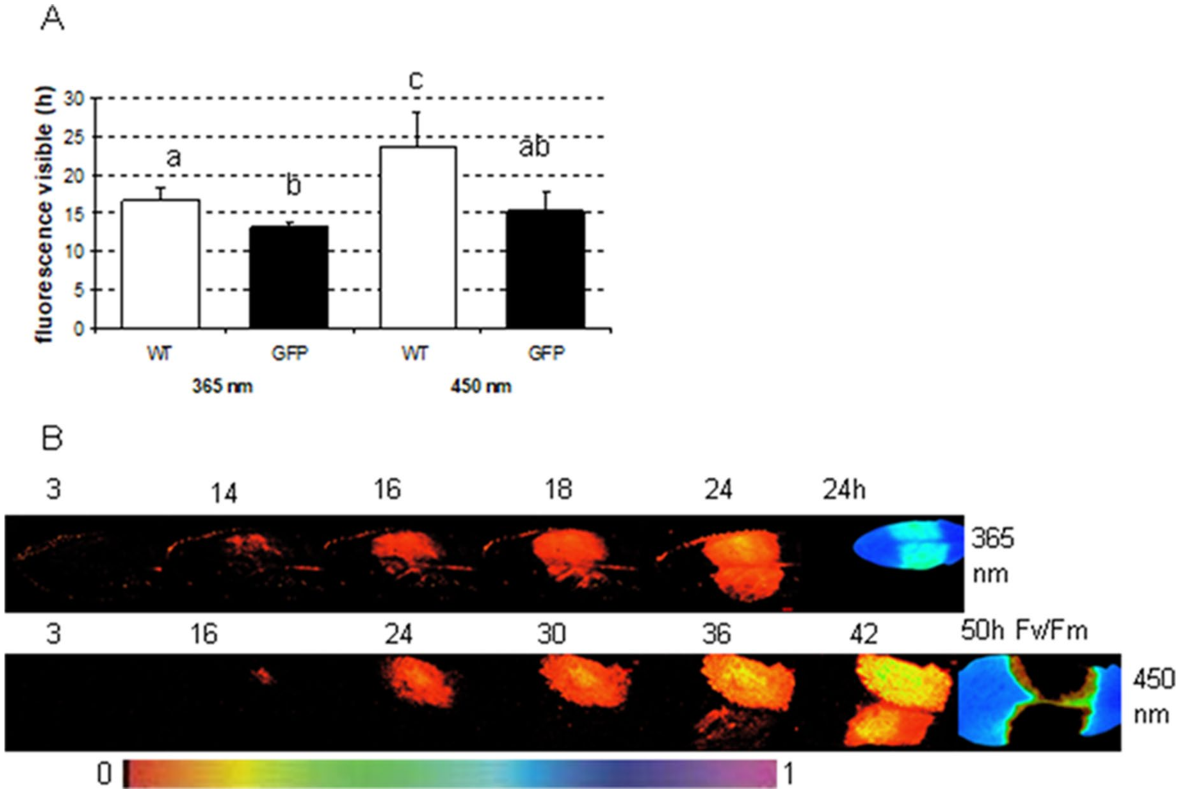


FIGURE 5 | The time in which fluorescence signal is detectable from *Arabidopsis* leaves infected either with wild-type or GFP-labeled *P. syringae* DC3000. **(A)** The fluorescence signal detection limit with either excitation by imaging PAM (450 nm) or by ROS head (365 nm) when *Arabidopsis* leaves were infected with 1×10^7 cfu ml^{-1} type (WT) or GFP-labeled (GFP) *P. syringae* DC3000. Different letters represent the significant differences between different treatments and methods ($P < 0.01$; Student *t* test, n = 6 ± 1). **(B)** An example of the fluorescence signal detection by ROS head (upper row) and imaging PAM (lower row) over time when one side of the *Arabidopsis* leaves was infected with GFP synthesizing (upper site) and the other side with wild-type (lower site) *P. syringae* DC3000. Maximum efficiency of PSII (Fv/Fm) in the end of the experiment is shown for each experiment on the right. False color scale is given in the bottom of the figure.

In the present study, phenolic compounds could be easily spatially located without damaging the leaf. In addition to spatial location, the time course of the induction of secondary metabolites after a pathogen attack was followed over several days with very short time intervals. With the help of ROS head, the measurements were recorded automatically over a longer period, which reduced the need of labor. A certain, minimum level of *P. syringae* is necessary before a detection of fluorescing bacterial cells is possible by imaging PAM, which is a disadvantage of the application. Another limitation is that the fluorescence signal is a mixture of GFP and plant phenolic fluorescence signals already approximately 24 h after the infection. Thus, the period in which GFP can be detected at 450 nm without background signals is relatively small, and the appearance of other signals has to be always excluded by nontransformed controls. Interestingly, however, by ROS head, phenolic fluorescence signal can be detected in the same time frame as GFP by imaging PAM. Instead of concentrating only to GFP detection, it might be of interest to use the detection of phenolic compounds to suspect a pathogen attack in several research applications. The initiation of plant defense metabolism could be proven well before physiological changes in the leaves are visible. Most interestingly, detecting phenolic fluorescence does not acquire transgenic bacterial lines.

Taken together, the different wavelengths between phenolic fluorescence, GFP fluorescence, and Chl-F allow the detection of all three from a single leaf within minutes. These imaging techniques enable a novel kind of insight to the plant pathogen interactions and could be applied for diverse research purposes. It is our interest to further optimize the technique to simultaneous image the plant performance and pathogen progeny *in situ*. Should this be possible in microscopic scale, an even more sensitive technique can be developed for visualization of plant–pathogen interactions. Further, this proof-of-concept study needs to be tested, verified, and validated with other pathosystem and expanded to other parameters, which can be determined in a noninvasive way. For initializing and valuing the technique, *A. thaliana* and *P. syringae* are used as model systems. *Pseudomonas syringae* is a hemibiotrophic pathogen that can invade several, also economically important plant species. Recent advancement and current development are facilitating the detection of pathogen in plants on multiple scales. Although it is challenging regarding the diverse type of pathogen, we must explore multiscale approach by possibility to combine this technique with other types of noninvasive analysis either in combination with reporter construct or fluorescent dye–like monitoring pH changes and ROS. Finally, within a truly holistic functional phenomics approach, the image-based, noninvasive phenotyping needs to be complemented by physiological phenotyping (Großkinsky et al., 2017). Thus, the optical signals need to be related to cell and ecophysiological parameters by methods such as the determination of enzyme activity signatures (Jammer et al., 2015) and phytohormone profiles (Großkinsky et al., 2014).

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

SH, MR, and KB conducted the experimental work and analyzed the data. MR drafted a preliminary version of the manuscript, and CP made an updated and revised version of the manuscript and finalized the manuscript for publication. TR designed the project and contributed to the preliminary and final version of the manuscript. All authors discussed the results.

FUNDING

The work of TR was supported by the Ministry of Education, Youth and Sports of CR within the National Sustainability Program I (NPU I), grant LO1415.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the essential support by Ulrich Schreiber (Julius-Sachs-Institute, University of Würzburg) by the design of a dual GFP and PAM fluorescence IMAGING PAM system, providing a prototype of a UV light measuring head, help with the measurement and data evaluation. We would like to thank Gwyn A. Beattie (Department of Plant Pathology, Iowa State University, Iowa, USA) for providing the plasmid pPNptGreen. This article is dedicated to Svend Christensen (Department of Plant and Environmental Sciences, University of Copenhagen), on the occasion of his 60th birthday.

REFERENCES

Aydi-Ben-Abdallah, R., Jabnoun-Khiareddine, H., Nefzi, A., and Daami-Remadi, M. (2019). Growth promotion and *Fusarium* wilt suppression in tomato using endophytic bacteria recovered from two wild solanaceous species. *Int. J. Adv. Agric. Sci.* 4 (2), 01–16.

Barón, M., Pineda, M., and Pérez-Bueno, M. L. (2016). Picturing pathogen infection in plants. *Z. Naturforsch. C* 71 (9–10), 355–368. doi: 10.1515/znc-2016-0134

Berger, S., Sinha, A. K., and Roitsch, T. (2007). Plant physiology meets phytopathology: plant primary metabolism and plant–pathogen interactions. *J. Exp. Bot.* 58, 4019–4026. doi: 10.1093/jxb/erm298

Bonfig, K. B., Schreiber, U., Gabler, A., Roitsch, T., and Berger, S. (2006). Infection with virulent and avirulent *P. syringae* strains differentially affects photosynthesis and sink metabolism in *Arabidopsis* leaves. *Planta* 225, 1–12. doi: 10.1007/s00425-006-0303-3

Bonfig, K. B., Gabler, A., Simon, U. K., Luschin-Ebengreuth, N., Hatz, M., Berger, S., et al. (2010). Post-translational derepression of invertase activity in source leaves *via* down-regulation of invertase inhibitor expression is part of the plant defense response. *Mol. Plant* 3 (6), 1037–1048. doi: 10.1093/mp/ssq053

Bohnenkamp, D., Kuska, M. T., Mahlein, A. K., and Behmann, J. (2019). Utilising pure fungal spore spectra as reference for a hyperspectral signal decomposition and symptom detection of wheat rust diseases on leaf scale. *Plant Pathol.* 68, 1188–1195. doi: 10.1111/ppa.13020

Chaliha, C., Rugen, M. D., Field, R. A., and Kalita, E. (2018). Glycans as modulators of plant defense against filamentous pathogens. *Front. Plant Sci.* 9, 928. doi: 10.3389/fpls.2018.00928

Carvalho, C. P., Cardoso-Gustavson, P., Rodrigues, E., Braga, M. R., Mercier, H., and Nievola, C. C. (2019). Low temperature acclimation and de-acclimation of the subtropical bromeliad *Nidularium minutum*: implications of changes in the NO, sugar content and NR activity. *Environ. Exp. Bot.* 1159, 34–43. doi: 10.1016/j.envexpbot.2018.12.004

Dong, X., Wang, M., Ling, N., Shen, Q., and Guo, S. (2016). Potential role of photosynthesis-related factors in banana metabolism and defense against *Fusarium oxysporum* f. sp. cubense. *Environ. Exp. Bot.* 129, 4–12. doi: 10.1016/j.envexpbot.2016.01.005

Dong, S., and Beckles, D. M. (2019). Dynamic changes in the starch–sugar interconversion within plant source and sink tissues promote a better abiotic stress response. *J. Plant Physiol.* 234–235, 80–93. doi: 10.1016/j.jplph.2019.01.007

Dobos, O., Horvath, P., Nagy, F., Danka, T., and Viczián, A. (2019). A deep learning-based approach for high-throughput hypocotyl phenotyping. *bioRxiv*, 1, 651729. doi: 10.1101/651729

Garavaglia, B. S., Thomas, L., Gottig, N., Zimaro, T., Garofalo, C. G., Gehring, C., et al., (2010). Shedding light on the role of photosynthesis in pathogen colonization and host defense. *Commun. Integr. Biol.* 3, 382–384. doi: 10.4161/cib.3.4.12029

Ge, Y., Wei, M., Li, C., Chen, Y., Lv, J., Meng, K., et al. (2018). Reactive oxygen species metabolism and phenylpropanoid pathway involved in disease resistance against *Penicillium expansum* in apple fruit induced by ε-poly-L-lysine. *J. Sci. Food Agric.* 98, 5082–5088. doi: 10.1002/jsfa.9046

Ghosal, S., Blystone, D., Singh, A. K., Ganapathysubramanian, B., Singh, A., and Sarkar, S. (2018). An explainable deep machine vision framework for plant stress phenotyping. *Proc. Natl. Acad. Sci. U. S. A.* 115 (18), 4613–4618. doi: 10.1073/pnas.1716999115

Großkinsky, D. B., Albacete, A., Jammer, A., Remele, K., v. d. Graaff, E., Pfeifhofer, H., et al. (2014). A rapid phytohormone and phytoalexin screening method for physiological phenotyping. *Mol. Plant* 7, 1053–1056. doi: 10.1093/mp/ssu015

Großkinsky, D. K., Tafner, R., Moreno, M. V., Stenglein, S. A., De Salamone, I. E. G., Nelson, L. M., et al. (2016a). Cytokinin production by *Pseudomonas fluorescens* G20-18 determines biocontrol activity against *Pseudomonas syringae* in *Arabidopsis. Sci. Rep.* 6, 23310. doi: 10.1038/srep23310

Großkinsky, D. K., van der Graaff, E., and Roitsch, T. (2016b). Regulation of abiotic and biotic stress responses by plant hormones. *Plant pathogen resistance biotechnology* 131, 131–147. doi: 10.1002/9781118867716.ch7

Großkinsky, D. K., Syaifullah, S. J., and Roitsch, T. (2017). Integration of multi-omics techniques and physiological phenotyping within a holistic phenomics approach to study senescence in model and crop plants. *J. Exp. Bot.* 69 (4), 825–844. doi: 10.1093/jxb/erx333

Gutierrez-Arellano, C., and Mulligan, M. (2018). A review of regulation ecosystem services and disservices from faunal populations and potential impacts of agriculturalisation on their provision, globally. *Nat. Conserv.* 30, 1. doi: 10.3897/natureconservation.30.26989

Giorgetti, L., Giorgi, G., Cherubini, E., Gervasi, P. G., Della Croce, C. M., Longo, V., et al. (2018). Screening and identification of major phytochemical compounds in seeds, sprouts and leaves of Tuscan black kale *Brassica oleracea* (L.) ssp *acephala* (DC) var. *sabellica* L. *Nat. Prod. Res.* 32 (14), 1617–1626. doi: 10.1080/14786419.2017.1392953

Hideg, E., and Schreiber, U. (2007). Parallel assessment of ROS formation and photosynthesis in leaves by fluorescence imaging. *Photosyn. Res.* 92, 103–108. doi: 10.1007/s11120-007-9146-4

Jammer, A., Gasperl, A., Luschin-Ebengreuth, N., Heyneke, E., Chu, H., Cantero-Navarro, E., et al. (2015). Simple and robust determination of the activity signature of key carbohydrate metabolism enzymes for physiological phenotyping in model and crop plants. *J. Exp. Botany.* 22, 5531–5542. doi: 10.1093/jxb/erv228

Jing, X., Wang, H., Gong, B., Liu, S., Wei, M., Ai, X., et al. (2018). Secondary and sucrose metabolism regulated by different light quality combinations involved in melon tolerance to powdery mildew. *Plant Physiol. Biochem.* 124, 77–87. doi: 10.1016/j.plaphy.2017.12.039

Kanwar, P., and Jha, G. (2018). Alterations in plant sugar metabolism: signatory of pathogen attack. *Planta* 28, 1–4. doi: 10.1007/s00425-018-3018-3

Kuska, M. T., Behmann, J., Grosskinsky, D. K., Roitsch, T., and Mahlein, A. K. (2018). Screening of barley resistance against powdery mildew by simultaneous high-throughput enzyme activity signature profiling and multispectral imaging. *Front. Plant Sci.* 9, 1074. doi: 10.3389/fpls.2018.01074

Kuska, M. T., Behmann, J., Namini, M., Oerke, E. C., Steiner, U., and Mahlein, A. K. (2019). Discovering coherency of specific gene expression and optical reflectance properties of barley genotypes differing for resistance reactions against powdery mildew. *PLoS One* 14 (3), e0213291. doi: 10.1371/journal.pone.0213291

Lu, Y., and Yao, J. (2018). Chloroplasts at the crossroad of photosynthesis, pathogen infection and plant defense. *Int. J. Mol. Sci.* 19, 3900. doi: 10.3390/ijms19123900

Liu, X., Chen, Y., Zhong, M., Chen, W., Lin, Q., and Du, H. (2019). Isolation and pathogenicity identification of bacterial pathogens in bleached disease and their physiological effects on the red macroalga *Gracilaria lemaneiformis. Aquat. Bot.* 153, 1–7. doi: 10.1016/j.aquabot.2018.11.002

Lozoya-Pérez, N. E., Casas-Flores, S., Martínez-Álvarez, J. A., López-Ramírez, L. A., Lopes-Bezerra, L. M., Franco, B., et al. (2018). Generation of *Sporothrix schenckii* mutants expressing the green fluorescent protein suitable for the study of host–fungus interactions. *Fungal Biol.* 122, 1023–1030. doi: 10.1016/j.funbio.2018.07.004

Mahlein, A. K., Alisaac, E., Al Masri, A., Behmann, J., Dehne, H. W., and Oerke, E. C. (2019). Comparison and combination of thermal, fluorescence, and hyperspectral imaging for monitoring *Fusarium* head blight of wheat on spikelet scale. *Sensors* 19, 2281. doi: 10.3390/s19102281

Maxwell, K., and Johnson, G. N. (2000). Chlorophyll fluorescence—a practical guide. *J. Exp. Bot.* 51, 659–668. doi: 10.1093/jexbot/51.345.659

Murchie, E. H., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64 (13), 3983–3998. doi: 10.1093/jxb/ert208

Ökmen, B., and Doehlemann, G. (2014). Inside plant: biotrophic strategies to modulate host immunity and metabolism. *Curr. Opin. Plant Biol.* 20, 19–25. doi: 10.1016/j.pbi.2014.03.011

Parente, A. F., Silva-Pereira, I., Baldani, J. I., Tibúrcio, V. H., Báo, S. N., and De-Souza, M. T. (2008). Construction of *Bacillus thuringiensis* wild-type S76 and Cry-derivatives expressing a green fluorescent protein: two potential marker organisms to study bacteria-plant interactions. *Can. J. Microbiol.* 54, 786–790. doi: 10.1139/W08-061

Pérez-Bueno, M. L., Pineda, M., Francisco, M. C., and Barón, M. (2016). Multicolor fluorescence imaging as a candidate for disease detection in plant phenotyping. *Front. Plant Sci.* 7, 1790. doi: 10.3389/fpls.2016.01790

Pineda, M., Pérez-Bueno, M. L., and Barón, M. (2018). Detection of bacterial infection in melon plants by classification methods based on imaging data. *Front. Plant Sci.* 9, 164. doi: 10.3389/fpls.2018.00164

Pineda, M., Bautista, R., Martínez-Cruz, J., Pérez-Bueno, M. L., Barón, M., and Pérez-García, A. (2019). RNA-seq analysis and fluorescence imaging of melon powdery mildew disease reveal an orchestrated reprogramming of host physiology. *Sci. Rep.* 9 (1), 7978. doi: 10.1038/s41598-019-44443-5

Polonio, Á., Pineda, M., Bautista, R., Martínez-Cruz, J., Pérez-Bueno, M. L., Barón, M., et al. (2019). RNA-seq analysis and fluorescence imaging of melon powdery mildew disease reveal an orchestrated reprogramming of host physiology. *Sci. Rep.* 9 (1), 7978. doi: 10.1038/s41598-019-44443-5

Preston, G. M. (2000). *Pseudomonas syringae* pv. *tomato*: the right pathogen, of the right plant, at the right time. *Mol. Plant Pathol.* 1, 263–275. doi: 10.1046/j.1364-3703.2000.00036.x

Proestos, C., Zoumpoulakis, P., and Sinanoglou, V. J. (2018). Isolation and characterization of phenolic compounds from selected foods of plant origin using modern spectroscopic approaches. *Stud. Nat. Prod. Chem.* 57, 203–220. doi: 10.1016/B978-0-444-64057-4.00007-7

Quemada, M., Gabriel, J., and Zarco-Tejada, P. (2014). Airborne hyperspectral images and ground-level optical sensors as assessment tools for maize nitrogen fertilization. *Remote Sens. (Basel)* 6, 2940–2962. doi: 10.3390/rs6042940

Rosa, E., Woestmann, L., Biere, A., and Saastamoinen, M. (2018). A plant pathogen modulates the effects of secondary metabolites on the performance and immune function of an insect herbivore. *Oikos* 127, 1539–1549. doi: 10.1111/oik.05437

Raacke, I. C., von Rad, U., Mueller, M. J., and Berger, S. (2006). Yeast increases resistance in *Arabidopsis* against *Pseudomonas syringae* and *Botrytis cinerea* by salicylic acid–dependent as well as –independent mechanisms. *Mol. Plant Microbe Interact.* 19, 1138–1146. doi: 10.1094/MPMI-19-1138

Riedel, M., Calmin, G., Belbahri, L., Lefort, F., Götz, M., Wagner, S., et al. (2009). Green fluorescent protein (GFP) as a reporter gene for the plant pathogenic oomycete *Phytophthora ramorum. J. Eukaryot. Microbiol.* 56, 130–135. doi: 10.1111/j.1550-7408.2008.00376.x

Rolfe, S. A., and Scholes, J. D. (2010). Chlorophyll fluorescence imaging of plant–pathogen interactions. *Protoplasma* 247 (3–4), 163–175. doi: 10.1007/s00709-010-0203-z

Sperschneider, J. (2019). Machine learning in plant–pathogen interactions: empowering biological predictions from field-scale to genome-scale. *New Phytol.* doi: 10.1111/nph.15771

Su, T., Han, M., Min, J., Chen, P., Mao, Y., Huang, Q., et al. (2018). Genome-wide survey of invertase encoding genes and functional characterization of an extracellular fungal pathogen-responsive invertase in *Glycine max. Int. J. Mol. Sci.* 19 (8), 2395. doi: 10.3390/ijms19082395

Tischler, Y. K., Thiessen, E., and Hartung, E. (2018). Early optical detection of infection with brown rust in winter wheat by chlorophyll fluorescence excitation spectra. *Comput. Electron. Agr.* 146, 77–85. doi: 10.1016/j.compag.2018.01.026

Thakur, M., Bhattacharya, S., Khosla, P. K., and Puri, S. (2018). Improving production of plant secondary metabolites through biotic and abiotic elicitation. *J. Appl. Res. Med. Aromat. Plants*. 12, 1–12. doi: 10.1016/j.jarmap.2018.11.004

Trouvelot, S., Héloir, M. C., Poinssot, B., Gauthier, A., Paris, F., Guillier, C., et al. (2014). Carbohydrates in plant immunity and plant protection: roles and potential application as foliar sprays. *Front. Plant Sci.* 5, 592. doi: 10.3389/fpls.2014.00592

Torti, S. D., Dearing, M. D., and Kursar, T. A. (1995). Extraction of phenolic compounds from fresh leaves: a comparison of methods. *J. Chem. Ecol.* 21, 117–125. doi: 10.1007/BF02036646

Wang, K., Kang, L., Anand, A., Lazarovits, G., and Mysore, K. S. (2007). Monitoring in planta bacterial infection at both cellular and whole-plant levels using the green fluorescent protein variant GFPuv. *New Phytol.* 174, 212–223. doi: 10.1111/j.1469-8137.2007.01999.x

West, J. S., Canning, G. G., Perryman, S. A., and King, K. (2017). Novel technologies for the detection of *Fusarium* head blight disease and airborne inoculum. *Trop. Plant Pathol.* 42, 203–209. doi: 10.1007/s40858-017-0138-4

Xue, C., Liu, Z., Dai, L., Bu, J., Liu, M., Zhao, Z., et al. (2018). Changing host photosynthetic, carbohydrate, and energy metabolisms play important roles in *Phytoplasma* infection. *Phytopathology* 108 (9), 1067–1077. doi: 10.1094/PHYTO-02-18-0058-R

Yang, Y., Jiang, N., Lai, Y. T., Chang, Y. Y., Yang, X., Sun, H., et al. (2019). Green fluorescent probe for imaging His6-tagged proteins inside living cells. *ACS Sens.* 4, 1190–1196. doi: 10.1021/acssensors.8b01128

Zaynab, M., Fatima, M., Abbas, S., Sharif, Y., Umair, M., Zafar, M. H., et al. (2018). Role of secondary metabolites in plant defense against pathogens. *Microb. Pathog.* 124, 198–202. doi: 10.1016/j.micpath.2018.08.034

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

OPINION
published: 22 October 2019
doi: 10.3389/fpls.2019.01339

Effective Phenotyping Applications Require Matching Trait and Platform and More Attention to Theory

Victor O. Sadras [1,2]*

[1] South Australia Research and Development Institute, Adelaide, South Australia, Australia, [2] School of Agriculture, Food and Wine, The University of Adelaide, Australia

Keywords: Scaling, yield, density-dependence, phenotype, context-dependence

> *We're in a maze, not a highway; there is nowhere that speed alone can take us … I'm not being held back by the university's computers … I'm being held back by my own lack of insight into the problems I'm addressing…*
>
> **Julie Dehghani**

Edited by:
Michelle Watt,
Julich Research Centre,
Germany

Reviewed by:
Richard Richards,
Commonwealth Scientific and
Industrial Research Organisation,
Australia

***Correspondence:**
Victor O. Sadras
victor.sadras@sa.gov.au

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *15 May 2019*
Accepted: *25 September 2019*
Published: *22 October 2019*

Citation:
Sadras VO (2019) Effective
Phenotyping Applications Require
Matching Trait and Platform and
More Attention to Theory.
Front. Plant Sci. 10:1339.
doi: 10.3389/fpls.2019.01339

In the spring of 2018, the city of Adelaide hosted the 5th International Plant Phenotyping Symposium (IPPS), titled "From plant, to data, to impact", as well as the OECD-sponsored workshop "Making science useful to agriculture".[1] Here I share a personal account of salient findings of these two meetings, with a focus on more effective phenotyping, namely, phenotyping that is more likely to deliver tangible outcomes to plant breeding. To improve the effectiveness of the large phenotyping effort worldwide, I suggest asking two questions:

Are we really limited by data? Or better, to what extent we are limited by data, and to what extent we are, in the terms of Julie Dehghani, being held back by our own lack of insight into the problems we're addressing? I would argue our theories are lagging, and the theoretical model of the phenotype has become a bottleneck.

Can we improve the matching of trait and phenotyping platform? Individual plants and populations are fundamentally different biological entities; yield is a population-level attribute where plant-plant interactions are important. I would propose that explicit consideration of scaling, density- and context-dependence can help to better match agronomic traits with phenotyping platform and method, avoid expensive distractions and improve return on R&D investment.

The underlying assumption of 'omics' technologies is that large datasets on genes, their expression (transcriptomics) and products (proteomics and metabolomics) would resolve the complexities of key traits (Langridge, 2018). These 'omics' efforts have been primarily driven by technology—it was possible and therefore done—and largely failed to resolve the agronomically relevant phenotype, particularly crop yield (Langridge, 2018). This failure was ascribed to lagging phenotyping technologies (Cobb et al., 2013; Araus and Cairns, 2014), which in turn motivated the phenotyping effort displayed in the 5th IPPS. The IPPS showed a similar technology-driven approach that leads to large datasets of variable quality and relevance. Once again, we are assuming we are limited by data—but are we? Big data (and associated technologies) is the next promise, and it might deliver in some areas. However, we need hypothesis-driven science at the

---

[1] Papers and presentations from this meeting are at: https://msua.aweb.net.au/.

very least for proper experimental design, and importantly, to match trait and phenotyping platform.

The complete lack of attention to the theoretical aspects of the construction of the phenotype in the 5th IPPS was striking, and reinforces the notion of a technology-driven effort. The oversimplified model based on the unidirectional arrow from genotype to phenotype is unjustified but remains influential (Vinocur and Altman, 2005; Pickett, 2016). More nuanced theories of the phenotype need consideration (West-Eberhard, 2003; Piersma and van Gils, 2011; Noble, 2012; Félix, 2016). The metaphors that genes "control" development, and that genomes embody "programs" for development are particularly misplaced (Noble, 2012; Félix, 2016). A unified theory of phenotypic development and evolution emphasizes that "the individual's genotype can never be said to control development. Development depends at every step on the pre-existent structure of the phenotype, a structure that is complexly determined by a long history of both genomic and environmental influences" (West-Eberhard, 2003). This perspective is more broadly captured in the concept of downward causation (Noble, 2012; Flack, 2017; Noble, 2017). A panel discussion on what it takes for phenotypic data to be useful focused on technical aspects along the lines of the FAIR principles (Wilkinson et al., 2016). A lonely voice advanced the importance of asking questions that matter biologically and agronomically for data to be useful, but after a lukewarm collective nod the discussion reverted to technical aspects of data management. Future editions of this IPPS will benefit from sessions devoted to the theory of the phenotype.

The 5th IPPS, and the literature more broadly, show that the mismatch between trait and phenotyping approach is a major source of inefficiency, e.g. phenotyping for density-dependent traits such as yield or nutrient uptake in isolated plants in the glasshouse, or in single rows in the field. For different reasons, research in plant biology frequently simplifies to potted plants in chamber or glasshouse. If we are interested in a biological process of an isolated plant in an artificial condition lacking agronomic context, this simplification is fine and needs no justification. However, we must ask rigorous questions about scaling, density- and context-dependence if we want to achieve agronomic relevance.

How does a trait measured in an individual plant relate to the trait in a crop stand, where yield and other agronomic attributes are resolved? Sadras and Richards (2014) classified

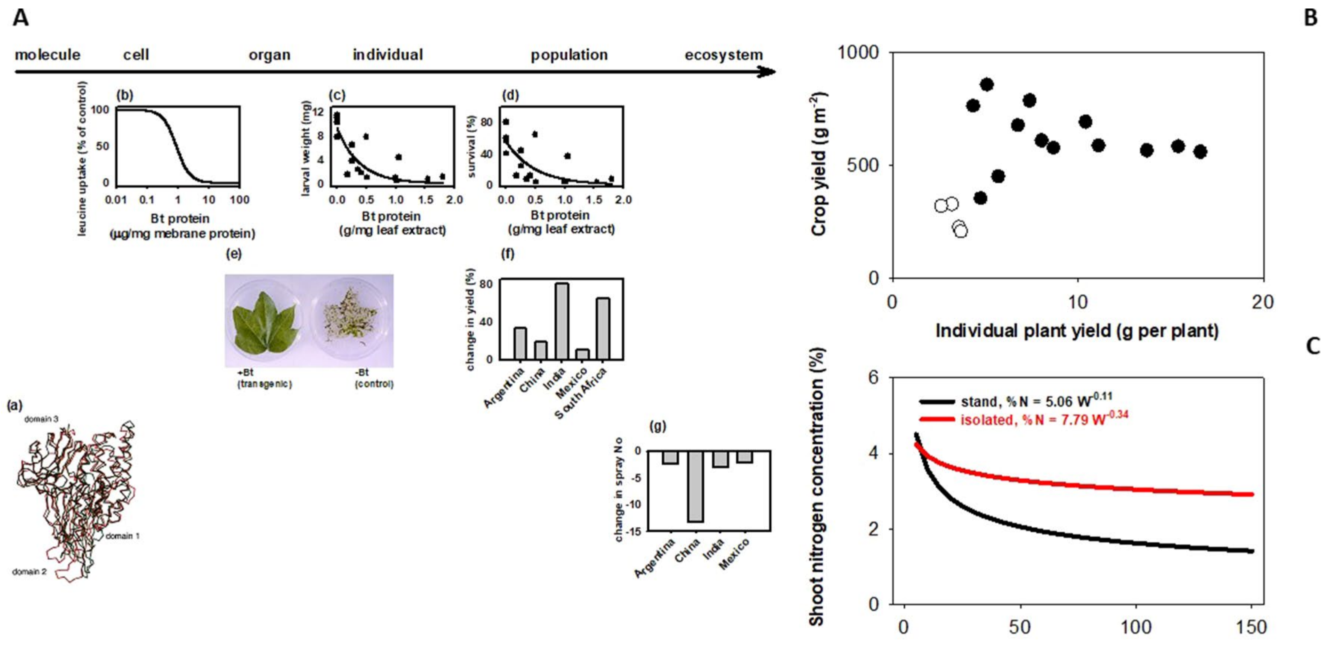


FIGURE 1 | Bt scales, but yield and plant allometry do not because they are density dependent. **(A)** Bt scales from molecular to ecosystem level. (a) *Bacillus turingensis* (Bt) Cry proteins have deleterious effects on Lepidoptera larvae at the (b) cell, (c) individual, and (d) population levels. Bt proteins expressed in transformed plant tissues confer protection at the (e) organ and (f) population level. This protection allows for (g) less reliance on synthetic insecticides for pest control, with consequences at the ecosystem level. Sources: (a) C^a traces of individual domains of CryIA(a) (red) and CryIIIA (black) proteins (Grochulski et al., 1995); (b) inhibitory effect of CryIA(a) on leucine uptake by membrane vesicles prepared from *Bombyx mori* larvae midgut; (c) weight and (d) survival of 7-day-old, 3rd instar larvae of *Elsmopalpus lignosellus* after a week of feeding treatments with peanut leaf containing variable concentration of CryIA(c) protein (Singsit et al., 1997); (f) comparison of cotton leaves expressing CryIA(c) proteins and non-transformed controls exposed to *Helicoverpa* spp larva (Lewis Wilson, unpublished); differences between Bt and conventional cotton in (g) lint yield. **(B)** Yield does not scale from plant to population. Yield of wheat stands (200–400 plants m^{-2}) does not correlate with yield of isolated plants (<20 plants m^{-2}) grown under the same conditions in the field. Source: Pedró et al. (2012). **(C)** Nitrogen-biomass allometry does not scale from plant to population. Owing to plant-plant interactions, dilution curves relating shoot nitrogen concentration and shoot dry matter show a sharp dilution for plants in stands, and a highly buffered dilution in isolated plants. Source: Lemaire and Millard (1999).

traits in three groups depending on their scaling. Some traits generally scale, like herbicide tolerance or Bt (**Figure 1A**). There are traits for which scaling is strictly dependent on experimental protocols; for example, leaf expansion rate seems to scale from glasshouse to field in well designed and carefully executed experiments (Reymond et al., 2003). There are traits that rarely scale, including yield (**Figure 1B**), photosynthesis, capture and efficiency in the use of water and nutrients, allometric relationships (**Figure 1C**), and architecture traits such as tillering, shoot and root branching (Sadras and Richards, 2014). Photosynthesis is strongly buffered from molecular to stand level (Pettigrew et al., 1989; Sinclair et al., 2004). Lack of both wind and canopy structure in plant-based settings (glasshouse, chamber) generates unrealistic canopy-atmosphere coupling, hence the difficulty in scaling gas exchange (Jarvis and McNaughton, 1986; Fereres et al., 2014). Artifacts from growing plants in pots are well established (Ben-Porath and Baker, 1990; Passioura, 2002; Passioura, 2006; Poorter et al., 2012; Ryan and Graham, 2018), and yet potted plants are used in phenotyping traits related to water and nutrient uptake and efficiency. Ryan and Graham (2018) critically revised the agronomic relevance of plant-mycorrhiza studies in controlled environments. Density-dependence (Donald, 1963; Harper, 1977) is a useful criterion to predict scaling; tillering responds to neighboring plants as related to both light quality and resources (Casal et al., 1986; Skinner and Simmons, 1993; Peltonen-Sainio and Järvinen, 1995; Robertson et al., 2009), hence is less likely to scale. Nitrogen-biomass allometry—the keystone for quantifying crop nitrogen status (Sadras and Lemaire, 2014; Gastal et al., 2015)—is strongly density dependent, and therefore does not scale from plant to crop (**Figure 1C**). Density-dependence leads to hierarchies in stands and is an artifact commonly overlooked in the glasshouse (Chen et al., 2018) and, to a lesser extent in the field (Rebetzke et al., 2014; Fischer and Rebetzke, 2018). For example, transformed tobacco plants with superior leaf-level photosynthesis were grown with a single-row buffer of shorter, untransformed wild-types leading to a gross misinterpretation of the effect of molecular manipulations on the crop phenotype (South et al., 2019). Gene expression is density-dependent in locust (Pener and Simpson, 2009), and in Arabidopsis (Geisler et al., 2012). In chickpea, Fst genome scan revealed a mismatch in the top genomic regions under selection for yield in border rows under relaxed competition and inner rows under full competition (Lake et al., 2016).

Consideration of biological context is critical for effective phenotyping. Environmental factors such as radiation, photoperiod and temperature are spatially and temporally correlated, and plants evolved receptors coupled with molecular signaling that extract information from these joint multivariate properties of the environment. Hence responses dependent on these properties of the environment are encoded in the genome, epigenome and phenome of plants (Karban, 2015). For this reason, phenotyping in an unnatural context, i.e. where correlations between environmental variables have been unrealistically altered, are often of little agronomic relevance because biased relationships among the states of different environmental variables disturb the information decoded by the plant, and hence the phenotype. For example, diurnal profiles of carbon and nitrogen metabolites of plants grown with a step-change in radiation do not match those for plants grown with both regular (day-night sinusoidal cycle) and irregular (due to clouds) fluctuations in radiation (Annunziata et al., 2017). Sowing date trials to screen for heat adaptation in the field bias the relative state of key environmental factors (radiation, photoperiod, temperature, vapor pressure deficit), and confound developmental and thermal-stress responses (Sadras et al., 2015).

The sequence "From plant, to data, to impact" resonates with both the directional "pipeline" in biotechnology, from lab to field (Nuccio et al., 2018) and with the simplistic genotype-to-phenotype model. These directional perspectives have delivered improvements in crop protection but have largely failed to improve yield and adaptation to drought (Dalal et al., 2017; Nuccio et al., 2018). Conceptually, Denis Noble concludes that there is no privileged level of causation (Noble, 2012). Renee Laffite and colleagues have advanced a robust approach that focuses on agronomically rigorous field phenotyping of grain yield and high-level secondary traits such as anthesis-silking interval, with complementary work under controlled conditions—working from field to lab ensures agronomic relevance (Habben et al., 2014; Shi et al., 2015; Brugiere et al., 2017; Lafitte et al., 2018).

To improve the effectiveness of the large phenotyping effort worldwide, I suggest asking two questions:

Are we really limited by data? Or better, to what extent we are limited by data, and to what extent we are, in the terms of Julie Dehghani, being held back by our own lack of insight into the problems we're addressing? I would argue our theories are lagging, and the theoretical model of the phenotype needs attention.

Can we improve the matching of trait and phenotyping platform? Individual plants and populations are fundamentally different biological entities; yield is a population attribute. I would propose that explicit consideration of scaling, density- and context-dependence can help to better match agronomic traits with phenotyping platform and method, avoid expensive distractions and improve return on R&D investment.

AUTHOR CONTRIBUTIONS

VS wrote the paper.

ACKNOWLEDGMENTS

The workshop "Making science useful to agriculture" was supported by OECD, Grains Research and Development Corporation, and South Australian Grains Industry Trust.

REFERENCES

Annunziata, M. G., Apelt, F., Carillo, P., Krause, U., Feil, R., Mengin, V., et al. (2017). Getting back to nature: a reality check for experiments in controlled environments. *J. Exp. Bot.* 68 (16), 4463–4477. doi: 10.1093/jxb/erx220

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19 (1), 52–61. doi: 10.1016/j.tplants.2013.09.008

Ben-Porath, A., and Baker, D. N. (1990). Tap-root restriction effects on growth, earliness, and dry weight partitioning of cotton. *Crop Sci.* 30, 809–814. doi: 10.2135/cropsci1990.0011183X003000040009x

Brugiere, N., Zhang, W. J., Xu, Q. Z., Scolaro, E. J., Lu, C., Kahsay, R. Y., et al. (2017). Overexpression of RING domain E3 ligase ZmXerico1 confers drought tolerance through regulation of ABA homeostasis. *Plant Physiol.* 175 (3), 1350–1369. doi: 10.1104/pp.17.01072

Casal, J. J., Sanchez, R. A., and Deregibus, V. A. (1986). The effect of plant-density on tillering - the involvement of r/fr ratio and the proportion of radiation intercepted per plant. *Environ. Exp. Bot.* 26 (4), 365–371. doi: 10.1016/0098-8472(86)90024-9

Chen, T.-W., Cabrera-Bosquet, L., Alvarez Prado, S., Perez, R., Artzet, S., Pradal, C., et al. (2018). Genetic and environmental dissection of biomass accumulation in multi-genotype maize canopies. *J. Exp. Bot.* 70, ery309–ery309. doi: 10.1093/jxb/ery309

Cobb, J. N., Declerck, G., Greenberg, A., Clark, R., and McCouch, S. (2013). Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* 126 (4), 867–887. doi: 10.1007/s00122-013-2066-0

Dalal, A., Attia, Z., and Moshelion, M. (2017). To produce or to survive: how plastic is your crop stress physiology? *Front. Plant Sci.* 8 (2067). doi: 10.3389/fpls.2017.02067

Donald, C. M. (1963). "Competition among crop and pasture plants," in *Advances in Agronomy*. Ed. A. G. Norman (Academic Press), 1–118. doi: 10.1016/S0065-2113(08)60397-1

Félix, M. A. (2016). Phenotypic evolution with and beyond genome evolution. *Curr. Topics Dev. Biol.* 119, 291–347. doi: 10.1016/bs.ctdb.2016.04.002

Fereres, E., Orgaz, F., Gonzalez-Dugo, V., Testi, L., and Villalobos, F. J. (2014). Balancing crop yield and water productivity tradeoffs in herbaceous and woody crops. *Funct. Plant Biol.* 41 (11), 1009–1018. doi: 10.1071/FP14042

Fischer, R. A., and Rebetzke, G. J. (2018). Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: a review. *Crop Pasture Sci.* 69 (5), 439–459. doi: 10.1071/CP17409

Flack, J. C. (2017). "Life's information hierarchy," in *From matter to life: information and casualty*. Eds. S. I. Walker, P. C. W. Davies, and G. F. R. Ellis (Cambridge: Cambridge University Press), 283–302. doi: 10.1017/9781316584200.012

Gastal, F., Lemaire, G., Durand, J. L., and Louarn, G. (2015). "Quantifying crop responses to nitrogen and avenues to improve nitrogen-use efficiency," in *Crop Physiology: Applications for genetic improvement and agronomy*. Eds. V. O. Sadras and D. F. Calderini. (San Diego: Academic Press), 161–206. doi: 10.1016/B978-0-12-417104-6.00008-X

Geisler, M., Gibson, D. J., Lindsey, K. J., Millar, K., and Wood, A. J. (2012). Upregulation of photosynthesis genes, and down-regulation of stress defense genes, is the response of Arabidopsis thaliana shoots to intraspecific competition. *Bot. Stud.* 53 (1), 85–96.

Grochulski, P., Masson, L., Borisova, S., Pusztai-Carey, M., Schwartz, J.-L., Brousseau, R., et al. (1995). Bacillus thuringiensisCrylA(a) insecticidal toxin: crystal structure and channel formation. *J. Mol. Biol.* 254 (3), 447–464. doi: 10.1006/jmbi.1995.0630

Habben, J. E., Bao, X. M., Bate, N. J., DeBruin, J. L., Dolan, D., Hasegawa, D., et al. (2014). Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. *Plant Biotechnol. J.* 12 (6), 685–693. doi: 10.1111/pbi.12172

Harper, J. L. (1977). *The population biology of plants*. London: Academic Press.

Jarvis, P. G., and McNaughton, K. (1986). Stomatal control of transpiration: scaling up from leaf to region. *Adv. Ecol. Res.* 15 (1), 49. doi: 10.1016/S0065-2504(08)60119-1

Karban, R. (2015). *Plant sensing and communication*. (Chicago: The University of Chicago Press). doi: 10.7208/chicago/9780226264844.001.0001

Lafitte, H. R., Habben, J. E., and Simmons, C. R. (2018). Searching for transgenes that improve yield: promise and reality. *Making Sci. Useful Agric. Adelaide Nov. 26-29 2018*; *https://msua.aweb.net.au/*.

Lake, L., Li, Y., Casal, J. J., and Sadras, V. O. (2016). Negative association between chickpea response to competition and crop yield: phenotypic and genetic analysis. *Field Crops Res.* 196, 409–417. doi: 10.1016/j.fcr.2016.07.021

Langridge, P. (2018). State-of-the-art in genetic resources. *Making Sci. Useful Agric. Adelaide Nov 26-29 2018*; *https://msua.aweb.net.au/*.

Lemaire, G., and Millard, P. (1999). An ecophysiological approach to modelling resource fluxes in competing plants. *J. Exp. Bot.* 50 (330), 15–28. doi: 10.1093/jexbot/50.330.15

Noble, A. D. (2017). "Digital and analogue information in organisms," in *From matter to life: information and casualty*. Eds. S. I. Walker, P. C. W. Davies, and G. F. R. Ellis Cambridge: (Cambridge University Press), 114–129. doi: 10.1017/9781316584200.006

Noble, D. (2012). A theory of biological relativity: no priviliged level of causation. *Interface Focus* 2, 55–64. doi: 10.1098/rsfs.2011.0067

Nuccio, M. L., Paul, M., Bate, N. J., Cohn, J., and Cutler, S. R. (2018). Where are the drought tolerant crops? An assessment of more than two decades of plant biotechnology effort in crop improvement. *Plant Sci.* 273, 110–119. doi: 10.1016/j.plantsci.2018.01.020

Passioura, J. B. (2002). Soil conditions and plant growth. *Plant Cell Environ.* 25 (2), 311–318. doi: 10.1046/j.0016-8025.2001.00802.x

Passioura, J. B. (2006). The perils of pot experiments. *Funct. Plant Biol.* 33 (12), 1075–1079. doi: 10.1071/FP06223

Pedró, A., Savin, R., and Slafer, G. A. (2012). Crop productivity as related to single-plant traits at key phenological stages in durum wheat. *Field Crops Res.* 138 (0), 42–51. doi: 10.1016/j.fcr.2012.09.016

Peltonen-Sainio, P., and Järvinen, P. (1995). Seeding rate effects on tillering, grain yield, and yield components of oat at high latitude. *Field Crops Res.* 40, 49–56. doi: 10.1016/0378-4290(94)00089-U

Pener, M. P., and Simpson, S. J. (2009). "Locust Phase Polyphenism: an update," in *Advances in Insect Physiology, Vol 36*. Eds. S. J. Simpson and J. Casas. (Academic Press), 1–272. doi: 10.1016/S0065-2806(08)36001-9

Pettigrew, W. T., Hesketh, J. D., Peters, D. B., and Woolley, J. T. (1989). Characterisation of canopy photosynthesis of chlorophyll-deficient soybean isolines. *Crop Sci.* 29, 1025–1029. doi: 10.2135/cropsci1989.0011183X002900040040x

Pickett, J. A. (2016). The essential need for GM crops. *Nat. Plants* 2 (6), 2. doi: 10.1038/nplants.2016.78

Piersma, T., and van Gils, J. A. (2011). *The flexible phenotype*. New York: Oxford University Press.

Poorter, H., Bühler, J., van Dusschoten, D., Climent, J., and Postma, J. A. (2012). Pot size matters: a meta-analysis of the effects of rooting volume on plant growth. *Funct. Plant Biol.* 39, 839–850. doi: 10.1071/FP12049

Rebetzke, G. J., Fischer, R. A., van Herwaarden, A. F., Bonnett, D. G., Chenu, K., Rattey, A. R., et al. (2014). Plot size matters: interference from intergenotypic competition in plant phenotyping studies. *Funct. Plant Biol.* 41, 107–118. doi: 10.1071/FP13177

Reymond, M., Muller, B., Leonardi, A., Charcosset, A., and Tardieu, F. (2003). Combining quantitative trait loci analysis and an ecophysiological model to analyze the genetic variability of the responses of maize leaf growth to temperature and water deficit. *Plant Physiol.* 131 (2), 664–675. doi: 10.1104/pp.013839

Robertson, D., Zhang, H., Palta, J. A., Colmer, T., and Turner, N. C. (2009). Waterlogging affects the growth, development of tillers, and yield of wheat through a severe, but transient, N deficiency. *Crop Pasture Sci.* 60 (6), 578–586. doi: 10.1071/CP08440

Ryan, M. H., and Graham, J. H. (2018). Little evidence that farmers should consider abundance or diversity of arbuscular mycorrhizal fungi when managing crops. *New Phytol.* 220, 1092–1107. doi: 10.1111/nph.15308

Sadras, V. O., and Lemaire, G. (2014). Quantifying crop nitrogen status for comparisons of agronomic practices and genotypes. *Field Crops Res.* 164, 54–64. doi: 10.1016/j.fcr.2014.05.006

Sadras, V. O., and Richards, R. A. (2014). Improvement of crop yield in dry environments: benchmarks, levels of organisation and the role of nitrogen. *J. Exp. Bot.* 65 (8), 1981–1995. doi: 10.1093/jxb/eru061

Sadras, V. O., Vadez, V., Purushothaman, R., Lake, L., and Marrou, H. (2015). Unscrambling confounded effects of sowing date trials to screen for crop adaptation to high temperature. *Field Crops Res.* 177, 1–8. doi: 10.1016/j.fcr.2015.02.024

Shi, J., Habben, J. E., Archibald, R. L., Drummond, B. J., Chamberlin, M. A., Williams, R. W., et al. (2015). Overexpression of ARGOS genes modifies plant sensitivity to ethylene, leading to improved drought tolerance in both *Arabidopsis* and maize. *Plant Physiol.* 169 (1), 266–26+. doi: 10.1104/pp.15.00780

Sinclair, T. R., Purcell, L. C., and Sneller, C. H. (2004). Crop transformation and the challenge to increase yield potential. *Trends Plant Sci.* 9 (2), 70–75. doi: 10.1016/j.tplants.2003.12.008

Singsit, C., Adang, M. J., Lynch, R. E., Anderson, W. F., Wang, A., Cardineau, G., et al. (1997). Expression of a Bacillus thuringiensis cryIA(c) gene in transgenic peanut plants and its efficacy against lesser cornstalk borer. *Transgenic Res.* 6 (2), 169–176. doi: 10.1023/A:1018481805928

Skinner, R., and Simmons, S. R. (1993). Modulation of leaf elongation, tiller appearance and tiller senescence in spring barley by far-red light. *Plant Cell Environ.* 16, 555–562. doi: 10.1111/j.1365-3040.1993.tb00903.x

South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 363, eaat9077. doi: 10.1126/science.aat9077

Vinocur, B., and Altman, A. (2005). Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr. Opin. Biotechnol.* 16 (2), 123–132. doi: 10.1016/j.copbio.2005.02.001

West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. New York: Oxford University Press.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Sci. Data* 3, 160018. doi: 10.1038/sdata.2016.18

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 30 October 2019
doi: 10.3389/fpls.2019.01380

The Development of Hyperspectral Distribution Maps to Predict the Content and Distribution of Nitrogen and Water in Wheat (*Triticum aestivum*)

Brooke Bruning[1]*, Huajian Liu[1], Chris Brien[1], Bettina Berger[1], Megan Lewis[2] and Trevor Garnett[1]

[1] Australian Plant Phenomics Facility, The Plant Accelerator, School of Agriculture, Food & Wine, University of Adelaide, Urrbrae, SA, Australia, [2] Ecology and Evolutionary Biology, School of Biological Sciences, University of Adelaide, Adelaide, SA, Australia

OPEN ACCESS

Edited by:
Matti Mõttus,
University of Helsinki, Finland

Reviewed by:
Martin Gnyp,
Yara International (Germany),
Germany
Lea Hallik,
University of Tartu, Estonia

****Correspondence:***
Brooke Bruning
brooke.bruning@adelaide.edu.au

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *31 May 2019*
Accepted: *07 October 2019*
Published: *30 October 2019*

Citation:
Bruning B, Liu H, Brien C, Berger B, Lewis M and Garnett T (2019) The Development of Hyperspectral Distribution Maps to Predict the Content and Distribution of Nitrogen and Water in Wheat (Triticum aestivum). Front. Plant Sci. 10:1380. doi: 10.3389/fpls.2019.01380

Quantifying plant water content and nitrogen levels and determining water and nitrogen phenotypes is important for crop management and achieving optimal yield and quality. Hyperspectral methods have the potential to advance high throughput phenotyping efforts by providing a rapid, accurate, and nondestructive alternative for estimating biochemical and physiological plant traits. Our study (i) acquired hyperspectral images of wheat plants using a high throughput phenotyping system, (ii) developed regression models capable of predicting water and nitrogen levels of wheat plants, and (iii) applied the regression coefficients from the best-performing models to hyperspectral images in order to develop prediction maps to visualize nitrogen and water distribution within plants. Hyperspectral images were collected of four wheat (*Triticum aestivum*) genotypes grown in nine soil nutrient conditions and under two water treatments. Five multivariate regression methods in combination with 10 spectral preprocessing techniques were employed to find a model with strong predictive performance. Visible and near infrared wavelengths (VNIR: 400–1,000nm) alone were not sufficient to accurately predict water and nitrogen content (validation $R^2 = 0.56$ and $R^2 = 0.59$, respectively) but model accuracy was improved when shortwave-infrared wavelengths (SWIR: 1,000–2,500nm) were incorporated (validation $R^2 = 0.63$ and $R^2 = 0.66$, respectively). Wavelength reduction produced equivalent model accuracies while reducing model size and complexity (validation $R^2 = 0.69$ and $R^2 = 0.66$ for water and nitrogen, respectively). Developed distribution maps provided a visual representation of the concentration and distribution of water within plants while nitrogen maps seemed to suffer from noise. The findings and methods from this study demonstrate the high potential of high-throughput hyperspectral imagery for estimating and visualizing the distribution of plant chemical properties.

Keywords: nitrogen, water, hyperspectral, wheat, PLSR, plant phenotyping

INTRODUCTION

Wheat (*Triticum aestivum*) is the major winter crop in Australia and sustainable improvement of yields is a major research focus. The availability of nitrogen and water are widely recognized as two of the main factors limiting crop growth and production (Garnett and Rebetzke, 2013). Nitrogen is essential for crops but nitrogen use efficiency is generally low (Raun and Johnson, 1999). Understanding the nitrogen dynamics within the plant is key to improving fertilization practices and breeding more efficient crops (Garnett et al., 2015) which in turn may help reach the increased yields required for a growing population. Water also profoundly influences plant health and potential yield. Water is a strong driver of photosynthesis, respiration, absorption, and the translocation of nutrients and metabolites throughout the plant (Lambers et al., 2008). The accurate assessment of water content also has importance for fertilization, irrigation practices, and drought assessment (Peñuelas et al., 1997; Torres et al., 2019).

High throughput plant phenotyping is an emerging approach for plant breeding and crop improvement studies (Awada et al., 2018; Hansen et al., 2018). Image-based phenotyping offers nondestructive techniques which can significantly reduce the cost, time, and labor involved in larger-scale screening trials. Collecting information from multiple sensors allows near-simultaneous data collection for the measurement of many plant traits (Li et al., 2014; Humplik et al., 2015). Since such methods are nondestructive, they are repeatable across a plant's lifecycle, thereby allowing for changes to be detected over time which otherwise could not be determined with traditional destructive analyses (Berger et al., 2012; Fahlgren et al., 2015).

Hyperspectral cameras are becoming more common in the plant research environment. Hyperspectral imaging combines the benefits of both spectroscopy and traditional imaging; it is able to quantify light reflectance across hundreds of narrow spectral bands for distinct spatial pixels (ElMasry et al., 2012; Liu et al., 2014). Hyperspectral imaging has successfully been used in the prediction of many plant traits such as early drought stress in barley (Behmann et al., 2014), macronutrient content and distribution in oilseed rape (Zhang et al., 2013), and nitrogen distribution in cucumber leaves (Yu et al., 2014). Chemical properties, including water content, micronutrient, and macronutrient concentrations, have also been quantified in maize and soybean plants using hyperspectral imaging in a high throughput phenotyping greenhouse (Ge et al., 2016; Pandey et al., 2017).

Multivariate analysis techniques are required to relate the spectral information gained from hyperspectral instruments to the chemical or physical traits of interest. Many different multivariate methods exist for model development but these techniques require a measured reference sample to use in the development of calibration models. Although ideal, this is impossible to achieve at a pixel-scale resolution; wet chemistry reference methods often require minimum amounts of tissue from an entire leaf or plant. Therefore, regression is often performed using mean spectra extracted from the entire plant or a region thereof (Ge et al., 2016; Pandey et al., 2017). Regression coefficient vectors based on these mean spectra can then be applied to the image at the individual pixel scale to make predictions at a resolution equivalent to that of the images acquired. These developed prediction maps, or distribution maps, provide a visual interpretation of the content and spatial variation of the predicted component which otherwise cannot be visualized by the hyperspectral data alone.

The present study focused on wheat, the major winter crop in Australia. The specific objectives were (i) to acquire hyperspectral images of wheat plants using a high throughput phenotyping system, (ii) develop regression models capable of predicting water and nitrogen levels of wheat plants, and (iii) apply the regression coefficients from the best-performing models to hyperspectral images in order to develop prediction maps to visualize nitrogen and water distribution within plants.

MATERIALS AND METHODS

The experiment was carried out in an automated phenotyping platform (LemnaTec GmbH, Aachen, Germany) at The Plant Accelerator (Australian Plant Phenomics Facility, University of Adelaide, Adelaide, Australia; longitude: 138.64, latitude: -34.97). The platform houses a hyperspectral imaging chamber (WIWAM, Ghent, Netherlands) which contains two individual cameras, a Specim FX10 (Specim, Oulu, Finland) operating in the VNIR (visible and near infrared: 400–1,000 nm) range and Specim SWIR (Specim, Oulu, Finland) operating at the longer SWIR (shortwave infrared: 1,300–2,500 nm) wavelengths. The VNIR FX10 has spatial sampling of 1024 pixels and a spectral interval of approximately 1.3 nm, capturing a total of 448 individual spectral measurements for each image acquisition. The SWIR camera has spatial sampling of 640 pixels and a spectral interval of 5.7 nm, capturing 288 bands. The hyperspectral imaging chamber is illuminated by 18 halogen lights to ensure a consistent light source across the wavelengths.

Experimental Design

Four soil nutrient factors at two levels each and two watering treatments were applied to four varieties of wheat- cv Gladius, Kukri, Mace and RAC875. The nutrient factors were nitrogen (N:25, 100 mg/kg), phosphorous (P:15, 40 mg/kg), potassium (K:20, 60 mg/kg), and Micromax (5, 10 g/150 g). Half of the full 2^4(= 16) combinations of the soil nutrient treatments, in addition to a control treatment where no nutrients were added, were included in the design, resulting in nine different soil nutrients treatments (**Table 1**). The base soil was a 1:1:1 mixture of UC (University of California)-mix:coco-peat:clay-loam without any nutrients but balanced to pH 6.4 using dolomite lime. The base soil was divided and different levels of nutrients were added and mixed by hand. Nitrogen was added as Polyon urea (Polyon, 42% N, Koch, Melbourne, Australia), phosphorous as Superphosphate (20.1% P, Incitec Pivot Ltd., Melbourne, Australia), potassium as Potash sulphate (Greenskote, 41% K), and secondary nutrients as Micromax (Scotts Micromax micronutrients; Scotts-Sierra Horticultural Products Co., Marysville, Ohio).

TABLE 1 | Nutrient levels of the different soils used in this study. Half of the full 2^4(= 16) combinations of the soil nutrient treatments were used in addition to a control treatment, where no nutrients were added.

Soil	N (mg/kg)	P (mg/kg)	K (mg/kg)	secondary nutrients (g/150kg)
1	0	0	0	0
2	25	15	20	5
3	100	15	20	10
4	25	40	20	10
5	100	40	20	5
6	25	15	60	10
7	100	15	60	5
8	25	40	60	5
9	100	40	60	10

For the drought treatment, pots were dried down to 10% (g/g) gravimetric water content 23 days after sowing (DAS) to ensure initial plant establishment in well-watered conditions. The well-watered pots were watered to 20% (g/g) water content which was then increased to 23% (g/g) from 34 DAS as growth increased and additional water reserves were required. Each soil, variety and drought combination was replicated six times resulting in a total of 432 pots. Plants were grown in 150 mm pots containing 2.5 kg of dry soil. The greenhouse maintained an average daytime temperature of 23.8°C and average night temperature of 17.5°C.

When first sown, pots were kept on greenhouse benches and hand watered daily until plant emergence. The pots were transferred to the conveyor system 17 DAS where the daily watering became automated. Pots were arranged in a criss-cross design with split-plots, randomized using "dae" (Brien, 2017), a package for the R statistical computing environment (R Core Team, 2017). The pots occupied 24 lanes divided into six zones, each containing a single replicate of allocated factors. The four varieties were randomized to the four lanes in each zone. The layout was split into two sides (west and east) and the drought treatments randomized to the combination of zones by sides using three 2x2 Latin squares. The nine soil nutrient treatments were assigned to the nine carts with a lane-side combination using an 8x9 Youden square.

Data Collection

Hyperspectral System and Data Measurements

Hyperspectral images were collected weekly from 31–61 DAS from a position above the plant. The data collected for each plant consisted of a raw hyperspectral datacube, a white reference and dark reference image. MATLAB (2017b, The MathWorks, Natick, MA) was used to write a function to convert the raw and reference image data into calibrated 3D datacubes of the plants (Equation 1):

$$I_{\text{calibrated}} = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{white}} - I_{\text{dark}}} \tag{1}$$

where $I_{\text{calibrated}}$ is the calibrated datacube image, I_{raw} is the raw, unprocessed datacube, I_{dark} is the dark current image, and I_{white} is the white reference image.

Plant Sampling and Chemical Analysis

At the end of the experiment (61 DAS), the flag leaf from the main tiller and the remainder of the plant were harvested separately. The plant (minus a flag leaf) was immediately weighed to measure plant fresh weight. Samples were then placed in an oven at 60°C for 72 hours until a constant weight had been achieved and were then reweighed to obtain dry weight. Plant water content was calculated by (Equation 2):

$$water\ content = \frac{W_F - W_D}{W_F} \text{ x } 100\% \tag{2}$$

where W_F is the fresh weight of the harvested sample and W_D is the dried sample weight.

The flag leaf samples were reserved for nitrogen analysis. Nitrogen was measured using a "rapid N exceed" N analyzer (Elementar Analysensysteme GmbH, Langenselbold, Germany) with the Dumas combustion method. Samples were dried using the same methods as the whole plant samples and were then ground using a Geno/Grinder (SPEX SamplePrep, NJ, USA). Although different nutrient factors of N, P, and K were selected in the experimental design, only nitrogen values were predicted; there was insufficient flag-leaf tissue to also allow for the analysis of phosphorous and potassium. Nitrogen was therefore selected due to its importance to crop health.

Data Analysis

Extraction of Mean Plant Spectra

In order to extract only the spectral information corresponding to the plant shoot, the background pixels of the images were identified and excluded. This was achieved in the VNIR (400–1,000 nm) images by establishing an enhanced vegetation index (EVI) threshold mask in order to segment the plant tissue from other pixels (Huete et al., 2002). The band at 670 nm was assigned as the red band, 800 nm as the NIR band, and 470 nm as the blue band. These bands were extracted from the datacubes and used to calculate the EVI output for each image (Equation 3):

$$EVI = \frac{2.5\ R_{\text{NIR}} - R_{\text{R}}}{(R_{\text{NIR}} + 6R_{\text{R}} - 7.5R_{\text{B}}) + 1} \tag{3}$$

where R_{NIR} is the reflectance value in the near infrared band, R_{R} is the reflectance value in the red band, and R_{B} is the reflectance value in the blue band. The EVI output image was found to be very effective at identifying and segmenting plant pixels from background pixels when a threshold of 0.25 was applied. A binary mask based on the EVI values above 0.25 was therefore built and applied to the calibrated images. The reflectance values of all pixels identified as vegetation were then averaged to obtain the average reflectance spectrum of each plant.

For the SWIR images (1,000–2,500 nm), a different segmentation method was adopted (Liu et al., 2019). The SWIR data was first transformed from the original space of the hypercube to a hyper-hue space. A SVM model was then trained with a radial basis function kernel using the svm.OneClassSVM function in the Python programming language (Python Software Foundation) with the sklearn toolbox (Pedregosa et al., 2011) and optimal parameter tuning. The extracted mean spectra from both the VNIR and SWIR cameras were then combined for each plant.

Multivariate Regression

The average plant reflectance spectrum extracted from each image was used to develop models capable of predicting water content (%) and nitrogen content (%) in wheat plants. Spectral pre-processing techniques were used to remove noise, transform spectra, emphasis features, and to extract useful information in order to develop multivariate prediction models. The preprocessing techniques selected were adapted from those of Dotto et al. (2018) and Gholizadeh et al. (2015) and included the first-order Savitzky-Golay derivative (see further detail below) on reflectance spectra (SGD1), second-order Savitzky-Golay derivate on reflectance spectra (SGD2), the first-order Savitzky-Golay derivative of the absorbance-transformed spectra (ASGD1), the second-order Savitzky-Golay derivative of the absorbance-transformed spectra (ASGD2), multiplicative scatter correction (MSC), extended multiplicative scatter correction (EMSC), normalization by range (NBR), standard normal variate spectra (SNV), and smoothed spectra (SMO). In addition, the raw spectra were also used as model input to evaluate whether preprocessing actually improved the regression results.

Savitzky-Golay (SG) derivatives are used to reduce baseline shifts and linear trends: the first derivative removes baseline drifts whereas second derivatives can remove both baseline and linear effects (Martens and Næs, 2011). SG derivatives perform least squares linear regression fits of a polynomial around each point in the spectrum to smooth the data. They involve a spectral smoothing method prior to derivation to reduce the signal-to-noise ratio and to determine how many adjacent variables will be used for their calculation. Consideration is required when selecting the tuning parameters for derivatives: the polynomial order, window size, and order of differentiation can strongly influence resulting spectra (Zimmermann and Kohler, 2013). The first-order derivative was calculated using Savitzky-Golay filtering in the "prospectr" package (Stevens and Ramirez-Lopez, 2015) in the R open-source statistical environment (R Core Team, 2007). The SGD1 treatment was fitted with a first order differentiation, second-order polynomial, and a window size of 11. Similarly, the SGD2 treatment was fitted with a second-order polynomial, a window size of 11 but with a second order differentiation. For the ASGD1 treated data, the raw reflectance was first converted to absorbance values before a first-order Savitzky-Golay derivative was applied with a second-order polynomial and a window size of 11 nm.

Multiplicative scatter correction (MSC) involves regressing each spectrum in a dataset against a reference spectrum (quite commonly, the mean spectrum) in order to estimate the intercept and slope of the equation representing the scattering component (Geladi et al., 1985). Each spectrum in the dataset is corrected by subtracting the intercept and dividing by the slope. Extended multiplicative scatter correction (EMSC) is an improvement to MSC that allows the physical light scattering effects to be separated from chemical light effects in spectra (Martens and Stark, 1991).

Standard normal variate (SNV) removes scatter effects from spectral data by performing a row-oriented transformation which centers and scales each spectrum (Barnes et al., 1989). Performing SNV will produce similar results to MSC. The main difference is that SNV does not use the mean spectrum for standardisation but relies only on the data of each individual spectrum. Normalization by range (NBR) is a simple normalisation which adjusts values that are measured on different scales. Neither SNV nor normalization involve least squares fitting and are therefore quite sensitive to spectral noise (Rinnan et al., 2009). The MSC treatment was applied using the "pls" (Mevik and Wehrens, 2007) package in the R open-source statistical environment (R Core Team, 2007). The CRR and SNV treatments were applied using the "prospectr" package (Stevens and Ramirez-Lopez, 2015), the EMSC transformation was applied through the "EMSC" package (Liland, 2017) using a 6-degree polynomial, and the NBR was applied using the "clusterSim" package (Walesiak and Dudek, 2017).

The five different multivariate regression methods applied in this study were partial least square regression (PLSR), Principal Component Regression (PCR), Multiple Linear Regression (MLR), Support Vector Machines (SVM), and Random Forest (RF). PLSR is a multivariate calibration method that uses data compression in order to reduce the full spectrum into a smaller number of noncorrelated components while maintaining the majority of the information contained in the data (Axelsson et al., 2013). PLSR is a popular and widely-used regression method because it performs well when variables contain high correlation or colinearity, as is the case with hyperspectral data, due to its ability to minimise redundancy and overfitting of models (Wold et al., 2001). PLSR was applied in the "pls" package (ncomp = 10, validation = "CV"). The optimal number of components to include in the model was determined by visual inspection of the root mean square error of prediction (RMSEP) graph; the number of components which gave the lowest RMSEP was selected.

PCR is a technique very similar to PLSR in that it is able to model variables when there are a large number of highly correlated predictors present (Wold et al., 1984). Furthermore, both techniques construct new predictor variables which are linear combinations of the original predictor variables. As with the PLSR, PCR was applied in the "pls" package (ncomp = 10, validation = "CV"). The optimal number of components to include in the model was determined by the number of components which gave the lowest RMSEP.

MLR is one of the most common forms of linear regression analysis which has had some success with hyperspectral data (Pan et al., 2016). However, it may not be particularly applicable to all high-dimensional hyperspectral data as many of the data assumptions are not met. For example, MLR assumes that a linear relationship between the independent and dependent variables exists as well as that variables are normally distributed (Montgomery et al., 2012; Osborne and Waters, 2002). The "MASS" package (Venables and Ripley, 2002) was used to fit a linear model by ridge regression (lambda = 0.1-10 by 0.1, repl = 10).

SVM is a group of supervised learning methods originally developed for classification and has recently been adapted for regression models. SVM create hyperplanes that maximize the margins between different classes by reducing the cost function and therefore enabling high prediction performance (Do et al., 2012; Karatzoglou et al., 2005). SVM models are able to fit both linear and nonlinear relationships between variables and are able to handle large datasets (Dotto et al., 2018). Support vector machine (type = "eps", kernel = "radial", cost = 50) was employed using the "e1071" package (Meyer et al., 2017).

RF is a machine learning technique that enhances the performance of a single decision tree by averaging the predictions from multiple trees, each of which is generated from a random selection of the input variables (Belgiu and Dragut, 2016). RF is flexible with small or large datasets but their validation performance is usually poor compared to their calibration performance. The results from a RF model can also be difficult to interpret so their implementation has not been widespread in past spectroscopic vegetation studies. The "randomForest" package was used to implement a RF algorithm (ntree = 1000, type = "regression") (Liaw and Wiener, 2002).

The dataset of mean spectra was split based on experimental replicates into a training set (n = 236 for nitrogen and n = 250 for water) and a validation (n = 109 for nitrogen and n = 105 for water) set to allow both the development and independent validation of the models. Only the training data set was used in the development of the nutrient prediction models; the validation set was used as an independent test set. The total number of samples used in model development deviated from the total number of pots in the experiment (n = 432) due to technical issues during image acquisition. The number of samples in the calibration sets varied between the water and nitrogen regressions because plants grown in soil 1 (no nutrients) did not produce a flag leaf and therefore were not sampled for nitrogen analysis. The accuracy of the developed models was assessed using the independent testing set excluded from the model calibration. Three statistical measures were calculated to evaluate the models developed by each of the multivariate methods: the coefficient of determination (R^2), root mean square error (RMSE), and the ratio of performance to deviation (RPD) (Equations 4–6).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2}{n}} \tag{4}$$

$$R^2 = \frac{\Sigma \left(\hat{y}_i - \bar{y}\right)^2}{\Sigma \left(y_i - \bar{y}\right)^2} \tag{5}$$

$$RPD = \frac{SD}{RMSE} \tag{6}$$

where $\hat{y}_i$ are the predicted values, $\overline{y_i}$ is the mean of the observed value, y is the observed value, n is the number of samples in the validation or calibration set, and SD is the standard deviation of the reference values.

Since VNIR cameras are more affordable than SWIR cameras, we first tested to see whether models based on VNIR wavelengths alone were able to predict nitrogen and water content in wheat. Full-spectra models incorporating both VNIR and SWIR wavelengths were then trialled to see if models were improved. While adding SWIR data increases the number of variables, it may contribute noise to the original VNIR models rather than adding useful information. SWIR wavelengths alone were not exclusively trialled because of the camera's coarser spatial resolution. The larger pixel size of the SWIR images would not allow subtle spatial variations to be visualized in the distribution maps.

Wavelength Selection

Wavelength selection is widely performed with hyperspectral data because a large number of wavelengths are often redundant and contribute to noise. In this study, wavelength selection was approached in two ways: firstly, using regression coefficients from the full-spectra models and secondly, using only wavelengths located at known-absorption features for both water and nitrogen. For the regression coefficient approach, the vector of regression coefficients (β), a measure of association between each wavelength and the response, was output from the original full-spectra PLSR models. Wavelengths with small absolute regression coefficients, and therefore low association to water or nitrogen, were removed. Both the nitrogen and water datasets were reduced to 132 wavelengths based on the top 30% of regression coefficient values.

A second method, a feature-selection method, was also used for wavelength selection. This is a more traditional approach in which only the wavelengths in a particular region of the spectrum known to be associated with responses (i.e., water or nitrogen content) are selected. Wavelengths known to be related to either water and nitrogen content of vegetation were identified from previous studies (**Tables 2** and **3**). A 40 nm range centered on each of these wavelengths was then included in the models to ensure that the identified wavelengths were encompassed.

RESULTS

Prior to destructive harvest of the wheat plants grown in varying nutrient and water regimes, hyperspectral images were acquired in the VNIR and SWIR regions. For each corrected hyperspectral image, the pixels corresponding to vegetation were identified and the mean reflectance spectra of those pixels were calculated to develop prediction models for both water content and nitrogen levels. The mean spectra exhibited the typical reflectance properties of vegetation with low reflectance across visible wavelengths, a dramatic increase in reflectance at the transition from visible to near infrared wavelengths, and maximum reflectance values throughout the near infrared domain (**Figure 1**). An obvious feature of the spectral graph was the offset at 1,000 nm. This was caused by the two separate cameras: the FX10 operating at 400–1,000 nm and the SWIR camera operating at 1,000–2,500 nm. The two cameras, while focally and geometrically aligned, had different optical systems and spectral and spatial resolutions, which attribute to the offset between cameras. While overlapping regions common to both cameras were removed (996–1,006 nm), appropriate

TABLE 2 | Wavelengths known to be associated with water content in vegetation. A broad range (40 nm) centered on these values were included in the feature-reduced models to ensure that each feature was captured.

Wavelength (nm)	Range Included (nm)	Assignation	Reference
600	580–620	O-H Hydrogen Bonding	Hunt and Rock, 1989
680	660–700	Electron transition	Hong et al., 2017
810	790–830	C-H	Hong et al., 2017
820	800–840	C-H	Hunt and Rock, 1989
860	840–880	C-H	Gao, 1996; Eitel et al., 2006
900	880–920	C-H	Peñuelas et al., 1993; Peñuelas et al., 1995
970	950–990	O-H bend	Curran, 1989; Peñuelas et al., 1993; Peñuelas et al., 1995
1240	1,220–1,260	C-H	Gao, 1996; Eitel et al., 2006
1530	1,510–1,550	N-H secondary amines	Foutry and Baret, 1997; Curran, 1989
1550	1,530–1,570	N-H secondary amines	Musick and Pelletier 1986; Musick and Pelletier, 1988
1720	1,700–1,740	C-H	Foutry and Baret, 1997
1750	1,730–1,770	C-H secondary overtones	Musick and Pelletier 1986; Musick and Pelletier, 1988
2080	2,060–2,100	N-H and C-H, O-H stretch and deformation	Musick and Pelletier 1986; Musick and Pelletier, 1988; Curran, 1989
2350	2,330–2,370	C-H combinations	Musick and Pelletier 1986; Musick and Pelletier, 1988; Curran, 1989
1400-1450	1,400–1,450	O-H bend and stretch	Curran, 1989; De Bei et al., 2011

TABLE 3 | From Ecarnot et al. (2013). Wavelengths known to be associated with nitrogen content in vegetation. A broad range (40 nm) centered on these values were included in the feature-reduced models to ensure that each feature was captured.

Wavelength (nm)	Range Included (nm)	Assignation
460	440–480	Electron transition, chlorophyll a,b
530	510–550	Electron transition, carotenoids
670	650–690	Electron transition, chlorophyll a,b
1440	1,420–1,460	O-H bend, first overtone, starch
1500	1,480–1,520	N-H stretch
1680	1,660–1,700	C-H stretch, aromatic
1712	1,692–1,732	C-H stretch, CH3
1770	1,750–1,790	C-H stretch, CH2
1900	1,880–1,920	O-H stretch, C = O, starch, CO2H
1960	1,940–1,980	N-H, CONH2
2080	2,060–2,100	N-H stretch, proteins
2115	2,095–2,135	N-H stretch, CONH2, CONHR
2140	2,120–2,160	Amide, proteins
2230	2,210–2,250	N-H stretch, C = H stretch, amino acid
2300	2,280–2,320	N-H stretch, C = O stretch, amino acid
2400	2,380–2,420	CH2 bend, C-H deformation, cellulose

preprocessing should be applied in order to correct the spectral jump in future studies. Baseline shifts across the spectra were also apparent which suggests that a preprocessing technique which is able to correct for additive scatter, such as EMSC or MSC, may improve subsequent regression results.

VNIR Wavelengths Alone Are Not Sufficient to Predict Water Content

Initially, only VNIR (400–1,000 nm) wavelengths were used as input into the regression models. The different multivariate regression techniques and preprocessing methods resulted in various predictive performances for quantifying water content. Only three preprocessing methods (ASGD1, SMO, and raw) achieved R^2 values above 0.50 (**Table 4**). Of the preprocessing methods, the smoothed and raw data consistently achieved the highest overall prediction performance with validation R^2 values of up to 0.62 and 0.57 for smoothed and raw data, respectively (when used in combination with multiple linear regression). For the regression methods, PLSR models produced the highest accuracy performance for the prediction of water content in fresh wheat leaves; they consistently returned models with higher R^2 values compared to other techniques with relatively high RPD values and low RMSE values. PLSR validation R^2 values varied from 0.33 to 0.56, RMSE varied from 3.61–2.85, and RPD varied from 1.18–1.5 (**Supplementary Table 1**). PLSR achieved the best results when used in conjunction with raw spectra, achieving validation statistics of $R^2 = 0.56$, RMSE = 2.85, and RPD = 1.50. However, even this model does not provide a sufficient level of accuracy for predicting water content. According to the model classification of Li et al. (2018), these models can only be considered as “acceptable” for the prediction of water content based on their R^2 and RPD values ($0.50 \leq R^2 \leq 0.75$ and $1.40 \leq \text{RPD} \leq 2.00$).

VNIR Wavelengths Are Not Sufficient for Nitrogen Prediction

As with water prediction, PLSR and MLR showed the strongest prediction of nitrogen (**Table 5**). The calibration R^2 values ranged from 0.23 to 0.90 with the calibration RMSE and RPD values varying from 0.55 to 0.22 and 1.13 to 2.84, respectively. On the other hand, the validation R^2 values ranged from 0.06 to 0.59 with the validation RMSE and RPD values varying from 0.97 to 0.41 and 0.66 to 1.56, respectively (**Supplementary Table 2**). As with the water models, the smoothed and raw data achieved the highest overall prediction performance for estimating nitrogen, considering both calibration and validation data. PLSR models also produced the highest accuracy performance for nitrogen prediction compared to other multivariate methods trialled.

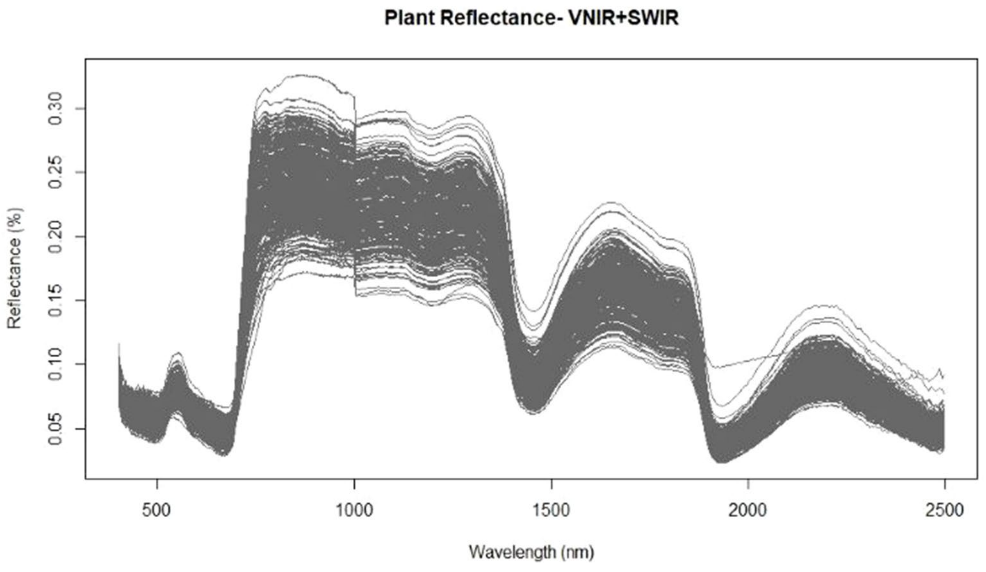


FIGURE 1 | Mean spectra extracted from the vegetation pixels for each image using the FX10 visible and near infrared wavelength (VNIR) camera for wavelengths from 400–1,000nm and the shortwave-infrared wavelength (SWIR) camera for the 1,000–2,500nm wavelengths.

TABLE 4 | Performance results for the prediction of water content in wheat using VNIR (400–1,000 nm) spectra.

Calibration						
		PLSR	PCR	MLR	RF	SVM
ASGD1	R^2	0.63	0.42	0.50	0.47	0.84
	RMSE	2.92	3.61	3.42	3.46	1.96
	RPD	1.63	1.32	1.39	1.38	2.43
SMO	R^2	0.64	0.60	0.61	0.44	0.67
	RMSE	2.90	3.04	2.96	3.57	2.76
	RPD	1.64	1.57	1.61	1.33	1.73
Raw	R^2	0.67	0.60	0.57	0.44	0.69
	RMSE	2.79	3.04	3.13	3.55	2.70
	RPD	1.71	1.56	1.52	1.34	1.77
Validation						
		PLSR	PCR	MLR	RF	SVM
ASGD1	R^2	0.51	0.41	0.54	0.49	0.52
	RMSE	3.01	3.30	2.94	3.06	2.96
	RPD	1.42	1.29	1.45	1.39	1.44
SMO	R^2	0.56	0.54	0.62	0.47	0.55
	RMSE	2.90	2.88	2.70	3.11	2.89
	RPD	1.47	1.48	1.58	1.37	1.48
Raw	R^2	0.56	0.54	0.57	0.49	0.56
	RMSE	2.85	2.88	2.82	3.05	2.86
	RPD	1.50	1.48	1.51	1.40	1.49

R^2, coefficient of determination, RMSE, root mean square error; RPD, ratio of performance to deviation. PLSR, partial least-squares regression; PCR, principal components regression; MLR, multiple linear regression; RF, random forest; SVM, support vector machine; SMO = smoothed.

TABLE 5 | Performance results (validation $R^2 \geq 0.5$) of trialled preprocessing and multivariate methods for the prediction of nitrogen in wheat.

Calibration						
		PLSR	PCR	MLR	RF	SVM
SMO	R^2	0.56	0.50	0.53	0.42	0.60
	RMSE	0.42	0.45	0.43	0.48	0.40
	RPD	1.49	1.40	1.46	1.31	1.58
Raw	R^2	0.59	0.49	0.48	0.42	0.62
	RMSE	0.41	0.45	0.45	0.47	0.39
	RPD	1.53	1.38	1.39	1.32	1.60
Validation						
		PLSR	PCR	MLR	RF	SVM
SMO	R^2	0.59	0.54	0.57	0.33	0.43
	RMSE	0.41	0.44	0.42	0.52	0.48
	RPD	1.56	1.47	1.53	1.22	1.33
Raw	R^2	0.57	0.52	0.58	0.33	0.43
	RMSE	0.42	0.44	0.42	0.52	0.48
	RPD	1.53	1.44	1.54	1.23	1.33

R^2 = coefficient of determination, RMSE; root mean square error; RPD; ratio of performance to deviation. PLSR; partial least-squares regression; PCR; principal components regression; MLR; multiple linear regression; RF; random forest; SVM; support vector machine; SMO; smoothed.

For different preprocessing methods, validation R^2 values varied from 0.26 to 0.59, RMSE varied from 0.55–0.41 and RPD varied from 1.16–1.56 (**Supplementary Table 2**). PLSR achieved the best results when used in conjunction with smoothed spectra, achieving validation statistics of $R^2 = 0.59$, RMSE = 0.41, and RPD = 1.56 (**Table 5**). However, no model developed with VNIR wavelengths alone can be considered as accurate for predicting nitrogen content. As with the VNIR water content models, these regressions are only considered as "acceptable" ($0.50 \leq R^2 \leq 0.75$ and $1.40 \leq \text{RPD} \leq 2.00$) (Li et al., 2018).

Full-Spectra (VNIR+SWIR) Regressions Improve Accuracies for Water and Nitrogen

Since both water and nitrogen also express strongly in the SWIR region, SWIR wavelengths were also included in the regression models to see whether prediction accuracies were improved compared to VNIR models alone. The incorporation of the SWIR wavelengths in the full-spectra models improved the prediction accuracies of both the water and nitrogen models (**Table 6**).

TABLE 6 | Validation prediction accuracies for full-spectra (VNIR+SWIR: 400-2500nm) regression models for predicting water content and nitrogen.

Water full-spectra (VNIR+SWIR) validation

		PLSR	**PCR**	**MLR**	**RF**	**SVM**
ASGD1	R^2	0.56	0.59	0.56	**0.60**	**0.61**
	RMSE	3.11	2.95	3.25	**2.92**	**2.88**
	RPD	1.48	1.56	1.41	**1.58**	**1.60**
ASGD2	R^2	0.58	0.50	0.34	0.59	0.55
	RMSE	3.07	3.24	5.22	3.00	3.10
	RPD	1.50	1.42	0.88	1.53	1.48
EMSC	R^2	0.57	0.59	0.54	0.55	0.59
	RMSE	3.10	3.01	3.27	3.08	2.94
	RPD	1.49	1.53	1.41	1.49	1.57
MSC	R^2	0.57	0.56	0.54	0.49	0.55
	RMSE	3.09	3.10	3.19	3.37	3.09
	RPD	1.49	1.49	1.44	1.37	1.49
SGD1	R^2	0.58	0.58	0.54	**0.60**	0.59
	RMSE	3.01	3.00	3.25	**2.92**	2.94
	RPD	1.53	1.54	1.42	**1.58**	1.57
SGD2	R^2	0.52	0.57	0.37	0.57	0.58
	RMSE	3.19	3.03	4.42	3.07	3.00
	RPD	1.44	1.52	1.04	1.50	1.54
SNV	R^2	0.57	0.57	0.54	0.48	0.55
	RMSE	3.10	3.08	3.19	3.37	3.10
	RPD	1.48	1.49	1.44	1.37	1.48
SMO	R^2	**0.63**	**0.62**	**0.61**	**0.61**	0.59
	RMSE	**2.80**	**2.83**	**2.96**	**2.89**	2.97
	RPD	**1.64**	**1.62**	**1.56**	**1.59**	1.55
Raw	R^2	**0.63**	**0.62**	0.58	**0.62**	**0.60**
	RMSE	**2.81**	**2.83**	3.01	**2.85**	**2.95**
	RPD	**1.64**	**1.62**	1.53	**1.61**	**1.56**

Nitrogen full-spectra (VNIR+SWIR) validation

		PLSR	PCR	MLR	RF	SVM
ASGD1	R^2	0.58	0.57	0.55	0.56	0.55
	RMSE	0.45	0.49	0.46	0.45	0.47
	RPD	1.46	1.36	1.43	1.46	1.42
ASGD2	R^2	0.51	0.14	0.29	0.54	0.59
	RMSE	0.48	0.63	0.63	0.47	0.45
	RPD	1.39	1.06	1.04	1.41	1.48
EMSC	R^2	0.56	**0.61**	0.54	0.48	0.53
	RMSE	0.45	**0.46**	0.46	0.49	0.47
	RPD	1.47	**1.44**	1.45	1.36	1.40
MSC	R^2	**0.61**	0.53	0.54	0.39	0.43
	RMSE	**0.43**	0.49	0.45	0.52	0.52
	RPD	**1.54**	1.35	1.48	1.26	1.28
SGD1	R^2	0.57	0.58	0.55	0.48	0.55
	RMSE	0.46	0.47	0.45	0.52	0.47
	RPD	1.45	1.41	1.46	1.27	1.40
SGD2	R^2	0.48	0.43	0.26	0.51	0.56
	RMSE	0.48	0.51	0.63	0.49	0.45
	RPD	1.38	1.31	1.05	1.36	1.46
SNV	R^2	0.57	0.50	0.55	0.36	0.42
	RMSE	0.45	0.51	0.45	0.54	0.52
	RPD	1.46	1.30	1.48	1.23	1.27
SMO	R^2	**0.66**	**0.63**	**0.63**	0.37	0.43
	RMSE	**0.41**	**0.44**	**0.41**	0.52	0.50
	RPD	**1.61**	**1.52**	**1.64**	1.26	1.32
Raw	R^2	**0.60**	**0.61**	0.59	0.36	0.43
	RMSE	**0.43**	**0.44**	0.43	0.53	0.50
	RPD	**1.54**	**1.52**	1.56	1.25	1.33

R^2 = coefficient of determination; RMSE, root mean square error; RPD, ratio of performance to deviation. PLSR, partial least-squares regression; PCR, principal components regression; MLR, multiple linear regression; RF, random forest; SVM, support vector machine; ASGD1, absorbance transformation then Savitzky-Golay first derivative; ASGD2, absorbance transformation then Savitzky-Golay second derivative; EMSC, extended multiplicative scatter-correction; MSC, multiplicative scatter-correction; SGD1, Savitzky-Golay first derivative; SGD2, Savitzky-Golay second derivative; SNV, standard normal variate; SMO, smoothed. Models with validation R^2≥0.6 are in bold.

For the prediction of water, the validation R^2 values ranged from 0.34 to 0.63 with the validation RMSE and RPD values varying from 5.22 to 2.8 and 0.88 to 1.64, respectively. Similarly, for the prediction of nitrogen, the validation R^2 values ranged from 0.14 to 0.66 with the validation RMSE and RPD values varying from 0.63 to 0.41 and 1.04 to 1.64, respectively. As was found with the VNIR models, the smoothed and raw data had the highest overall prediction performance for both water and nitrogen, except for the RF and SVM prediction of nitrogen which was improved with preprocessing. PLSR models again had the most consistently high performance compared to other multivariate methods trialled, particularly when used with raw or smoothed data (**Table 6**).

Wavelength Selection Using Regression Coefficients Reduces Data Without Compromising Accuracy

Wavelength selection was performed in order to remove some of the redundant information inherent in hyperspectral data. Two approaches to wavelength selection were used: through the regression coefficients from the full-spectra models and by selecting only wavelengths located at known-absorption features for both water and nitrogen.

The wavelengths retained in the reduced β coefficient water model were 398–404, 410–415, 426–432, 440–443, 637–705, 724–752, and 950–1,000 nm. Direct associations between reflectance spectra and water content are those features related specifically to the water content in the leaves. These are known to occur at 1450 and 1940 nm with smaller features at 980, 1150, and 1400 nm associated with the bonding and stretching of O-H molecules (Curran, 1989; Ollinger, 2011). However, reflectance properties can also be influenced by indirect effects, those associated with other traits that vary alongside water or nitrogen status. These indirect effects can be caused by properties such as varying leaf pigments or vegetation architecture, which are represented by reflectance changes throughout visible and near-infrared wavelengths (Ollinger, 2011). The fact that the model produced predictions based on wavelengths within only the visible and near-infrared regions suggests that rather than directly quantifying water content, it is detecting indirect associations between water and secondary traits influencing the overall reflectance.

The nitrogen wavelengths with the highest β coefficients were located at 510–637 nm and 693–739 nm. The 693–739 nm region is likely representing changes in the red-edge position, which have previously been shown to vary with crop chlorophyll concentration (Horler et al., 1983; Curran et al., 1990). However, the identified wavelengths from 510–637 nm do not correspond to a direct association with nitrogen. Foliar nitrogen is predominantly found within proteins, largely chlorophyll or Rubisco (Evans, 1989; Elvidge, 1990). Protein absorption features only occur in the infrared region around 910, 1,020, 1,510, 1,690, 1,980, 2,060, 2,130, 2,180, 2,240, and 2,300 nm while chlorophyll absorptions are found in the visible region of the spectrum near 430, 460, 640, and 660 nm (Curran, 1989). Therefore, the 510–637 nm wavelengths are likely detecting an indirect association between the spectra and nitrogen content.

The second approach to wavelength reduction was to use only wavelengths that have previously been related to either water and nitrogen content. The wavelengths identified by previous studies included both direct and indirect associations to water (**Table 2**) and nitrogen (**Table 3**). The wavelengths selected for the reduced nitrogen models excluded some of the direct spectral asssociations known to exist with nitrogen. Wavelengths with direct association to nitrogen occurring within the SWIR region were not included in the model because those absorption features are generally weaker (Curran, 1989; Ollinger, 2011). Therefore, wavelengths were selected (from Ecarnot et al., 2013) which are known to be correlated with nitrogen content in vegetation, even if they are detecting secondary traits relating to changes in water content. Similarly, some of the major water absorption features, those located at 1,940 and 1,150 nm were excluded from the reduced water models while other wavelengths that don't have direct associations to water were included (e.g., those located within the visible region). A number of plant functions are related to changes in water content. Therefore, even if the exact compound or process is not determined, spectral features of several secondary traits and processes have been used, and included in these reduced water models, for indirect associations to water content.

Since PLSR consistently showed the strongest performance in the previous model development, it was the only multivariate technique considered in the reduced-wavelength models. The predictive performances of the wavelength-selection models developed using the regression coefficient method had higher accuracy than the full-spectra or VNIR model (**Figure 2**). For the prediction of water content, the regression coefficient method generated a PLSR model with validation $R^2 = 0.69$, RMSE = 2.53, and RPD = 1.78 while the feature method gave accuracies of $R^2 = 0.64$, RMSE = 2.74, and RPD = 1.76. (**Table 7**). This is compared to a maximum validation R^2 value of 0.63 using the full-spectra PLSR model (**Table 8**). For nitrogen, the regression coefficient method had a validation accuracy of $R^2 = 0.66$, RMSE = 0.41, and RPD = 1.66 (**Table 7**), compared to a validation R^2 value of 0.60 using the same methods (PLSR with raw data) from the full-spectra models (**Table 8**).

The reduced wavelength models, while greatly simplifying the models also reduced the degree of overfitting. The calibration and validation prediction values (**Table 7**) are much closer together than the models without wavelength selection (**Tables 4** and **5**). The feature method produced accuracies of $R^2 = 0.52$, RMSE = 0.47, and RPD = 1.39 for predicting nitrogen. The feature reduction method for wavelength refinement did not improve either the water or nitrogen models.

Distribution Maps Can Visualize Water and Nitrogen Distribution

Distribution maps, or prediction maps, were developed in order to provide a visual representation of the concentration of water and nitrogen content within the plants as well as to show any spatial variability within individual plants. Distribution maps were created by applying the regression coefficients of the highest performing VNIR models to the calibrated hyperspectral images (**Figure 3**). In the case of water content, the coefficients from the VNIR PLSR model of raw spectra were used. For nitrogen, the

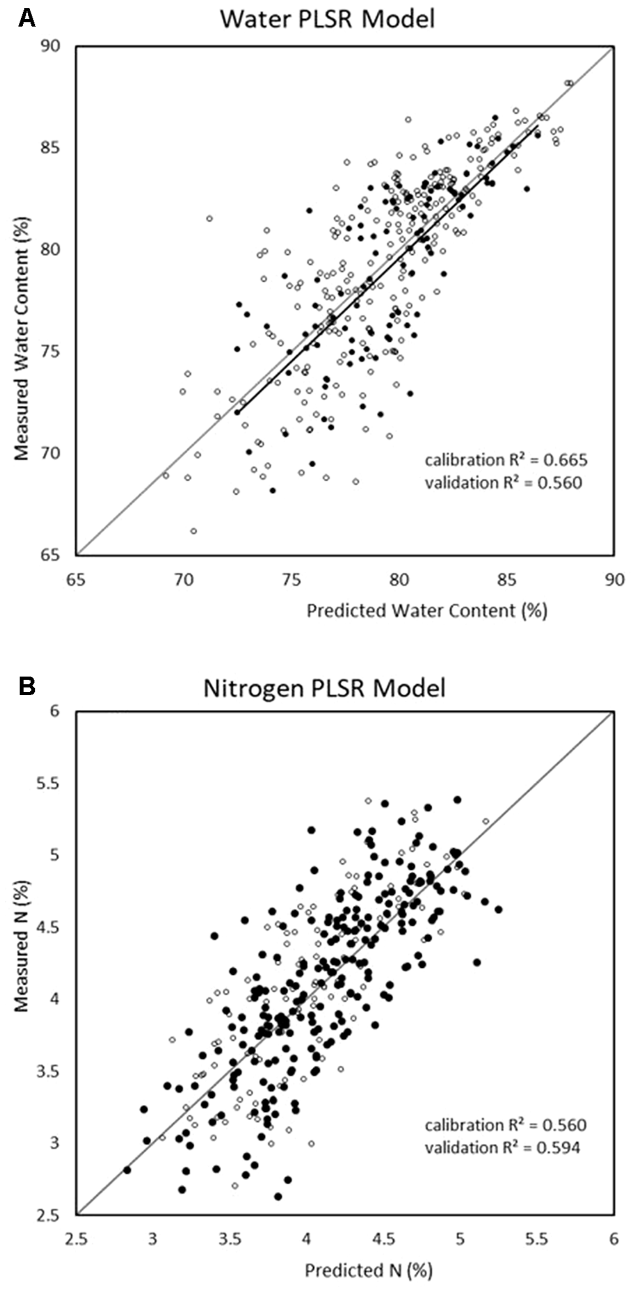


FIGURE 2 | Water (top) and nitrogen (bottom) regression graphs for the models showing the best validation predictive performance from all trialled methods (regression coefficient reduction method with raw spectra and partial least square regression (PLSR)).

coefficients from the VNIR PLSR model with smoothed spectra were applied. Only the VNIR models were used to develop the distribution maps; the combined full-spectra models were not considered. This is because the two cameras (FX10 and SWIR) operate at different spatial resolutions, so the combined models cannot be visualized on a single image.

The resulting prediction maps revealed the spatial variation in biochemical properties, particularly water content, and allowed for a visual comparison between and within the plants which is otherwise impossible with the raw hyperspectral data (**Figure 4**). There were noticeable differences between the maps of the watered (**Figure 4A**) and drought (**Figure 4C**) plants as indicated by their color scale; the watered plants appear yellow-red (72–88% water content) whereas the drought plants are predominantly green (64–72% water content). This is also the case for the low (**Figure 4B**) and high (**Figure 4D**) nitrogen plants, however this difference is less obvious.

In general, water content was higher at the base of the leaves and decreased toward the tips (**Figures 4A, C**). Higher levels of water were also apparent around the midrib region as opposed to the outsides of the leaves. These clear patterns and gradual degradation between neighboring pixels suggests that the water distribution maps provide a visually plausible indication of spatial variability within the plants. The nitrogen distribution maps (**Figures 4B, D**) do not appear to follow a plausible spatial pattern; each pixel is a different colour with no clear gradation to the neighboring pixels. This "noise" in the image is likely the result of model overfitting (Gowen et al., 2014).

DISCUSSION

Hyperspectral imagery combines the spatial information of traditional RGB imagery with the benefits of high-resolution spectral reflectance data. This reflectance data, an indication of how light is interacting with the target, provides a unique spectral fingerprint of the chemical nature of each plant. However, hyperspectral images alone are not visually intuitive or easily interpretable. The development of distribution maps has allowed for a visual comparison between and within plants and revealed the spatial variation in water and nitrogen.

Of the five different multivariate regression methods and 10 different spectral preprocessing techniques trialled, the smoothed and raw data consistently achieved the highest overall prediction performances for both water and nitrogen. This suggests that preprocessing may not have been required for this dataset and that applying preprocessing techniques may have dampened some of the important features of the spectra. Without accounting for the physical influence of scatter or noise, it would normally be expected that a poorer performance is achieved for the smoothed data in comparison to pre-processed data, however, this was found not to be the case.

The high accuracies achieved with PLSR in comparison to other techniques is due to its ability to handle both the high dimensionality and collinearity inherent in hyperspectral data and its effectiveness when a large number of collinear predictor variables are present (Wold et al., 1984). The strong performance of PLSR and its ability to handle spectral data explains why it has been so widely adopted in past hyperspectral studies quantifying plant biochemical properties. MLR and SVM also performed well.

One apparent and perhaps significant finding was the large difference between the calibration and validation performances for the VNIR models (**Supplementary Tables 1** and **2**). The inconsistencies between the sets, particularly in the case where validation statistics are considerably lower than calibration

TABLE 7 | Predictive performances for the wavelength refined models. Cal = calibration. Val = validation. Results are shown for the two different approaches used for variable selection: using the top 30% of wavelengths based on the regression coefficients from the full-spectra models and using wavelengths previously determined to be associated with nitrogen and water.

Water wavelength selection models				Nitrogen wavelength selection models			
	PLSR	Cal	Val		PLSR	Cal	Val
Regression Coefficients	R^2	0.81	0.69	Regression Coefficients	R^2	0.74	0.66
	RMSE	2.05	2.53		RMSE	0.32	0.41
	RPD	2.31	1.78		RPD	1.89	1.66
Known Absorption Features	R^2	0.71	0.64	Known Absorption Features	R^2	0.54	0.52
	RMSE	2.57	2.74		RMSE	0.42	0.47
	RPD	1.85	1.76		RPD	1.45	1.39

TABLE 8 | Validation results for the strongest performing models for water and nitrogen prediction across the different pre-processing and regression methods trialled-PLSR in combination with raw input data. Wavelength selection using the regression coefficient method produced the strongest models.

		VNIR	Full spectra (VNIR+SWIR)	Wavelength selection regression coefficients	Wavelength selection feature positions
Water	R^2	0.56	0.63	0.69	0.64
	RMSE	2.85	2.81	2.53	2.74
	RPD	1.50	1.64	1.78	1.76
Nitrogen	R^2	0.57	0.60	0.66	0.52
	RMSE	0.42	0.43	0.41	0.47
	RPD	1.53	1.54	1.66	1.39

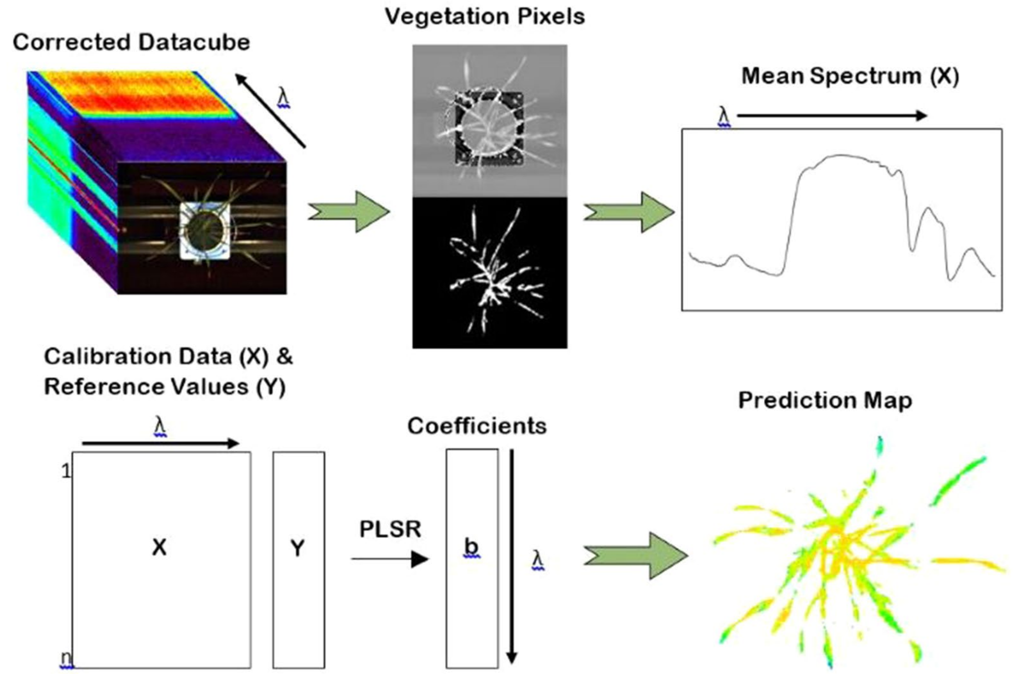


FIGURE 3 | The procedure for the development of water and nitrogen distribution maps. Mean spectra were used as input to generate a PLSR model. The coefficients of the partial least square regression (PLSR) model were then applied to the unfolded datacube at the individual pixel level providing a spatial visualisation of water and nitrogen distribution within the plants.

statistics, may be attributed to model overfitting. This discrepancy is not as severe in the wavelength-selection models suggesting that the wavelength selection methods were able to remove the noise and irrelevant wavelengths in the data. The prediction accuracies between the nitrogen and water models were quite similar; there was no drastic difference in the ability to predict one property over the other. This is true for all models developed: VNIR, full-spectra, and wavelength-refined models (besides perhaps the absorption feature method for wavelength-selection which yielded validation $R^2 = 0.64$ for water and $R^2 = 0.52$ for nitrogen).

The stronger performing models, e.g., PLSR, were not drastically improved in the full-spectra models in comparison to the VNIR models. However, the regression and preprocessing methods that did not perform as well in the VNIR models, e.g., EMSC and multiple linear regression, were improved significantly. By considering only the visible and near infrared wavelengths initially, it might have been expected that nitrogen and water, both which have major spectral expressions in the shortwave infrared region, would not be predicted accurately. The fact that the full-spectra (VNIR+SWIR) models did not significantly improve the VNIR models suggests that perhaps neither water nor nitrogen content is being directly detected, rather secondary processes within the plant spectrally expressed in the VNIR region.

The applied wavelength reduction techniques, both the "regression coefficient" and "known absorption feature" methods, while successfully reducing the volume of data, did not significantly improve the model accuracies. The wavelengths incorporated in the two approaches contained both direct and indirect associations with nitrogen and water. Since the regression coefficient approach is likely detecting both direct and indirect associations, selecting known absorption features also based on both direct and indirect associations does not necessarily provide additional or new information

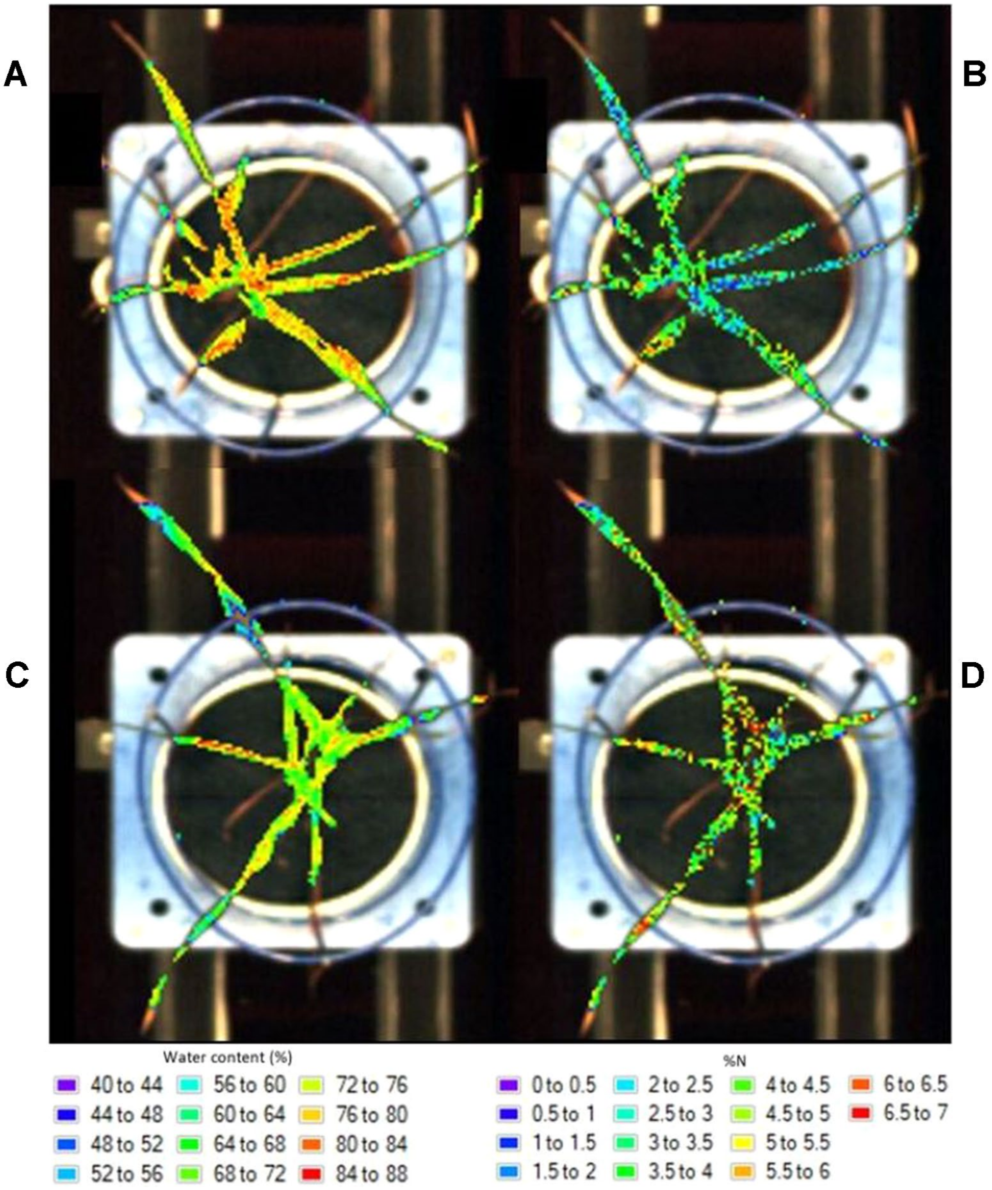


FIGURE 4 | Distribution maps showing the prediction of water content in a watered **(A)** and drought **(C)** plant and nitrogen levels in a low **(B)** and high **(D)** nitrogen soil plant.

(i.e., there is some degree of duplication in the two approaches since they both target direct and indirect associations). If only direct features of nitrogen or water were selected from previous literature during the known absorption feature approach, then the wavelength selection methods could be considered as more independent approaches; one approach, the known absorption feature approach, is targeting only the direct associations to water or nitrogen while the regression coefficient approach is targeting both direct and indirect associations. Targeting only direct associations would mean that the developed models would likely have a stronger predictive performance for independent datasets. Due to the indirect associations included, the wavelengths identified as being sensitive to water and nitrogen cannot be generalised and may not hold true for all future studies.

Previous studies have utilized high-throughput phenotyping platforms equipped with hyperspectral cameras for investigating crop traits. Pandey et al. (2017) imaged maize and soybean leaves using a hyperspectral camera with a spectral range of 550–1700 nm to predict leaf water content and nutritional status. They showed that both water and nitrogen content could be predicted accurately using PLSR (R^2 = 0.93 and 0.92 with RPD = 3.80 and 3.60, for water and nitrogen respectively). Ge et al. (2016) used high throughput hyperspectral imaging to characterize the temporal dynamics of the leaf water content of maize. They used PLSR to accurately predict leaf water content using a hyperspectral camera with a wavelength range of 550–1750 nm and a spectral bandwidth of 5 nm (R^2 = 0.87, RMSE = 3.0%, and RPD = 2.82). Although different species and smaller datasets were used (Ge et al. (2016) used 80 plants while Pandey et al. (2017) used 120), these previous studies achieved much greater accuracies for predicting water and nitrogen content than presented here. Compared to wheat, which has narrow, twisting leaves with irregular structure, maize has considerably broader leaves with a more regular leaf architecture. There are fewer pixels to work with when analyzing wheat plants with a greater proportion suffering from mixed signals (i.e., edge pixels of background and plant material). The fact that maize plants have more pixels to work with, and proportionally fewer mixed-pixels, may have contributed to the higher accuracies achieved in these previous studies.

The generally poor R^2 values achieved in this study can be attributed to a number of factors. The main factor influencing the poor performance of the nitrogen models is likely the different scales at which the spectral and reference measurements were made. The spectra were extracted and averaged over the entire plant whereas the reference laboratory measurements of nitrogen were taken on a single leaf. Nitrogen is considered a mobile element within plants and can be retranslocated to younger leaves from older leaves (Taiz and Zeiger, 1998). Its distribution is not homogenous throughout wheat plants but is generally lower in older leaves (Wang et al., 2005). Therefore, the nitrogen measured in a single leaf may not be representative of the entire plant's nitrogen status. Regressions may be improved if either spectra were extracted from only the leaf which was sampled for nitrogen (i.e., the flag leaf), or alternatively, reference nitrogen measurements were determined on the plant scale. This could be achieved by calculating masks which identify only the flag leaf within the image. However, the flag leaf could not be identified automatically in the images. Manually extracting the flag leaf would have been too time-consuming and would have diminished the benefits of using a high-throughput system. By separating the flag leaf from the remainder of the plant material and using only the leaf pixels corresponding to those analysed for nitrogen, regression results are likely to improve. Similarly, water regressions could also be improved by reversing the mask and removing the flag leaves which were not used in the calculation of water content. A smaller subset of the data could be used to this affect to assess how much of the noise and error can be attributed to different measurement scales.

Alternatively, the poor results for the water regressions may be attributed to measurement protocols used to obtain reference values. The relative water content (RWC), or relative turgidity of a leaf or plant, is a measure of its hydration status (actual water content) relative to its maximal water holding capacity at full turgidity (Weatherley, 1950; Mullan and Pietragalla, 2012). Determining the RWC may have provided a better indication of plant water status than the method used here which did not consider turgid weight. Alternatively, regressions may have been stronger if leaf thickness was also considered. Light reflected from leaves or transmitted through leaves depends on both the RWC of leaf cells as well as the thickness of the leaves (Seelig et al., 2008). Therefore, models may have proved more accurate if leaf thickness or even total plant biomass were taken into consideration.

A further contribution to the poor R^2 values for both the water and nitrogen regressions may be the canopy geometry. The geometry of plants and inclination of individual leaves has a strong influence on the spectral information acquired (Behmann et al., 2015; Huang et al., 2018). The cameras took a top-view image from directly above the plant. Lower leaves may have been hidden by leaves above and this information consequently missed. Incorporating 3-D structural information alongside the hyperspectral images may reduce the effect of plant geometry on the acquired spectra.

Prediction accuracies may also improve if the spectral reflectance jump between the two sensors, caused by the different properties and measurement principles of the cameras and detectors, was corrected. Overlapping wavelength regions (996–1006 nm) were removed from the spectra, however, correcting the spectral jumps with appropriate preprocessing techniques, such as splice or parabolic correction, may improve confidence in the data when combining information from separate cameras. In this case, combining the data from the two cameras still provided meaningful information without removing the jump. The spectral signatures with the jump arevstill representative of the spectral properties of the plants, providing the jumps are consistent in the calibration and validation datasets. A further consideration in combining the two datasets is the difference in the spatial resolution of the two cameras. The issue of different spatial resolutions was reduced by extracting the mean spectrum of the entire plant rather than using pixel-level information. Combining the information from two cameras gives meaningful information at the plant level but the different spatial resolutions of the cameras may prove an issue when performing pixel-level analysis.

As well as providing information related to the biochemical properties of plants, the developed prediction maps provide a visual representation of model performance. This is apparent when comparing the distribution maps of nitrogen to those of water content; the nitrogen maps appear "noisier" in comparison to the water maps. The spatial distribution and patterns within the plant do not appear feasible and may not be caused by real spatial variation in nitrogen. This is likely an effect of model overfitting (Gowen et al., 2014). This noise could be reduced by selecting a different PLSR model with which to create the maps, likely one with fewer components.

One of the main advantages of developing distribution maps is their potential to be used in time-series analyses in order to track the dynamics of biochemical parameters throughout the growing season. Images can be non-destructively acquired at multiple times throughout the season and used to predict changes or remobilisation of biochemical components over time. While hyperspectral images were acquired weekly, reference measurements of nitrogen and water were only collected at the conclusion of the experiment. Therefore, models were developed using plants of only one age (61 DAS). The growth stage of the plant has a strong influence on the developed models (Li et al., 2010; Haiying and Hongchun, 2016; Wen et al., 2019). As such, time-series analysis would first require the development of robust models using a variety of plant ages and growth stages.

Using hyperspectral phenotyping methods to estimate the content and spatial distribution of nitrogen and water in wheat shows strong promise but models could be improved by incorporating additional data from a larger range of growing conditions, e.g., seasons, soils, and genotypes. Using the techniques and methods developed here, it may be possible to measure other plant biochemical and structural properties, such as other nutrients, salts, lignin, cellulose, and water-soluble carbohydrates. Such methods may lead to advances in high-throughput phenotyping and subsequent improvements in the way that breeding trials are conducted and their biochemical properties analysed.

CONCLUSION

The development of distribution maps through hyperspectral imaging was demonstrated as a nondestructive, *in vivo* tool for estimating the concentration, and spatial distribution of water content and nitrogen levels in wheat. Hyperspectral images were collected of wheat plants and multivariate regression was performed in order to relate the spectral information to the measured water and nitrogen levels. Both plant water content and nitrogen level could be predicted with "acceptable" accuracy using PLSR models developed with the mean reflectance from the full-spectra wavelengths. Wavelength selection using a regression coefficient approach slightly improved model accuracy while significantly reducing model complexity.

The regression coefficients from the best-performing VNIR models were applied to the calibrated images to develop distribution maps. The water distribution maps provided a plausible visual representation of the water distribution within the plant, however, the nitrogen maps appeared to suffer from noise likely due to model overfitting. The findings and methods from this study demonstrate the high potential of hyperspectral imagery, multivariate regression, and distribution maps have for estimating the level and distribution of plant chemical properties.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

The Plant Accelerator®, Australian Plant Phenomics Facility, is funded under the National Collaborative Research Infrastructure Strategy (NCRIS). BrBr acknowledges the University of Adelaide for research support through the provision of an Australian Government Research Training Program Scholarship and The Plant Accelerator® for a PhD stipend top-up. Financial support from the Grains Research and Development Corporation (GRDC) and The AW Howard Memorial Trust (Department of Primary Industries and Regions, South Australia) is also acknowledged.

ACKNOWLEDGMENTS

The authors would like to thank The Plant Accelerator® staff, particularly Lidia Mischis, Nicole Bond and Fiona Groskreutz, for their assistance with the glasshouse experiments and Sanjiv Satija for his assistance with the nitrogen measurements.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01380/full#supplementary-material

REFERENCES

Awada, L., Phillips, P. W. B., and Smyth, S. J. (2018). The adoption of automated phenotyping by plant breeders. *Euphytica* 214, 148. doi: 10.1007/s10681-018-2226-z

Axelsson, C., Skidmore, A. K., Schlerf, M., Fauzi, A., and Verhoef, W. (2013). Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *Int. J. Remote Sens.* 34 (5), 1724–1743. doi: 10.1080/01431161.2012.725958

Barnes, R. J., Dhanoa, M. S., and Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectroscopy* 43 (5), 772–777. doi: 10.1366/0003702894202201

Behmann, J., Mahlein, A.-K., Paulus, S., Kuhlmann, H., Oerke, E.-C., and Plumer, L. (2015). Calibration of hyperspectral close-range pushbroom cameras for plant phenotyping. *ISPRS J. Photogrammetry Remote Sens.* 106, 172–182. doi: 10.1016/j.isprsjprs.2015.05.010

Behmann, J., Steinrücken, J., and Plümer, L. (2014). Detection of early plant stress responses in hyperspectral images. *ISPRS J. Photogrammetry Remote Sens.* 93, 98–111. doi: 10.1016/j.isprsjprs.2014.03.016

Belgiu, M., and Dragut, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogrammetry Remote Sens.* 114, 24–31. doi: 10.1016/j.isprsjprs.2016.01.011

Berger, B., de Regt, B., and Tester, M. (2012). "High-Throughput Phenotyping in Plants. Methods in Molecular Biology (Methods and Protocols)," in *High-Throughput Phenotyping of Plant Shoots*, vol. 918. Ed. Normanly, J. (Totowa, NJ: Humana Press). doi: 10.1007/978-1-61779-995-2_2

Brien, C. J. (2017) dae: Functions useful in the design and ANOVA of experiments. Version 3.0-12. https://cran.r-project.org/package=dae.

Curran, P. J. (1989). Remote Sensing of Foliar Chemistry. *Remote Sens. Environ.* 30, 271–278. doi: 10.1016/0034-4257(89)90069-2

Curran, P. J., Dungan, J. L., and Gholz, H. L. (1990). Exploring the relationship between reflectance red edge and chlorophyll content in slash pine. *Tree Physiol.* 7, 33–48. doi: 10.1093/treephys/7.1-2-3-4.33

De Bei, R., Cozzolino, D., Sullivan, W., Cynkar, W., Fuentes, S., Dambergs, R., et al. (2011). Non-destructive measurement of grapevine water potential using near infrared spectroscopy. *Australian J. Grape Wine Res.* 17, 62–71. doi: 10.1111/j.1755-0238.2010.00117.x

Do, H., Kalousis, A., Wang, J., and Woznica, A. (2012). "*A metric learning perspective of SVM: on the relation of LMNN and SVM.*" In: *Proceedings of the 15th International Intelligence and Statistics (AISTATS)*, JMLR Proceedings. 308–317.

Dotto, A. C., Dalmolin, R. S. D., Caten, A., and Grunwald, S. (2018). A systematic study on the application of scatter-corrective and spectral derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. *Geoderma* 314, 262–274. doi: 10.1016/j.geoderma.2017.11.006

Ecarnot, M., Compan, F., and Roumet, P. (2013). Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer. *Field Crops Res.* 140, 44–50. doi: 10.1016/j.fcr.2012.10.013

Eitel, J. U. H., Gessler, P. E., Smith, A. M. S., and Robberecht, R. (2006). Suitability of existing and novel spectral indices to remotely detect water stress in Populus spp. *For. Ecol. Management* 229, 170–182. doi: 10.1016/j.foreco.2006.03.027

ElMasry, G., Sun, D. W., and Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *J. Food Eng.* 110 (1), 127–140. doi: 10.1016/j.jfoodeng.2011.11.028

Elvidge, C. D. (1990). Visible and near infrared reflectance characteristics of dry plant materials. *Int. J. Remote Sens.* 11 (10), 1775–1795. doi: 10.1080/01431169008955129

Evans, J. R. (1989). Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia* 78, 9–19. doi: 10.1007/BF00377192

Fahlgren, N., Gehan, M. A., and Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

Foutry, T., and Baret, F. (1997). Vegetation water and dry matter contents estimated from top of the atmosphere reflectance data: a simulation study. *Remote Sens. Environ.* 61, 34–45. doi: 10.1016/S0034-4257(96)00238-6

Gao, B.-C. (1996). NDWI A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sens. Environ.* 58, 257–266. doi: 10.1016/S0034-4257(96)00067-3

Garnett, T., Plett, D., Heuer, S., and Okamoto, M. (2015). Genetic approaches to enhancing nitrogen-use efficiency (NUE) in cereals: challenges and future directions. *Funct. Plant Biol.* 42, 10, 921–941. doi: 10.1071/FP15025

Garnett, T. P., and Rebetzke, G. J. (2013). "Improving water and nutrient-use efficiency in food production systems," in *Improving crop nitrogen use in dryland farming*. Eds. Rengel, Z., and Wiley, N. J., 123–144. doi: 10.1002/9781118517994.ch8

Ge, Y., Bai, G., Stoerger, V., and Schnable, J. C. (2016). Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. *Computers Electronics Agric.* 127, 625–632. doi: 10.1016/j.compag.2016.07.028

Geladi, P., MacDougall, D., and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectroscopy* 39 (3), 491–500. doi: 10.1366/0003702854248656

Gholizadeh, A., Boruvka, L., Saberioon, M. M., Kozak, J., Vasat, R., and Nemecek, K. (2015). Comparing different data preprocessing methods for monitoring soil heavy metals based on soil spectral features. *Soil Water Res.* 10 (4), 218–227. doi: 10.17221/113/2015-SWR

Gowen, A., Burger, J., Esquerre, C., Downey, G., and O'Donnell, C. (2014). Near infrared hyperspectral image regression: on the use of prediction maps as a tool for detecting model overfitting. *J. Near Infrared Spectroscopy* 22, 261–270. doi: 10.1255/jnirs.1114

Haiying, L., and Hongchun, Z. (2016). Hyperspectral characteristic analysis for leaf nitrogen content in different growth stages of winter wheat. *Appl. Optics* 55 (34), 151–161. doi: 10.1364/AO.55.00D151

Hansen, N. J. S., Plett, D., Berger, B., and Garnett, T. (2018). "Engineering Nitrogen Utilization in Crop Plants," in *Tackling Nitrogen Use Efficiency in Cereal Crops Using High-Throughput Phenotyping*. Eds. Shrawat, A., Zayed, A., and Lightfoot, D. (Cham, Switzerland: Springer). doi: 10.1007/978-3-319-92958-3_7

Hong, M., Zeng, W., Ma, T., Lei, G., Zha, Y., Fang, Y., et al. (2017). Determination of growth stage-specific crop coefficients (Kc) of sunflowers (Helianthus annuus L.) under salt stress. *Water* 9, 1–17. doi: 10.3390/w9030215

Horler, D. N. H., Dockray, M., and Barber, J. (1983). The red edge of plant leaf reflectance. *Int. J. Remote Sens.* 4 (2), 278–288. doi: 10.1080/01431168308948546

Huang, P., Luo, X., Jin, J., Wang, L., Zhang, L., Liu, J., et al. (2018). Improving high-throughput phenotyping using fusion of close-range hyperspectral camera and low-cost depth sensor. *Sensors* 18 (8), 2711. doi: 10.3390/s18082711

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., and Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83 (1), 195–213. doi: 10.1016/S0034-4257(02)00096-2

Humplik, J. F., Lazar, D., Husickova, A., and Spichal, L. (2015). Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses- a review. *Plant Methods* 11 (29), 1–10. doi: 10.1186/s13007-015-0072-8

Hunt, E. R., and Rock, B. N. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. *Remote Sens. Environ.* 30 (1), 43–54. doi: 10.1016/0034-4257(89)90046-1

Karatzoglou, A., Meyer, D., and Hornik, K. (2005). "Research Report Series / Department of Statistics and Mathematics, Department of Statistics and Mathematics," in *Support vector machines in R*, vol. 21. (Vienna: WU Vienna University of Economics and Business).

Lambers, H., Chapin, F. S., and Pons, T. L. (2008). "Plant physiological ecology," in *Plant water relations*. Eds. Lambers, H., Chapin, F. S., and Pons, T. L. (New York: Springer), 163–223. doi: 10.1007/978-0-387-78341-3_5

Li, F., Miao, Y., Hennig, S. D., Gnyp, M. L., Chen, X., Jia, L., et al. (2010). Evaluating hyperspectral vegetation indices for estimating nitrogen concentration of winter wheat at different growth stages. *Precision Agric.* 11, 335–357. doi: 10.1007/s11119-010-9165-6

Li, L., Wan Li, L., Wang, S., Ren, T., Wei, Q., Ming, J., et al. (2018). Ability of models with effective wavelengths to monitor nitrogen and phosphorus status of winter oilseed rape leaves using in situ canopy spectroscopy. *Field Crops Res.* 215, 173–186. doi: 10.1016/j.fcr.2017.10.018

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078

Liaw, A., and Wiener, M. (2002). Classification and Regression by random Forest. *R. News* 2, 3, 18–, 22.

Liland, K. H. (2017) "Package 'EMSC': extended multiplicative signal correction" R package Vignette version 0.9.0.

Liu, D., Sun, D. W., and Zeng, X. A. (2014). Recent advances in wavelength selection techniques for hyperspectral image processing in the food industry. *Food Bioprocess Tech.* 7 (2), 307–323. doi: 10.1007/s11947-013-1193-6

Liu, H., Bruning, B., Berger, B., and Garnett, T. (2019). "7th Edition of the International Workshop on Image Analysis Methods for the Plant Sciences," in *Green plant segmentation in hyperspectral images using SVM and hyper-hue* (France: Lyon).

Martens, H., and Næs, T. (2011). *Pretreatment and linearization, Multivariate Calibration*. Hoboken, NJ: John Wiley&Sons Ltd.

Martens, H., and Stark, E. (1991) "Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy". *J. Phar. Biomed. Anal.* 9 (8), 625–635

Mevik, B. H., and Wehrens, R. (2007). The pls package: Principal component and partial least squares regression in R. *J. Statistical Software* 18 (2), 1–23. doi: 10.18637/jss.v018.i02

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., et al. (2017) "Package 'e1071': Misc functions of the department of statistics, Probability Theory Group (Formerly: E1071), TU Wien" R package Vignette version 1.6-8.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2012). *Introduction to Linear Regression Analysis*. Hoboken, New Jersey: A John Wiley & Sons, Inc.

Mullan, D., and Pietragalla, J. (2012). "Physiological Breeding II: a field guide to wheat phenotyping," in *Leaf relative water content*. Eds. Pask, A. J. D., Pietragalla, J., Mullan, D. M., and Reynolds, M. P. (Mexico, D.F.: CIMMYT).

Musick, H. B., and Pelletier, R. E. (1986). Response of some Thematic Mapper band ratios to variation in soil water content. *Photogrammetric Eng. Remote Sens.* 52 (10), 1661–1668.

Musick, H. B., and Pelletier, R. E. (1988). Response to soil moisture of spectral indexes derived from bidirectional reflectance in thematic mapper wavebands. *Remote Sens. Environ.* 25 (2), 167–184. doi: 10.1016/0034-4257(88)90099-5

Ollinger, S. V. (2011). Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytolog.* 189, 375–394. doi: 10.1111/j.1469-8137.2010.03536.x

Osborne, J. W., and Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment Res. Evaluation* 8 (2), 1–5.

Pan, T. T., Sun, D. W., Cheng, J. H., and Pu, H. (2016). Regression algorithms in hyperspectral data analysis for meat quality detection and evaluation. *Comprehensive Rev. Food Sci. Food Safety* 15, 529–541. doi: 10.1111/1541-4337.12191

Pandey, P., Ge, Y., Stoerger, V., and Schnable, J. C. (2017). High throughput *in vivo* analysis of plant leaf chemical properties using hyperspectral imaging. *Front. Plant Sci.* 8, 1348. doi: 10.3389/fpls.2017.01348

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Machine Learn. Res.* 12, 2825–2830.

Peñuelas, J., Filella, I., Biel, C., Serrano, L., and Save, R. (1993). The reflectance at the 950-970nm region as an indicator of plant water status. *Int. J. Remote Sens.* 14 (10), 1887–1905. doi: 10.1080/01431169308954010

Peñuelas, J., Filella, I., Lloret, P., Munoz, F., and Vilajeliu, M. (1995). Reflectance assessment of mite effects on apple trees. *Int. J. Remote Sens.* 16, 2727–2733. doi: 10.1080/01431169508954588

Peñuelas, J., Piñol, J., Ogaya, R., and Filella, I. (1997). Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *Int. J. Remote Sens.* 18 (13), 2869–2875. doi: 10.1080/014311697217396

R Core Team. (2017). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raun, W. R., and Johnson, G. V. (1999). Improving nitrogen use efficiency for cereal production. *Agronomy J.* 91 (3), 357–363. doi: 10.2134/agronj1999.00021962009100030001x

Rinnan, A., van den Berg, F., and Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends Analytical Chem.* 28 (10), 1201–1222. doi: 10.1016/j.trac.2009.07.007

Seelig, H. D., Hoehn, A., Stodieck, L. S., Klaus, D. M., Adams, W. W., and Emery, W. J. (2008). Relations of remote sensing leaf water indices to leaf water thickness in cowpea, bean, and sugarbeet plants. *Remote Sens. Environ.* 112, 445–455. doi: 10.1016/j.rse.2007.05.002

Stevens, A., and Ramirez-Lopez, L. (2015) "Package 'prospectr'" [online] Available: https://cran.r-project.org/web/packages/prospectr/prospectr.pdf.

Taiz, L., and Zeiger, E. (1998). "Plant Physiology," in *Mineral Nutrition*, Eds. Taiz, L., and Zeiger, E. (Sunderland, Massachusetts: Sinauer Associates Publishers), 103–124.

Torres, I., Sánchez, M.-T., Benlloch-González, M., and Pérez-Marín, D. (2019). Irrigation decision support based on leaf relative water content determination in olive grove using near infrared spectroscopy. *Biosyst. Eng.* 180, 50–58. doi: 10.1016/j.biosystemseng.2019.01.016

Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer. doi: 10.1007/978-0-387-21706-2

Walesiak, M., and Dudek, A. (2017) "Package 'clusterSim': Searching for optimal clustering procedure for a data set" R package Vignette version 0.47-1.

Wang, Z., Wang, J., Zhao, C., Zhao, M., Huang, W., and Wang, C. (2005). Vertical distribution of nitrogen in different layers of leaf and stem and their relationship with grain quality of winter wheat. *J. Plant Nutrition* 28 (1), 73–91. doi: 10.1081/PLN-200042175

Weatherley, P. E. (1950). Studies in the water relations of the cotton plant. I. The field measurement of water deficits in leaves. *New Phytolog.* 49, 81–97. doi: 10.1111/j.1469-8137.1950.tb05146.x

Wen, P.-F., He, J., Ning, F., Wang, R., Zhang, Y.-H., and Li, J. (2019). Estimating leaf nitrogen concentration considering unsynchronized maize growth stages with canopy hyperspectral technique. *Ecological Indicators* 107, 1–16. doi: 10.1016/j.ecolind.2019.105590

Wold, S., Sjöström, M., and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics Intelligent Lab. Syst.* 58, 109–130. doi: 10.1016/S0169-7439(01)00155-1

Wold, S., Ruhe, A., Wold, H., and Dunn, W. J., III (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Int. J. Sci. Statistical Comput.* 5 (3), 735–743. doi: 10.1137/0905052

Yu, K.-Q., Zhao, Y.-R., Li, X.-L., Shao, Y.-N., Liu, F., and He, Y. (2014). Hyperspectral imaging for mapping of total nitrogen spatial distribution in pepper plant. *PLoS One* 9 (12), 1–19. doi: 10.1371/journal.pone.0116205

Zhang, X., Liu, F., He, Y., and Gong, X. (2013). Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosyst. Eng.* 115, 56–65. doi: 10.1016/j.biosystemseng.2013.02.007

Zimmermann, B., and Kohler, A. (2013). Optimizing Savitzky-Golay parameters for improving spectral resolution and quantification in infrared spectroscopy. *Appl. Spectroscopy* 67 (8), 892–902. doi: 10.1366/12-06723

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 31 October 2019
doi: 10.3389/fpls.2019.01404

Learning Semantic Graphics Using Convolutional Encoder–Decoder Network for Autonomous Weeding in Paddy

Shyam Prasad Adhikari*[1], *Heechan Yang*[2] *and Hyongsuk Kim*[1,2]

[1] Division of Electronics Engineering, Intelligent Robots Research Center (IRRC), Chonbuk National University, Jeonju, South Korea, [2] Division of Electronics and Information Engineering, Chonbuk National University, Jeonju, South Korea

Weeds in agricultural farms are aggressive growers which compete for nutrition and other resources with the crop and reduce production. The increasing use of chemicals to control them has inadvertent consequences to the human health and the environment. In this work, a novel neural network training method combining semantic graphics for data annotation and an advanced encoder–decoder network for (a) automatic crop line detection and (b) weed (wild millet) detection in paddy fields is proposed. The detected crop lines act as a guiding line for an autonomous weeding robot for inter-row weeding, whereas the detection of weeds enables autonomous intra-row weeding. The proposed data annotation method, semantic graphics, is intuitive, and the desired targets can be annotated easily with minimal labor. Also, the proposed "extended skip network" is an improved deep convolutional encoder–decoder neural network for efficient learning of semantic graphics. Quantitative evaluations of the proposed method demonstrated an increment of 6.29% and 6.14% in mean intersection over union (mIoU), over the baseline network on the task of paddy line detection and wild millet detection, respectively. The proposed method also leads to a 3.56% increment in mIoU and a significantly higher recall compared to a popular bounding box-based object detection approach on the task of wild–millet detection.

Keywords: semantic graphics, convolutional neural network, autonomous weeding, crop line extraction, encoder–decoder network

Edited by:
Kioumars Ghamkhar,
AgResearch (New Zealand),
New Zealand

Reviewed by:
Christopher James Bateman,
Lincoln Agritech Ltd,
New Zealand
Dong Xu,
University of Missouri,
United States

****Correspondence:***
Hyongsuk Kim
hskim@jbnu.ac.kr

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *04 June 2019*
Accepted: *10 October 2019*
Published: *31 October 2019*

Citation:
Adhikari SP, Yang H and Kim H
(2019) Learning Semantic Graphics
Using Convolutional Encoder–
Decoder Network for Autonomous
Weeding in Paddy.
Front. Plant Sci. 10:1404.
doi: 10.3389/fpls.2019.01404

INTRODUCTION

The resurgence of neural networks in the form of "deep" neural networks (DNNs) (Krizhevsky et al., 2012) has dramatically improved the performance of various computer vision tasks such as image classification (Simonyan and Zisserman, 2014; Szegedy et al., 2015; He et al., 2016; Huang et al., 2017), object detection and localization (Ren et al., 2015; Redmon et al., 2016; He et al., 2017), and semantic segmentation (Long et al., 2015; Ronneberger et al., 2015; Badrinarayanan et al., 2017).

Recently, DNNs have also been used extensively for problems in agriculture. Researchers have applied deep learning in agriculture to automate different tasks such as plant recognition (Grinblat et al., 2016), crop type classification (Kussul et al., 2017), plant disease classification (Mohanty et al., 2016; Fuentes et al., 2018), weed identification (Dyrmann et al., 2016; Dyrmann et al., 2017), and land cover classification (Kussul et al., 2017; Ienco et al., 2017). Agricultural farm is a semi-constrained environment which is easier than unconstrained natural environments for the adoption of DNN. However, application of DNN to agriculture has its own challenges because of confusion due to low variation between the target classes. Crops and weeds are similar in shape, texture, color, and position, which results in significant reduction in accuracy of DNN systems (Mohanty et al., 2016; Dyrmann et al., 2016). Furthermore, severe overlapping and occlusion, a common phenomenon in the farm, also poses serious challenges to the application of DNN in agriculture. Among the different areas for the use of DNN in agriculture, plant and weed identification has received much attention in the literature due to its enormous practical impact. This study is focused on the use of DNN in rice fields.

Rice is a widely eaten staple food by billions of people around the world. It is considered the lifeline of the Asia-Pacific region where 90% of the world's rice is consumed. With increasing population, the demand for rice is expected to grow, and the challenge is to increase the production of rice using limited land, water, and manpower and less use of agrochemicals. One of the factors responsible for reduced rice yield is weeds. Weeds are aggressive growers which compete for nutrition and other resources and thus reduce production. Moreover, weeds serve as hosts to pests and diseases that are otherwise harmful for the crop. Various weed control methods like hand weeding, mechanical weeding, chemical weeding, and biological control are available for weed management. Herbicides are used extensively to manage weeds; however, their increasing use has inadvertent consequences to the human health and the environment. Though mechanical weeding saves farmers from the drudgery of hand weeding, it is nonetheless labor-intensive. With a decline in interest among the younger generation to join agriculture, the available manpower for labor is limited. Biological control methods using fish, insects, and birds are environmentally friendly and used for effective weed management in organic rice cultivation.

With the advancements in robotics, autonomous agricultural robots have been widely adopted to increase crop productivity and improve labor efficiency. Machine vision-based systems have been used in autonomous agricultural robots for weed management in row crops like rice and maize (Guerrero et al., 2017; Ma et al., 2019). Navigation systems are a crucial part of such autonomous robots where a guidance line has to be computed to guide the robot for weed control. Vision sensor-based autonomous guidance systems have been widely researched for extracting the crop lines to guide the robot (Choi et al., 2015; Zhang et al., 2017).

In this work, we used data from a row-transplanted organic rice field in the Republic of Korea where the golden apple snail (*Pomacea canaliculata*) was used for biological control of weeds. The golden apple snail is effective in controlling most of the weeds except for the wild millet. Wild millet being similar in appearance to the rice plant makes it difficult for hand weeding. Towards the end goal of an autonomous weeding system for paddy, we present a DNN-based system to (a) automatically detect rows of crop and (b) detect weed (willet millet) in row-sown (transplanted) paddy field. The detected crop lines act as a guiding line for an autonomous weeding robot for inter-row weeding, whereas the detection of weeds enables autonomous intra-row weeding.

RELATED WORK

Crop Line Detection

Previous works on detecting crop rows using vision-based systems primarily detect the position of the crops using different handcrafted features like living tissue indicators (Søgaard and Olsen, 2003), vegetation index (Bakker et al., 2008; Montalvo et al., 2012), morphological features (Choi et al., 2015), and extraction of the crop line using different pattern recognition and machine learning techniques like distribution of pixel values, vanishing point detection, Hough transform, and linear regression (Søgaard and Olsen, 2003; Bakker et al., 2008; Montalvo et al., 2012; Choi et al., 2015; Jiang et al., 2016).

Methods based on handcrafted features work well under controlled conditions; however, they can fail to work in real farm conditions, as it is practically infeasible to hand-engineer features which capture the extensive diversity found in real farm environments. The methods based on color index work well in the absence of weeds in between the rows, as the vegetation index or living tissue index of weeds is similar to that of crops. The presence of weeds and different natural conditions like shades or light reflection affects the extraction of binary morphological features, which ultimately affects the accuracy of the extracted crop line.

Recent advancements in neural networks have demonstrated that automatic feature learning using convolutional neural networks (CNNs) are more successful than hand-engineered features. Methods based on CNNs have produced state-of-the-art results in different computer vision and pattern recognition problems like object detection and classification (Ren et al., 2015; Redmon et al., 2016; Huang et al., 2017) and semantic segmentation (He et al., 2016).

In this work, we use CNN to extract the crop lines. Unlike prior works which segment the input into different regions and extract the crop lines, we propose to train a CNN to directly learn the concept of a crop line using "semantic graphics" as shown in **Figure 1**.

Weed Detection

Recently, DNN-based algorithms for classification of weeds and crops have attracted much attention. Two different CNNs were used to segment and classify image pixels into crop and weeds (Potena et al., 2016). A method based on *K*-means feature learning combined with CNN was used for weed identification in soybean seedlings (Tang et al., 2017). A fully CNN was used to detect single weed instances in image from winter wheat fields with leaf occlusion (Dyrmann et al., 2017). CNN-based semantic segmentation approaches to separate crops, weeds, and background have also been studied (Milioto et al., 2018; Ma et al., 2019). While semantic segmentation-based approaches are helpful for widely spaced crops and weeds, these approaches are difficult to adopt in fields with heavy overlap and occlusion owing to the difficulty in obtaining per-pixel ground truth annotations. Moreover, the difficulty in obtaining ground truth labels is compounded for crop and weeds, like rice and wild millet, which have similar appearances.

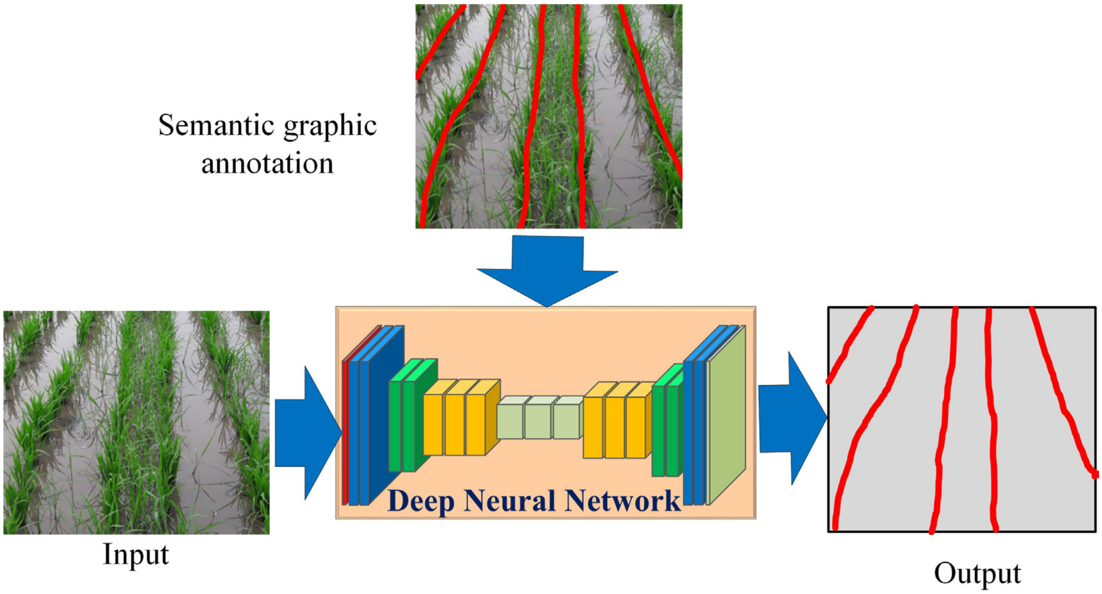


FIGURE 1 | The proposed approach of training deep neural networks to learn the concept of crop line using semantic graphics.

In this work, we propose to learn "semantic graphics" using CNN for the identification of rice and wild millet.

Semantic Graphics

One of the factors enabling the increase in performance of DNNs is the availability of a huge amount of data for training. However, for supervised training of DNNs, the data has to be annotated manually with ground truth. It is expensive and time-consuming to prepare large-scale ground truth annotations (Bearman et al., 2016), and hence, there is a bottleneck in extending the application of DNN to new applications which require the network to be trained on custom datasets. Manual annotation is particularly time-consuming for semantic segmentation where per-pixel annotation is required. Per-pixel semantic labeling is also economically not viable without employing methods which reduce human labor.

To reduce the dependency on large-scale detailed annotations, weakly or semi-supervised learning techniques have been explored in the literature. In the weakly supervised setting, the training images are annotated only at the image level or sparsely annotated at the pixel level, thus requiring less time and effort for annotation. Different forms of weak supervision have also been explored in the literature such as image-level labels (Pinheiro and Collobert, 2015), bounding boxes (Papandreou et al., 2015), and point annotations and free-form scribbles (Bearman et al., 2016; Lin et al., 2016). However, much of the focus in the literature has been towards detecting or segmenting "objects" with a well-defined shape, appearance, and boundary. Less attention has been paid towards understanding complex scenes that are difficult even to annotate correctly due to similar appearance and ambiguous boundaries.

To simplify the process of annotating such complex scenes, we introduce the notion of semantic graphics. Semantic graphics is a graphical sketch where a target concept is expressed in the form of a figure for easy learning by neural networks. Semantic graphics can encode human knowledge directly in intuitive graphics which can be annotated with considerable ease even for complex scenes. For example, in the image of a line-transplanted paddy field shown in **Figure 2**, the lines of paddy have been rendered indistinguishable due to high weed pressure. However, humans can easily figure out the actual rows of paddy in the image, including in those regions where the actual demarcation does not exist due to weeds. One of the meaningful ways to mark the rows is by sketching a line as shown at the bottom of **Figure 2**.

Semantic graphics is different from semantic segmentation as pixels belonging to the same semantic region or super-pixel may not be necessarily labeled with the same target category. Semantic graphics is particularly useful for tasks which are otherwise challenging for existing pixel-based semantic segmentation methods. For example, the rows of paddy and the wild millet in between the rows, as shown in **Figure 2**, are semantically similar; therefore, it is difficult and time-consuming to prepare dense per-pixel annotation to be used for semantic segmentation. However, it is easier to figure out the actual crop rows and represent those using semantic graphics. In this work, we demonstrate that semantic graphics are an effective way towards training CNNs to learn higher-order concepts like the crop line and to differentiate between crops and weeds.

Convolutional Encoder–Decoder Network

A convolutional encoder–decoder network is a standard network used for tasks requiring dense pixel-wise predictions like semantic segmentation (Badrinarayanan et al., 2017), computing optical flow and disparity maps (Mayer et al., 2016), and contour detection (Yang et al., 2016). The encoder in the network computes progressively higher-level abstract features as the receptive fields

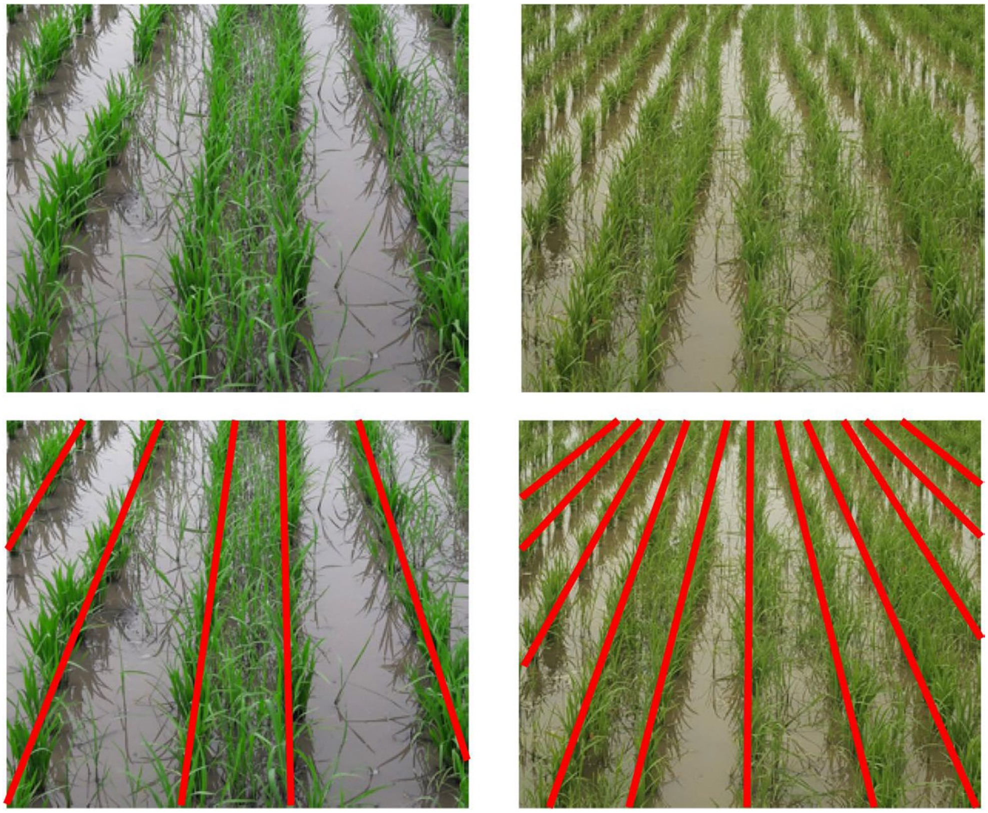


FIGURE 2 | Semantic graphics: (top) images of row-transplanted paddy field. (bottom) Manually marked semantic graphics representing the rows of paddy is superimposed on the original images. Even at places where the paddy lines are rendered indistinguishable due to the heavy presence of weeds, humans can easily figure out the actual lines and represent those using semantic graphics. (Best viewed in color).

in the encoder increase with the depth of the encoder. The spatial resolution of the feature maps is reduced progressively *via* a down-sampling operation, whereas the decoder computes feature maps of progressively increasing resolution *via* un-pooling (Zeiler and Fergus, 2014) or up-sampling. The network has the ability not only to model features like shape or appearance of different classes but also to model long-range spatial relationships. This attribute of modeling local and global features makes this architecture suitable for learning semantics graphics, as shown in **Figure 1**.

Different variations of the encoder–decoder network have been explored in the literature for improved performance. Skip connections (Ronneberger et al., 2015) have been used to recover the fine spatial details during reconstruction which get lost due to successive down-sampling operations involved in the encoder. Addition of larger context information using image-level features (Liu et al., 2015), recurrent connections (Pinheiro and Collobert, 2014; Zheng et al., 2015), and larger convolutional kernels (Peng et al., 2017) has also significantly improved the accuracy of semantic segmentation. Other methods studied for improving semantic segmentation accuracy include hierarchical supervision (Chen et al., 2016) and iterative concatenation of feature maps (Jégou et al., 2017).

In this work, we design an enhanced encoder–decoder network, named "extended skip network" (*ESNet*), to learn the semantic graphics. We demonstrate that the enhanced network exhibits significant performance improvement over the baseline network on the problem of crop line detection and weed detection. We also demonstrate that the proposed method has improved performance on the task of weed detection over a popular bounding box-based object detection method.

MATERIALS AND METHODS

Dataset

Paddy Line Dataset

The focus of this dataset is to extract the rows of paddy, as shown in **Figure 2**. The detected crop lines will enable the navigation of an autonomous agent in the field to accomplish different agricultural tasks like mechanical weeding and precision spraying of herbicides, pesticides, nutrients, etc. *Paddy line dataset* was prepared to evaluate the proposed method. This dataset consists of 350 images of line-transplanted paddy field captured with a handheld camera while walking between the rows of the crop. The dataset contains different scenarios like unevenly spaced rows, weed-infested fields rendering crop rows indistinguishable, and missing crops in a row which make the problem of detecting rows challenging. The images were captured in three different fields at different geographical locations but during the same phonological stage; tillering. Out of the total 350 images, 300 images were used for training and 50 images were set aside for the test. Due to perspective, the rows of rice appear to converge at the horizon and are indistinguishable. In this study we consider only the near-field view for ease of annotation. The rows of rice were annotated with few-pixel-thick lines as shown in **Figure 2**.

The images were down-sampled to a uniform size of 600 × 600 pixels to reduce computation time and memory requirement. Though this dataset has less number of training images, extensive data augmentation was carried out during training by scaling the original image by a factor sampled randomly in the range [0.5, 1.5], rotating the image by an angle sampled randomly from [−15, 15] degrees, mirroring the image randomly along the vertical axis, randomly distorting the image brightness and saturation, and generating random crops of size 512 × 512.

Paddy–Millet Dataset

Paddy and wild millet are similar in appearance; therefore, they are difficult to discriminate. Wild millet are aggressive growers which compete for resources and therefore have to be weeded out for better yield of paddy. The goal is to identify and localize the "weed" wild millet present among the paddy so that an autonomous agricultural robot can eliminate the "weed" while keeping the crop intact.

A dataset, namely, *paddy–millet dataset*, consisting of 760 images of row-transplanted paddy field captured with a handheld camera while walking between rows of the crop, as shown in **Figure 3**, was prepared for the experiments. Out of the total 760 images, 660 images were used for training and 100 images were set aside for testing. Semantic graphics was used to annotate the ground truth data and the base of the respective plant categories; namely, paddy and wild millet were the target key-points to be detected. These key-points were annotated with solid circles, and all unmarked pixels were considered as background. The key-points near the camera viewpoint were annotated with bigger radius circles which could extend well beyond the boundary of the key-point whereas the key-points farther away from the viewpoint were annotated with progressively smaller circles. The semantic graphics used to annotate this dataset can represent multiple higher-level meanings such as category of the plant, location of the key-point, and their distance from a viewpoint. However, only the plant category and location of the key-point are considered in this work.

Each high-resolution image was down-sampled to a uniform size of 288 × 288 pixels to reduce computation time and memory requirement. The data were augmented by mirroring the images randomly along the vertical axis and generating random crops of size 256 × 256 during training.

Architecture of Extended Skip Network

An enhanced fully convolutional encoder–decoder network, called "enhanced skip network" (*ESNet*), as shown in **Figure 4A**, is proposed for end-to-end learning of semantic graphics. The network consists of a contracting encoder and an expanding decoder. The detailed network architecture is given as **Supplementary Material**. The encoder consists of multiple VGGNet-like (Simonyan and Zisserman, 2014) blocks, where each block consists of multiple 3 × 3 convolution followed by batch normalization (Ioffe and Szegedy, 2015) and a nonlinear activation. Each VGG-style block in the encoder, except the last block, is followed by max pooling to reduce the spatial resolution of the feature maps. These blocks are followed by two convolution blocks (with large kernels) → batch normalization → nonlinear activation blocks, which are used at the tail of the encoder to capture a wider context. To reduce the computation overhead, these large convolutions are computed using separable kernels (Jin et al., 2014). The rectified linear unit (ReLU)is used as the nonlinear activation throughout the network.

The decoder is similar in architecture to the encoder but with fewer feature maps for optimized computation and memory requirements. Each block in the decoder is also a repeating structure of up-sampling, followed by multiple 3 × 3 deconvolution, batch normalization, and nonlinear activation operations. The number of feature maps at each level in the decoder is kept constant except for the output layer where it is equal to the number of target classes. The network contains extended skip connections where the feature maps from the encoder are concatenated to the corresponding feature maps in the decoder. The extended skip module consists of a bank of

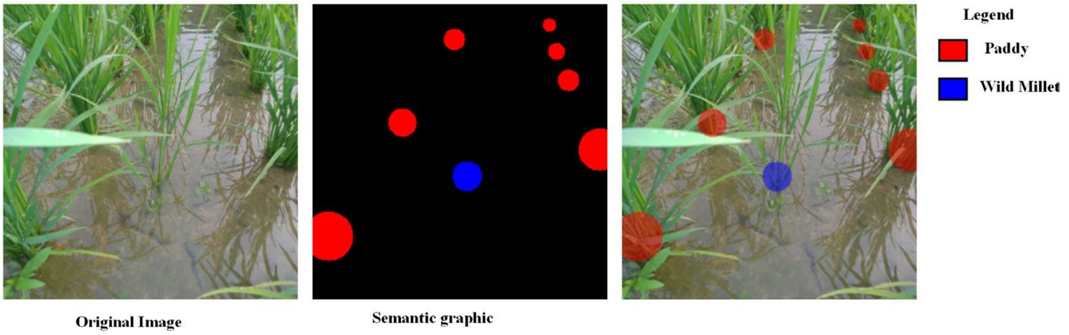


FIGURE 3 | Semantic graphics for paddy–millet dataset. **(A)** Image of a paddy field with wild millet. **(B)** Semantic graphics annotation. The color-filled circles used to annotate the base of the plants indicate multiple meanings such as category of the plant, location of the key-point, and their distance from a viewpoint. **(C)** Annotation superimposed on the source image. (Best viewed in color)

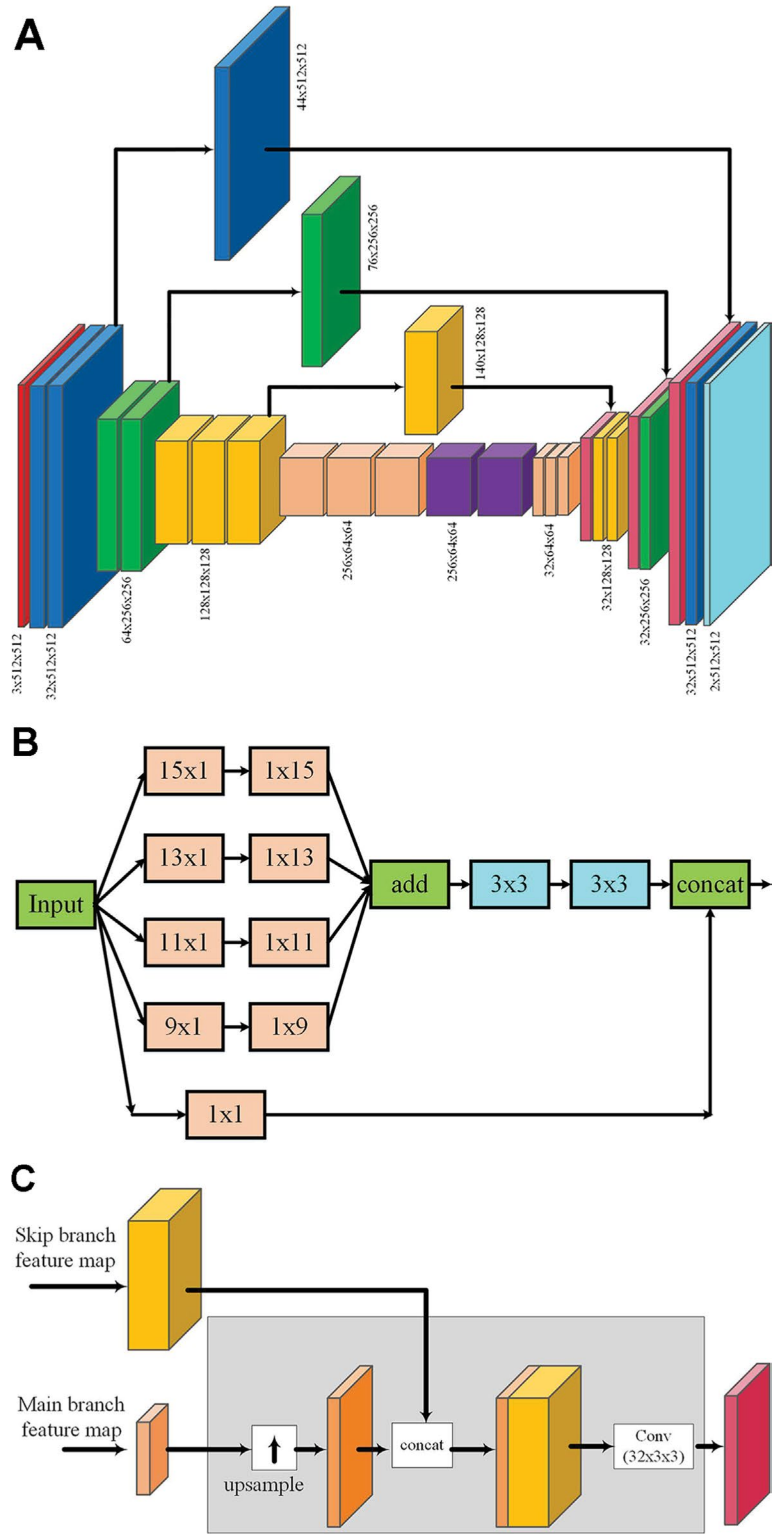


FIGURE 4 | *ESNet*: The proposed extended skip network for end-to-end learning of semantic graphics **(A)**. Diagram representing the output feature maps at each stage of the network **(B)**. Extended skip module: If *Cin* is the number of channels in the input, the output after 1 × 1 convolution has the same number of feature maps as the input, whereas to keep the computational complexity minimal, the number of feature maps at the output after other multi-scale filter banks is kept constant at 12. Hence, the total number of output feature maps of the skip module is (*Cin* + 12). **(C)** The combined operations (up-sampling, concatenation, and convolution with fixed number of output feature maps) involved while merging feature maps from the extended skip module with the decoder. For simplicity, this detailed structure is not shown in **(A)**.

multi-scale filters as shown in **Figure 4B**. The output feature maps of the extended skip module are merged with the corresponding feature maps of the decoder as shown in **Figure 4C**.

The proposed *ESNet* is inspired from and exploits the elements of three different DNNs in a single network, namely, (a) skip layers to recover fine spatial details (Ronneberger et al., 2015), (b) larger convolutional kernels to incorporate a wider image context (Peng et al., 2017), and (c) multi-scale filter bank or "inception" module (Szegedy et al., 2015). However, unlike Ronneberger et al. (2015), whose study used skip layers that are fixed identity connections (copy and concatenate), we propose to make the architecture more general by learning these connections using multi-scale convolution. Large convolutional kernels (Peng et al., 2017) are used to increase the effective receptive field of the network for learning semantic graphics. However, the large kernels are used only at the tail of the encoder and the skip layers.

Finally, the large convolutional kernels in the skip layers are arranged in a multi-scale filter bank module (Szegedy et al., 2015),as shown in **Figure 4B**, to incorporate the required input context during learning without having to empirically find an appropriate kernel size. This module provides multi-scale features which are more efficient for learning semantic graphics than selecting a single-scale context, as will be shown in the results presented in Section Ablation Experiments.

Training Parameters and Evaluation Metrics

The proposed approach is evaluated on the two problems, paddy line detection and wild millet detection, by training the models from scratch. The DNN models for both the tasks are trained by minimizing the pixel-wise cross-entropy loss given as

$$CELoss = -\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\sum_{c=1}^{L} 1_{y_i \in L_c} \log \mathrm{p}[\mathrm{y}_i \in \mathrm{L}_c] \qquad (1)$$

where N is the total number of pixels, L is the number of semantic categories, $1_{\mathrm{y_i \in L_c}}$ is a binary indicator function if category c is the ground truth label for the ith observation, and $\mathrm{p}[\mathrm{y_i} \in \mathrm{L_c}]$ is the predicted probability of the model for that category.

The network shown in **Figure 4** was used for learning the semantic lines. The details of the network are included as **Supplementary Material**. The network was initialized using Xavier initialization (Glorot and Bengio, 2010) and trained on mini-batches of five using the Adam method (Kingma and Ba, 2014) with an exponential decaying learning rate of 10^{-4} for a total of 100 epochs, with all the training images being processed per epoch. As the paddy lines and background pixels are highly imbalanced in each mini-batch, the loss for the two categories is weighted by the class proportion of pixels computed on the training set. The paddy–millet dataset was trained on mini-batches of size 10, with a learning rate of 10^{-4} and a decay factor of 0.94 after successive 10,000 iterations. The network was trained for a total of 60,000 iterations.

The performance of the trained model for both the datasets are evaluated using an intersection-over-union (IoU) metric,

$$IoU = \frac{T \cap P}{T \cup P} \qquad (2)$$

where T is the target and P is the predicted category. In addition to the IoU metric, the precision and recall values for wild millet detection and the average pixel deviation of the predicted line from the ground truth for paddy line detection are also reported. The experiments were conducted in TensorFlow (Abadi et al., 2016) using an NVIDIA Titan-X graphics processing unit (GPU).

Comparison Models

The proposed *ESNet* is compared to other commonly used CNN architectures which produce image-like outputs like the *UNet* (Ronneberger et al., 2015), *FCN8* (Long et al., 2015), and *DeepLabV3* (Chen et al., 2017). The problem of paddy and wild millet detection can be addressed as a bounding box-based object detection and localization approach of *Faster-RCNN* (Ren et al., 2015) also. Therefore, the proposed network is compared with *Faster-RCNN* on the task of paddy and wild millet detection. We also implement a basic encoder–decoder network (*EDNet*) with a comparatively large number of parameters for comparison on the paddy–millet dataset. The details of the networks used in this study are included as **Supplementary Material**.

In the *Faster-RCNN* setting, the paddy–millet dataset was annotated by replacing the semantic graphics with minimum bounding boxes and the problem was solved as a detection and localization problem. The IoU was then computed on the predicted bounding boxes [proposals with class scores $p > 0.8$ with a non-maximum suppression (NMS) threshold of 0.2] and the ground truth annotation. For a fair comparison with the semantic graphics method, the IoU was computed after substituting each bounding box with a maximal circle that fit the box. The detection accuracy was also evaluated using precision and recall values. Any prediction whose center lay within a distance of *d_thresh* (= 15) pixels from the center of its corresponding ground truth was deemed correct (true positive). The *VGG16* (Simonyan and Zisserman, 2014) model pre-trained on ImageNet (Deng et al., 2009) was used to initialize the *Faster-RCNN* and *EDNet* and fine-tuned on the paddy–millet dataset.

Post-Processing: Dominant Semantic Line Extraction

The proposed method of detecting crop lines outputs semantic lines for every visible row of paddy. However, for practical purposes, it is often sufficient and meaningful to detect only a few dominant rows, for example, the host rows and a few of its neighbors. Therefore, a simplified random sample consensus (RANSAC) (Fischler and Bolles, 1981) like post-processing step is employed to extract only the four dominant rows. The output semantic graphics is binarized, and the line segments are sorted according to their length. The longest line segment is chosen as a seed, and a straight line is fit to this segment. All the points within a distance of *d_thresh* (= 15) pixels are assigned as inliers to the initial line, and a new estimate of the line is computed. The resultant line after the second iteration is the first dominant line.

After the first line is detected, all the pixels that are inliers to this line and any other line segments with more than 50% pixel inliers to this line are excluded from processing the remaining dominant line segments. The next longest line segment in the binarized output is then chosen as the second seed, and the above procedure is repeated until the required number of dominant lines are extracted.

The accuracy of the extracted dominant lines is computed using mean pixel deviation (*mpd*) from the ground truth line. The *mpd* is computed as the average of the row-wise difference between the predicted line and its corresponding ground truth. Let (x_p, y) be a point on the predicted line and (x_g, y) be its corresponding point on the ground truth line; then the row-wise pixel deviation (*pd*) is given as $pd = |x_p - x_g|$. Then *mpd* is the mean of *pd* computed across all the lines in the test set.

RESULTS AND DISCUSSION

Crop Line Detection

The quantitative results of the proposed network, *ESNet*, along with the results of *UNet*, *FCN8*, and *DeepLabV3* on the paddy line test set are presented in **Table 1**. From **Table 1**, we see that the proposed network achieves the highest mean intersection over union (mIoU, 62.73%) among all the models considered in this study. The mIoU of the proposed method is 6.29%, 4.56%, and 2.38% higher than that of *UNet*, *DeepLabV3*, and *FCN8*, respectively.

However, the mIoU of the detected semantic lines is less than the mIoU reported on the task of semantic segmentation using similar networks. This is because, unlike the per-pixel ground truth labels used in semantic segmentation, the annotations used for semantic lines are abstract and can be subjective; *i.e.*, annotation of the same line of crop by two human annotators can differ significantly with little overlap between the two. This subjective nature of annotation affects network training and test accuracy. The quantitative analysis on the effect of the subjective nature of annotating semantic graphics is a subject of our future research.

From **Table 1**, we see that the proposed method is slow during inference. Even on a Titan-X GPU, the method runs at 10 fps. This is due to the large-sized kernels used in the network. The bulky Titan-X GPU may not be an optimal choice for use in field robots, and lightweight and more power-efficient GPUs like the Jetson TX2 are more practical. We can expect a considerable slowdown in inference time using the Jetson TX2. However, for a carefully designed system, we can limit the field of view of the vision sensor and restrict the region of interest (ROI) to gain inference speed. From our experiments, it was observed that the proposed network can process 5 fps for an input ROI of 192 × 256. This inference time is expected to be sufficient for any practical application of a slow-moving robot like a tractor running in a flooded rice field.

Some qualitative results on the paddy line test set are presented in **Figure 5.** While the proposed method is able to successfully detect paddy line in well-separated crop rows (first and fifth rows), the crop rows are delineated in high-weed-pressure areas also (second, third, and fourth rows). We also see that the line detection accuracy is higher for rows near the principal axis of the camera lens, whereas it is low for rows lying further away. Training the network on a larger dataset is expected to increase the accuracy of the detected lines throughout the image.

However, as explained in Section Post processing: Dominant Semantic Lines Extraction , for the practical purpose of navigating the field, it is often not necessary to detect crop rows lying further away from the principal axis. Some qualitative results of the extracted dominant lines are presented in **Figure 5D**. The detected dominant lines are in close agreement with the ground truth line, which is also evident from the *mpd* values presented in **Table 2**. Though the difference in mIoU of *UNet* and *ESNet* is high, no significant difference in *mpd* is observed between these two networks. The random sample consensus (RANSAC)-based post-processing compensates for the low mIoU of *UNet*.

Wild Millet Detection

The quantitative results of the proposed method on the paddy–millet dataset along with results of *Faster-RCNN*, *EDNet*, *UNet*, *FCN8*, and *DeepLabV3* are presented in **Table 3**. From our experiments, it was observed that initializing *Faster-RCNN* and *EDNet* with *VGG16* weights pre-trained on ImageNet and fine-tuning only the last few layers resulted in low-accuracy networks. However, a significant increase in mIoU was observed when all the layers were fine-tuned. The lower accuracy of the networks with few layers fine-tuned is due to the difference in the type of classes used in the pre-trained *VGG16* model. The generic "object" features extracted by the pre-trained *VGG16* are not optimal to discriminate between the categories used for this dataset.

Though *EDNet* has a fraction of the parameters, it exhibits an mIoU higher than that of *Faster-RCNN*. This shows that the proposed method can be used to solve the problem of discriminating paddy and wild millet with higher accuracy, fewer parameters, and a simple end-to-end training compared to the existing bounding box approach of object detection. From **Table 3**, we see that the proposed *ESNet* leads to a 0.44% increment in mIoU with significantly less number of parameters than did *EDNet*. We also see that the mIoU of *ESNet* is 22.17%, 6.14%, and 2.42% higher than that of *DeepLabV3*, *UNet*, and *FCN8*, respectively.

TABLE 1 | Comparison of different networks on the paddy line dataset.

Method	#parameters (million)	mIoU (%)	fps (512 × 512 pixels, Titan-X GPU)
UNet	**~2.14**	56.44	21.28
FCN8	~38.16	60.35	21.60
DeepLabV3	~4.14	58.17	**31.30**
***ESNet* (proposed)**	~5.74	**62.73**	10.97

ESNet, enhanced skip network; GPU, graphics processing unit; mIoU, mean intersection over union. The performance is quantified using mIoU. *For Methods, bold is used to highlight the proposed method, whereas bold numbers are used to highlight the best results.*

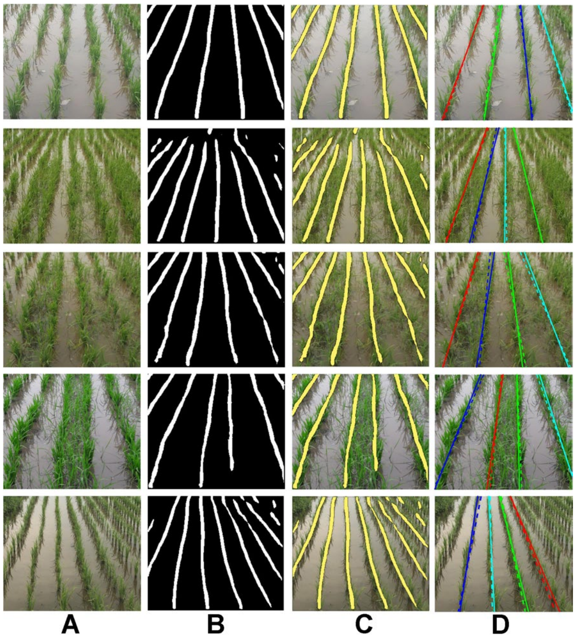


FIGURE 5 | Qualitative results of learning semantic lines using the proposed extended skip network on the paddy line dataset. **(A)** Test images, **(B)** corresponding outputs of the proposed network, **(C)** output superimposed on test image, and **(D)** extracted dominant paddy lines along with the ground truth lines (dotted). (Best viewed in color).

Though *Faster-RCNN* has the highest precision, it has poor recall values. On the other hand, *ESNet* has balanced precision and recall values. From an application perspective, though *Faster-RCNN* is less likely to mistake a rice plant as millet, it is more likely to leave a significant number of weeds in the field undetected. However, *ESNet* detects most of the millets present in the field and is also less likely to mistake rice for millet.

TABLE 2 | Comparison of different networks on the paddy line dataset.

Method	Mean pixel deviation	Deviation [–max, max]
UNet	3.39	[–27, 48]
ESNet (proposed)	**2.89**	**[–24, 24]**

ESNet, enhanced skip network. The performance is quantified using mean pixel deviation of the predicted line from the ground truth line. *For Methods, bold is used to highlight the proposed method, whereas bold numbers are used to highlight the best results.*

Some qualitative results on the paddy–millet dataset are presented in **Figure 6**. While *ESNet* detects most of the millets in the field, it also produces some false positives (second and third rows). Some failure cases (third and fourth rows) are also observed where there is overlap between the two classes. Training the network with a larger dataset is expected to increase the accuracy of the system and reduce the number of failure cases. Though no post-processing has been implemented in the current study, these failure cases can also be reduced by using morphology-based post-processing operations like erosion and filtering.

ABLATION EXPERIMENTS

The effectiveness of the proposed *ESNet* is evaluated by comparing it with different ablated versions. The paddy line test set is used for evaluation, and the results are presented in **Table 4**. From **Table 4**, we see that the addition of large convolutional kernels, at the tail of the encoder of *UNet* to capture a wider image context, improves the mIoU by 3.29%. Further, replacing the *UNet*-style fixed skip connections with the proposed multi-scale filter bank leads to an additional 3% improvement in mIoU.

The motive behind using the multi-scale filters in the skip layers was to incorporate multi-scale features for reconstructing the output without having to rigidly set the convolutional kernel size. To verify this intuition, the multi-scale filter bank module is replaced with single-scale filters of size $k \times k$. Different values of k ranging from 7 to 15 were evaluated, and the results are presented in **Table 5**.

From **Table 5**, we see that the network with $k = 7$ shows the best performance among the different single-scale filters evaluated. It can be observed that there is no straightforward relationship between the size of the kernel and network performance. From **Tables 4** and **5**, we see that the network with the proposed multi-scale filter bank outperforms all other networks with single-scale filters. The increased network capacity of the proposed filter bank may have led to increased

TABLE 3 | Comparison of different variants of *Faster-RCNN* and the proposed method on the paddy–millet dataset.

Method	#parameters (million)	Paddy	Millet	mIoU (%)	Precision (%) (*d_thresh* = 15)		Recall (%) (d_thresh = 15)		F1 score	
					Paddy	Millet	Paddy	Millet	Paddy	Millet
Faster-RCNN	~136	50.07	46.37	48.22	**95.42**	94.69	74.87	68.58	83.90	79.54
EDNet	~15.27	**57.15**	45.52	51.34	90.0	86.29	**92.30**	68.59	**92.19**	76.42
UNet	**~2.14**	48.65	42.62	45.64	91.86	84.37	81.02	69.23	86.10	76.05
FCN8	~38.16	53.30	45.40	49.36	89.29	77.07	89.74	77.56	89.51	77.31
DeepLabV3	~4.14	15.93	43.27	29.61	51.58	**95.69**	33.33	57.05	40.49	71.48
ESNet (proposed)	~5.74	56.53	**47.02**	**51.78**	87.80	84.56	**92.30**	**80.76**	89.99	**82.16**

EDNet, encoder–decoder network; ESNet, enhanced skip network; mIoU, mean intersection over union. The performance is quantified using intersection over union (IoU), precision, and recall. *For Methods, bold is used to highlight the proposed method, whereas bold numbers are used to highlight the best results.*

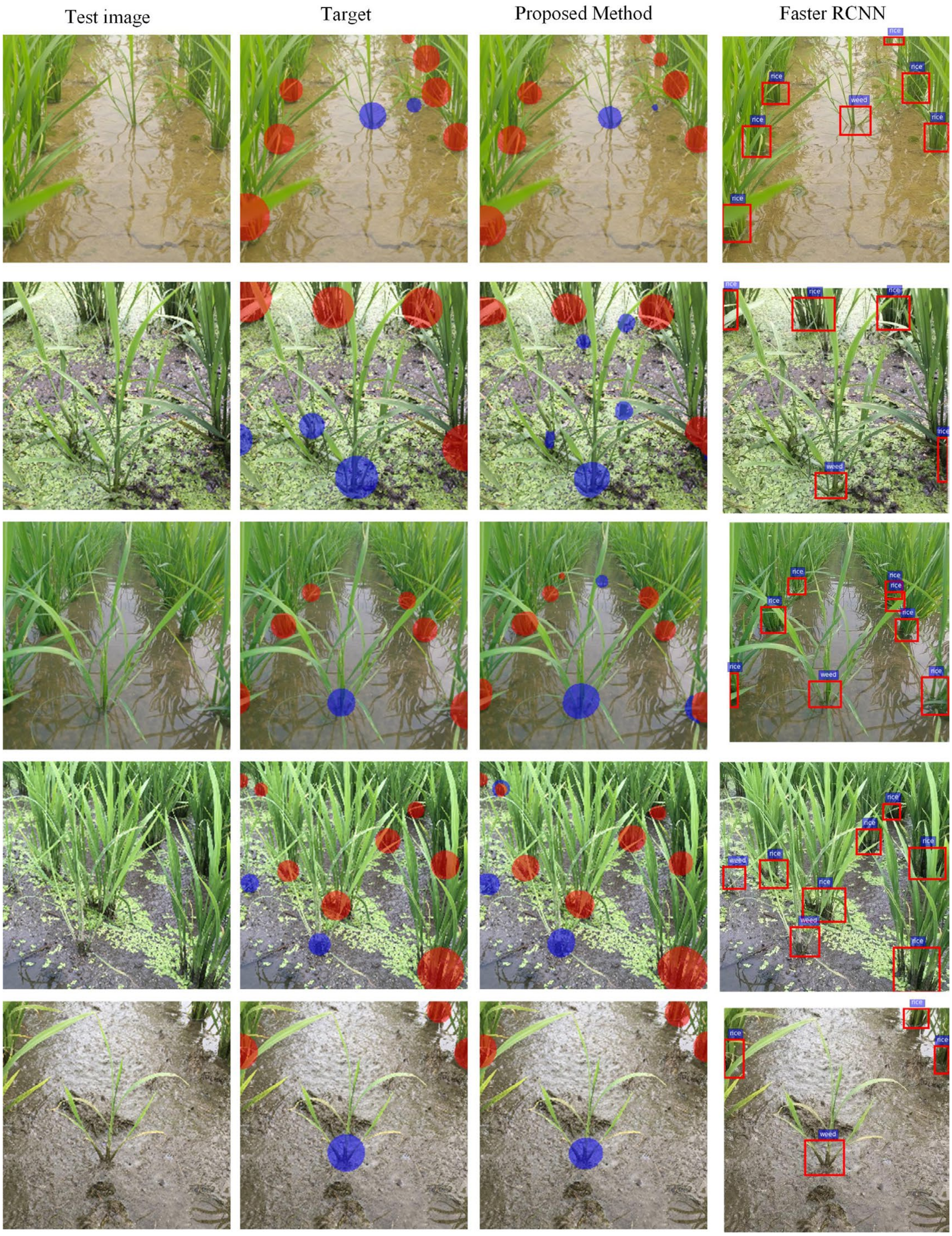


FIGURE 6 | Qualitative results of learning semantic graphics using the proposed convolutional encoder-decoder network on paddy–millet dataset. (Best viewed in color).

accuracy. However, from **Table 5**, we see that increasing the network capacity by simply increasing the number of parameters does not necessarily improve the accuracy. The proposed structure allows the learning algorithm to choose either single-scale features or a combination of multi-scale features, whichever are efficient, and leads to better accuracy.

CONCLUSION

In this study, we proposed a convolutional encoder–decoder network-based system to (a) extract the crop line and (b) differentiate between weeds and crops, in row-transplanted paddy fields. Different from the conventional methods of training DNNs,

TABLE 4 | Ablation experiments to evaluate the effectiveness of the proposed extended skip connections.

Method	Baseline	Skip layer (fixed)	Large conv	Skip layer (multi-scale filters)	Paddy line IoU (%)
UNet	√	√			56.44
*UNet*_WC	√	√	√		59.73
ESNet	√		√	√	62.73

ESNet, enhanced skip network; IoU, intersection over union.

TABLE 5 | Performance comparison using different scales of filter in the skip layer.

K	7	9	11	13	15
IoU (%)	60.68	58.25	59.22	60.43	59.26

IoU, intersection over union.

a novel method of training DNN using "semantic graphics" was proposed. Semantic graphics was introduced to annotate the target functional key-points, semantic regions, or other higher-level concepts which are otherwise challenging to annotate using existing bounding box-based or dense per-pixel-based approaches. An enhanced convolutional encoder–decoder network was then trained to directly learn the concept of crop line and discriminate between weeds and crop using semantic graphics.

Results demonstrating enhanced performance of the proposed method on the paddy line detection problem compared to other existing networks were presented. Experiments demonstrating enhanced performance of the proposed method on detecting paddy and wild millet compared to the more commonly used bounding box-based object detection approach were also presented.

The proposed crop line detection system can be easily extended to extract the rows of different types of crops. While the traditional handcrafted feature-based crop row extraction methods can fail to generalize well in real farm environments, the proposed crop line extraction system exhibits robust performance in real farm environments as demonstrated by the results. Though only wild millet detection is considered in this study, the proposed method can be extended easily to detect any other species of weeds. The crop lines extracted by the proposed method are accurate and can act as a reliable guiding line for an autonomous robot for inter-row weeding, whereas the detection of individual plants and weeds enables autonomous intra-row weeding. A combination of these two approaches for inter-row and intra-row weeding can be used to realize a comprehensive autonomous weeding system.

In the future, we plan to use the semantic graphics-based crop row detection method for vision-based control of an autonomous tractor for unmanned inter-row weeding in paddy and extend the system for intra-row weeding.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

SA designed the study, collected the data, performed the experiments, analyzed the data, and wrote the paper. HY collected the data and performed the experiments. HK supervised and administered the overall project and reviewed and edited the writing.

FUNDING

This work was supported in part by the Korea Research Fellowship Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT (NRF-2015H1D3A1062316), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1A6A1A09031717 and NRF-2019R1A2C1011297).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01404/full#supplementary-material

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "Tensorflow: a system for large-scale machine learning," in *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, (USENIX The Advanced Computing Systems Association) 265–283.

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (12), 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Bakker, T., Wouters, H., Van Asselt, K., Bontsema, J., Tang, L., Müller, J., et al. (2008). A vision based row detection system for sugar beet. *Comput. Electron. Agric.* 60 (1), 87–95. doi: 10.1016/j.compag.2007.07.006

Bearman, A., Russakovsky, O., Ferrari, V., and Fei-Fei, L. (2016). "What's the point: semantic segmentation with point supervision," in *European conference on computer vision* (Cham: Springer), 549–565. doi: 10.1007/978-3-319-46478-7_34

Chen, H., Qi, X. J., Cheng, J. Z., and Heng, P. A. (2016). "Deep contextual networks for neuronal structure segmentation," in *Thirtieth AAAI conference on artificial intelligence*. Association for the Advancement of Artificial Intelligence (AAAI) Publications.

Chen, L. C., Papandreou, G., Schroff, F., and Adam, H. (2017). *Rethinking atrous convolution for semantic image segmentation*. (Cornell University, USA: arXiv preprint arXiv:1706.05587).

Choi, K. H., Han, S. K., Han, S. H., Park, K. H., Kim, K. S., and Kim, S. (2015). Morphology-based guidance line extraction for an autonomous weeding robot in paddy fields. *Comput. Electron. Agric.* 113, 266–274. doi: 10.1016/j.compag.2015.02.014

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: a large-scale hierarchical image database. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 248–255. doi: 10.1109/CVPR.2009.5206848

Dyrmann, M., Jørgensen, R. N., and Midtiby, H. S. (2017). RoboWeedSupport-detection of weed locations in leaf occluded cereal crops using a fully convolutional neural network. *Adv. Anim. Biosci.* 8 (2), 842–847. doi: 10.1017/S2040470017000206

Dyrmann, M., Karstoft, H., and Midtiby, H. S. (2016). Plant species classification using deep convolutional neural network. *Biosyst. Eng.* 151, 72–80. doi: 10.1016/j.biosystemseng.2016.08.024

Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24 (6), 381–395. doi: 10.1145/358669.358692

Fuentes, A. F., Yoon, S., Lee, J., and Park, D. S. (2018). High-performance deep neural network-based tomato plant diseases and pests diagnosis system with refinement filter bank. *Front. Plant Sci.* 9, 1–15. doi: 10.3389/fpls.2018.01162

Glorot, X., and Bengio, Y. (2010). "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. (JMLR, Proceedings of Machine Learning Research). 249–256.

Grinblat, G. L., Uzal, L. C., Larese, M. G., and Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. doi: 10.1016/j.compag.2016.07.003

Guerrero, J. M., Ruz, J. J., and Pajares, G. (2017). Crop rows and weeds detection in maize fields applying a computer vision system based on geometry. *Comput. Electron. Agric.* 142, 461–472. doi: 10.1016/j.compag.2017.09.028

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. . 2961–2969. doi: 10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. . 770–778. doi: 10.1109/CVPR.2016.90

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. . 4700–4708. doi: 10.1109/CVPR.2017.243

Ienco, D., Gaetano, R., Dupaquier, C., and Maurel, P. (2017). *Land cover classification via multi-temporal spatial data by recurrent neural networks*. (Cornell University, USA: arXiv preprint arXiv:1704.04055). doi: 10.1109/LGRS.2017.2728698

Ioffe, S., and Szegedy, C. (2015). *Batch normalization: accelerating deep network training by reducing internal covariate shift*. (Cornell University, USA: arXiv preprint arXiv:1502.03167).

Jégou, S., Drozdzal, M., Vazquez, D., Romero, A., and Bengio, Y. (2017). The one hundred layers tiramisu: fully convolutional densenets for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,* 11–19.

Jiang, G., Wang, X., Wang, Z., and Liu, H. (2016). Wheat rows detection at the early growth stage based on Hough transform and vanishing point. *Comput. Electron. Agric.* 123, 211–223. doi: 10.1016/j.compag.2016.02.002

Jin, J., Dundar, A., and Culurciello, E., (2014). *Flattened convolutional neural networks for feedforward acceleration*. (Cornell University, USA: arXiv preprint arXiv:1412.5474).

Kingma, D. P., and Ba, J. (2014). *Adam: a method for stochastic optimization*. (Cornell University, USA: arXiv preprint arXiv:1412.6980).

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*. (Neural Information Processing Systems Foundation, Inc.), 1097–1105.

Kussul, N., Lavreniuk, M., Skakun, S., and Shelestov, A. (2017). Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* 14 (5), 778–782. doi: 10.1109/LGRS.2017.2681128

Lin, D., Dai, J., Jia, J., He, K., and Sun, J. (2016). Scribblesup: scribble-supervised convolutional networks for semantic segmentation. 3159–3167. doi: 10.1109/CVPR.2016.344

Liu, W., Rabinovich, A., and Berg, A. C. (2015). Parsenet: looking wider to see better. arXiv preprint arXiv:1506.04579.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 3431–3440. doi: 10.1109/CVPR.2015.7298965

Ma, X., Deng, X., Qi, L., Jiang, Y., Li, H., Wang, Y., et al. (2019). Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields. *PloS One* 14 (4), e0215676. doi: 10.1371/journal.pone.0215676

Mayer, N., Ilg, E., Hausser, P., Fischer, P., Cremers, D., Dosovitskiy, A., et al. (2016). A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. . 4040–4048. doi: 10.1109/CVPR.2016.438

Milioto, A., Lottes, P., and Stachniss, C. (2018). Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNs. . 2229–2235. doi: 10.1109/ICRA.2018.8460962

Mohanty, S. P., Hughes, D. P., and Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7, 1419. doi: 10.3389/fpls.2016.01419

Montalvo, M., Pajares, G., Guerrero, J. M., Romeo, J., Guijarro, M., Ribeiro, A., et al. (2012). Automatic detection of crop rows in maize fields with high weeds pressure. *Expert Syst. Appl.* 39 (15), 11889–11897. doi: 10.1016/j.eswa.2012.02.117

Papandreou, G., Chen, L. C., Murphy, K. P., and Yuille, A. L. (2015). Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. . 1742–1750. doi: 10.1109/ICCV.2015.203

Peng, C., Zhang, X., Yu, G., Luo, G., and Sun, J. (2017). Large kernel matters–improve semantic segmentation by global convolutional network. . 4353–4361. doi: 10.1109/CVPR.2017.189

Pinheiro, P. O., and Collobert, R. (2014). "Recurrent convolutional neural networks for scene labeling," in *Proceedings of the 31st International Conference on International Conference on Machine Learning-Volume.* (JMLR, Proceedings of Machine Learning Research). *32*, I–82. JMLR. org.

Pinheiro, P. O., and Collobert, R. (2015). From image-level to pixel-level labeling with convolutional networks. 1713–1721. doi: 10.1109/CVPR.2015.7298780

Potena, C., Nardi, D., and Pretto, A. (2016). "Fast and accurate crop and weed identification with summarized train sets for precision agriculture," in *International Conference on Intelligent Autonomous Systems* (Cham: Springer), 105–121. doi: 10.1007/978-3-319-48036-7_9

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: unified, real-time object detection. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788. doi: 10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-cnn: towards real-time object detection with region proposal networks. *Adv. Neural Inf. Processing Syst.* 91–99. doi: 10.1109/tpami.2016.2577031

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention* (Cham: Springer), 234–241. doi: 10.1007/978-3-319-24574-4_28

Simonyan, K., and Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. (Cornell University, USA: arXiv preprint arXiv:1409.1556).

Søgaard, H. T., and Olsen, H. J. (2003). Determination of crop rows by image analysis without segmentation. *Comput. Electron. Agric.* 38 (2), 141–158. doi: 10.1016/S0168-1699(02)00140-0

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15 (1), 1929–1958.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, et al. (2015). Going deeper with convolutions. . 1–9. doi: 10.1109/CVPR.2015.7298594

Tang, J., Wang, D., Zhang, Z., He, L., Xin, J., and Xu, Y. (2017). Weed identification based on K-means feature learning combined with convolutional neural network. *Comput. Electron. Agric.* 135, 63–70. doi: 10.1016/j.compag.2017.01.001

Yang, J., Price, B., Cohen, S., Lee, H., and Yang, M. H. (2016). Object contour detection with a fully convolutional encoder-decoder network. . 193–202. doi: 10.1109/CVPR.2016.28

Zeiler, M. D., and Fergus, R. (2014). "Visualizing and understanding convolutional networks," in *European conference on computer vision* (Cham: Springer), 818–833. doi: 10.1007/978-3-319-10590-1_53

Zhang, Q., Chen, M. S., and Li, B. (2017). A visual navigation algorithm for paddy field weeding robot based on image understanding. *Comput. Electron. Agric.* 143, 66–78. doi: 10.1016/j.compag.2017.09.008

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., et al. (2015). Conditional random fields as recurrent neural networks. 1529–1537. doi: 10.1109/ICCV.2015.179

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 05 November 2019
doi: 10.3389/fpls.2019.01372

Edited by:
Roland Pieruschka,
Julich Research Centre,
Germany

Reviewed by:
Honghong Wu,
Huazhong Agricultural University,
China
Vijay Gahlaut,
University of Delhi, India

***Correspondence:**
Giao N. Nguyen
giao.nguyen@agriculture.vic.gov.au
Surya Kant
surya.kant@agriculture.vic.gov.au

Some results of this article were originally from the poster presentations at the 5th International Plant Phenotyping Symposium, 2nd- 5th October 2018, Adelaide, Australia

†Present address:
Lance Maphosa,
NSW Department of Primary
Industries, Yanco Agricultural
Institute, Yanco,
NSW, Australia

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 28 June 2019
Accepted: 04 October 2019
Published: 05 November 2019

Citation:
Nguyen GN, Maharjan P, Maphosa L, Vakani J, Thoday-Kennedy E and Kant S (2019) A Robust Automated Image-Based Phenotyping Method for Rapid Vegetative Screening of Wheat Germplasm for Nitrogen Use Efficiency.
Front. Plant Sci. 10:1372.
doi: 10.3389/fpls.2019.01372

A Robust Automated Image-Based Phenotyping Method for Rapid Vegetative Screening of Wheat Germplasm for Nitrogen Use Efficiency

Giao N. Nguyen[1], Pankaj Maharjan[1], Lance Maphosa[1†], Jignesh Vakani[1], Emily Thoday-Kennedy[1] and Surya Kant[1,2]**

[1] *Agriculture Victoria, Grains Innovation Park, Horsham, VIC, Australia* [2] *Centre for Agricultural Innovation, The University of Melbourne, Melbourne, VIC, Australia*

Nitrogen use efficiency (NUE) in crops is generally low, with more than 60% of applied nitrogen (N) being lost to the environment, which increases production costs and affects ecosystems and human habitats. To overcome these issues, the breeding of crop varieties with improved NUE is needed, requiring efficient phenotyping methods along with molecular and genetic approaches. To develop an effective phenotypic screening method, experiments on wheat varieties under various N levels were conducted in the automated phenotyping platform at Plant Phenomics Victoria, Horsham. The results from the initial experiment showed that two relative N levels—5 mM and 20 mM, designated as low and optimum N, respectively—were ideal to screen a diverse range of wheat germplasm for NUE on the automated imaging phenotyping platform. In the second experiment, estimated plant parameters such as shoot biomass and top-view area, derived from digital images, showed high correlations with phenotypic traits such as shoot biomass and leaf area seven weeks after sowing, indicating that they could be used as surrogate measures of the latter. Plant growth analysis confirmed that the estimated plant parameters from the vegetative linear growth phase determined by the "broken-stick" model could effectively differentiate the performance of wheat varieties for NUE. Based on this study, vegetative phenotypic screens should focus on selecting wheat varieties under low N conditions, which were highly correlated with biomass and grain yield at harvest. Analysis indicated a relationship between controlled and field conditions for the same varieties, suggesting that greenhouse screens could be used to prioritise a higher value germplasm for subsequent field studies. Overall, our results showed that this phenotypic screening method is highly applicable and can be applied for the identification of N-efficient wheat germplasm at the vegetative growth phase.

Keywords: high-throughput phenotyping, digital imaging, controlled environment, plant growth analysis, broken-stick model

INTRODUCTION

Over the past five decades, there has been a significant increase in global food production resulting, in part, from the major contribution of substantial nitrogen (N) fertilizer application. Nevertheless, food production must be increased to sufficiently meet the projected world population of 9 billion people by 2050 (Godfray et al., 2010). However, with current agricultural practices this means that more than 240 million metric tons of additional N fertilizer would be utilized (Good et al., 2004). Approximately 110 million metric tons of synthetic N fertilizers are used annually for farming and food crop production globally (International Fertilizer Industry Association, 2013). However, nitrogen use efficiency (NUE) is generally low, with only 40% of applied N being taken up by the crop plants, while the remainder is lost to the environment resulting in increased production costs and environmental pollution (Good and Beatty, 2011; Nguyen et al., 2015), as well as affecting human health (Ahmed et al., 2017). Annually, excessive N application is estimated to cost up to €320 billion of damage to the environment in Europe (Brink et al., 2011). Proper N fertilization management practices are expected to reduce N fertilizer application while maintaining stable crop production (Good et al., 2004; Good and Beatty, 2011). To achieve this, improving NUE in crops is one of the most effective ways to ensure current crop yields can be maintained while N supply is reduced, or increasing crop yields with an optimum N input (Cormier et al., 2013).

NUE is a complex concept, but can be defined as the function of two varying components: N uptake efficiency (NUpE)—the plant's ability to obtain N from the soil—and N utilisation efficiency (NUtE)—the plant's ability to assimilate and remobilize absorbed N into the grain (Moll et al., 1982; Craswell and Godwin, 1984; Xu et al., 2012). In simplistic terms, NUE is determined by a plant's ability to utilise supplied N into biomass and grain yield, and can be calculated as the ratio of biomass or grain yield to the amount of N inputs (Nguyen et al., 2016; Hawkesford, 2017). Multiple approaches have been proposed for NUE improvement in crops that include applications of advanced agronomical practices, genetic improvement through molecular breeding, and genetic engineering (Hirel et al., 2007; Hawkesford, 2014; Beatty and Good, 2018; Nguyen and Kant, 2018).

Among these, molecular breeding for N-efficient varieties is considered the most effective method to lift NUE in wheat (Cormier et al., 2016), although this approach depends on the availability of reliable and accurate molecular markers linked to N-efficient genes for marker assisted and genomic selection (Cabrera-Bosquet et al., 2012; Garnett et al., 2015; Han et al., 2015). However, molecular breeding for N-efficient varieties is still a daunting task, given that NUE is a polygenic trait with complex interactions, and associated genes are heavily influenced by environmental conditions such as varying soil N, soil type, rainfall pattern and soil water availability (Cormier et al., 2013; Lammerts Van Bueren and Struik, 2017; Nguyen et al., 2017). A large volume of high-quality phenotypic data is needed to dissect NUE's genetic influences into smaller manageable and measurable components, and to derive reliable and accurate molecular markers or genomic estimated breeding values (Araus and Cairns, 2014; Nadeem et al., 2018). Thus, molecular breeding goals for N-efficient varieties rely heavily on the deployment of effective phenotyping methods (Araus and Kefauver, 2018; Araus et al., 2018). However, the absence of a robust, high-throughput, and reliable phenotyping method that is powerful enough to break down genetic components is currently limiting breeders' efforts to make a breakthrough in the genetic improvement of NUE traits (Nguyen and Kant, 2018). Effective screening methods for identifying N-efficient germplasm that performs consistently in the greenhouse and field conditions are required to facilitate breeding outcomes (Garnett et al., 2015; Nguyen and Kant, 2018). High-throughput phenotyping methods which can effectively differentiate the performance of germplasm at early growth stages and predict their performance at harvest are urgently required (Sharma and Bali, 2018).

Over the last two decades, proximal sensing technology has become one of the most promising high-throughput phenotyping approaches that can provide key non-destructive support in measuring performance and predicting crop yield in controlled and field environments (Fiorani and Schurr, 2013; Araus and Cairns, 2014; Araus and Kefauver, 2018). This technology was fundamentally developed on the principle that the light reflectance from the interaction between the natural light spectrum with plant components could provide accurate information on the morphological and physiological status of plants (Homolová et al., 2013; Fahlgren et al., 2015). The light reflectance captured by specially designed optical instruments can then be used to generate vegetation indices (VIs) and digital plant objects. Once validated, these derived VIs or digital plant objects can be used as proxies of various plant traits, such as shoot biomass, leaf area or N content, to compare the performance of individual varieties (Nguyen and Kant, 2018). For instance, the most common vegetation index i.e. normalised difference vegetation index (NDVI), is often used to assess a plant response to varying N supplies (Nguyen et al., 2016). Moreover, the application of optical devices such as Red-Green-Blue (RGB) digital cameras has also been used for measuring crop growth, phenology and yield components, as well as the development of VIs (Casadesús et al., 2007; Casadesús and Villegas, 2014; Nguyen et al., 2018). Unlike conventional spectral indices, RGB indices are not affected at long wavelengths by elements such as crop architecture and soil cover, and they were shown to outperform the conventional spectral indices in measuring crop growth and N use to some extent (Araus and Kefauver, 2018). There have been a few reports on the application of automated RGB imaging platforms to study phenotypic responses of grass species (Poiré et al., 2014), and sorghum (Neilson et al., 2015; Berry et al., 2018) to N fertilizer under controlled environments. Recently, ground and aerial based RGB imaging has been successfully used to study NUE in wheat and maize under field conditions (Prey et al., 2018; Buchaillot et al., 2019). However, a robust method using imaging technology for screening of crop germplasm for NUE under controlled conditions is still to be reported. To the best of our knowledge, this is the first study describing a vegetative phenotypic screening method for NUE improvement in wheat using automated imaging phenotyping technology under controlled environments.

The aim of this work was to develop a high-throughput and high-resolution phenotyping protocol that can effectively screen wheat varieties at the vegetative growth phase for NUE improvement in controlled environments, and then to compare the performance of wheat varieties under controlled and field conditions with respect to NUE. The perspectives of applying sensing technologies for phenotypic screens of wheat germplasm for NUE under field conditions are also discussed.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Fifteen genotypically diverse wheat (*Triticum aestivum* L.) varieties used in our previous field trial (Nguyen et al., 2016), where they were shown to be differentially responsive to N, were studied in two separate experiments at Plant Phenomics Victoria, Horsham, described in detail by Nguyen et al. (2018). Briefly, the automated phenotyping platform consists of a conveyor belt system, a watering and weighing station (**Figure 1A**), and an imaging chamber with a Scannalyzer 3D imaging system (LemnaTec GmbH, Aachen, Germany; **Figure 1B**).

In the first experiment, two varieties, Bobwhite and Chara, were used for the identification of appropriate low and optimal N levels for further screens. White plastic pots (200 mm diameter x 190 mm deep, Garden City Plastics Pty Ltd, Victoria, Australia) were filled with 3.5 litres of cereal standard soil mix without added fertilizers (Biogro, South Australia, Australia). To ensure that each pot was filled with an equal amount of soil, pots were weighed prior to sowing. Three seeds were sown per pot on rolling benches and thinned to one plant at 3 leaf stage (~ 2 weeks old). To avoid water leaking, pots were placed on white saucers for the duration of the experiment. On a weekly or fortnightly basis, between 100 ml and 200 ml of nutrient solution, components listed in **Supplementary Table 1**, was supplied depending on the crop growth stages (vegetative or reproductive). Ferrous fertilizer (Fe^{3+}) was supplied as Librel® Fe-LO (CW Pacific Specialties Pty Ltd, NSW, Australia) and phosphorus fertilizer (PO_4^{-3}) was supplied as a phosphate buffer with pH 6. Six relative N levels using KNO_3 as the sole N source were applied; 2 mM, 5 mM, 10 mM, 15 mM, 20 mM, and 25 mM N. Water was supplied adequately and equally among pots to keep plants growing healthily. Due to some greenhouse conditions, we could not continue the experiment until crop maturity. However, the crop growth and canopy development were carefully observed and used as guidance for the second experiment.

In the second experiment, 15 wheat varieties were screened for their responsiveness to two N levels, 5 mM and 20 mM, designated as low and optimum N. Under our observation, the total amount of N supplied for these two levels was equivalent to 147 mg and 588 mg N per pot, respectively, which were similar to the N levels supplied in the previous study by Malik et al. (2016). Pot preparation and plant growth management were conducted similarly to the first experiment at the Plant Phenomics Victoria, Horsham. Three weeks after sowing, pots were loaded and laid out on the conveyor belt system in a split-plot design with 15 replicates per N treatment, where N was the main treatment and variety was the sub-treatment. The growth conditions in the greenhouse were 24°C during the day and 16°C during the night with a 12 h photoperiod. To keep plants upright for imaging, a cage was placed into each pot, which was painted blue so that its images could easily be segmented and removed as background noise during image analysis.

Wheat plants at both N treatments were divided into two sets. The first set of plants was harvested at 49 days after sowing (DAS) for the vegetative growth evaluation. The second set continued to grow until maturity and was harvested for grain yield and yield attributes assessment. Heading date was recorded as the day the first head in each pot completely emerged as described previously by Guedira et al. (2016). Physiological maturity date was recorded as the day when the lower glumes of spikes completely lost all the green colour.

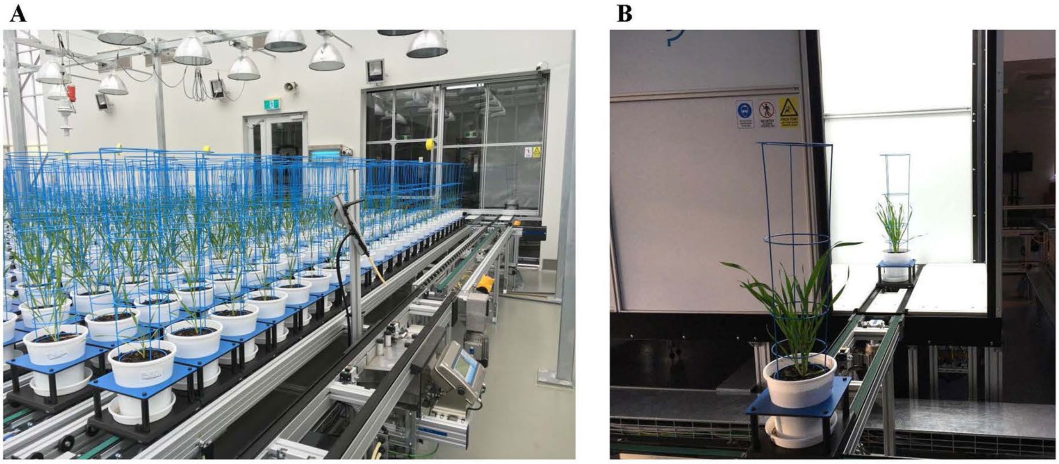


FIGURE 1 | The automated system in Plant Phenomics Victoria, Horsham. **(A)** Pots laid out on conveyor system with a watering and weighing station. **(B)** Plants moving to imaging booth containing a side and a top RGB camera for image acquisition.

RGB Image Acquisition and Analysis

Digital image capture and analysis were implemented following the procedure described previously by Nguyen et al. (2018). In brief, after loading plants on the automated system, they were imaged twice a week by the Scannalyzer 3D plant-to-sensor imaging unit which consists of two 28.8 megapixel RGB cameras (a side and a top camera), and model Prosilica GT6600C (Allied Vision Technologies, Stadtroda, Germany) (**Figure 1B**). Three colour images were acquired from 3 sides of the plant after consecutive rotations of 0, 120, 240 degrees, and a top-view image of the plant was also acquired. Images taken were automatically recorded in the database server which is managed by LemnaBase software (LemnaTec GmbH, Aachen, Germany). **Figure 2A** illustrates a simplified image processing algorithm, containing key steps and devices of LemnaGrid software (LemnaTec GmbH, Aachen, Germany). To analyse the images, the region of interest consisting of whole plant parts in the raw images (**Figure 2B**, i) was separated from the background by Vessel Cofig Marker device. In subsequent steps, the noise was removed from the region of interest and purified by LabtoGrey Converter and Threshold devices (**Figure 2A**) and finally the digital plant objects were determined (**Figure 2B**, ii). This object was used to estimate morphological and physiological traits of the plant (the bright green objects, **Figure 2B**, iii). **Table 2** lists traits measured by digital plant objects and conventionally destructive methods.

Destructive Phenotyping

The first set of plants was destructively harvested at 49 DAS by cutting plants above the soil level in the pots. Whole plants were immediately weighed to determine fresh biomass (MB) per pot. All leaves from the plant were then detached from stems and used to determine leaf area (LA) using a Portable Area Meter, model LI-3050A (LI-COR Inc, Lincoln, Nebraska, USA).

The second set of plants was harvested at physiological maturity to determine yield and yield attributes as described in **Table 1**. All above ground parts of the plant were removed from pots and oven-dried at 65°C for 5 days. After the measurement of total dry biomass (DW), the spikes were separated from stems and counted to determine number of spikes per pot (SN), as well as number of grains per spike (GN), total grain yield (GY) and 1000-grain weight (1000-GW).

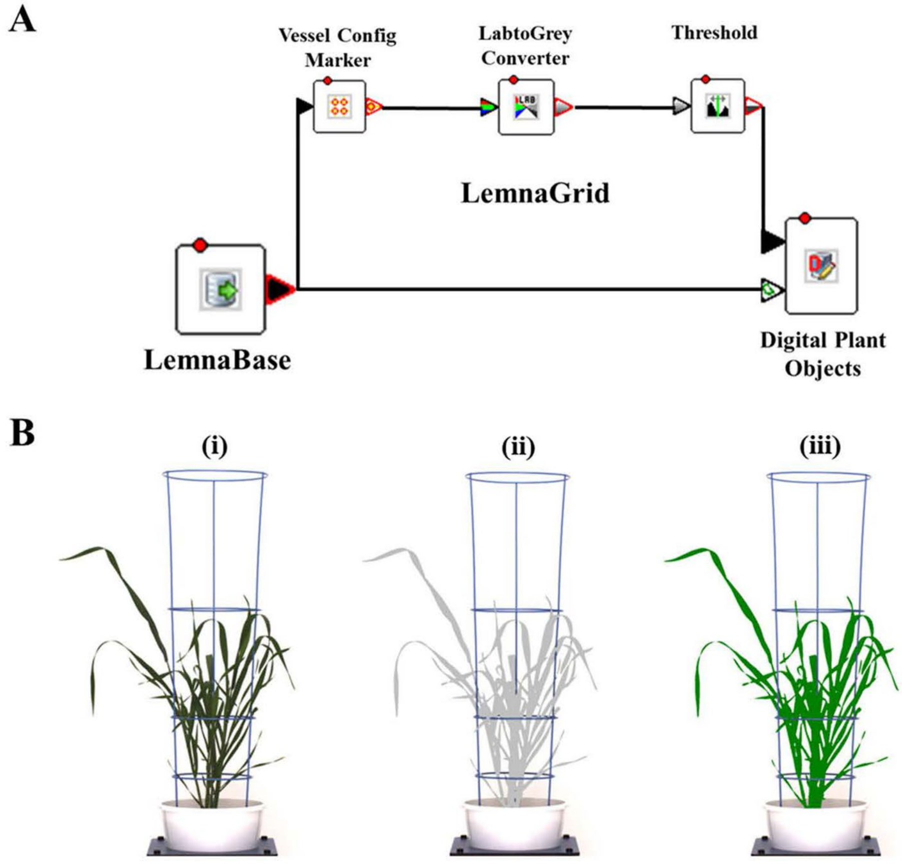


FIGURE 2 | Image acquisition and analysis by the automated plant phenotyping system, Plant Phenomics Victoria, Horsham. Panel **(A)**, a simplified image processing algorithm comprising key steps and devices of LemnaGrid software. Panel **(B)**, i) a raw side-view image of a wheat plant cv. Bobwhite at 49 DAS; ii) the color classification and identification of corresponding object (the grey plant); iii) the processed image object (the bright green plant). Estimated shoot biomass is the pixel sum of highlighted green objects in processed images.

TABLE 1 | Wheat traits measured by digital RGB imaging and destructive methods.

Traits	Abbreviation	Unit	Description
Estimated shoot area	EB	kilopixel (kPix)	The pixel sums of three side-views and top-view image of the plant
Top-view area	TVA	kPix	The pixel sums of the top-view image
Measured shoot bimass	MB	gram (g)	Destructive biomass harvest at 49 DAS
Measured leaf area	LA	cm^2	Total leaf area per plant per pot
1000-grain weight	1000-GW	gram (g)	Manually count and weigh 1000 grains
Number of spikes	SN	spike	Manually count the number of spikes per pot
Number of grain per spike	GN	grain	Manually count the number of grain per spike
Dry biomass	DW	gram (g)	Determined by manually weighing total dry biomass per pot
Grain yield	GY	gram (g)	Determined by manually weighing total seed yield per pot

Water Soluble Carbohydrate Assay

Water soluble carbohydrate (WSC) concentration of plant shoots was determined by near-infrared reflectance (NIR) spectroscopy. The ground plant samples were measured using FOSS XDS Rapid Content Analyser (FOSS, Hillerød, Denmark). The WSC composition was predicted using the WINISI 4 NIR calibration with standard error of prediction of 1% and R^2 of 0.98. The reference WSC method used for validating NIR data was adopted from Maharjan et al. (2018).

Shoot and Grain N Concentration Measurement

The procedure for shoot and grain N concentration measurement was described previously by Nguyen et al. (2016). Briefly, a subset of samples was randomly collected from shoot dry biomass or grain samples and ground to fine powder by a grinder (Cyclotec; Foss, Hillerød, Denmark). Total N concentration in the shoot and grain samples were determined by NIR spectroscopy (Foss XDS Rapid Content Analyser) (AACC method 39-25) and calculated on a dry weight basis.

Nitrogen Use Efficiency

The NUE of wheat plant biomass and grain per pot was calculated according to the formula adapted from Crasswell and Godwin (1984) with modifications.

$$\mathrm{NUE_b} = \mathrm{W_{biomass}} / \mathrm{W_{N\ inputs}} \tag{1}$$

$$\mathrm{NUE_g} = \mathrm{W_{grain}} / \mathrm{W_{N\ inputs}} \tag{2}$$

where $\mathrm{NUE_b}$ (1) and $\mathrm{NUE_g}$ (2) are the nitrogen use efficiency of wheat plants per pot in regard to biomass and grain yield, respectively; $\mathrm{W_{biomass}}$ and $\mathrm{W_{grain}}$ are the weight (grams) of plant biomass and grain yield per pot at harvest, respectively; $\mathrm{W_{N\ inputs}}$ is the amount (grams) of nitrogen inputs.

Comparison of the Performance of Wheat Varieties Under Greenhouse and Field Conditions

To compare the performance of 15 wheat varieties grown under greenhouse and field conditions, we utilized the published data set from our previous field trial (Nguyen et al., 2016). Harvest plant dry biomass and grain yield from this study were compared with the performance of the same wheat genotypes grown in the field trial.

Plant Growth Model and Statistical Analysis

Imaging-derived and manually measured data were initially checked for outliers by using GENSTAT statistical software version 18.0 statistical software (VSN International Ltd, Hemel Hempstead, UK). Two-way analysis of variance (ANOVA) was performed to determine varietal effects by using the same software. The procedure for the selection of the best fit nonlinear regression plant growth model based on the estimated biomass and statistical analyses was adopted from Nguyen et al. (2018). In brief, the biomass accumulation of wheat plants over the growth period follows a sigmoidal growth pattern (Malhi et al., 2006; Archontoulis and Miguez, 2015) and the “broken-stick” statistical model fitting two straight lines using GENSTAT was used to identify the linear growth phase of wheat plants. Linear regressions and Pearson’s correlation coefficient (*r*) were used to determine the correlations between estimated and measured plant traits by using R statistical software (version R-3.5.0) (R Core Team, 2017).

RESULTS

Wheat Varietal Response to Various N Supplies

The initial experiment screened six N concentrations, 2 mM; 5 mM; 10 mM; 15 mM; 20 mM and 25 mM on two bread wheat cultivars, Bobwhite and Chara. Overall, data showed that increased N concentrations resulted in a larger canopy and higher biomass accumulation (**Figure 3**). Plants did not grow well at 2 mM N, whereas they appeared over grown at 25 mM N with very large canopies. Leaf overlap due to large canopies can reduce the correlation of the digital image to actual biomass, especially when leaf area index > 3 (Serrano et al., 2000). Therefore, the two N concentrations, 5 mM and 20 mM were chosen and designated as low and optimum N levels, for subsequent screens of wheat genotypes for NUE traits.

In the second experiment, 15 genetically diverse wheat cultivars including Bobwhite and Chara were grown under the

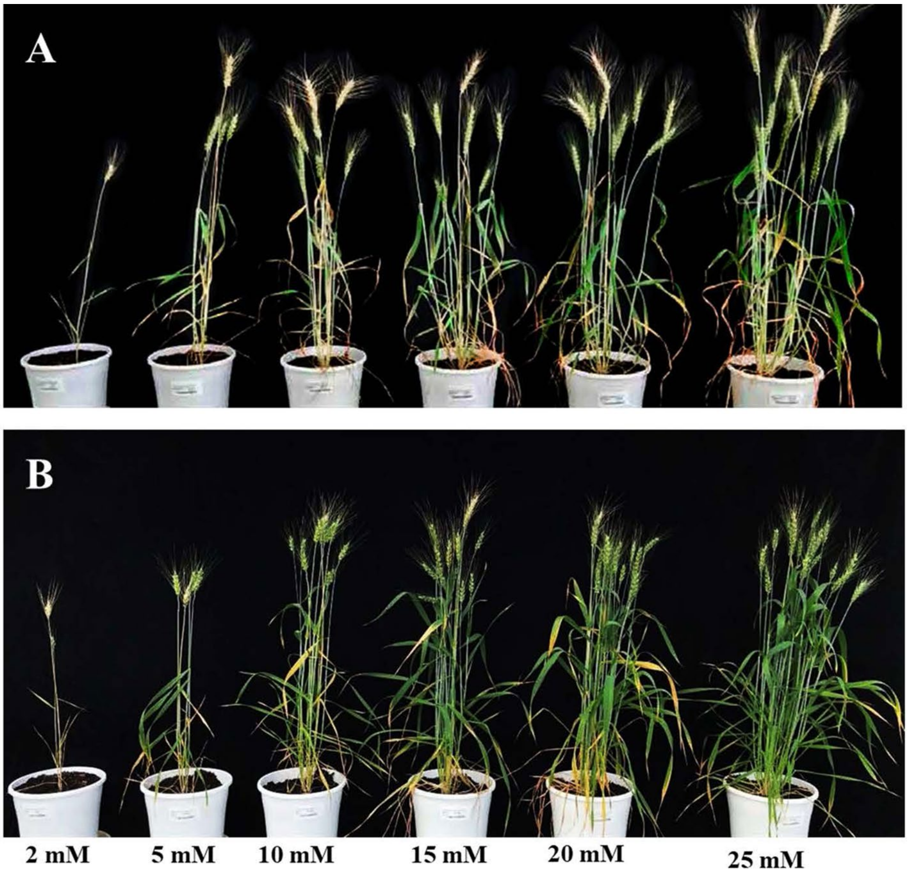


FIGURE 3 | Wheat varieties Bobwhite **(A)** and Chara **(B)** grown under six levels N levels: 2 mM, 5 mM, 10 mM, 15 mM, 20 mM, and 25 mM. Photos taken at 90 DAS.

low and optimum N levels. Overall, all varieties showed positive responses to the increased N supply, resulting in longer growth duration, higher DW, GY, shoot and grain N concentration (**Supplementary Figure 1** and **Table 2**). Data showed that wheat varieties responded differentially to the supplied N leading to a significant N and variety interaction (**Table 2**). Varieties such as Yitpi, Chara and Alsen had high DW accumulation, in contrast to Westonia, Kennedy and Drysdale, which accumulated less DW at both N levels (**Table 2**). Grain yield showed a highly positive association with DW accumulation at both N levels (**Table 2**). Greater DW accumulators such as Yitpi, Chara, and Alsen also had higher GY than Westonia, Kennedy and Drysdale (**Table 2**). However, the shoot and grain N concentrations of wheat varieties showed a highly negative trend with DW accumulation, GY and WSC at both N levels. Varieties with the lowest DW, GY, and WSC such as Kennedy and Drysdale had shoot and grain N concentration higher than Yitpi, Chara and Gladius at both N levels (**Table 2**). WSC concentrations of plants at maturity were higher under optimum N than low N for all the varieties (**Table 2**). Interestingly, WSC concentrations corresponded more with DW and GY at low N than optimum N (**Table 2**). On average, wheat varieties at optimum N level had higher SN, but slightly lower harvest index than those at the low N level (**Table 3**). However, the 1000-GW and GN were not changed significantly (**Table 3**).

The boxplots showed significant variations in NUE of biomass (NUEb) and grain (NUEg) per pot between varieties and N levels, confirming a significant interaction between these two factors (**Figure 4**). Varieties such as Alsen, Chara and Yitpi had high NUEs compared to Drysdale, Kennedy and Volcani DDI. Within N levels, 11 wheat varieties had significant variations in NUEb (**Figure 4A**), while only 3 varieties i.e. Alsen, Gladius and Excalibur were significantly different in NUEg (**Figure 4B**). This suggests that N treatments resulted in a more stable NUE for grain than biomass.

Validation of Imaging Phenotyping

To validate the application of the image-based phenotyping technology used to study the responses of wheat varieties to N supplies, we determined the association between the morphological and physiological parameters of 15 wheat varieties derived from digital imaging and conventionally destructive sampling methods (**Figure 5**). Manually harvested samples were collected at 49 DAS and used to measure MB and LA. The EB and TVA were derived from digital RGB images collected the night before the destructive harvest. Results from the correlation analysis at low N level showed that EB and TVA were highly correlated with important NUE traits such as measured fresh

TABLE 2 | Shoot dry biomass, grain yield, shoot water soluble carbohydrates, shoot and grain N concentration of wheat varieties grown under low and optimum N treatments at maturity. Varieties are ranked in descending order of DW at low N level. In a column: dark green cells, the highest values; dark red cells, the lowest values; DW, plant dry biomass; GY, grain yield; WSC, shoot water soluble carbohydrate; s.e.d., standard error difference of the means.

Variety	DW (g pot^{-1})		GY (g pot-1)		WSC (%)		Shoot N (%)		Grain N (%)	
	Low N	Optimum N	Low N	Optimum N	Low N	Optimum N	Low N	Optimum N	Low N	Optimum N
Yitpi	19.21	77.67	7.34	27.77	8.52	10.52	0.40	0.44	1.65	2.26
Excalibur	15.74	57.95	7.74	24.62	5.55	8.77	0.36	0.70	1.94	2.86
Chara	13.43	61.13	6.52	29.05	6.18	8.32	0.31	0.56	2.11	2.51
Pastor	13.34	60.53	6.46	28.13	3.03	6.58	0.33	0.57	1.90	2.51
Bobwhite	12.66	50.02	6.31	24.54	4.30	8.05	0.28	0.74	2.07	2.61
Alsen	12.59	61.37	6.13	28.23	5.18	9.33	0.35	0.49	2.30	2.73
Gladius	12.30	58.60	6.13	28.26	6.00	11.32	0.31	0.50	2.08	2.42
Wyalkatchem	12.08	48.71	5.95	22.30	5.83	8.90	0.48	0.71	1.98	2.68
RAC875	12.06	56.21	5.59	24.84	6.93	12.83	0.31	0.61	2.28	2.68
Kukri	11.99	54.14	6.49	25.35	5.10	10.12	0.47	0.66	2.13	2.72
Baxter	11.21	50.98	5.32	23.91	3.93	5.63	0.41	0.86	2.17	2.76
Volcani DDI	10.53	45.16	4.95	20.69	4.35	10.45	0.29	0.55	2.45	3.17
Westonia	10.52	49.45	5.72	23.99	4.15	8.08	0.33	0.57	2.11	2.74
Kennedy	8.70	41.33	4.79	21.97	3.40	5.48	0.33	0.76	2.32	2.92
Drysdale	8.26	38.78	4.54	18.87	2.97	7.00	0.34	0.90	2.31	2.78

ANOVA	N	V	N x V	N	V	N x V	N	V	N x V	N	V	N x V	N	V	N x V
s.e.d.	0.54	1.03	1.50	0.187	0.733	1.018	0.194	0.675	0.942	0.014	0.047	0.066	0.005	0.03	0.042
p	<0.001	<0.001	<0.001	<0.001	< 0.001	<0.001	<0.001	<0.001	0.02	< 0.001	< 0.001	<0.001	<0.001	<0.001	<0.001
l.s.d. ($P = 0.05$)	2.903	3.006		2.076	2.038		0.954	0.942		0.13	0.13		0.084	0.086	

TABLE 3 | Mean value of yield components of 15 wheat varieties. GW, grain weight; SN, number of spikes per pot; GN, number of grains per spike; CV, coefficient of variation; n.s., not significant difference ($p > 0.05$).

Component	Low N	Optimum N	l.s.d. ($p = 0.05$)	CV%	p
1000-GW (g)	41.35	44.26	3.11	9.7	ns
SN	4.21	15.38	1.83	25	< 0.001
GN	34.8	37.88	3.84	14.1	ns
Harvest index	49.41	46.36	3.02	8.4	0.05

MB and LA for all 15 wheat varieties with statistically significant coefficients of 0.94 and 0.82, respectively (**Figure 5A**). Likewise, the correlation analysis at optimum N level showed similar results; EB and TVA were also highly correlated with these NUE traits with correlation coefficients of 0.93 and 0.90, respectively (**Figure 5B**). The results also showed that there were intercorrelations within estimated and measured NUE traits at both N levels. For instance, EB was highly correlated with TVA and MB was highly correlated with LA (r ≥ 0.89, **Figure 5**). Similarly, EB was highly correlated with LA and TVA was highly correlated with MB (r ≥ 0.84, **Figure 5**). This suggests that these traits can be used interchangeably for an effective NUE assessment.

Association Between Vegetative Performance and Harvest of Wheat Varieties

Biomass accumulation of wheat and other grain crops over the growth period follow a sigmoidal growth pattern (Malhi et al., 2006; Archontoulis and Miguez, 2015; Nguyen et al., 2018). In the current study, our data showed that biomass accumulation (as indicated by EB) of wheat varieties under both N levels followed a sigmoidal growth pattern (**Figures 6A, B**). We determined the breakpoints where varieties commence or complete their linear growth phase using the “broken-stick” model (**Table 4**). Breakpoints are the reference points where each linear regression was “broken”, or the slope changed, given as X, Y coordinates, with X being the DAS and Y being the EB. The broken-stick model fitted well with the EB of wheat varieties, as indicated by high values of adjusted coefficients of determination (adjusted $R^2 > 0.99$; **Table 4**). Regression slopes before (slope 1) and after (slope 2) the breakpoint indicate the commencement or completion of the linear growth phase, while the exact values can indicate the number of days spent in either the linear or lag growth phases during the imaging period (**Table 4**). At the low N level, ten wheat varieties, i.e. Alsen, Baxter, Bobwhite, Drysdale, Excalibur, Kennedy, Kukri, RAC875, Volcani DDI and Westonia, commenced their linear growth phase early, with the breakpoint identified in **Table 4** being at the end of linear growth rather than the start, seen by a higher slope 1 and lower slope 2. Out of these, Drysdale completed its linear growth phase the earliest at 66.3 DAS (**Table 4**). Meanwhile, the remaining five varieties commenced their linear growth phase later, with 48.7 DAS being the latest time point when one of the varieties Yitpi commenced linear growth (**Table 4**). Thus, the period between 48–66 DAS is the duration when all wheat varieties were in their linear

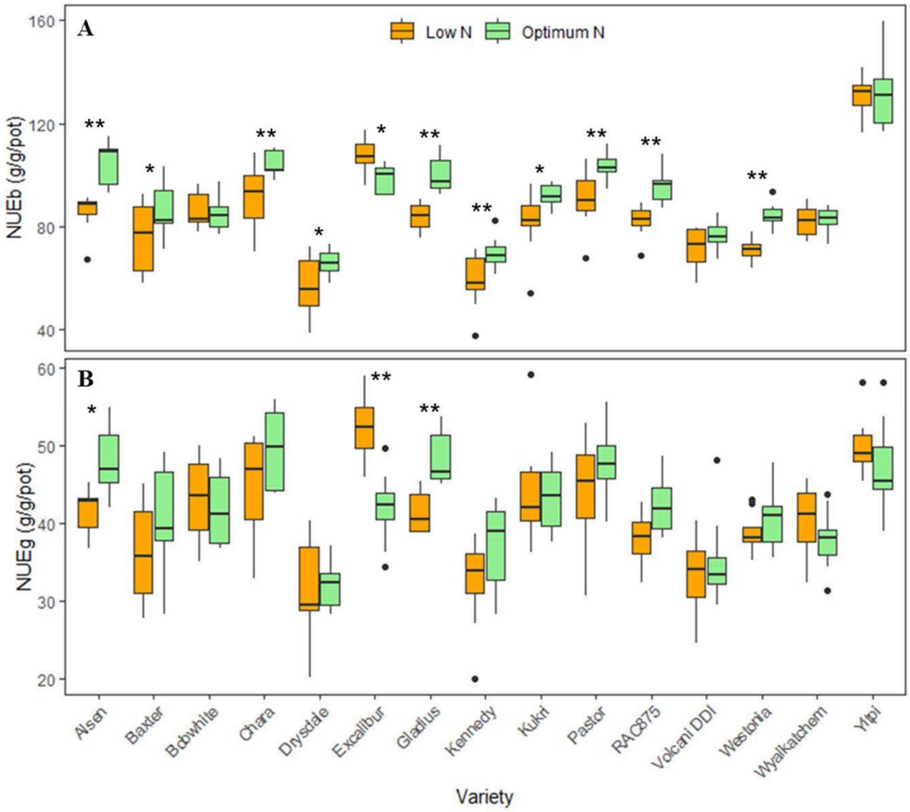


FIGURE 4 | Boxplots of nitrogen use efficiencies (NUEs) of 15 wheat varieties at low and optimum N levels. **(A)** NUE_b is nitrogen use efficiency calculated by harvested biomass (equation 1); **(B)** NUE_g is nitrogen use efficiency calculated by grain yield (equation 2). The asterisks indicate the statistically significant levels of ANOVA, comparing the NUE of a variety within N levels (* $p \leq 0.05$; ** $p \leq 0.01$).

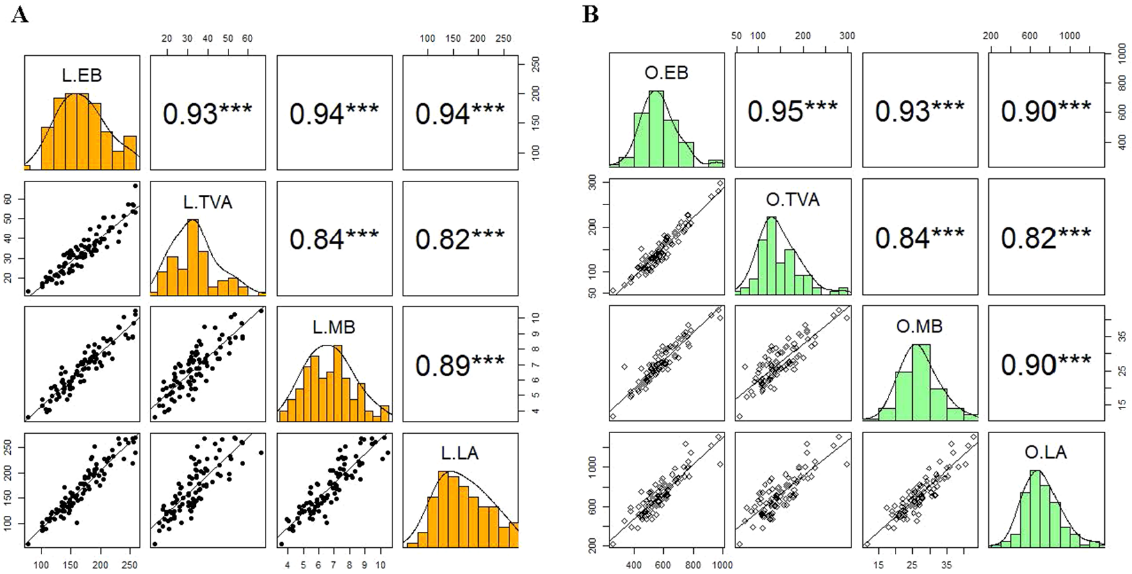


FIGURE 5 | Validation of the relationships between estimated and measured NUE traits of 15 wheat varieties. Panels **A**) and **B**) represent the correlation between estimated biomass and top-view area with measured fresh shoot biomass and leaf area collected at 49 DAS at low and optimum N supplies, respectively. In each panel, the coloured windows are the histograms of individual traits. The windows above and below the diagonal of the coloured windows are Pearson's correlation coefficients (r) and bivariate scatter plots with trend lines, respectively. L, low N; O, optimum N. EB, estimated biomass; TVA, top-view area, MB, measured biomass; LA, leaf area. The asterisks are the statistically significant levels (*** $p \leq 0.001$). Sample number = 90.

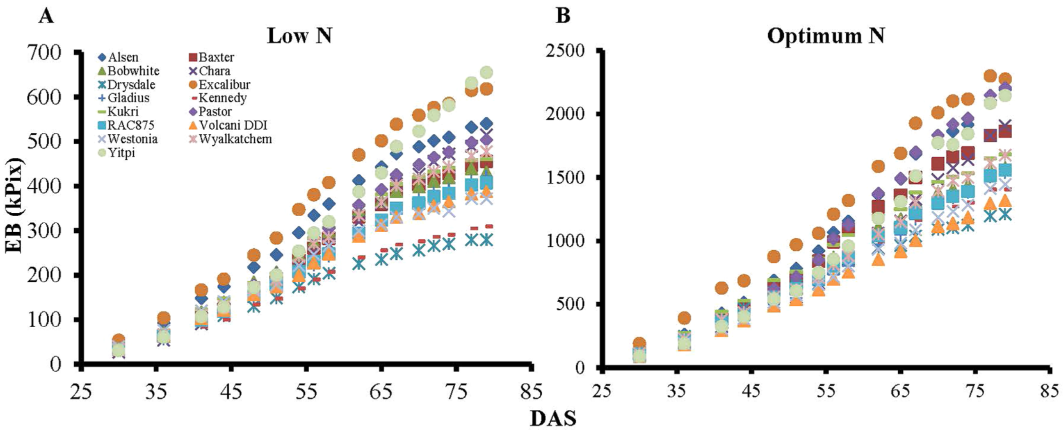


FIGURE 6 | Dynamic growth of wheat genotypes under low and optimum N levels. Graph **A**) and **B**) represent biomass accumulation over the growth period for 15 wheat genotypes at low and optimum N level, respectively. EB, estimated biomass; DAS, days after sowing. Specific marker shapes and colors indicate particular genotypes.

TABLE 4 | Regression parameters as determined by the split-line linear regression model of 15 wheat genotypes*. X is DAS and Y is EB, which represent the coordinates of the breakpoint where the linear regression was split or broken; slope 1 and slope 2, slopes of the regression before and after the breakpoint, respectively.

Variety	Low N				Optimum N			
	Breakpoint X	Breakpoint Y	Slope 1	Slope 2	Breakpoint X	Breakpoint Y	Slope 1	Slope 2
Alsen	68.7	480.8	11.9	5.9	46.2	570.9	28.1	49.8
Baxter	71.2	413.4	10.0	5.4	43.6	454.0	26.0	41.8
Bobwhite	68.1	393.6	10.0	4.1	37.8	242.0	18.1	33.8
Chara	45.1	113.7	6.0	12.3	52.4	578.7	22.1	50.2
Drysdale	66.3	247.4	6.0	2.8	62.0	939.1	25.9	16.5
Excalibur	69.1	554.5	13.6	6.9	52.3	1024.0	37.6	53.7
Gladius	42.5	97.3	5.6	9.5	49.9	514.5	22.3	36.5
Kennedy	67.8	268.9	6.7	3.7	37.7	241.2	17.6	30.4
Kukri	69.3	419.1	10.4	4.6	36.1	273.0	22.6	34.3
Pastor	38.2	77.3	5.4	11.2	46.4	501.0	24.8	54.2
RAC875	71.8	374.5	8.8	4.8	45.7	457.3	22.8	33.8
Volcani DDI	68.6	384.4	8.3	5.4	39.2	232.7	15.5	27.6
Westonia	66.5	326.1	8.1	3.7	45.5	408.1	21.2	31.2
Wyalkatchem	38.6	81.6	5.7	10.4	51.0	619.7	24.7	39.1
Yitpi	48.7	170.4	7.9	16.4	51.9	620.7	25.5	57.4

**Adjusted $R^2 > 99\%$.*

growth phase under low N (**Table 4**). At the optimum N level, all varieties commenced their linear growth phase later between 36.1–52.4 DAS, compared to the same varieties at the low N level. However, like the results seen at low N level, Drysdale once again completed its linear growth phase earlier than all other varieties, at 62 DAS (**Table 4**). Therefore, the period between 52–62 DAS is when all wheat varieties were in their linear growth phase under optimum N (**Table 4**).

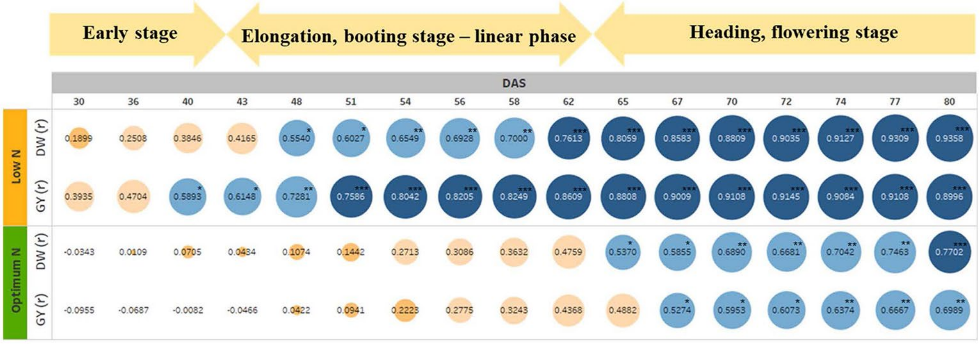
To validate whether EB during the linear growth phase can be used to evaluate the performance of wheat varieties for improved NUE, we compared the EB of the second plant set captured at 54 DAS against fresh shoot biomass and leaf area of the first plant set harvested 49 DAS for all wheat varieties (**Table 5**). The fifteen wheat varieties are ranked in descending order of MB accumulation at the low N level, where the dark green cells denote the higher values and in contrast the dark red cells represent the lower values (**Table 5**). Under low N, varieties such as Bobwhite and Yitpi accumulated higher MB in contrast to Drysdale and Volcani DDI (**Table 5**). However, Pastor and Alsen showed a stronger response to optimum N. Interestingly, Excalibur showed strong responses to N at both levels. Data also showed that EB and TVA were well correlated with MB and LA at both N levels for all varieties, showed by similar ranking patterns across the three traits (**Table 5**).

To test if the estimated parameters at the vegetative stage could predict biomass accumulation and grain yield at harvest, we determined the association between EB on different DAS with final DW and GY (**Figure 7**). The heat map shows increasing positive correlations between EB and DW or GY at low N level from 30 DAS onwards, with the correlation peaking at 80 DAS coinciding with the start of the flowering period (**Figure 7** and **Supplementary Figure 1**). In contrast, the positive correlation

TABLE 5 | Comparative performance of 15 wheat varieties at vegetative stage. Data are means of fresh shoot biomass, leaf area at 49 DAS (n = 6) and estimated biomass, top-view area at 54 DAS (n = 9) under low and optimum N levels. Varieties are ranked in descending order of fresh biomass accumulation at low N level. In a column: dark green cells, the highest values; dark red cells, the lowest values. MB, measured fresh shoot biomass; LA, measured leaf area; EB, estimated shoot biomass; TVA, top-view area; n = sample number.

Variety	MB (g pot^{-1})		LA (cm^2 pot^{-1})		EB (kPix pot^{-1})		TVA (kPix pot^{-1})	
	Low N	Optimum N	Low N	Optimum N	Low N	Optimum N	Low N	Optimum N
Bobwhite	8.33	25.52	191.01	699.92	245.86	759.47	57.22	215.81
Excalibur	8.04	36.63	238.46	1162.61	347.36	1059.88	74.14	296.44
Yitpi	7.66	25.94	217.56	693.79	254.35	751.63	48.70	166.91
Alsen	7.66	31.61	214.09	906.27	295.29	921.56	66.91	249.65
Kukri	7.63	29.21	191.82	768.57	251.00	859.99	51.20	223.55
RAC875	7.49	26.79	178.45	636.31	207.91	699.55	39.89	180.93
Wyalkatchem	6.67	28.38	190.34	792.61	242.24	711.04	46.49	169.47
Chara	6.59	21.97	174.86	599.51	218.43	663.29	39.50	153.32
Baxter	6.29	25.97	173.00	784.56	233.44	840.74	49.58	218.80
Pastor	6.10	32.06	157.68	831.87	251.98	849.85	51.37	224.07
Kennedy	6.00	27.37	136.43	723.73	170.16	721.79	33.02	174.04
Gladius	5.87	23.14	142.66	587.06	199.04	647.07	39.25	163.88
Westonia	5.62	24.99	125.66	569.03	221.09	656.73	45.03	162.15
Volcani DDI	5.49	19.68	123.79	413.69	199.42	614.73	43.58	163.12
Drysdale	5.01	25.74	96.90	635.07	173.51	712.40	33.09	181.50

ANOVA	N	V	N x V	N	V	N x V	N	V	N x V	N	V	N x V
s.e.d	0.083	1.052	1.440	4.76	40.27	52.22	13.30	25.00	36.65	5.59	7.15	11.25
p	<0.001	<0.001	0.003	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
l.s.d (*p* = 0.05)	4.17	4.13		113.05	116.66		73.42	70.82		22.75	20.25	



		DAS																
		30	36	40	43	48	51	54	56	58	62	65	67	70	72	74	77	80
Low N	DW (r)	0.1899	0.2508	0.3846	0.4165	0.5540*	0.6027*	0.6549**	0.6928**	0.7000**	0.7613***	0.8059***	0.8583***	0.8809***	0.9035***	0.9127***	0.9309***	0.9358***
	GY (r)	0.3935	0.4704	0.5893*	0.6148*	0.7281**	0.7586***	0.8042***	0.8205***	0.8249***	0.8609***	0.8808***	0.9009***	0.9108***	0.9145***	0.9084***	0.9108***	0.8996***
Optimum N	DW (r)	-0.0343	0.0109	0.0705	0.0434	0.1074	0.1442	0.2713	0.3086	0.3632	0.4759	0.5370*	0.5855*	0.6890**	0.6681**	0.7042**	0.7463**	0.7702***
	GY (r)	-0.0955	-0.0687	-0.0082	-0.0466	0.0422	0.0941	0.2223	0.2775	0.3243	0.4368	0.4882	0.5274*	0.5953*	0.6073*	0.6374**	0.6667**	0.6989**

FIGURE 7 | The relationship between estimated shoot biomass at different growth stages with harvest dry biomass and grain yield. The numbers inside the circle are correlation coefficients (r). Colour and circle size indicate r magnitude. DW, plant dry biomass per pot; GY, grain yield per pot. The asterisks are the statistically significant levels (* p ≤ 0.05; ** p ≤ 0.01; *** p ≤ 0.001).

between these parameters at optimum N only started at 48 DAS, also peaking at 80 DAS, even though the correlations were much lower compared to those under low N (**Figure 7**). Data also showed that EB observed at 60 DAS could possibly explain 50% and 70% of the variations in DW and GY at low N, respectively. Whereas, these figures were only approximately 18% and 14% at optimum N (**Figure 7**).

Comparative Performance of Wheat Varieties Under Controlled and Field Conditions

To determine if better performing varieties in the greenhouse perform well in the field, we analysed the correlation between DW and GY of wheat plants grown under greenhouse and field conditions. We compared the performance of the 15 wheat varieties from this greenhouse trial to the same varieties grown in the field at Horsham, Victoria, Australia in 2013 (Nguyen et al., 2016). There were several correlations between DW and GY in the greenhouse and the field, with some trends identifiable (**Supplementary Table 2**). These results showed low to moderate positive correlation for DW, especially the DW of wheat plants in the greenhouse at both N levels was significantly correlated with that of field plants at 80 N, the N level gives optimum NUE (**Supplementary Table 2**). The GY at low N in the greenhouse and the three field N levels also demonstrated a positive correlation trend with each other (**Supplementary Table 2**) (Nguyen et al., 2016).

DISCUSSION

An Advanced Phenotypic Screening Method for NUE Improvement in Wheat Under Controlled Environments

This work describes the development of a robust, high-throughput, reliable plant phenotyping method using automated digital imaging technology, that can be used to effectively screen a diverse range of wheat germplasm for NUE improvement at the vegetative stage in a controlled environment. The development of N-efficient wheat varieties through molecular breeding will undoubtedly contribute to the more effective use of N fertilizer, which is currently causing significant production and environmental costs (Cormier et al., 2016). However, NUE is a multi-genic trait and the lack of reliable phenotyping methods is currently a rate determining factor in NUE genetic improvement programs. These methods will aid in effectively screening breeding populations, phenotyping training populations for genomic selection, and evaluating breeding lines (Cabrera-Bosquet et al., 2012; Araus et al., 2018; Campbell et al., 2018). Thus, the availability of efficient phenotyping methods that are capable of characterising and quantifying multiple NUE traits, will provide useful tools to wheat breeders. In the present study, our results have demonstrated an applicable and reproducible wheat growth assessment system in a controlled environment for NUE studies. The results also revealed corresponding performance between wheat varieties screened by the digital RGB imaging unit of automated plant phenotyping platform, compared to field conditions for NUE traits.

An optimal plant growth system that can precisely apply and manipulate nutrient supplies in a timely way will play an important role in N studies. Several potting systems using pre-fertilized mixes have been reported to screen wheat for improved NUE in greenhouse studies (Tian et al., 2015; Malik et al., 2016; Veres et al., 2017). Although effective, these systems pose challenges such as the ability to timely adjust the amount of N supply for a range of varieties with unknown and diverse N responsiveness. The gradually supplied liquid fertilizer method used here, has advantages over pre-fertilized potting mixes, especially when used in conjunction with an automated watering system, which can also accurately dispense a set volume of fertilizer solution (Nguyen et al., 2018). Since water highly interacts with N availability in the growth media (Nguyen et al., 2017), this system will ensure an adequate supply of water for plants so that NUtE is not affected by either a shortage or excess of water, and will help reduce the time and labour costs associated with manual watering. More importantly, it was demonstrated that the current screening method could effectively produce significant variations in NUE traits (NUEb and NUEg) among wheat varieties, which will facilitate selection in screening processes.

Vegetative Screens by RGB Imaging for N-Efficient Wheat Genotypes

Early vigour, biomass accumulation, grain yield and grain protein are key criteria for the selection of N-efficient wheat materials (Nguyen and Kant, 2018). Since wheat grain yield is largely determined by the availability of carbohydrate reserves in the leaves and stems pre-anthesis, the higher the biomass accumulation during vegetative growth, the higher DW and GY at harvest (Reynolds et al., 2009; Li et al., 2013). Improving wheat yield potential by increasing DW at harvest has been demonstrated as an achievable and feasible strategy in breeding programs (Aparicio et al., 2002; Hawkesford, 2017). The primary objective of vegetative screens is to save time and costs, while still being able to effectively identify wheat genotypes which perform better for DW and GY at harvest in controlled environments and the field; ultimately speeding up breeding outcomes (Aparicio et al., 2000; Aparicio et al., 2002). In the present study, results confirmed that the automated RGB imaging unit was effective in estimating biomass accumulation from early vegetative to heading stages with a high degree of accuracy. The high correlations between EB and TVA and important traits such as MB and LA confirmed that the formers can be used as surrogates of the latter to evaluate the performance of wheat varieties for NUE at vegetative stage without destructive samplings. Our results also demonstrated that the performance of wheat varieties could be assessed effectively at early vegetative stages, as plant status during the linear growth phase truly reflects the potential at maturity regarding DW and GY. Being able to phenotype traits early in a high-throughput, precise and reproducible manner is a clear advantage in greenhouse screens and crucial for accelerating the breeding of improved NUE varieties (Ly et al., 2017). Vegetative screens, using conventional phenotyping methods, have also been successfully used for quantitative trait loci (QTL) mapping of early growth traits such as seedling height or shoot biomass for improved NUE in wheat (An et al., 2006; Guo et al., 2012). Highly accurate estimates of vegetation coverage of field grown wheat at booting stage using digital RGB images have been reported previously (Lukina et al., 1999). Moreover, the automated RGB imaging unit used in this study, was also advantageous against other proximal sensing tools, since it was not negatively influenced by genotypic variations or N levels observed elsewhere (Babar et al., 2006; Nguyen et al., 2016).

However, the causal relationship between vegetative performance and both DW and GY isn't fully understood. Under low N conditions, the EB of wheat plants was better correlated with GY than DW at early growth (40 DAS) and with correlation strength increasing until 74 DAS. Whereas, the correlation started at 67 DAS under optimum N conditions. In wheat, the remobilization of N reserved in vegetative parts, such as shoots and roots, before flowering, contributes up to 95% of grain N content at maturity (Palta and Fillery, 1995). Previous studies suggested that higher grain yield and grain NUtE in wheat were determined by a higher N remobilization efficiency, which was subject to genotypic assimilation efficiency and the availability of stored N in vegetative parts of the plants (Barbottin et al., 2005; Tian et al., 2015). Additionally, the remobilization of reserved carbohydrate pre-anthesis in wheat contributes up to 20% of grain yield under favourable conditions and up to 60% under stressful conditions, including N stress (Li et al., 2013), which was supported by our WSC assay results. Since NUE-related traits, e.g. remobilization efficiency, were highly expressed under low N conditions (Barbottin et al., 2005; Lammerts Van Bueren and Struik, 2017), it is likely that most of the reserved N and WSC in the wheat plants were translocated to grain yield before maturity due to N stress, leading to higher correlations between vegetative EB, DW and GY. Whereas, abundant N supplies meant that a large portion of the N and WSC reserved in shoots and roots were leftover in the DW, resulting in a lower correlation between EB from early stages and DW and GY (Barbottin et al., 2005; Gaju et al., 2014). Thus, selection might focus on the performance of wheat genotypes under low N rather than optimum N conditions for vegetative screens. The small variations in NUEg within N levels among varieties again supports this hypothesis. Based on growth analysis in the current study of 15 wheat varieties, it is recommended that EB collected at 60 DAS, that coincides with booting stages, can be used to compare vegetative performance of wheat varieties.

Perspective of Image-Based Phenotyping for NUE Improvement in Wheat Under Field Conditions

One of the biggest challenges in the development of N-efficient wheat varieties is necessity of developing an effective screening system in controlled environments that can effectively foresee the performance of wheat varieties under field conditions (Nguyen and Kant, 2018). In the present study, we compared the performance of wheat varieties for NUE under controlled and field conditions. We observed moderate and low-level correlations between greenhouse and field data for the DW and GY of identical wheat varieties, respectively (**Supplementary Table 2**). Interestingly, the DW of wheat varieties appears more consistent under both N levels in greenhouse compared to the 80 N, the optimum level under field conditions (**Supplementary Table 2**). The inconsistent performance of varieties under greenhouse and field has been well documented (Poorter et al., 2012; Junker et al., 2015). Quite often, genotypes selected from the controlled environments do not substantiate their performance under field conditions, because of significant competition among plants within plots under field conditions, which is not present for individual plants in pots in controlled environments (Araus and Cairns, 2014; Fischer and Rebetzke, 2018). Associations between greenhouse and field trials using the same varieties is further complicated by other environmental factors such as soil type, microorganisms, N and water availability (Cormier et al., 2016; Nguyen et al., 2017). However, several studies have reported a causal relationship between greenhouse screens and the field performance of crops (Chapuis et al., 2012; Pardo et al., 2015; Peirone et al., 2018). Therefore, results from vegetative screens in greenhouses, like those in the current study, can still be useful indicators of the performance of genotypes for NUE, which can help reduce the time and cost of developing new breeding materials.

Non-invasive remote sensing and imaging, using sensors and cameras, has been successfully applied to field crop phenotyping for NUE improvement (Nguyen and Kant, 2018). Simple to set up and cost effective conventional digital cameras have been effectively used as assessment tools for leaf area index and biomass in cereals (Casadesús and Villegas, 2014). The advent of various ground-based and aerial-based plant phenotyping platforms has made the estimation of final biomass and grain yield in wheat faster, more accurate and economical (Reyniers et al., 2006; Wang et al., 2014; Kefauver et al., 2017). Vegetation indices have been used to estimate biomass and grain yield under varying N supplies with high accuracy (Serrano et al., 2000). However, all the above-mentioned phenotyping platforms were deployed at the booting and heading stages to predict yield and final biomass, since biomass accumulation peaks at anthesis (Aparicio et al., 2000; Chang et al., 2005; Malhi et al., 2006); but, none of them were designed to predict final biomass and grain yield at early vegetative stages. Early vegetative prediction of N-efficient genotypes by high-throughput phenotyping will be especially useful, particularly when applied with genomic selection for NUE (Ly et al., 2017). This is particularly helpful for sensor and image based phenotyping in the field because vegetation indices will possibly become saturated if the leaf area index of the canopy is > 3 (Aparicio et al., 2000; Serrano et al., 2000). Several recent reports showed the potential of RGB imaging technology to study early crop growth and yield for NUE improvement in the field. Prey et al. (2018) used canopy cover from RGB imaging as a criterion to assess early vigour in wheat. In a similar approach, Buchaillot et al. (2019) used vegetation indices generated by both ground and aerial based RGB sensors at the vegetative stage in combination with crop's agronomic parameters, to successfully develop regression models for yield prediction of maize genotypes. Since TVA, observed here, was highly correlated with other traits, an avenue for further investigation is the deployment of digital RGB cameras, either handheld or mounted on ground or aerial vehicles (Araus and Kefauver, 2018; Fernandez-Gallego et al., 2019; Gracia-Romero et al., 2019) to capture and assess the performance of wheat genotypes for NUE traits at the linear growth phase under field conditions.

CONCLUSIONS

Here, we have described the development of a robust, high-throughput and reliable screening method at vegetative growth phases to investigate NUE improvements in wheat under controlled environment. Our results have shown that this digital RGB imaging method is strongly correlated to important NUE traits such as MB of wheat varieties. The observed relationship between controlled and field conditions for the same varieties indicates that greenhouse screening could be used to prioritise germplasm for subsequent field studies. Therefore, the application of this designated wheat growth system in conjunction with the digital imaging will provide breeders with an excellent assessment tool to enable the rapid phenotyping of diverse wheat genotypes to select N-efficient germplasm. This screening method may also provide a basis for the rapid phenotyping method of other crop species to identify germplasm efficient to a range of nutrients and stresses.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript, except raw RGB images which are available upon request.

AUTHOR CONTRIBUTIONS

SK and GN designed the experiments. GN, PM, LM, JV, and ET-K conducted the experiment. GN performed data curation and statistical analysis and wrote the first draft of the manuscript. SK, ET-K, PM, LM, JV, and GN contributed to manuscript editing and revision, and read and approved the submitted version.

ACKNOWLEDGMENTS

We thank Agriculture Victoria and the Grains Research & Development Corporation for providing infrastructure funding for the establishment of the Plant Phenomics Victoria facility at Horsham. We also thank Dr Sameer Joshi, Mandy Briggs and Rod Dickerson for their technical assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01372/full#supplementary-material

SUPPLEMENTARY FIGURE 1 | Phenology of wheat varieties under two N levels. **(A)** days from sowing to heading; **(B)** days from sowing to physiological maturity. Bars are mean of nine replicates and capped lines are standard errors.

REFERENCES

Ahmed, M., Rauf, M., Mukhtar, Z., and Saeed, N. A. (2017). Excessive use of nitrogenous fertilizers: an unawareness causing serious threats to environment and human health. *Environ. Sci. Pollut. Res.* 24, 26983–26987. doi: 10.1007/s11356-017-0589-7

An, D., Su, J., Liu, Q., Zhu, Y., Tong, Y., Li, J., et al. (2006). Mapping QTLs for nitrogen uptake in relation to the early growth of wheat (*Triticum aestivum* L.). *Plant Soil* 284, 73–84. doi: 10.1007/s11104-006-0030-3

Aparicio, N., Villegas, D., Araus, J. L., Casadesús, J., and Royo, C. (2002). Relationship between growth traits and spectral vegetation indices in Durum wheat. *Crop Sci.* 42, 1547–1555. doi: 10.2135/cropsci2002.1547

Aparicio, N., Villegas, D., Casadesus, J., Araus, J. L., and Royo, C. (2000). Spectral vegetation indices as nondestructive tools for determining Durum wheat yield. *Agron. J.* 92, 83–91. doi: 10.2134/agronj2000.92183x

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Araus, J. L., and Kefauver, S. C. (2018). Breeding to adapt agriculture to climate change: affordable phenotyping solutions. *Curr. Opin. Plant Biol.* 45, 237–247. doi: 10.1016/j.pbi.2018.05.003

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Archontoulis, S. V., and Miguez, F. E. (2015). Nonlinear regression models and applications in agricultural research. *Agron. J.* 107, 786–798. doi: 10.2134/agronj2012.0506

Babar, M. A., Reynolds, M. P., Van Ginkel, M., Klatt, A. R., Raun, W. R., and Stone, M. L. (2006). Spectral reflectance to estimate genetic variation for in-season biomass, leaf chlorophyll, and canopy temperature in wheat. *Crop Sci.* 46, 1046–1057. doi: 10.2135/cropsci2005.0211

Barbottin, A., Lecomte, C., Bouchard, C., and Jeuffroy, M.-H. (2005). Nitrogen remobilization during grain filling in wheat. *Crop Sci.* 45, 1141–1150. doi: 10.2135/cropsci2003.0361

Beatty, P. H., and Good, A. G. (2018). "Engineering Nitrogen Utilization in Crop Plants," in *Improving nitrogen use efficient in crop plants using biotechnology approaches*. Eds. Shrawat, A., Zayed, A., and Lightfoot, D. A. (Cham: Springer International Publishing), 15–35. doi: 10.1007/978-3-319-92958-3_2

Berry, J. C., Fahlgren, N., Pokorny, A. A., Bart, R. S., and Veley, K. M. (2018). An automated, high-throughput method for standardizing image color profiles to improve image-based plant phenotyping. *PeerJ* 6, e5727. doi: 10.7717/peerj.5727

Brink, C., Van Grinsven, H., Jacobsen, B., Rabl, A., Gren, I., Holland, M., et al, (2011). "The European nitrogen assessment," in *Costs and benefits of nitrogen in the environment*. Eds. Sutton, M., Howard, C., Erisman, J., Billen, G., Bleeker, A., Grennfelt, P., Van Grinsven, H., and Grizzetti., B. (Cambridge, UK: Cambridge University Press), 513–540. doi: 10.1017/CBO9780511976988.025

Buchaillot, M. L., Gracia-Romero, A., Vergara-Diaz, O., Zaman-Allah, M. A., Tarekegne, A., Cairns, J. E., et al. (2019). Evaluating maize genotype performance under low nitrogen conditions using RGB UAV phenotyping techniques. *Sensors (Basel Switzerland)* 19, 1815. doi: 10.3390/s19081815

Cabrera-Bosquet, L., Crossa, J., Von Zitzewitz, J., Serret, M. D., and Araus, J. L. (2012). High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.* 54, 312–320. doi: 10.1111/j.1744-7909.2012.01116.x

Campbell, Z. C., Acosta-Gamboa, L. M., Nepal, N., and Lorence, A. (2018). Engineering plants for tomorrow: how high-throughput phenotyping is contributing to the development of better crops. *Phytochem. Rev.* 17, 1329–1343. doi: 10.1007/s11101-018-9585-x

Casadesús, J., Kaya, Y., Bort, J., Nachit, M. M., Araus, J. L., Amor, S., et al. (2007). Using vegetation indices derived from conventional digital cameras as selection criteria for wheat breeding in water-limited environments. *Ann. Appl. Biol.* 150, 227–236. doi: 10.1111/j.1744-7348.2007.00116.x

Casadesús, J., and Villegas, D. (2014). Conventional digital cameras as a tool for assessing leaf area index and biomass for cereal breeding. *J. Integr. Plant Biol.* 56, 7–14. doi: 10.1111/jipb.12117

Chang, K. W., Shen, Y., and Lo, J. C. (2005). Predicting rice yield using canopy reflectance measured at booting stage. *Agron. J.* 97, 872–878. doi: 10.2134/agronj2004.0162

Chapuis, R., Delluc, C., Debeuf, R., Tardieu, F., and Welcker, C. (2012). Resiliences to water deficit in a phenotyping platform and in the field: How related are they in maize? *Eur. J. Agron.* 42, 59–67. doi: 10.1016/j.eja.2011.12.006

Cormier, F., Faure, S., Dubreuil, P., Heumez, E., Beauchêne, K., Lafarge, S., et al. (2013). A multi-environmental study of recent breeding progress on nitrogen use efficiency in wheat (Triticum aestivum L.). *Theor. Appl. Genet.* 126, 3035–3048. doi: 10.1007/s00122-013-2191-9

Cormier, F., Foulkes, J., Hirel, B., Gouache, D., Moënne-Loccoz, Y., and Le Gouis, J. (2016). Breeding for increased nitrogen-use efficiency: a review for wheat (T. aestivum L.). *Plant Breed.* 135, 255–278. doi: 10.1111/pbr.12371

Craswell, E., and Godwin, D. (1984). "Advances in Plant Nutrition," in *The efficiency of nitrogen fertilizers applied to cereals in different climates*. Eds. Tinker, P. B., and Lauchli, A. (New York: Praeger), 1–55.

Fahlgren, N., Gehan, M. A., and Baxter, I. (2015). Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

Fernandez-Gallego, J. A., Kefauver, S. C., Vatter, T., Aparicio Gutiérrez, N., Nieto-Taladriz, M. T., and Araus, J. L. (2019). Low-cost assessment of grain yield in durum wheat using RGB images. *Eur. J. Agron.* 105, 146–156. doi: 10.1016/j.eja.2019.02.007

Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137

Fischer, R. A., and Rebetzke, G. J. (2018). Indirect selection for potential yield in early-generation, spaced plantings of wheat and other small-grain cereals: a review. *Crop Pasture Sci.* 69, 439–459. doi: 10.1071/CP17409

Gaju, O., Allard, V., Martre, P., Le Gouis, J., Moreau, D., Bogard, M., et al. (2014). Nitrogen partitioning and remobilization in relation to leaf senescence, grain yield and grain nitrogen concentration in wheat cultivars. *Field Crops Res.* 155, 213–223. doi: 10.1016/j.fcr.2013.09.003

Garnett, T., Plett, D., Heuer, S., and Okamoto, M. (2015). Genetic approaches to enhancing nitrogen-use efficiency (NUE) in cereals: challenges and future directions. *Funct. Plant Biol.* 42, 921–941. doi: 10.1071/FP15025

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Good, A. G., and Beatty, P. H. (2011). Fertilizing nature: a tragedy of excess in the commons. *PloS Biol.* 9, e1001124. doi: 10.1371/journal.pbio.1001124

Good, A. G., Shrawat, A. K., and Muench, D. G. (2004). Can less yield more? Is reducing nutrient input into the environment compatible with maintaining crop production? *Trends Plant Sci.* 9, 597–605. doi: 10.1016/j.tplants.2004.10.008

Gracia-Romero, A., Kefauver, S. C., Fernandez-Gallego, J. A., Vergara-Díaz, O., Nieto-Taladriz, M. T., and Araus, J. L. (2019). UAV and ground image-based phenotyping: a proof of concept with Durum wheat. *Remote Sens.* 11, 1244. doi: 10.3390/rs11101244

Guedira, M., Xiong, M., Hao, Y. F., Johnson, J., Harrison, S., Marshall, D., et al. (2016). Heading date QTL in winter wheat (Triticum aestivum L.) coincide with major developmental genes VERNALIZATION1 and PHOTOPERIOD1. *PloS One* 11, e0154242. doi: 10.1371/journal.pone.0154242

Guo, Y., Kong, F.-M., Xu, Y.-F., Zhao, Y., Liang, X., Wang, Y.-Y., et al. (2012). QTL mapping for seedling traits in wheat grown under varying concentrations of N, P and K nutrients. *Theor. Appl. Genet.* 124, 851–865. doi: 10.1007/s00122-011-1749-7

Han, M., Okamoto, M., Beatty, P. H., Rothstein, S. J., and Good, A. G. (2015). The genetics of nitrogen use efficiency in crop plants. *Annu. Rev. Genet.* 49, 269–289. doi: 10.1146/annurev-genet-112414-055037

Hawkesford, M. J. (2014). Reducing the reliance on nitrogen fertilizer for wheat production. *J. Cereal Sci.* 59, 276–283. doi: 10.1016/j.jcs.2013.12.001

Hawkesford, M. J. (2017). Genetic variation in traits for nitrogen use efficiency in wheat. *J. Exp. Bot.* 68, 2627–2632. doi: 10.1093/jxb/erx079

Hirel, B., Le Gouis, J., Ney, B., and Gallais, A. (2007). The challenge of improving nitrogen use efficiency in crop plants: towards a more central role for genetic variability and quantitative genetics within integrated approaches. *J. Exp. Bot.* 58, 2369–2387. doi: 10.1093/jxb/erm097

Homolová, L., Malenovský, Z., Clevers, J. G. P. W., García-Santos, G., and Schaepman, M. E. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecol. Complex.* 15, 1–16. doi: 10.1016/j.ecocom.2013.06.003

International Fertilizer Industry Association, I. (2013). IFA Data. (International Fertilizer Industry Association: Paris, France) Available at http://ifadata.fertilizer.org/ucSearch.aspx [Verified 11 June 2019].

Junker, A., Muraya, M. M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A. E., et al. (2015). Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant Sci.* 5, 770. doi: 10.3389/fpls.2014.00770.

Kefauver, S. C., Vicente, R., Vergara-Díaz, O., Fernandez-Gallego, J. A., Kerfal, S., Lopez, A., et al. (2017). Comparative UAV and field phenotyping to assess yield and nitrogen use efficiency in hybrid and conventional barley. *Front. Plant Sci.* 8, 1733. doi: 10.3389/fpls.2017.01733

Lammerts Van Bueren, E. T., and Struik, P. C. (2017). Diverse concepts of breeding for nitrogen use efficiency. *A Rev. Agron. Sustain. Dev.* 37, 50. doi: 10.1007/s13593-017-0457-3

Li, H., Cai, J., Jiang, D., Liu, F., Dai, T., and Cao, W. (2013). Carbohydrates accumulation and remobilization in wheat plants as influenced by combined waterlogging and shading stress during grain filling. *J. Agron. Crop Sci.* 199, 38–48. doi: 10.1111/j.1439-037X.2012.00532.x

Lukina, E. V., Stone, M. L., and Raun, W. R. (1999). Estimating vegetation coverage in wheat using digital images. *J. Plant Nutr.* 22, 341–350. doi: 10.1080/01904169909365631

Ly, D., Chenu, K., Gauffreteau, A., Rincent, R., Huet, S., Gouache, D., et al. (2017). Nitrogen nutrition index predicted by a crop model improves the genomic prediction of grain number for a bread wheat core collection. *Field Crops Res.* 214, 331–340. doi: 10.1016/j.fcr.2017.09.024

Maharjan, P., Jacobs, J. L., Deighton, M., and Panozzo, J. (2018). A high-throughput method using Ultra-Performance Liquid Chromatography to determine water-soluble carbohydrate concentrations in pasture plants. *Grass Forage Sci.* 73, 562–571. doi: 10.1111/gfs.12315

Malhi, S., Johnston, A., Schoenau, J., Wang, Z., and Vera, C. (2006). Seasonal biomass accumulation and nutrient uptake of wheat, barley and oat on a Black Chernozem soil in Saskatchewan. *Can. J. Plant Sci.* 86, 1005–1014. doi: 10.4141/P05-116

Malik, A. I., Veres, S., and Rengel, Z. (2016). Differential nitrogen-use efficiency in wheat parents of doubled-haploid mapping populations. *Plant Soil* 408, 311–325. doi: 10.1007/s11104-016-2943-9

Moll, R. H., Kamprath, E. J., and Jackson, W. A. (1982). Analysis and interpretation of factors which contribute to efficiency of nitrogen utilization. *Agron. J.* 74, 562–564. doi: 10.2134/agronj1982.00021962007400030037x

Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., et al. (2018). DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnol. Biotechnol. Equip.* 32, 261–285. doi: 10.1080/13102818.2017.1400401

Neilson, E. H., Edwards, A. M., Blomstedt, C. K., Berger, B., Møller, B. L., and Gleadow, R. M. (2015). Utilization of a high-throughput shoot imaging system to examine the dynamic phenotypic responses of a C4 cereal crop plant to nitrogen and water deficiency over time. *J. Exp. Bot.* 66, 1817–1832. doi: 10.1093/jxb/eru526

Nguyen, G., Rothstein, S., Spangenberg, G., and Kant, S. (2015). Role of MicroRNAs involved in plant response to nitrogen and phosphorous limiting conditions. *Front. Plant Sci.* 6, 629. doi: 10.3389/fpls.2015.00629

Nguyen, G. N., Joshi, S., and Kant, S. (2017). "Plant macro-nutrient use efficiency: molecular and genomic perspectives in crop plants," in *Water availability and nitrogen use in plants: effects, interaction and underlying molecular mechanisms.* Eds. Hossain, M. A., Kamiya, T., Burritt, D., Tran, L.-S. P., and Fujiwara, T. (San Diego, CA, USA: Academic Press), 233–243. doi: 10.1016/B978-0-12-811308-0.00013-2

Nguyen, G. N., and Kant, S. (2018). Improving nitrogen use efficiency in plants: effective phenotyping in conjunction with agronomic and genetic approaches. *Funct. Plant Biol.* 45, 606–619. doi: 10.1071/FP17266

Nguyen, G. N., Norton, S. L., Rosewarne, G. M., James, L. E., and Slater, A. T. (2018). Automated phenotyping for early vigour of field pea seedlings in controlled environment by colour imaging technology. *PloS One* 13, e0207788. doi: 10.1371/journal.pone.0207788

Nguyen, G. N., Panozzo, J., Spangenberg, G., and Kant, S. (2016). Phenotyping approaches to evaluate nitrogen-use efficiency related traits of diverse wheat varieties under field conditions. *Crop Pasture Sci.* 67, 1139–1148. doi: 10.1071/CP16173

Palta, J., and Fillery, I. (1995). N application enhances remobilization and reduces losses of pre-anthesis N in wheat grown on a duplex soil. *Aust. J. Agric. Res.* 46, 519–531. doi: 10.1071/AR9950519

Pardo, E. M., Vellicce, G. R., Aguirrezaba, L. L., Pereyra, I. G., Rocha, C. M. L., García, M. G., et al. (2015). Drought tolerance screening under controlled conditions predicts ranking of water-limited yield of field-grown soybean genotypes. *J. Agron. Crop Sci.* 201, 95–104. doi: 10.1111/jac.12106

Peirone, L. S., Pereyra Irujo, G. A., Bolton, A., Erreguerena, I., and Aguirrezábal, L. a. N. (2018). Assessing the efficiency of phenotyping early traits in a greenhouse automated platform for predicting drought tolerance of soybean in the field. *Front. Plant Sci.* 9, 587. doi: 10.3389/fpls.2018.00587

Poiré, R., Chochois, V., Sirault, X. R. R., Vogel, J. P., Watt, M., and Furbank, R. T. (2014). Digital imaging approaches for phenotyping whole plant nitrogen and phosphorus response in Brachypodium distachyon. *J. Integr. Plant Biol.* 56, 781–796. doi: 10.1111/jipb.12198

Poorter, H., Fiorani, F., Stitt, M., Schurr, U., Finck, A., Gibon, Y., et al (2012). The art of growing plants for experimental purposes: a practical guide for the plant biologist. *Funct. Plant Biol.* 39, 821–838. doi: 10.1071/FP12028

Prey, L., Von Bloh, M., and Schmidhalter, U. (2018). Evaluating RGB imaging and multispectral active and hyperspectral passive sensing for assessing early plant vigor in winter wheat. *Sensors* 18, 2931. doi: 10.3390/s18092931

R Core Team. (2017). R: a language and environment for statistical computing. https://www.r-project.org/

Reyniers, M., Vrindts, E., and De Baerdemaeker, J. (2006). Comparison of an aerial-based system and an on the ground continuous measuring device to predict yield of winter wheat. *Eur. J. Agron.* 24, 87–94. doi: 10.1016/j.eja.2005.05.002

Reynolds, M., Foulkes, M. J., Slafer, G. A., Berry, P., Parry, M. a. J., Snape, J. W., et al. (2009). Raising yield potential in wheat. *J. Exp. Bot.* 60, 1899–1918. doi: 10.1093/jxb/erp016

Serrano, L., Filella, I., and Peñuelas, J. (2000). Remote sensing of biomass and yield of winter wheat under different nitrogen supplies. *Crop Sci.* 40, 723–731. doi: 10.2135/cropsci2000.403723x

Sharma, L., and Bali, S. (2018). A review of methods to improve nitrogen use efficiency in agriculture. *Sustainability* 10, 51. doi: 10.3390/su10010051

Tian, H., Fu, J., Drijber, R. A., and Gao, Y. (2015). Expression patterns of five genes involved in nitrogen metabolism in two winter wheat (Triticum aestivum L.) genotypes with high and low nitrogen utilization efficiencies. *J. Cereal Sci.* 61, 48–54. doi: 10.1016/j.jcs.2014.09.007

Veres, S., Malik, A. I., and Rengel, Z. (2017). Differential nitrogen supply causes large variability in photosynthetic traits in wheat germplasm. *Crop Pasture Sci.* 68, 703–712. doi: 10.1071/CP17126

Wang, L., Tian, Y., Yao, X., Zhu, Y., and Cao, W. (2014). Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. *Field Crops Res.* 164, 178–188. doi: 10.1016/j.fcr.2014.05.001

Xu, G., Fan, X., and Miller, A. J. (2012). Plant nitrogen assimilation and use efficiency. *Annu. Rev. Plant Biol.* 63, 153–182. doi: 10.1146/annurev-arplant-042811-105532

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 26 November 2019
doi: 10.3389/fpls.2019.01516

Convolutional Neural Net-Based Cassava Storage Root Counting Using Real and Synthetic Images

John Atanbori*[1]*, Maria Elker Montoya-P*[2]*, Michael Gomez Selvaraj*[2]*, Andrew P. French*[1] ***and Tony P. Pridmore*[1]**

[1] Agrobiodiversity Research Area, School of Computer Science, University of Nottingham, Nottingham, United Kingdom,
[2] Agrobiodiversity Research Area, International Center for Tropical Agriculture (CIAT), Cali, Colombia

Cassava roots are complex structures comprising several distinct types of root. The number and size of the storage roots are two potential phenotypic traits reflecting crop yield and quality. Counting and measuring the size of cassava storage roots are usually done manually, or semi-automatically by first segmenting cassava root images. However, occlusion of both storage and fibrous roots makes the process both time-consuming and error-prone. While Convolutional Neural Nets have shown performance above the state-of-the-art in many image processing and analysis tasks, there are currently a limited number of Convolutional Neural Net-based methods for counting plant features. This is due to the limited availability of data, annotated by expert plant biologists, which represents all possible measurement outcomes. Existing works in this area either learn a direct image-to-count regressor model by regressing to a count value, or perform a count after segmenting the image. We, however, address the problem using a direct image-to-count prediction model. This is made possible by generating synthetic images, using a conditional Generative Adversarial Network (GAN), to provide training data for missing classes. We automatically form cassava storage root masks for any missing classes using existing ground-truth masks, and input them as a condition to our GAN model to generate synthetic root images. We combine the resulting synthetic images with real images to learn a direct image-to-count prediction model capable of counting the number of storage roots in real cassava images taken from a low cost aeroponic growth system. These models are used to develop a system that counts cassava storage roots in real images. Our system first predicts age group ('young' and 'old' roots; pertinent to our image capture regime) in a given image, and then, based on this prediction, selects an appropriate model to predict the number of storage roots. We achieve 91% accuracy on predicting ages of storage roots, and 86% and 71% overall percentage agreement on counting 'old' and 'young' storage roots respectively. Thus we are able to demonstrate that synthetically generated cassava root images can be used to supplement missing root classes, turning the counting problem into a direct image-to-count prediction task.

Keywords: convolutional neural networks, generative adversarial networks, cassava phenotyping, machine learning, root counting

Edited by:
Jose Antonio Jimenez-Berni,
Spanish National Research Council
(CSIC), Spain

Reviewed by:
Anton Paul Wasson,
Commonwealth Scientific and
Industrial Research Organisation
(CSIRO), Australia
Dong Xu,
University of Missouri, United States

***Correspondence:**
Tony P. Pridmore
tony.pridmore@nottingham.ac.uk

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *05 July 2019*
Accepted: *31 October 2019*
Published: *26 November 2019*

Citation:
Atanbori J, Montoya-P ME,
Selvaraj MG, French AP and
Pridmore TP (2019) Convolutional
Neural Net-Based Cassava Storage
Root Counting Using Real and
Synthetic Images.
Front. Plant Sci. 10:1516.
doi: 10.3389/fpls.2019.01516

INTRODUCTION

The tropical root crop, cassava (*Manihot esculenta Crantz*), is a staple food for more than a tenth of the world's population. However, a major obstacle reducing its industrial potential is it's long and variable growth cycle. Information on the development of the edible cassava storage root is therefore crucial for selecting high yielding, early bulking cassava root crops for industrial-scale production. Cassava root systems comprise two key types of root. Fibrous roots develop first, and only a small number of these go on to form the larger storage roots. It is these storage roots which become an important food source, in particular a major source of carbohydrates. Understanding the growth of these storage roots therefore becomes an important phenotyping task. Presently, phenotyping of cassava storage roots is carried out using manual, destructive sampling methods (Okogbenin et al., 2013; Belalcazar et al., 2016), which are labour-intensive and require many replications of each genotype. The physiological traits of the cassava crop are usually measured manually, often during harvesting, but also pre-harvest. Measurements begin in the third month and continue every month until harvested (Okogbenin et al., 2013). The important pre-harvest traits measured include the number of storage roots and the primary stems, while harvest traits include the above-ground biomass, stem diameter and number of storage roots, along with their length and volume."

Image-based software tool development and usage for plant phenotyping tasks have increased in recent years (Furbank and Tester, 2011). Ideally, such tools should be high-throughput and at least semi-automatic, making them capable of providing accurate, quantitative data on plant structure and function with minimal manual labour. Most current phenotyping installations require precisely-designed, automated image acquisition hardware matched to specialist software solutions to achieve the best quality data and throughput. Often, the function of the image analysis step is impeded if the images are not captured in a tightly controlled, systematic way. Nevertheless, these tools are gaining more attention due to their merits in providing large-scale plant phenotyping when compared with manual methods. Image-based phenotyping techniques have recently been used in plant segmentation (Aich and Stavness, 2017; Aich et al., 2018), leaf counting (Giuffrida et al., 2015; Aich et al., 2018; Aich and Stavness, 2017) and to automatically identify root and leaf tips (Pound et al., 2017a). Dedicated development frameworks are even available to make building custom systems easier. For example, PlantCV (Gehan et al., 2017) can support a number of plant phenotyping tasks *via* processing pipelines, and the Deep Plant Phenomics platform (Ubbens and Stavness, 2017) specifically supports deep learning development.

Recently, very deep Convolutional Neural Networks (CNNs) have been used to recover plant traits in an attempt to gain improved robustness and accuracy (Scharr et al., 2016; Minervini et al., 2016a; Minervini et al., 2016b; Aich and Stavness 2017; Pound et al., 2017b). Here, and in the broader computer vision community, these techniques have increased the accuracy of the image analysis, but require large numbers of data samples to make them sufficiently general. Deep networks can comprise very many parameters (in the millions), which in turn introduces expensive computations (Long et al., 2015; Yu and Koltun, 2015; Badrinarayanan et al., 2017; Lin et al., 2017; Zhao et al., 2017) and can make such models inefficient on low-cost, resource-limited devices.

To date, deep learning methods addressing feature counting tasks have focused almost exclusively on phenotyping the plant shoot system. Two broad approaches are in use. The first begins by segmenting the input image. Learning to segment requires individual annotation of the relevant objects to create a training data set, a task which is usually error prone and time-consuming for the plant biologist to undertake. The second approach learns a direct regression model. The regression approach solves this problem by using the total object count as its only supervision information, which is comparatively very easy to collect. A complete, pixel-by-pixel labeling of the training images is not required, only instead requiring a numerical label giving the count of the features of interest; e.g. a root count. The regression models which must be learned are, however, non-linear and of very high dimensionality (Aich and Stavness, 2017): Here, instead, we propose to develop a direct image-to-count prediction model instead.

We aim to develop a fully-automated, image-based phenotyping system to count storage roots in color images of aeroponically grown cassava, including and in particular counting early bulking storage roots (those appearing in the first 2.5 months of growth). The challenge here is that early storage roots are usually particularly difficult to detect. There are comparatively few such roots on any given plant, and they are often occluded by fibrous roots, which have similar color and texture (see example images in **Figure 1**, bottom row). We develop here a direct "image-to-count" prediction model, avoiding the complexity of the regression approach, and avoiding a pure segmentation approach which can be problematic when the boundaries of the objects involved (especially fibrous roots) are not well defined. This prediction approach effectively *classifies* each image according to the number of plant features, e.g. storage roots, present. This raises a further challenge: to successfully train a classifier a training set containing multiple images of each class is required. In the current context this means that we require images showing 0, 1, 2, 3, 4 etc, storage roots, up to the maximum number expected to be encountered. Though the number of roots that can be reasonably expected is not large—we do not need examples for every integer, of course—complete data sets of this type are often unavailable in existing repositories, and can take a significant amount of time to assemble. This is particularly true of more recently-studied species like cassava, for which limited image data exists. To remedy the problem of classes short on, or missing data, we first develop a conditional GAN to generate synthetic images of the storage roots classes which are not sufficiently well-represented with real image data.

FIGURE 1 | Sample real cassava root images from our datasets. The top row illustrates cassava roots that are 2.5 months and more, which we term "old" roots. The bottom row illustrates "young" roots, those less than 2.5 months.

The contributions of this work, then, are:

1. Design of a conditional Generative Adversarial Net (GAN) that can automatically generate synthetic cassava root images when presented with ground truth segmentation masks of the desired image classes.
2. Design of a deep CNN age-prediction model that predicts the age of cassava roots as either "old" (≥2.5 months) or "young" (≤2.5 months).
3. Design of a deep CNN-based storage root counting model, which given an input image and an age class will classify the image according to the number of storage roots present.
4. Combination of these components, which will create a cassava root counting tool to support an aeroponic phenotyping system (Selvaraj, 2019); this will be evaluated against a segmentation-based counting approach

The remainder of this paper is structured as follows. In *Background*, we review existing works that generate synthetic images using a GAN approach, and those that perform object counting using deep CNN systems. In *Datasets*, we introduce the cassava root datasets used in our experiments and proceed in *Image Prediction and Generation Methods* to describe our methods. We describe our experimental set-up, including a benchmark, in *Experimental Evaluation*. We then proceed to present and discuss our results in *Results* and draw conclusions in *Conclusion*.

BACKGROUND

As many current phenotyping techniques were initially developed in Europe and North America, where cereal crops dominate, comparatively few studies have phenotyped cassava (Subere et al., 2009; Okogbenin et al., 2013; Adu et al., 2018 (Polthanee et al., 2016). Though these studies have considered both shoot and root phenotyping, and some even examined roots regularly during their development (Okogbenin et al., 2013), measurement of cassava root traits is typically carried out only during harvest. Despite consideration of root numbers, size and length alongside shoot structural measures and biomass (Okogbenin et al., 2013; Adu et al., 2018), it has not yet been established which traits or variables provide the most detailed differentiation between distinctive genotypes (Adu et al., 2018).

Traditionally, cassava storage roots are phenotyped destructively using manual or semi-automated methods. This usually involves extracting the roots from the soil, losing a large number of small and fibrous roots in the process. Cassava roots are cut from their stem. They are then suspended in water to measure volume and spread on a black background to ease counting and measurement of their total length. Care is taken to avoid roots overlaying each other, as much as is practically possible. In semi-automated methods, a digital camera is used to capture root images. The controlled imaging conditions often allow simple thresholding methods to segment roots from background, and simple image analysis methods, controlled by the user, can be used to extract total root lengths and counts from the resulting binary images. This process is not only destructive to the plants but also time-consuming for the scientist.

Though there is existing literature presenting methods that *count* the plant features visible in an image, to our knowledge no study has considered automatically counting cassava storage roots. Automated counting methods have focussed on counting shoot features such as leaves (Giuffrida et al, 2015; Aich and Stavness, 2017; Ubbens and Stavness, 2017), plants (Ribera et al., 2017; Aich et al., 2018), seeds (Uzal et al., 2018), and fruits (Chen et al., 2017; Rahnemoonfar and Sheppard, 2017). Current approaches are all based on deep CNNs, and can be broadly divided into three categories: counting by segmentation, and direct image-to-count by class prediction or regression.

Methods that adopt a segmentation-based approach (Aich and Stavness, 2017; Aich et al., 2018) first identify pixels arising from the relevant plant component(s) in, usually, RGB images, using a CNN-based segmentation model. Aich et al. (Aich et al., 2018) counted leaves by summing the predictions of image patches from a deep CNN model. However, in Aich and Stavness, 2017, both the RGB image and the corresponding binary segmentation image produced by the CNN were used to estimate the number of leaves. The complexity of the images involved means that segmentation approaches may generate spurious segmentations that in turn lead to inaccurate counts.

Regression-based approaches usually pose the counting task as a non-linear regression problem, regressing the output of the final CNN to a single value which represents the object count. Giuffrida et al. (2015) used this approach, converting images to log-polar form to benefit from the information present in the natural radial structure of the plants. They extracted patches from the log-polar image to form a feature vector which was used to train a support vector regression network to predict leaf number. This study, however, uses perfect segmentation together with the image and it is not clear how robust the system is to segmentation errors.

Such a regression-based approach is effective, but introduces a non-linear regression problem of very high dimensionality, which can be avoided by image-to-count class prediction methods (Ubbens and Stavness, 2017; Ribera et al., 2017; Uzal et al., 2018). This approach treats the counting problem as one of classification. Direct image-to-count prediction methods that use a deep CNN typically have their final layer made up of a number of neurons equal to the maximum number of plant features to be counted. Ubbens and Stavness (Ubbens and Stavness, 2017) have shown that this method outperforms both the segmentation- and regression-based approaches. However, the problem with this approach is that samples representing all classes must be available. If some classes are not represented well, the network cannot be trained.

One way of overcoming this problem of missing data is to generate synthetic images for non-represented or under-represented classes. Various methods for generating synthetic data have been proposed in the computer vision literature. Recent state-of-the-art methods commonly use conditional GANs (Isola et al., 2017). A conditional GAN is a general-purpose solution to image-to-image translation problems. Conditional GANs learn the mapping from an input image to an output image, as well as a loss function to learn this mapping. Some previous works (Giuffrida et al., 2017; Ward et al., 2018; Zhu et al., 2018), have reported using a GAN for image-to-image mapping in plant phenotyping. These methods used the GAN to generate synthetic images to augment real data for leaf counting in the CVPPP 2017 LCC dataset (Scharr et al., 2017) and showed that the testing error is reduced compared with all the other state-of-the-art methods reported to date for the challenge.

DATASETS

We use two combined plant image datasets, which we refer to as the "old" and "young" cassava root sets, to perform all of our experiments. **Figure 1** shows samples taken from each of these. The top row comprises sample images drawn from the "old" dataset and the bottom from the "young." The "old" roots dataset is made up of cassava roots that are at least two and a half months old, while the young dataset contains images of roots that are less than two and a half months. Though they may have very similar size, color, and texture, some roots considered to be storage roots when seen in "young" cassava plants would not be classed as storage roots when they appear later in the plant's development, in "old" plants: storage root identification must therefore consider plant age. We therefore first classify an unseen image as containing "new" or "old" roots, then count storage roots taking that classification into account, as the analysis challenge presented by the two age groups is very different.

Three semi-aeroponic systems (fog, drip and spray) designed and constructed at the International Center for Tropical Agriculture (CIAT), Colombia, were used to grow and image cassava roots (Selvaraj, 2019). Semi-aeroponic growth made it is easier to record cassava root images at regular intervals without disturbing or damaging the plants. We captured images at a resolution of 960 × 720 pixels using Logitech C922 Pro Stream Webcams with a custom-developed capture software tool, built on the OpenCV library (Intel Corporation, 2017). When imaging, the cassava plants were taken from the semi-aeroponic chambers and their roots hung carefully over a black background. Cassava experts from CIAT segmented the cassava storage roots from the captured images using a further custom-built annotation tool, and at the same time provided a manual count of the storage roots to form our annotated cassava root datasets. Each dataset was divided randomly following an "80/20" train/test split and the training data subjected to a further "80/20" train/validation split. Images were normalized by scaling their RGB values to the range 0–1. Image annotations were converted to a class label and then to a binary class matrix (one-hot encoding) before passing them to the deep neural networks. Example real data and classes can be seen in **Figure 2**.

IMAGE PREDICTION AND GENERATION METHODS

The storage root counting system proposed, and experimental evaluation conducted, here relies upon variations of two popular CNN networks: SegNet and DenseNet. In each case we reduced the number of model parameters using separable convolution, before training the resulting networks on the datasets described above, and synthetic images generated by a conditional GAN.

Separable Convolution

MobileNet (Howard et al., 2017), MobileNetV2 (Sandler et al., 2018) and Xception (Chollet, 2017) previously used separable convolution to reduce the number of model parameters. Separable convolution reduces the number of multiplications and additions in the convolution operation, thus reducing the model's weight matrix and speeding up both the training and application of large CNNs.

A 2D convolution can be defined as in Equation 1.

$$y(m,n)=\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}h(i,j)x(m-i,n-j) \tag{1}$$

where x is the $(m \times n)$ matrix being convolved with a $(k \times k)$ kernel h. If the kernel h can be separated into two kernels, say h_1 of dimension $(m \times 1)$ and h_2 of dimension $(1 \times n)$, then the 2D convolution can be expressed in terms of two 1D convolutions as in Equation 2.

$$y(m,n)=\sum_{i=0}^{k-1}h_1(i)\left[\sum_{j=0}^{k-1}h_2(j)x(m-i,n-j)\right] \tag{2}$$

The 2D convolution requires $k \times k$ multiplications and additions. However, separable convolution has its kernels decomposed into two 1D kernels, which then reduces the multiplications and additions to $k + k$ and thus reduces the number of model parameters.

Age-Prediction and CNN-Based Count Models

Our Age-Prediction and CNN-Based count models both use the DenseNet (Huang et al., 2017) architecture but with some

FIGURE 2 | This figure shows sample real cassava root images drawn from our datasets. The top row shows the cassava root images and the bottom, the ground truth annotation (segmentation mask and number of storage roots). The challenge of identifying storage roots is particularly evident in the one storage root image.

minor changes. To decrease the model parameters, we reduced the 1 × 1 and 3 × 3 convolution blocks in the first dense block to 3, the second to 6, third to 12, and fourth to 8 and converted 2D convolutions to 2D separable convolutions. However, we used a classification layer similar to the original DenseNet: 7 × 7 global average pool, 7D fully-connected layer with a softmax activation (CNN-Based Count model only), and 2D fully-connected layer with a softmax activation (Age-Prediction model only). The reduced-parameter architecture is shown in **Figure 3**.

Synthetic Cassava Root Generation Model

Our synthetic cassava root generation model uses a conditional GAN (Isola et al., 2017). The GAN comprises a generative and discriminative network chained together to make a composite model for training end-to-end. The network learns a mapping from the input mask to an output image, as well as a loss function to train the mapping (see **Figure 4**). We adapted our network architectures from those in Isola et al. (2017) and deposited the code in the GitHub repository[1]. Similar to Isola et al. (2017), our generator uses a "U-Net"-based architecture and a "PatchGAN"classifier with a patch size of 60 × 45. Both generator and discriminator use modules of the form convolution-BatchNorm-ReLu (ie. a 2D convolution followed by a Batch Normalization and then Rectified Linear Unit respectively). We use the convolution block "Conv2D-LeakyReLU-BatchNorm" (ie. a 2D convolution followed by a leaky Rectifier Unit and then a Batch Normalization respectively) denoted by C_i and the "Convolution-Dropout-BatchNorm" block, CD_i, where i is the number of filters. The convolutions are 4 × 4 spatial filters applied with stride 2. The convolutions in the encoder of both generative and discriminative models are downsampled by a factor of 2 and the decoder in the generative model up-samples by a factor of 2. Our generative model's encoder has:

$$C_{64} \Rightarrow C_{128} \Rightarrow C_{256} \Rightarrow C_{512} \Rightarrow C_{512}$$
$$\Rightarrow C_{512} \Rightarrow C_{512} \Rightarrow C_{512} \Rightarrow C_{512} \Rightarrow C_{512}$$

The decoder has:

$$D_{512} \Rightarrow D_{512} \Rightarrow D_{512} \Rightarrow D_{512} \Rightarrow D_{512} \Rightarrow D_{512} \Rightarrow D_{256} \Rightarrow D_{128} \Rightarrow D_{64}$$

We used the BatchNorm to improve the model's training procedure, thus allowing us to use much higher learning rates. ReLU is a type of activation function, which is defined as $y = max(0, x)$, meaning for all negative values of x, $y = 0$ and $y = x$ otherwise. On the other hand, the Leaky ReLU has a small gradient for negative values, instead of zero (for example, $y = 0.01x$ when $x < 0$).

The difference between our model and that in Isola et al. (2017) is that our encoder and decoder are both two blocks deeper. Here, we replace the MSE-based content loss with a loss calculated over feature maps of the VGG network, which are more robust to changes in the pixel space. We used the pre-trained VGG-16 network instead of the VGG-19. The discriminator architecture has a 60 × 45 patchGAN:

$$D_{64} \Rightarrow D_{128} \Rightarrow D_{256} \Rightarrow D_{512}$$

Our approach, which we refer to as a "CNN-based count model," first predicts the age of the root from an image using the age-prediction model in *Age-Prediction and CNN-Based Count Models*. This prediction helps determine whether to use the "old" or "young" CNN-based model to count the number of storage roots in the given image. It should be stressed that while the architectures of these models are identical, they are trained on the distinct age datasets. **Figure 5** summarizes our method.

Segmentation-CNN Model

For our deep segmentation CNN model—used in our experimental evaluation of the proposed method as a comparison approach—we used a a VGG-16 style architecture similar to SegNet (Badrinarayanan et al., 2017). We followed the convolution layers in each encoder with a batch normalization and ReLU activation except for the last block, where we placed a max-pooling at the end of each encoder block. Max-pooling is a sample-based discretization process, which down-samples the input images by reducing their dimensionality. We used the same settings as Badrinarayanan et al. (2017), with max-pooling indices for up-sampling. 2D convolutions were replaced with 2D separable convolutions to reduce the number of model parameters and produce a model similar to the Lite CNN models in (Atanbori et al., 2018). The convolution layers in both the encoder and decoder were made separable, and batch normalization and ReLU activations applied to the separated convolutions. The first convolution of the network was, however, not separated, as this captures important, high-detail features. The reduced architecture is illustrated in **Figure 6**. To provide a point of comparison with the direct prediction method, a storage root count is obtained from the segmentation mask produced by the SegNet-based CNN, following noise removal using morphological operations, by counting the number of contours in the mask.

EXPERIMENTAL EVALUATION

We combined the cassava root datasets detailed in *Datasets*, and the CNN architectures presented in *Image Prediction and Generation Methods* to perform the following experiments:

- The conditional GAN architecture described in *Synthetic Cassava Root Generation Model* was used to generate synthetic cassava root images from storage root segmentation masks derived from the ground truth segmentations manually created at CIAT. The similarity of synthetic and natural images was quantitatively evaluated by comparing the results of segmenting those images with the segmentation network presented in *Segmentation-CNN Model*. The hypothesis is that if synthetic and real images can be automatically segmented

---

[1] https://github.com/Amotica/CNN-Based-Cassava-Storage-Roots-Counting/tree/master/rootGAN

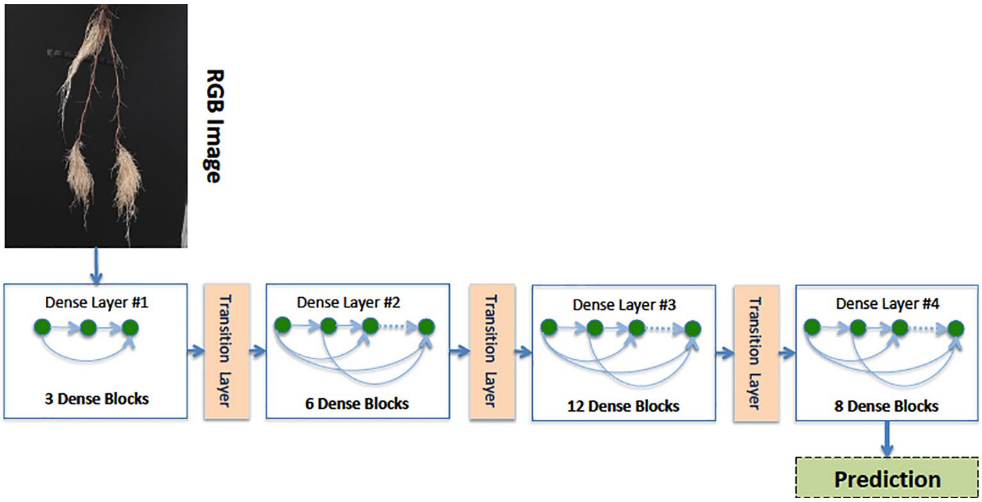


FIGURE 3 | Our reduced-parameter DenseNet architecture has four dense blocks. Each dense block is made up of a 2D separable Convolution (SeparableConv2D), Batch Normalization (BN), and ReLU activations.

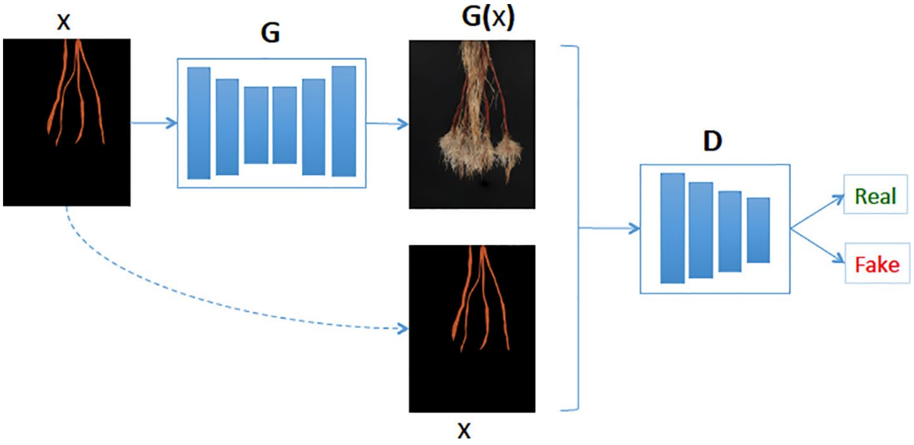


FIGURE 4 | Our generative model (G) learns the mapping between cassava storage root masks and cassava roots. Then given a cassava storage root masks (x), the generator predicts the cassava root image (G(x)) and the Discriminator (D) determines if the generated cassava storage root (G(x)) is fake or real.

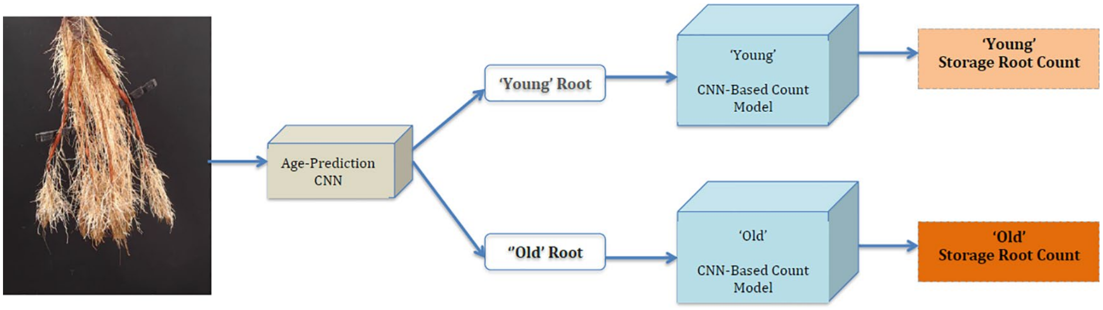


FIGURE 5 | Block diagram of our Convolutional Neural Network (CNN)-Based Count Model.

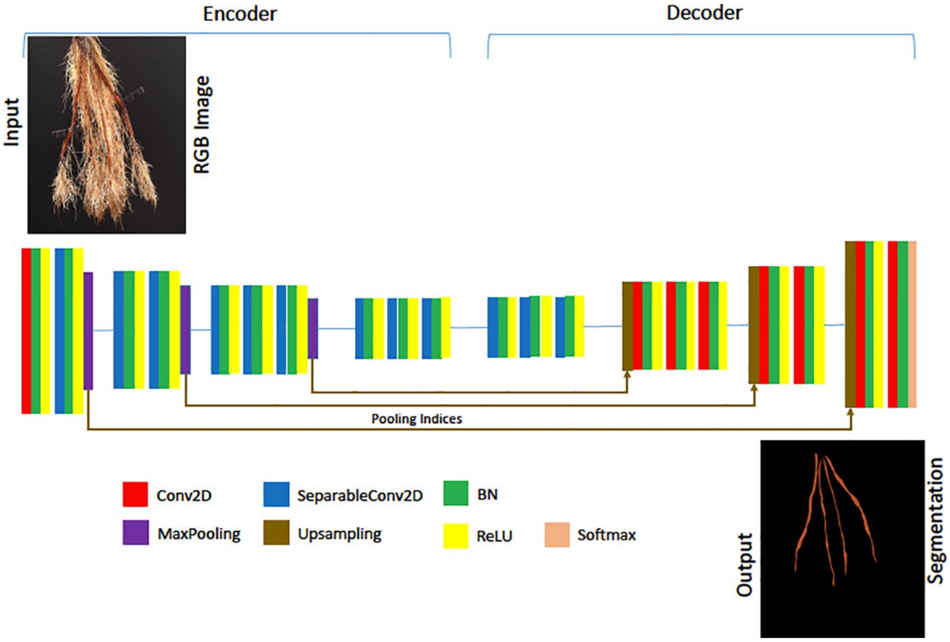


FIGURE 6 | The architecture of our Lite-SegNet model is a typical VGG-16 architecture with only four blocks. The building blocks comprise of a 2D Convolution (Conv2D), 2D separable Convolution (SeparableConv2D), Batch Normalization (BN), a ReLU activation, Max Pooling, and Upsampling. This network is used to generate a reference root segmentation to help evaluate our Generative Adversarial Network (GAN) and CNN counting models.

with similar accuracy, the synthetic data is likely suitable as additional training data when training the counting networks.

- The age-prediction network described in *Age-Prediction and CNN-Based Count Models* was used to classify real images of cassava roots as "young" or "old." The performance of this tool is evaluated in the usual way, as a classification task. The motivation for this test is to show we can identify root age with sufficient accuracy that a suitable counting model (ie. optimized for younger or older roots) can then be selected for the image in question.
- Two instantiations of the root counting network architecture described in *Age-Prediction and CNN-Based Count Models* were trained, both using real and synthetic images, to perform a direct image-to-count mapping. One of these models is trained for "young" root images, and one for "old" images. Performance of each of these networks was quantitatively compared to manually obtained ground truth. Results were also compared with identical measures of performance obtained using the segmentation-based approach described in *Segmentation-CNN Model*.

To train the image generation GAN we apply an Adam solver with a learning rate of 0.0002, and momentum parameters of $\beta1 = 0.5$ $\beta2 = 0.999$. For this experiment, we trained the network from scratch for 900 epochs and used a batch size of 2. Similar to Ledig et al. (2017), we minimize the MSE between features extracted with a pre-trained VGG19 model for real and synthetic root images. We did this because the pixel-wise loss functions such as MSE usually struggle to handle the uncertainty inherent in recovering lost high-frequency details such as texture. As input to the GAN, we supply a storage root mask, which can derive from a mask from an existing image with the same root count, or, in the case of missing data for a class (such as a lack of real data of plants with five storage roots), can be synthesized. To generate such a novel mask, masks of different root counts were combined before being passed into the GAN (see **Figure 4**).

We set the batch size of the segmentation CNN model to 2 and the age-prediction and CNN-based count models to 32. The input image resolution of the segmentation CNN is 640 × 480 pixels: large enough to maintain details of young cassava storage roots and small enough for the network to train reasonably quickly. The input image resolution of the Age-Prediction and CNN-based Count models is 256 × 256 pixels, large enough to maintain the storage root structure but again small enough for the model to train efficiently. Data augmentation applied consists of a zoom range of 0.2, brightness scaling ranging between 0.2 and 1.0, a rotation range of 10 degrees, and a horizontal flip.

We implemented all our models using Python 3.5.3 and Keras 2.0.6 with a Tensorflow backend, and trained them on a Linux server with three GeForce GTX TITAN X GPUs (12 GB memory each). Testing was carried out on a Windows 10 computer with 64GB RAM and a 3.6GHz processor.

Metrics Used

We used the “SegNet-score,” which is similar to the "FCN-score" used in Isola et al. (2017) to quantitatively evaluate our generative model. The SegNet-scores used in our evaluation include Precision, Recall, Pixel Accuracy, and MeanIoU, previously used in Atanbori et al. (2018) to evaluate competing segmentation models. We have detailed these below:

- Pixel accuracy: This tells us about the overall effectiveness of the classifier and is defined in Equation 3.

$$\frac{\sum_{i=1}^{c} n_{ii}}{\sum_{i=1}^{c}\left(\sum_{j=1}^{c} n_{ij}\right)} \tag{3}$$

- Mean intersection over Union (MeanIoU): This compares the similarity and diversity of the complete sample set and is defined in Equation 4:

$$\frac{1}{c} * \sum_{i=1}^{c} \frac{n_{ii}}{\sum_{j=1}^{c} n_{ij} + \left(\sum_{j=1}^{c} n_{ji}\right) - n_{ii}} \tag{4}$$

- Average Precision: This tells us about the class agreement of the data labels with the positive labels given by the classifier and is defined in Equation 5.

$$\frac{1}{c} * \sum_{i=1}^{c} \frac{n_{ii}}{\sum_{j=1}^{c} n_{ji}} \tag{5}$$

- Average Recall: This is the effectiveness of classifier to identify positive labels and is defined in Equation 6.

$$\frac{1}{c} * \sum_{i=1}^{c} \frac{n_{ii}}{\sum_{j=1}^{c} n_{ij}} \tag{6}$$

where n_{ij} is the number of pixels of class i predicted to belong to class j, n_{ji} is the number of pixels of class j predicted to belong to class i, and c is the total number of classes.

We used the metrics reported in Giuffrida et al. (2015) for evaluating the leaf counting challenge to compare our CNN-based count model with the count derived from the Seg-Based Model. We choose to use these metrics since they have been widely used by the plant phenotyping community when evaluating counting models. They are:

- PercentAgreement, indicating in how many cases the algorithmic estimation agrees with ground truth
- CountDiff, average difference between algorithmic estimation of the count and ground truth, reported as mean and SD
- AbsCountDiff, average of absolute count errors, and reported as mean (SD)
- Mean Squared Error (MSE), the average squared difference between the predicted and ground truth values.

RESULTS

As we need to generate synthetic images to fill in classes of root numbers which are missing training images, and to augment other data-poor classes, we first examine the success of our GAN-based synthetic image generation approach. We compared our synthetically-generated cassava images (see **Figure 7** and **Supplemental Data** for examples) against real images using the SegNet-scores, which we have detailed in *Metrics Used*, and present the results in **Table 1**. The SegNet-scores we considered are Precision, Recall, Pixel Accuracy, and MeanIoU.

The results reported were based on using the segmentation model trained on real images, for testing real and synthetically generated cassava images. The reasoning is that if the synthetic images are sufficiently similar to real images, then the model should be able to segment the synthetic images with comparable accuracy; if the generated images are visually different to the real images, segmentation will fail. From **Table 1**, it can be seen that although the SegNet-scores of the synthetically generated cassava images are lower than those associated with real images, the difference is small—only 4% based on the MeanIoU, and even smaller when considering the other metrics. From this, we observe that if the synthetic images can be segmented almost as well as the real images, they will be suitable as synthetic data to replace missing real images in the training data for the counting CNN procedure.

In order to choose an appropriate counting CNN model, we must estimate the age of the plant in an image as either young or old. Therefore, we next evaluate classifying cassava images from each dataset into featuring either old (≥ 2.5 months) or young (< 2.5 months) cassava roots using our age-prediction model. Results are presented in **Table 2**. These results are then used as input into our CNN-based count model to count storage roots. Storage roots of an image predicted as "young" are counted using the CNN-based model trained on "young" cassava roots whereas those predicted as "old" are counted by the model trained on old cassava roots.

We then evaluated our CNN-based cassava storage roots counting model, comparing counts generated by our CNN model with counts generated from an image processing pipeline deriving from the Seg-Based model. We evaluated using the test-split (data from train/test split), test data (taken from the field after building the models), and combined test data (both test-split and test data). The results are presented in **Tables 3**, **4**, and **5**, respectively, and some good quality example outputs are shown in **Figure 9** (error cases are raised in the discussion).

The CNN-based model (ours) outperformed the Seg-Based system based on all metrics and test data splits reported. The difference is more than 36% for "old" roots and 25% for "young" roots based on the test-split data only. However, based on the combined test data (both test and test-splits), this is 39% and 23%. In all cases our CNN-based approach outperforms the segmentation-based approach. We present a discussion of these results in *Discussion*.

Discussion

We generated synthetic images for only the "old" cassava storage roots (≥2.5 months). We did this since our model requires all classes of the dataset (ie. all possible numbers of storage roots) to

FIGURE 7 | Our generative model (GAN) results: The first row show results of generated synthetic cassava roots images, all having five storage roots; no images of this class are present in the real image dataset, hence requiring synthetic generation. The last row shows results of generated synthetic cassava root images featuring a variety of storage root numbers. More examples from our GAN network can be seen in **Supplemental Images S1–S3**.

TABLE 1 | Comparison of segmenting synthetically generated cassava images versus the real images using the SegNet-scores: Precision, Recall, Pixel Accuracy, and MeanIoU.

Data	Precision	Recall	Pixel accuracy	MeanIOU
Real	99.30%	99.30%	99.30%	70.35%
Synthetic	98.38%	98.36%	98.48%	65.56%

TABLE 2 | Percentage of correctly classified images for young and old root classes, using the age-prediction model.

	Test-split	Test data	Combined
Old	95%	71%	83%
Young	99%	100%	100%
Overall	**97%**	**86%**	**91%**

Test-split represents data from the train/test split, test data is taken from the field after building the models, and combined data comprises both test-split and test data.

TABLE 3 | Comparison of storage root-counting accuracy for our proposed Convolutional Neural Network (CNN)-based approach versus a more traditional segmentation-based approach.

		% Agreement	CountDiff	ABS CountDiff	MSE
CNN-Based	Old	90%	0.17 ± 0.79	0.26 ± 0.76	0.65
	Young	74%	0.08 ± 0.84	0.38 ± 0.76	0.72
Seg-Based	Old	54%	1.39 ± 1.92	1.39 ± 1.92	5.61
	Young	49%	1.04 ± 1.54	1.08 ± 1.52	3.45

Results from the Test-split data; this is the data from train/test split that was reserved for testing the model. We report: %Agreement: the higher the better, CountDiff and ABSCountDiff as mean ± SD: the smaller the better and MSE: the smaller the better.

TABLE 4 | Comparison of storage root-counting accuracy for our proposed CNN-based approach versus a more traditional segmentation-based approach.

		% Agreement	CountDiff	ABS CountDiff	MSE
CNN-Based (Ours)	Old	70%	0.27 ± 0.96	0.45 ± 0.89	1.00
	Young	62%	0.36 ± 1.59	0.79 ± 1.25	3.18
Seg-Based	Old	20%	0.09 ± 1.31	1.00 ± 0.85	1.72
	Young	46%	1.14 ± 1.25	1.14 ± 1.25	3.64

Results here are from new isolated test data after the model was built. We report: percent agreement: the higher the better, CountDiff and ABSCountDiff as mean ± SD: the smaller the better and MSE: the smaller the better.

be present, but this age category had some missing. Based on the "SegNet-Scores," our synthetically generated cassava images are comparable to real images and therefore should be able to be used to train our CNN-based count models. **Figure 7** shows the results of synthetic images generated using our generative model. Visually inspecting these images also show that they are comparable to the

TABLE 5 | The table shows results from combining the Test and test-split Data.

		% Agreement	CountDiff	ABS CountDiff	MSE
CNN-Based (Ours)	Old	86%	0.12 ± 0.77	0.24 ± 0.74	0.61
	Young	71%	0.14 ± 1.06	0.47 ± 0.96	1.14
Seg-Based	Old	47%	0.84 ± 1.63	1.04 ± 1.51	3.37
	Young	48%	1.08 ± 1.49	1.11 ± 1.47	3.38

We reported %Agreement—the higher the better, CountDiff and ABSCountDiff as mean ± SD—the smaller the better and MSE—the smaller the better.

real images (c.f. **Figure 1**). Even though the synthetically generated images were high resolution (960 × 720), they were less blurry, since our perceptual loss function, which uses the high-level feature maps of the VGG network adopted from Ledig et al. (2017) has been shown to produce less blurry images than an L1-Loss function, for example. We thus effectively use the generated images to supplement the real ones for training our count-model.

The correct prediction rate of our age-prediction model on our "old" roots is 83%, which by itself can be considered a good result. However, we observed that the model performs better in correctly predicting "young" cassava roots than the "old" ones. This is evidenced in the correct prediction rates reported in **Table 2**. The difference in the correct prediction rate between "young" and "old" roots is 4% on the test-split data, 29% on the test data collected after training the model, and 17% based on the combined test sets. We have also shown results of some correctly and incorrectly predicted ages of cassava storage root in **Figure 8**. Most of the incorrect predictions lie closer to the boundary age (2.5 months) of "old" and "young" roots. The second and third images in row two of **Figure 8** shows some of these types of incorrect age predictions. Other incorrectly classified images were varieties of cassava roots that were poorly represented in the training data (images 1 and 4 in row two of **Figure 8**).

The "old" cassava roots are usually correctly counted by the CNN-based model (ours) and Seg-Based model. "Young" roots with well-defined storage roots are also very likely to be correctly counted by both models. **Figure 9** shows the results of correctly counted storage roots. The top row shows images of the "old" cassava roots that are correctly counted and the bottom the "young" ones. Clearly, "young" roots with well-defined storage roots are correctly counted by both models [especially the CNN-Based model (ours)]. However, overall, counting the "old" cassava storage roots are more successful than the "young" ones as their storage roots are well-defined.

We have observed that our CNN-based counting model outperforms the Seg-based model substantially based on both datasets. The reason for this difference is that the Seg-based model uses the masks of cassava storage roots, which sometimes overlaps, thus making them difficult to count. However, because the CNN-Based model (ours) does not rely on segmented masks, it is usually more successful on this type of images. Furthermore, the "Seg-Based" model fails to correctly count the storage roots when there is incorrect segmentation from the segmentation-CNN model. There are also additional cases when both our "CNN-Based" and "Seg-Based" models incorrectly count roots (see **Figure 10**). Again,

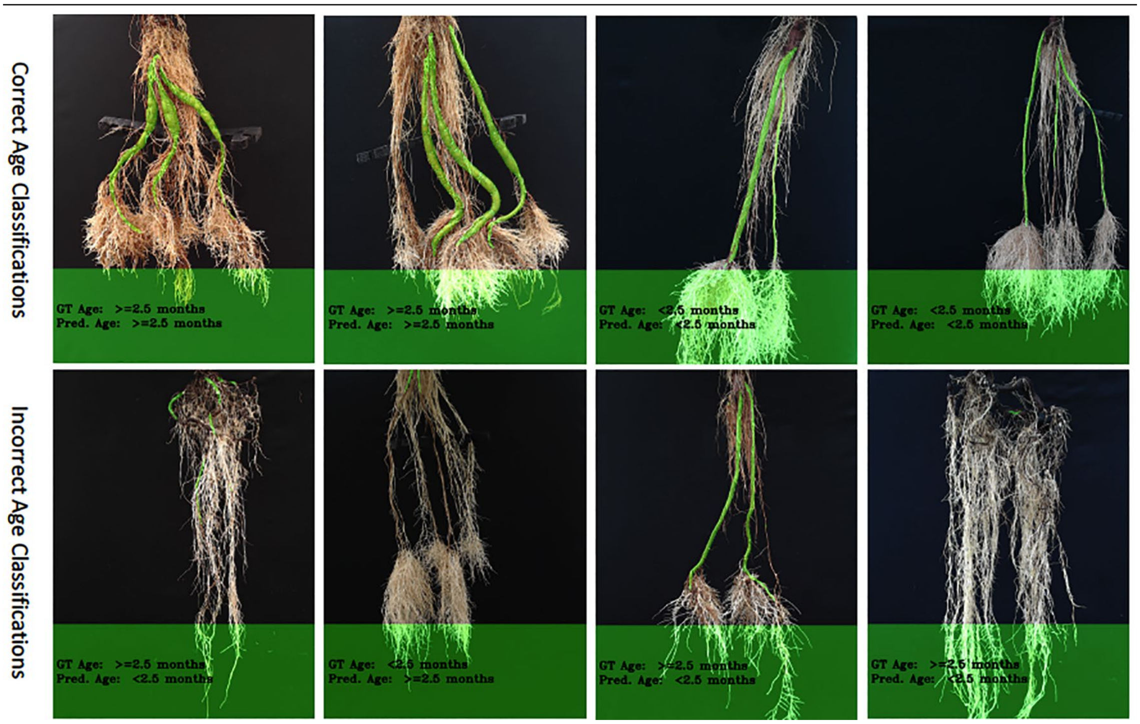


FIGURE 8 | Example results of our age-prediction model. Top row shows images of correctly predicted age; bottom row shows incorrect predictions. More age prediction example images can be seen in **Supplemental Images S4–S9**.

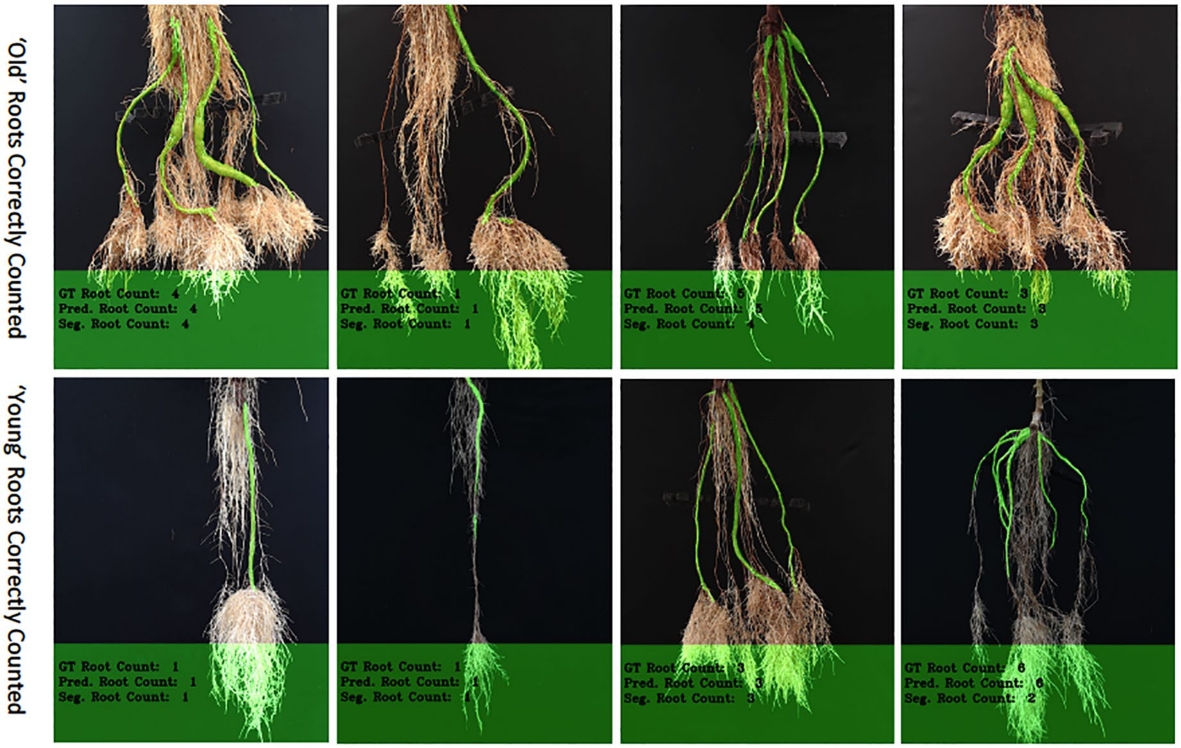


FIGURE 9 | The figure shows correctly counted storage roots, with the top row showing "old" storage roots and the bottom, "young" ones. More examples of outputs for the counting networks can be seen in **Supplemental Images S10–S15**.

FIGURE 10 | The figure shows incorrectly counted storage roots, with the top row showing results from the CNN-based model and the bottom, the Seg-based model.

this happens more in "young" roots where storage roots are visibly harder to pick out, and also varieties of cassava that are under-represented in the training set. Perhaps more insight could be gained into these errors by building in an explainable approach to the deep learning. To elucidate why decisions are made by the deep learning system, future systems will attempt to reveal to the user regions which are used in the counting process. Understanding the exact mechanism of GAN-based image generation is more challenging, and is a focus of current research.

CONCLUSION

We proposed two convolutional network architectures for counting "old" and "young" cassava storage roots, which we refer to as "old" and "young" CNN-Based count models respectively. Since we needed two models, we further proposed a CNN-based age-prediction architecture that first classifies storage roots as either "old" or "young" and then use the appropriate CNN-based count model to predict the number of storage roots. In our experiments, the age-prediction model achieved a state-of-the-art prediction accuracy on both datasets. We evaluated our CNN-based count model with a similar approach that uses a segmentation based method, and it outperformed it considerably.

We generated synthetic images for missing count classes in the "old" root dataset since our approach requires data for all classes, and found that they are comparable (both visually and when automatically segmented) to the real mages. We investigated incorrect counting by our model and found they were often caused by storage roots lying closer to the boundary age (2.5 months) used to separate "old" and "young" roots. We also found some incorrect classifications caused by testing varieties of cassava roots that were few or missing from the training data. As future work, we propose to collect more data for each variety of cassava roots in our dataset, which will help improve our models' performance. Even though we can generate these images with our conditional GAN, this also requires more data to produce realistic images. We also propose to design additional CNN-Based models that will predict the total length and volume of cassava storage roots, which will help us develop a complete image-based cassava root phenotyping system. The approach here has been developed to support cassava phenotyping work, in particular to support the development of a low-cost aeroponic phenotyping system. Future work will need to consider the ease of transfer from this system to other, similar systems, and a transfer learning approach may be required to update models for new image sets.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

JA was primary author and developer of the models and experiments. AF and TP supervised JA, and assisted with model development and manuscript writing. MM and MS provided biological expertise and data collection and annotation in Colombia.

FUNDING

This work was supported by the Biotechnology and Biological Sciences Research Council [BB/P022790/1].

ACKNOWLEDGMENTS

The authors want to thank Hernan Ceballos, Hershey Clair, and Joe Tohme, from the International Center for Tropical Agriculture (CIAT) for their guidance and support for cassava research. We would like to acknowledge Jorge Casas, CIAT for image annotations.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01516/full#supplementary-material

Supplementary Figure 1–15 are available. These detail more example images of GAN generation, age prediction, and storage root counting.

REFERENCES

Adu, M. O., Asare, P. A., Asare-Bediako, E., Amenorpe, G., Ackah, F. K., Afutu, E., et al. (2018). Characterising shoot and root system trait variability and contribution to genotypic variability in juvenile cassava (manihot esculenta crantz) plants. *Heliyon* 4, e00665. doi: 10.1016/j.heliyon.2018.e00665

Aich, S., and Stavness, I. (2017). Leaf counting with deep convolutional and deconvolutional networks, in The 2017 IEEE International Conference on Computer Vision Workshops (ICCVW). Venice, Italy. pp. 2080–2089. doi: 10.1109/ICCVW.2017.244

Aich, S., Josuttes, A., Ovsyannikov, I., Strueby, K., Ahmed, I., Duddu, H. S., et al. (2018). Deepwheat: Estimating phenotypic traits from crop images with deep learning: The 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (IEEE). pp. 323–332. doi: 10.1109/WACV.2018.00042

Atanbori, J., Chen, F., French, A. P., and Pridmore, T. (2018). "Towards low-cost image-based plant phenotyping using reduced-parameter cnn" in *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*. Eds. Tsaftaris, H. S. S. A., and Pridmore, T. (UK: BMVA Press).

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi: 10.1109/TPAMI.2016.2644615

Belalcazar, J., Dufour, D., Andersson, M. S., Pizarro, M., Luna, J., Londoño, L., et al. (2016). High-throughput phenotyping and improvements in breeding cassava for increased carotenoids in the roots. *Crop Sci.* 56, 2916–2925. doi: 10.2135/cropsci2015.11.0701

Chen, S. W., Shivakumar, S. S., Dcunha, S., Das, J., Okon, E., Qu, C., et al. (2017). Counting apples and oranges with deep learning: A data-driven approach. *IEEE Robotics Automation Lett.* 2, 781–788. doi: 10.1109/LRA.2017.2651944

Chollet, F. (2017). "Xception: Deep Learning with Depthwise Separable Convolutions", The 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, pp. 1800-1807. doi: 10.1109/CVPR.2017.195

Furbank, R. T., and Tester, M. (2011). Phenomics-technologies to relieve the phenotyping bottleneck. *Trends In Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ* 5, e4088–e4088. doi: 10.7717/peerj.4088

Giuffrida, M. V., Minervini, M., and Tsaftaris, S. (2015). “Learning to count leaves in rosette plants,” in *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*. Eds. Tsaftaris, H. S. S. A., and Pridmore, T. (BMVA Press, UK), 1.1–1.13. doi: 10.5244/C.29.CVPPP.1

Giuffrida, M. V., Scharr, H., and Tsaftaris, S. A. (2017). “Arigan: Synthetic arabidopsis plants using generative adversarial network,” in *Proceedings of the 2017 IEEE International Conference on Computer Vision Workshop (ICCVW)* (Venice, Italy), 2064–2071. doi: 10.1109/ICCVW.2017.242

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., et al. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. Available at: https://arxiv.org/abs/1704.04861

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks, in *Proceedings of the IEEE conference on computer vision and pattern recognition.* (pp. 4700-4708). doi: 10.1109/CVPR.2017.243

Intel Corporation. (2017). The OpenCV Library (Version 3.2.0) [Software]. Retrieved from https://opencv.org/

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). “Image-to-image translation with conditional adversarial networks”, *The 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, pp. 5967-5976. doi: 10.1109/CVPR.2017.632 arXiv preprint.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., et al. (2017). Photo-realistic single image super-resolution using a generative adversarial network, In *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 4681–4690.

Lin, G., Milan, A., Shen, C., and Reid, I. (2017). "Refinenet: Multi-path refinement networks for high-resolution semantic segmentation", in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi: 10.1109/CVPR.2017.549

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation, in: In Proceedings of the IEEE conference on computer vision and pattern recognition . pp. 3431–3440. doi: 10.1109/CVPR.2015.7298965

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016a). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recog. Lett.* 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Minervini, M., Giuffrida, M. V., and Tsaftaris, S. A. (2016b). In S. A. Tsaftaris, H. Scharr, and T. Pridmore, editors, Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP), pages 6.1-6.13. BMVA Press, UK, September 2015.

Okogbenin, E., Setter, T. L., Ferguson, M., Mutegi, R., Ceballos, H., Olasanmi, B., et al. (2013). Phenotypic approaches to drought in cassava. *Front. In Physiol.* 4, 93. doi: 10.3389/fphys.2013.00093

Polthanee, A., Taboonmuang, R., and Manaonok, J. (2016). Root yield and nutrient removal of four cassava cultivars planted in early rainy season of northeastern thailand: Crop experienced to drought at mid-growth stage. *Asian J. Crop Sci.* 8, 24–30. doi: 10.3923/ajcs.2016.24.30

Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., et al. (2017a). Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *GigaScience*. 6 (10), gix083. doi: 10.1093/gigascience/gix083

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017b). Deep learning for multi-task plant phenotyping. The 2017 IEEE Conference on Computer Vision and Pattern Recognition. pp. 2055–2063. doi: 10.1109/ICCVW.2017.241

Rahnemoonfar, M., and Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. *Sensors* 17, 905. doi: 10.3390/s17040905

Ribera, J., Chen, Y., Boomsma, C., and Delp, E. J. (2017). Counting plants using deep learning, 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, pp. 1344–1348. doi: 10.1109/GlobalSIP.2017.8309180

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. The 2018 IEEE Conference on Computer Vision and Pattern Recognition. pp. 4510–4520. doi: 10.1109/CVPR.2018.00474

Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., et al. (2016). Leaf segmentation in plant phenotyping: a collation study. *Mach. Vision Appl.* 27, 585–606. doi: 10.1007/s00138-015-0737-3

Scharr, H., Pridmore, T., and Tsaftaris, S. A. (2017). Computer vision problems in plant phenotyping, cvppp 2017: Introduction to the CVPPP 2017 workshop papers, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW) (IEEE). pp. 2020–2021. doi: 10.1109/ICCVW.2017.236

Selvaraj, M.G., Montoya-P, M.E., Atanbori, J. et al. (2019) A low-cost aeroponic phenotyping system for storage root development: unravelling the below-ground secrets of cassava *(Manihot esculenta)*. *Plant Methods* 15, 131. doi: 10.1186/s13007-019-0517-6

Subere, J. O. Q., Bolatete, D., Bergantin, R., Pardales, A., Belmonte, J. J., Mariscal, A., et al. (2009). Genotypic variation in responses of cassava (manihot esculenta crantz) to drought and rewatering: root system development. *Plant Production Sci.* 12, 462–474. doi: 10.1626/pps.12.462

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. In Plant Sci.* 8, 1190. doi: 10.3389/fpls.2017.01190

Uzal, L. C., Grinblat, G. L., Namías, R., Larese, M. G., Bianchi, J., Morandi, E., et al. (2018). Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. In Agric.* 150, 196–204. doi: 10.1016/j.compag.2018.04.024

Ward, D., Moghadam, P., and Hudson, N. (2018). “Deep Leaf Segmentation Using Synthetic Data”, In *Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*. Eds. Tsaftaris, H. S. S. A., and Pridmore, T. (BMVA Press, UK).

Yu, F., and Koltun, V. (2016). Multi-scale context aggregation by dilated convolutions. In *Proceedings of the International Conference on Learning Representations (ICLR) (2016),* Available at: https://arxiv.org/abs/1511.07122

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network, in: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). pp. 2881–2890. doi: 10.1109/CVPR.2017.660

Zhu, Y., Aoun, M., Krijn, M., Vanschoren, J., and Campus, H. T. (2018). “Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants”, *In Proceedings of the Computer Vision Problems in Plant Phenotyping (CVPPP)*. Eds. Tsaftaris, H. S. S. A., and Pridmore, T. (BMVA Press, UK).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Plant Science

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/fpls.2019.01629

Automated Methods Enable Direct Computation on Phenotypic Descriptions for Novel Candidate Gene Prediction

Ian R. Braun*[1,2] *and Carolyn J. Lawrence-Dill*[1,2,3]*

[1] Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, IA, United States, [2] Interdepartmental Bioinformatics and Computational Biology, Iowa State University, Ames, IA, United States, [3] Department of Agronomy, Iowa State University, Ames, IA, United States

Natural language descriptions of plant phenotypes are a rich source of information for genetics and genomics research. We computationally translated descriptions of plant phenotypes into structured representations that can be analyzed to identify biologically meaningful associations. These representations include the entity–quality (EQ) formalism, which uses terms from biological ontologies to represent phenotypes in a standardized, semantically rich format, as well as numerical vector representations generated using natural language processing (NLP) methods (such as the bag-of-words approach and document embedding). We compared resulting phenotype similarity measures to those derived from manually curated data to determine the performance of each method. Computationally derived EQ and vector representations were comparably successful in recapitulating biological truth to representations created through manual EQ statement curation. Moreover, NLP methods for generating vector representations of phenotypes are scalable to large quantities of text because they require no human input. These results indicate that it is now possible to computationally and automatically produce and populate large-scale information resources that enable researchers to query phenotypic descriptions directly.

Keywords: ontology, natural language processing, machine learning, semantic similarity, phenotype, phenologs

OPEN ACCESS

Edited by:
Ingo Ebersberger,
Goethe-Universität Frankfurt am Main, Germany

Reviewed by:
Matthias Lange,
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany
Gerhard Buck-Sorlin,
Agrocampus Ouest, France

****Correspondence:***
Carolyn J. Lawrence-Dill
triffid@iastate.edu

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *05 July 2019*
Accepted: *19 November 2019*
Published: *10 January 2020*

Citation:
Braun IR and Lawrence-Dill CJ (2020) Automated Methods Enable Direct Computation on Phenotypic Descriptions for Novel Candidate Gene Prediction. Front. Plant Sci. 10:1629. doi: 10.3389/fpls.2019.01629

BACKGROUND

Phenotypes encompass a wealth of important and useful information about plants, potentially including states related to fitness, disease, and agricultural value. They comprise the material on which natural and artificial selection act to increase fitness or to achieve desired traits, respectively. Determining which genes are associated with traits of interest and understanding the nature of these relationships is crucial for manipulating phenotypes. When causal alleles for phenotypes of interest are identified, they can be selected for in populations, targeted for deletion, or employed as transgenes to introduce desirable traits within and across species. The process of identifying candidate genes and specific alleles associated with a trait of interest is called candidate gene prediction.

Genes with similar sequences often share biological functions and therefore can create similar phenotypes. This is one reason sequence similarity search algorithms like BLAST (Altschul et al.,

1990) are so useful for candidate gene prediction. However, similar phenotypes can also be attributed to the function of genes that have no sequence similarity. This is how protein-coding genes that are involved in different steps of the same metabolic pathway or transcription factors involved in regulating gene expression contribute to shared phenotypes. For example, knocking out any one of the many genes involved in the maize anthocyanin pathway can result in pigment changes (reviewed in Sharma et al., 2011). This concept is modelled in **Figure 1**, where, notably, the sequence-based search with Gene 1 as a query can only return genes with similar sequences, but querying for similar phenotypes to those associated with Gene 1 returns many additional candidate genes.

High-throughput and computational phenotyping methods are largely sensor and image-based (Fahlgren et al., 2015). These methods can produce standardized datasets such that, for example, an image can be analyzed, data can be extracted, and those data can be interrogated (Green et al., 2012; Gehan et al., 2017; Miller et al., 2017). However, while such methods are adept at comparing phenotypic information between plants that are physically similar, they are limited in their ability to transfer this knowledge between physically dissimilar species. For example, traits such as leaf angle vary greatly among different species, and therefore cannot be compared directly. Moreover, where shared pathways and processes are conserved across broad evolutionary distances, it can be hard to identify equivalent phenotypes. McGary et al. (2010) call these non-obvious shared phenotypes phenologs. Between species, phenologs may present as equivalent properties in disparate biological structures (Braun et al., 2018). For example, *Arabidopsis* KIN-13A mutants and mouse KIF2A mutants both show increased branching in single-celled structures, but with respect to neurons in mouse (Homma et al., 2003) and with respect to trichomes in *Arabidopsis* (Lu et al., 2004). Taken together, the ability to compute on phenotypic descriptions to identify phenologs within and across species has the potential to aid in the identification of novel candidate genes that cannot be identified by sequence-based methods alone and that cannot be identified *via* image analysis.

In order to identify phenologs, some methods rely on searching for shared orthologs between causal gene sets (McGary et al., 2010; Woods et al., 2013). For example, McGary et al. (2010) identified a phenolog relationship between "abnormal heart development" in mouse and "defective response to red light" in *Arabidopsis* by identifying four orthologous genes between the sets of known causal genes in each species. However, these methods are not applicable when the known causal gene set for one phenotype or the other is small or non-existent. In these cases, using natural language descriptions to identify phenologs avoids this problem by relying only on characteristics of the phenotypes, per se. These phenotypic descriptions are a rich source of information that, if leveraged to identify phenolog pairs, can enable identification of novel candidate genes potentially involved in generating phenotypes beyond what has already been described.

Unfortunately, computing on phenotype descriptions is not straightforward. Text descriptions of phenotypes present in the literature and in online databases are irregular because natural language representations of even very similar phenotypes can be highly variable. This makes reliable quantification of phenotype similarity particularly challenging (Thessen et al., 2012; Braun et al., 2018). To represent phenotypes in a computable manner, researchers have recently begun to translate and standardize phenotype descriptions into entity–quality (EQ) statements composed of ontology terms, where an entity (e.g., "leaf") is modified by a quality (e.g., "increased length"; Mungall

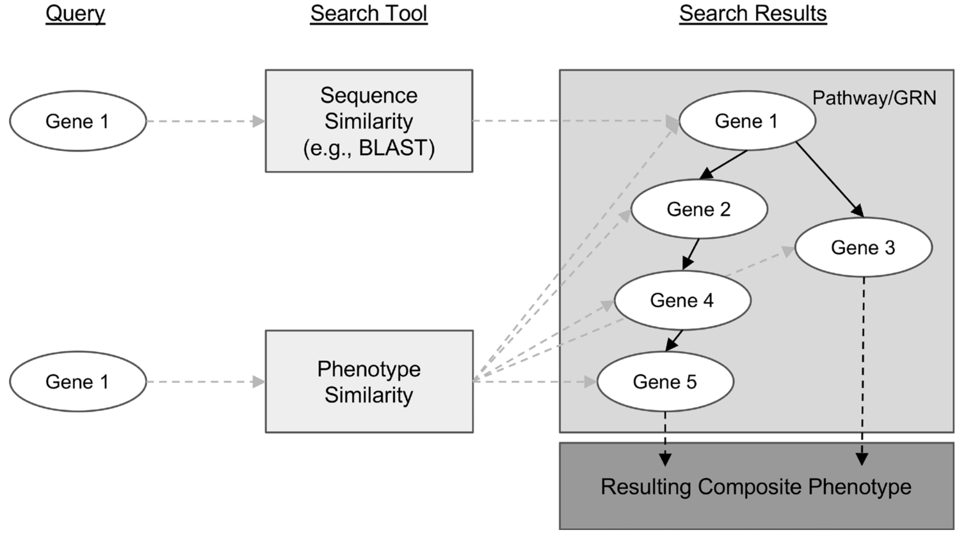


FIGURE 1 | Conceptual comparison of querying with a gene sequence or its associated phenotypic description. Genes are shown as *white ovals*. Methods of searching for related genes are shown as *light gray boxes*. *Gray dashed arrows* indicate the path from the query gene to the set of genes that are returned from the search. *Solid black arrows* indicate relationships between genes in a biological pathway or gene regulatory network. *Dashed black arrows* indicate relationships between the pathway or network and the resulting phenotype.

et al., 2010).[1] Using this formalism, complex phenotypes are represented by multiple EQ statements. For example, multiple EQ statements are required to represent dwarfism, where the entity and quality pairs ("plant height," "reduced") and ("leaf width," "increased") may be used, among others. Each of these phenotypic components of the more general phenotype is termed a "phene." Because both entities and qualities are represented by terms from biological ontologies (fixed vocabularies arranged as hierarchical concepts in a directed acyclic graph), quantifying the similarity between two phenotypes that have been translated to EQ statements can be accomplished using graph-based similarity metrics (Hoehndorf et al., 2011; Slimani, 2013). Such techniques for estimating semantic similarity based on arranging concepts hierarchically in a graph have long been employed in the field of natural language processing (NLP; e.g., Resnik, 1999) and, as applied to biological ontologies, have been useful in applications from clustering gene function annotations for data visualization (Supek et al., 2011) to assessing functional similarities between orthologous genes (Altenhoff et al., 2016).

Oellrich, Walls et al. (2015) developed Plant PhenomeNET, an EQ statement-based resource primarily consisting of a phenotype similarity network containing phenotypes across six different model plant species, namely, *Arabidopsis* (*Arabidopsis thaliana*), maize (*Zea mays* ssp. *mays*), tomato (*Solanum lycopersicum*), rice (*Oryza sativa*), *Medicago* (*Medicago truncatula*), and soybean (*Glycine max*). Their analysis demonstrated that the method developed by Hoehndorf et al. (2011) could be used to recover known genotype to phenotype associations for plants. The authors found that highly similar phenotypes in the network (phenologs) were likely to share causal genes that were orthologous or involved in the same biological pathways. In constructing the network, text statements comprising each phenotype were converted by hand into EQ statements primarily composed of terms from the Phenotype and Trait Ontology (PATO; Gkoutos et al., 2005), Plant Ontology (PO; Cooper et al., 2013), Gene Ontology (GO; Ashburner et al., 2000), and Chemical Entities of Biological Interest (ChEBI; Hastings et al., 2013) ontology.

The success of this plant phenotype pilot project was encouraging, but to scale up to computing on all available phenotypic data for each of the six species was not a reasonable goal given that curating data for this pilot project took approximately 2 years and covered only phenotypes of dominant alleles for 2,747 genes across the six species. More specifically, human translation of text statements into EQ statements is the most time-consuming aspect of generating phenotype similarity networks using this method. Automation of this translation promises to increase the rate at which such networks can be generated and expanded. Notable efforts to automate this process include Semantic Charaparser (Cui, 2012; Cui et al., 2015), which extracts characters (entities) and their corresponding states (qualities) after a curation step that involves assigning terms to categories and then mapping these characters and states to EQ statements constructed from input ontologies. Other existing annotation tools such as NCBO Annotator (Musen et al., 2012) and NOBLE Coder (Tseytlin et al., 2016) are fully automated, relying only on input ontologies. Both map words in the input text to ontology terms without imposing an EQ statement structure. State-of-the-art machine learning approaches to annotating text with ontology terms also have been developed (Hailu et al., 2019). These can be trained using a dataset such as the Colorado Richly Annotated Full-Text corpus (CRAFT; Bada et al., 2012), but are not readily transferable to ontologies that are not represented in the training set.

In addition to using ontology-based methods, similarity between text descriptions of phenotypes can also be quantified using NLP techniques such as treating each description as a bag-of-words and comparing the presence or absence of those words between descriptions, or using neural network-based tools such as Doc2Vec to embed descriptions into abstract high-dimensional numerical vectors between which similarity metrics can then be easily applied (Mikolov et al., 2013; Le and Mikolov, 2014). Conceptually, this process involves converting natural language descriptions into locations in space, such that descriptions that are near each other are interpreted as having high similarity and those that are distant have low similarity.

In this work, we demonstrate that automated techniques for generating computable representations of natural language can be applied to a dataset of phenotypic descriptions in order to generate biologically meaningful phenotype similarity networks. See **Figure 2** for an overview of how phenotype similarity networks are computationally generated as an output when text descriptions are provided as the input. We first show that these computational techniques are limited in their capability to exactly reproduce the annotations and corresponding phenotype similarity networks generated with hand-curation. However, we subsequently show that the hand-curated network does not outperform networks built with purely computational approaches on dataset-wide tasks of biological relevance, such as organizing genes by function and predicting membership in biochemical pathways. Most importantly, we discuss how we can now use these computational approaches to automatically generate new datasets necessary to identify phenotypic similarities and predict gene function within and across species without requiring the use of time-consuming and costly hand-curation.

METHODS

Dataset of Phenotypic Descriptions and Curated EQ Statements

The pairwise phenotype similarity network described in Oellrich, Walls et al. (2015) was built based on a dataset of phenotype descriptions across six different model plant species (*A. thaliana*, *Z. mays* ssp. *mays*, *S. lycopersicum*, *O. sativa*, *M. truncatula*, and *G. max*). In that work, each phenotype description was split into one or more atomized statements describing individual phenes, each of which mapped to exactly one curated EQ statement (**Table 1**). The EQ statements in this dataset were primarily built from terms present in PATO, PO, GO, and ChEBI. For this work, we used this existing dataset as the source of genes and associated

---

[1] In relation to sentence structure, the entity represents the subject and the quality represents the predicate. Qualities are also elsewhere referred to as attributes, features, or characteristics of a biological structure or process.

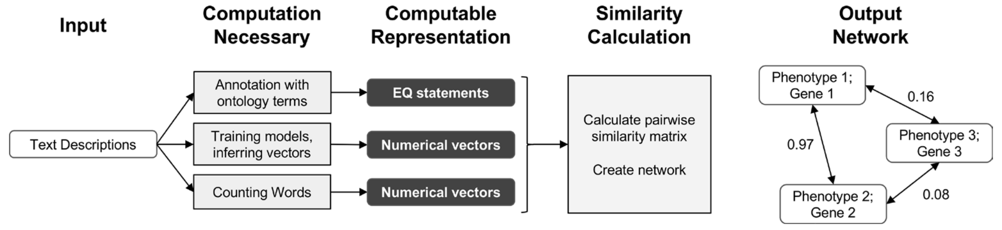


FIGURE 2 | Overview of computational pipelines used here to generate phenotype similarity networks from text descriptions of phenotypes. *Rounded white rectangles* represent data in the form of text descriptions as input or network nodes as output. *Rounded black rectangles* represent the intermediate data forms that are computable representations of text descriptions. These allow for quantitative similarity metrics to be applied. *Gray rectangles* represent computational methods carried out at each step. *Single-headed arrows* represent flow of data through each pipeline. *Double-headed arrows* represent edges between nodes in resulting similarity networks. Values *next to double-headed arrows* indicate magnitude of phenotype similarity. One output network is created for each computable representation, but only one example is shown here.

TABLE 1 | Description of the Oellrich, Walls et al. (2015) dataset in terms of number of phenotype descriptions, phene descriptions, and EQ statements.

Species	Phenotypes	Phenes[1]	EQ statements[2]
Arabidopsis	1385	5172	5172
Maize	117	373	373
Tomato	90	269	269
Rice	86	340	340
Medicago	40	149	149
Soybean	24	61	61
Example gene: Arabidopsis *PKS2* (ATIG14280.1)	Phenotype: short hypocotyl and expanded cotyledon under hourly far red pulses	Phene 1: short hypocotyl	PO:0020100 (hypocotyl) + PATO:0000574 (decreased length)
		Phene 2: xpanded cotyledon	PO:0020030 (cotyledon) + PATO:0000586 (increased size)

[1]Also referred to as 'atomized statements'.
[2]Each EQ statement represents a single specific phene.

phenotypic descriptions on which to test automated methods for assessing similarity networks between phenotypes and using the resulting phenotype similarity networks to perform comparative analyses across the whole dataset to predict gene function.

Computationally Generating EQ Statements From Phenotypic Descriptions

For each phenotype and phene description in the dataset, we computationally generated corresponding EQ statements without human interaction. To accomplish this, terms were first annotated to each text description and then combined to form complete EQ statements. Two different existing computational tools and a simple machine learning technique were used to map ontology terms to text descriptions. Specifically, these were NCBO Annotator and NOBLE Coder, which are tools for matching ontology terms to specific words in text, and a Naïve Bayes bag-of-words classifier, which assigns terms to descriptions based on the observed frequencies of term–word co-occurrence in a training dataset. The Oellrich, Walls et al. (2015) dataset of descriptions and curated EQ statements was split into four groups such that any three groups of the dataset were used to train a Naïve Bayes model that was then applied to the remaining group. The result of applying these three annotation methods was a set of ontology terms from PATO, PO, GO, and ChEBI assigned to each text description. Terms were then combined to form full EQ statements by assigning default root terms where none were matched, such as the entity term *whole plant* (PO:0000003), and organizing the matched terms into the different roles of the EQ statement by removing overlapping terms and automatically applying compositional rules used by curators in Oellrich, Walls et al. (2015). As an example, these rules include the fact that ChEBI terms cannot be the primary entity. The EQ statements were scored based on how well the terms aligned with the text description they were annotated to, so that the closest matching EQ statements for each text description were output and used downstream to generate phenotype similarity networks. See the *Supplemental Methods* section for a more detailed description of this process.

Computationally Generating Numerical Vectors From Phenotypic Descriptions

In addition to generating EQ statements for each phenotype and phene description in the dataset, Doc2Vec was used for generating numerical vectors for each description. A model pre-trained on Wikipedia was used (Lau and Baldwin, 2016). In these document embeddings, positions within the vector do not refer to the presence of specific words but rather abstract features learned by the model. A size of 300 was used for each vector representation, which is the fixed vector size of the pre-trained model. In addition, vectors were generated for each description using bag-of-words and set-of-words representations of the text. For these methods, each position within the vector refers to a particular word in the vocabulary. Each vector element with bag-of-words refers to the count of that word in the description, and each vector element with set-of-words is a binary value indicating presence or absence of the word. In cases where phene descriptions were used instead of phenotype descriptions, the descriptions were concatenated prior to embedding to obtain a single vector.

Creating Gene and Phenotype Networks

Oellrich, Walls et al. (2015) developed a network with phenotypes as nodes and similarity between them as edges for all the phenotypes in the dataset. For each type of text representations that we generated with computational methods, comparable networks were constructed. For EQ statement representations, Jaccard similarity either taking the structure and order of terms in the EQ statement into account (referred to as metric S_1) or ignoring the structure and treating the ontology terms in the EQ statement as an unordered set (referred to as metric S_2) were used to determine edge values. See the *Supplemental Methods* section for a more detailed description of these similarity metrics. For vector representations generated using Doc2Vec and bag-of-words, cosine similarity was used. For the vector representations generated using set-of-words, Jaccard similarity was used. These networks are considered to be simultaneously gene and phenotype similarity networks because each phenotype in the dataset corresponds to a specific causal gene and a node in the network represents both that causal gene and its cognate phenotype. However, two phenotype descriptions corresponding to the same gene are retained as two separate nodes in the network, so while each node represents a unique gene/phenotype pair, a single gene may be represented within more than one node.

RESULTS

Performance of Computational Methods in Reproducing Hand-Curated Annotations

We tested the ability of computational semantic annotation methods to assign ontology terms similar to those selected by curators to phenotype and phene descriptions in the Oellrich, Walls et al. (2015) dataset. Specifically, the ontology terms mapped by each method to a particular description were compared against the terms present in the EQ statement(s) that were created by hand-curation for that same description. Metrics of partial precision (PP) and partial recall (PR), as well as the harmonic mean of these values (PF_1) as a summary statistic, were used to evaluate performance (**Table 2**). Metrics PP and PR were applied as in Dahdul et al. (2018); see the *Supplemental Methods* section for a detailed description of these metrics.

NOBLE Coder and NCBO Annotator generally produced semantic annotations more similar to the hand-curated dataset using phenotype descriptions as inputs than using the set of phene descriptions as inputs, a result consistent across ontologies. We considered this to be counterintuitive because the phene descriptions are more directly related to the individual EQ statements in terms of semantic content. However, the set of target ontology terms considered correct is larger in the case of the phenotype descriptions because this set of terms includes all terms in any EQ statements derived from that phenotype rather than a single EQ statement, which could contribute to this measured increase in both partial recall and partial precision. Accounting for synonyms and related words generated through Word2Vec models increased PR in the case of specific annotation methods as the threshold for word similarity was decreased (from 1.0 to 0.5), but did not increase PF_1 in any instance due to the corresponding losses in PP (**Supplemental Figure 1**).

NOBLE Coder and NCBO Annotator performed comparably in the case of each type of text description and ontology, with NOBLE Coder using the precise matching parameter slightly outperforming the other annotation method with respect to these particular metrics for these particular descriptions. Both outperformed the Naïve Bayes classifier, for which performance dropped significantly for the ontologies with smaller relative representation in the dataset (GO and ChEBI), as might be expected. When the results were aggregated, the increase in partial recall for PATO, PO, and GO terms relative to the maximum recall

TABLE 2 | Performance metrics for semantic annotation methods.

Annotator	Ontology	*n*[1]	Phenotype Description			Phene Descriptions		
			PP[3]	*PR*[3]	*PF_1*[3]	*PP*[3]	*PR*[3]	*PF_1*[3]
NOBLE Coder (Precise)	PATO	7882	0.641	0.627	0.634	0.601	0.572	0.586
	PO	5634	0.622	0.380	0.472	0.546	0.294	0.382
	GO	1505	0.514	0.521	0.517	0.510	0.514	0.512
NOBLE Coder (Partial)	PATO	7882	0.412	0.748	0.532	0.375	0.689	0.486
	PO	5634	0.309	0.758	0.439	0.269	0.659	0.382
	GO	1505	0.102	0.846	0.182	0.091	0.839	0.165
NCBO Annotator	PATO	7882	0.640	0.619	0.629	0.598	0.563	0.580
	PO	5634	0.550	0.259	0.352	0.458	0.170	0.248
	GO	1505	0.478	0.433	0.454	0.480	0.424	0.450
	ChEBI	775	0.429	0.888	0.579	0.431	0.913	0.586
Naïve Bayes Classifier	PATO	7882	0.517	0.394	0.447	0.642	0.484	0.552
	PO	5634	0.474	0.258	0.334	0.636	0.429	0.512
	GO	1505	0.091	0.073	0.081	0.155	0.157	0.156
	ChEBI	775	0.035	0.031	0.033	0.001	0.001	0.001
Aggregate Annotations[2]	PATO	7882	0.412	0.798	0.543	0.383	0.815	0.522
	PO	5634	0.351	0.809	0.489	0.304	0.831	0.445
	GO	1505	0.107	0.839	0.190	0.090	0.839	0.163
	ChEBI	775	0.366	0.890	0.519	0.305	0.913	0.457

[1] *The number of terms from a given ontology in all curated EQ statements in the dataset.*

[2] *Annotation set formed by taking the union of the annotations from all other methods.*

[3] *Metrics are partial precision and recall (PP, PR; Dahdul et al., 2018) and their harmonic mean (PF_1).*

achieved by any individual method indicates that the curated terms that were recalled by each method were not entirely overlapping. This is as expected given that different methods used for semantic annotation recalled target (curated) ontology terms to different degrees, as measured by Jaccard similarity of a given target term to the closest predicted term annotated by that particular method. These sets of obtained similarities to target terms were comparable between NCBO Annotator and NOBLE Coder ($\rho = 0.84$ with phene descriptions and $\rho = 0.86$ with phenotype descriptions) and dissimilar between either of those methods and the Naïve Bayes classifier ($\rho < 0.10$ in both cases for either type of description) using Spearman rank correlation adjusted for ties.

These results indicate that automated annotation methods (NCBO Annotator, NOBLE Coder, and Naïve Bayes classifier) do not reproduce the exact same ontology term annotations selected by hand-curation for each phenotypic description, as expected. Given this result, we next assessed how these differences between the hand-curated annotations and computationally generated annotations translated into differences between the phenotype similarity networks based on these annotations.

Comparing Computational Networks to the Hand-Curated Network

Oellrich, Walls et al. (2015) developed a network with phenotype/gene pairs as nodes and similarity between them as edges for all phenotypes in the dataset. In this work, comparable networks were constructed for the same dataset using a number of computational approaches for representing phenotype and phene descriptions and for predicting similarity. For the purposes of this assessment, the network built from hand-curated EQ statements and described in Oellrich, Walls et al. (2015) is considered the gold standard against which each network we produced is compared. The computational and gold standard networks were compared using the F_1 metric to assess similarity in predicted phenolog pairs at a range of k values, where k is the allowed number of phenolog pairs predicted by the networks (the k most highly valued edges). Results are reported through $k = 583,971$, which is the number of non-zero similarities between phenotypes in the gold standard network, and were repeated using phenotype descriptions and phene descriptions as inputs to the computational methods (**Figure 3**). The simplest NLP methods for assessing similarity (set-of-words and bag-of-words) consistently recapitulated the gold standard network the best using phenotype descriptions, whereas the document embedding method using Doc2Vec outperformed these methods for values of $k \leq 200,000$ based on phene descriptions. The differences in the performance of each method are robust to 80% subsampling of the phenotypes present in the dataset.

These results illustrate that computational methods do not exactly reproduce the phenotype similarity network built from the hand-curated EQ statements. However, this does not necessarily mean that the hand-curated network is inherently more biologically meaningful. To assess how useful each network is in a biological context, we next compared how the hand-curated

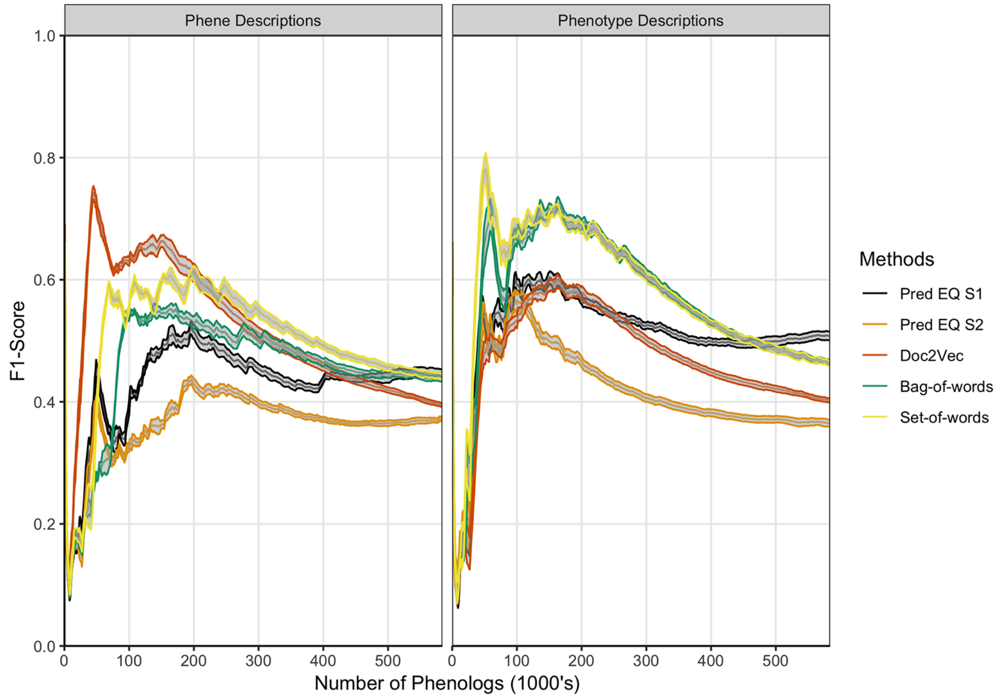


FIGURE 3 | Comparison of phenolog pairs identified by predictive methods in comparison to the Oellrich, Walls et al. (2015) dataset. The *x*-axis indicates the number of phenologs pairs (highest valued edges in the phenotype similarity network) at each point. The standard deviation of resampling with 80% of the phenotypes in the dataset (network nodes) are indicated by *ribbons* for each method. Phene descriptions (*left*) or phenotype descriptions (*right*) were used as the text input for each particular method.

network and each computational network performed on the task of sorting genes into functional groups.

Computational Methods Outperform Hand-Curation for Gene Functional Categorization in *Arabidopsis*

Lloyd and Meinke (2012) previously organized a set of *Arabidopsis* genes with accompanying phenotype descriptions into a functional hierarchy of groups (e.g., "morphological"), classes (e.g., "reproductive"), and finally subsets (e.g., "floral"), in order from most general to most specific. See Supplemental Table 1 in Lloyd and Meinke (2012) for a full specification of this hierarchy to which the genes were assigned, and Supplemental Table 2 in Lloyd and Meinke (2012) for a mapping between genes and this hierarchical vocabulary. Oellrich, Walls et al. (2015) later used this set of genes and phenotypes to validate the quality of their dataset of hand-curated EQ statements by reporting the average similarity of phenotypes (translated into EQ statements) that belonged to the same functional subset. We used this same functional hierarchy categorization and a similar approach to assess the utility of computationally generated representations of phenotypes towards correctly categorizing the functions of the corresponding genes and to compare this utility against that of the dataset of hand-curated EQ statements. For each class and subset in the hierarchy, the mean similarity between any two phenotypes related to genes within that class or subset ("within" mean) was quantified using each computable representation of interest and compared to the mean similarity between a phenotype related to a gene within that class or subset and one outside of it ("between" mean), quantified in terms of standard deviation of the distribution of all similarity scores generated for each given method. The difference between the "within" mean and "between" mean (referred to here as the Consistency Index) for each functional category for each method indicates the ability of that method to generate strong similarity signal for phenotypes in this dataset that share that function (**Figure 4**). In the case of these data, most computational methods using either phene or phenotype descriptions as the input text were able to recapitulate the signal present in the network Oellrich, Walls et al. (2015) generated from hand-curated EQ statements, and the simplest NLP methods (bag-of-words and set-of-words) produced the most consistent signal.

In order to more directly compare each method on a general classification task, networks constructed from curated EQ statements and those generated using each computational method were used to iteratively classify each *Arabidopsis* phenotype into classes and subsets. This was accomplished by removing one phenotype at a time and withholding the remaining phenotypes as training data, learning a threshold value from the training data, and then classifying the held-out phenotype by calculating its average similarity to each training data phenotype in each class or subset and classifying it as belonging to any category for which the average similarity to other phenotypes in that category exceeded the learned threshold. Performance on this classification task using each network was assessed using the F_1 metric, where the functional category assignments for each gene reported by Lloyd and Meinke (2012) were considered to be the correct classifications (**Table 3**). The simplest NLP methods (bag-of-words and set-of-words) outperformed the Oellrich, Walls et al. (2015) hand-curated EQ statement network on this classification task in all cases, while using the computationally

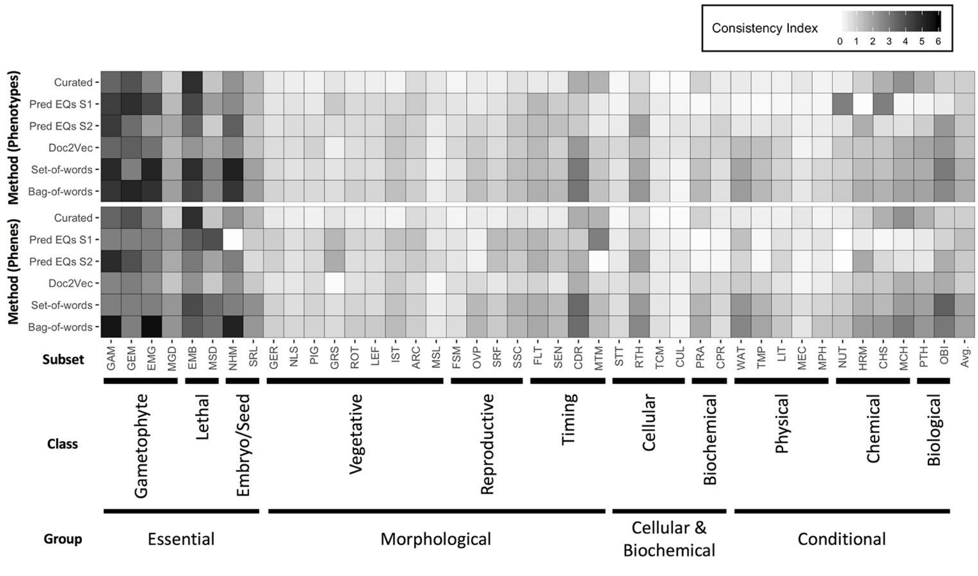


FIGURE 4 | Heatmap of Consistency Index. The difference between average similarity for two phenotypes within a subset and one phenotype within and one outside, for each functional subset defined in the dataset of *Arabidopsis* phenotypes, and for each method of quantifying similarity between phenotypes is shown, with *darker cells* indicating higher consistency within a subset. Differences are measured in standard deviations of the distributions of similarities obtained for each method. The meaning of subset abbreviations are specified in Supplemental Table 1 of Lloyd and Meinke (2012). Methods are listed at *left*. Input text for calculating similarities between the phenotypes were either derived from phenotype descriptions (*top*) or phene descriptions (*bottom*). The *far right column* in the heatmap refers to an average Consistency Index for a given method across all subsets.

TABLE 3 | Evaluation (F_1 scores) for each method used to categorize *Arabidopsis* genes by function.

Method	Phenes		Phenotypes	
	Class	Subset	Class	Subset
Curated EQ	0.470	0.359	0.470	0.359
Pred EQ S1	0.472	0.472	0.369	0.320
Pred EQ S2	0.504	0.413	0.437	0.368
Set-of-words	0.613	0.447	0.587	0.426
Bag-of-words	0.595	0.423	0.549	0.409
Doc2Vec	0.455	0.331	0.486	0.377

generated EQ statements or document embeddings generated with Doc2Vec only outperformed the curated EQ statement network in some cases.

Taken together, these results indicate that even though the computationally generated networks are significantly different than the hand-curated network (**Figure 3**), they generally perform equally well *or better* on tasks related to organizing *Arabidopsis* genes into functional groups. We next examined how these networks compare on the task of predicting biochemical pathway membership for specific genes, both within a single species and across multiple species.

Computational Methods Outperform Hand-Curation for Recovering Genes Involved in Anthocyanin Biosynthesis Both Within and Between Species

Oellrich, Walls et al. (2015) illustrated the utility of using EQ statement representations of phenotypes to provide semantic information necessary to recover shared membership of causal genes in regulatory and metabolic pathways. Specifically, they showed that by querying a six-species phenotype similarity network with the *c2* (*colorless2*) gene in maize, which is involved in anthocyanin biosynthesis, genes *c1*, *r1*, and *b1* (*colorless1*, *red1*, and *booster1*), which are also involved in anthocyanin biosynthesis in maize, are recovered. Querying in this instance is defined as returning other genes in the similarity network, ranked using the maximal value of the edges connecting a phenotype corresponding to the query gene and a phenotype corresponding to each other gene in the network. There are 2,747 genes in the dataset, so querying with one gene returns a ranked list of 2,746 genes. This result was included by Oellrich, Walls et al. (2015) as a specific example of the general utility of the phenotype similarity network to return other members of a pathway or gene regulatory network when querying with a single gene. See **Figure 1** for a general illustration of this concept.

To evaluate this same utility in the phenotype similarity networks we generated using computational methods and to compare their utility to that of the network from Oellrich, Walls et al. (2015) generated using hand-curated EQ statements, we first expanded the set of maize anthocyanin pathway genes to include those present in the description of the pathway given by Li et al. (2019), and listed in Supplementary Table 1 of that publication. Of those genes, 10 are present in the Oellrich, Walls et al. (2015) dataset (**Table 4**). Additionally, we likewise identified the set of *Arabidopsis* genes known to be involved in anthocyanin biosynthesis (listed in **Table 1** of Appelhagen et al., 2014) that were present in the Oellrich, Walls et al. (2015) dataset. This yielded a total of 16 *Arabidopsis* genes (**Table 5**).

Recovering Anthocyanin Biosynthesis Genes Within a Single Species

Using each phenotype similarity network, each anthocyanin biosynthesis gene from one species was iteratively used as a query against the network. The rank of each other gene in the set of anthocyanin biosynthesis genes corresponding to the same species as the query was quantified. We grouped the ranks into bins of width 10 for ranks less than or equal to 50 and combined all ranks greater than 50 into a single bin. For each phenotype similarity network, the mean and standard deviation of the number of anthocyanin biosynthesis genes in each bin were calculated (**Figure 5**). The average number of pathway genes ranked within the top 10 across all queries was greater for all computationally generated networks than for the network built from hand-curated EQ statements, although variance across the queries was high. In general, computational networks built from predicted EQ statements performed best for this task, whereas the network built using the hand-curated EQs performed the worst. The networks constructed using the numerical vector representations (set-of-words, bag-of-words, and Doc2Vec) were intermediate in performance as a group (**Figure 5**).

Recovering Anthocyanin Biosynthesis Genes Between Two Species

To determine whether the methods performed similarly both within and across species, we repeated the analysis described in the previous section (*Recovering Anthocyanin Biosynthesis Genes Within a Single Species*), but instead of quantifying the ranks of all anthocyanin biosynthesis genes from the same species as the query gene, we quantified the ranks of all anthocyanin genes that derived from the other species. In other words, *Arabidopsis* genes were used to query for maize genes, and maize genes were used to query for *Arabidopsis* genes. As shown in **Figure 6**, the phenotype similarity network constructed from hand-curated EQ statements did not recover (provide ranks of less than or equal to 50) any of the anthocyanin biosynthesis genes when queried with genes from the other species. Networks generated using the set-of-words and bag-of-words approaches, or with Doc2Vec, performed similarly, recovering on average less than one anthocyanin biosynthesis gene per query. Only networks built from computationally generated EQ statements recovered an appreciable number of anthocyanin biosynthesis genes on average across the queries between species (**Figure 6**).

DISCUSSION

Computationally Generated Phenotype Representations Are Useful

A primary purpose for generating representations of phenotypes that are easy to compute on (EQ statements, vector embeddings, etc.) is to construct similarity networks that enable the use of one

TABLE 4 | Maize genes involved in anthocyanin biosynthesis.

Gene name (Symbol)	Gene model ID[1]	Category[2]	Encoded protein[3]
colorless2 (c2)	**GRMZM2G422750**	Enzyme	naringenin-chalcone synthase
chalcone flavone isomerase1 (chi1)	GRMZM2G155329	Enzyme	chalcone isomerase
red aleurone1 (pr1)	GRMZM2G025832	Enzyme	flavonoid 3'-hydroxylase (flavonoid 3'-monooxygenase)
flavone 3-hydroxylase1 (fht1; F3H)	GRMZM2G062396	Enzyme	flavonone 3'-hydroxylase (flavonol synthase)
anthocyaninless1 (a1)	**GRMZM2G026930**	Enzyme	dihydroflavonol 4-reductase (flavonone 4-reductase)
anthocyaninless2 (a2)	**GRMZM2G345717**	Enzyme	anthocyanidin synthase (leucoanthocyanidin dioxygenase)
bronze1 (bz1)	**GRMZM2G165390**	Enzyme	flavonol 3-O-glucocyltransferase
bronze2 (bz2)	**GRMZM2G016241**	Enzyme	glutathione transferase (maleylacetoacetate isomerase)
multidrug resistance associated protein3 (mrpa3; ZmMrp4)	GRMZM2G111903	Transporter	multidrug-resistance-like-transporter
scutellar node color1 (sn1)	GRMZM5G822829	T F	bHLH
colorless1 (c1)	**GRMZM2G005066**	T F	R2 R3-MYB
pericarp color1 p1	**GRMZM2G084799**	T F	R2 R3-MYB
purple plant 1 (pl1)	**GRMZM2G701063**	T F	R2 R3-MYB
colored1 (r1)	**GRMZM5G822829**	T F	bHLH
colored plant (b1)	**GRMZM2G172795**	T F	bHLH
pale aleurone color1 (pac1)	GRMZM2G058432	T F	WD40

[1] *Gene model IDS in bold were present in the* Oellrich, Walls et al. (2015) *dataset.*

[2] *The abbreviation TF is short for transcription factor.*

[3] *Enzyme encoded protein names from the Plant Metabolic Network* (Schläpfer et al., 2017).

phenotype as a query to retrieve similar phenotypes. This process serves as a means of discovering relatedness between phenotypes (potential phenologs) within and across species, thus generating hypotheses about underlying genetic relatedness (reviewed in Oellrich, Walls et al., 2015).

The computational methods discussed in this work were demonstrated to only partially recapitulate the phenotype similarity network constructed by Oellrich, Walls et al. (2015) using hand-curated EQ statements (*Comparing Computational Networks to the Hand-Curated Network*). Despite the limited similarity between the network built from hand-curated annotations and the computationally generated networks, the computationally generated networks performed as well or better than the hand-curated network (based on curated EQ statements) in terms of correctly organizing phenotypes and their causal genes into functional categories at multiple hierarchical levels (*Computational Methods Outperform Hand-Curation for Gene Functional Categorization in* Arabidopsis). In addition, each computationally generated network performed better than the hand-curated network for querying with either maize or *Arabidopsis* anthocyanin biosynthesis genes to return other anthocyanin biosynthesis genes from the same species (*Recovering Anthocyanin Biosynthesis Genes Within a Single Species*), a task originally used to demonstrate the utility of the phenotype similarity network constructed in Oellrich, Walls et al. (2015).

Moreover, the networks built from computationally generated EQ statements were useful for recapturing anthocyanin biosynthesis genes from a species different than the species of origin for the queried gene/phenotype pair. None of the other networks, including the network built from curated EQ statements, exhibited this utility for this task (*Recovering Anthocyanin Biosynthesis Genes Between Two Species*). This particular result indicates that high accuracy of constructed EQ statements is not specifically necessary for tasks such as querying for related genes across species because potentially inaccurate (computationally predicted) EQ statements generated a more successful network for the task. Replicating these analyses with phenotype descriptions in a different biological domain, such as vertebrates, would determine whether these results generalize to additional species groups and datasets.

TABLE 5 | *Arabidopsis* genes involved in anthocyanin biosynthesis.

Gene Name (Symbol)	Locus Name[1]	Category[2]	Encoded Protein[3]
TRANSPARENT TESTA 4 (TT4)	**At5g13930**	Enzyme	naringenin-chalcone synthase
TRANSPARENT TESTA 5 (TT5)	**At3g55120**	Enzyme	chalcone isomerase
TRANSPARENT TESTA 6 (TT6)	**At3g51240**	Enzyme	flavanone 3'-hydroxylase (flavonol synthase)
TRANSPARENT TESTA 7 (TT7)	**At5g07990**	Enzyme	flavonoid 3'-hydroxylase (flavonoid 3'-monooxygenase)
TRANSPARENT TESTA 3 (TT3)	**At5g42800**	Enzyme	dihydroflavonol 4-reductase (flavonone 4-reductase)
TRANSPARENT TESTA 11 (TT11) *TRANSPARENT TESTA 17 (TT17)* *TRANSPARENT TESTA 18 (TT18)* *TANNIN-DEFICIENT SEED 4(TDS4)*	At4g22880	Enzyme	anthocyanidin synthase (leucoanthocyanidindioxygenase)
ARABIDOPSIS SIP1 CLADE TRIHELIX 1 (AST1) *BANYULS (BAN1)*	**At1g61720**	Enzyme	anthocyanidin reductase
TRANSPARENT TESTA 14 (TT14) *TRANSPARENT TESTA 19 (TT19)*	**At5g17220**	Enzyme	glutathione transerase (maleylacetoacetate isomerase)
AUTOINHIBITED H^+ - ATPASE (AHA)	At1g17260	Enzyme	ATP-ase
TRANSPARENT TESTA 10 (TT10)	**At5g48100**	Enzyme	laccase
TRANSPARENT TESTA 5 (TT15)	**At1g43620**	Enzyme	3β- hydroxy sterol glucosyltransferase
TRANSPARENT TESTA 11 (TT12)	**At3g59030**	Transporter	MATEefflux proton antiporter
TRANSPARENT TESTA 16 (TT16)	**At5g23260**	T F	K-box, MADS-box
TRANSPARENT TESTA 1 (TT1)	**At1g34790**	T F	C2H2
TRANSPARENT TESTA 2 (TT2)	**At5g35550**	T F	bHLH
TRANSPARENT TESTA 8 (TT8)	**At4g09820**	T F	bHLH
TRANSPARENT TESTA GLABRA 1 (TTG1)	**At5g24520**	T F	WD40
TRANSPARENT TESTA GLABRA 1(TTG2)	**At2g37260**	T F	WRKY

[1] Locus names in bold were present in the Oellrich, Walls et al. (2015) *dataset.*
[2] The abbreviation TF is short for transcription factor.
*[3] Enzyme encoded protein names from the Plant Metabolic Network (*Schläpfer et al., 2017*).*

Taken together, these results over this particular dataset of phenotype descriptions suggest that while the EQ statements generated through manual curation are likely the most accurate and informative computable representation of a given phenotype in specific cases, other representations generated entirely computationally with no human intervention are capable of meeting or exceeding the performance of the hand-curated annotations on dataset-wide tasks such as sorting phenotypes and genes into functional categories, as well as in the case of specific tasks such as querying with particular genes to recover other genes involved in the same pathway. Therefore, in cases where the volume of data is large, the results are understood to be predictive, and manual curation is impractical, using automated annotation methods to generate large-scale phenotype similarity networks is a worthwhile goal and can provide biologically relevant information that can be used for hypothesis generation, including novel candidate gene prediction.

Multiple Approaches to Representing Natural Language Are Useful

EQ statement annotations comprising ontology terms allow for interoperability with compatible annotations from varied data sources. They are also a human-readable annotation format, meaning that a knowledgeable human could fix an incorrect annotation by selecting a more appropriate ontology term (a process that is not possible using abstract vector embeddings). Their uniform structure also provides a means of explicitly querying for phenotypes involving a biological entity that is similar to some structure or process (e.g., trichomes) or matches some quality (e.g., an increase in physical size). Ontology-based annotations have the potential to increase the information attached to a phenotype (through inferring ancestral terms which are not specifically referred to in the phenotype description), but do not necessarily fully capture the detail and semantics of the natural language description.

For this reason, future representations of phenotypes in relational databases for the purpose of generating phenotype similarity networks across a large volume of phenotypes described in literature and in databases likely should include both ontology-based annotations describing the phenotypes, as well as the original natural language descriptions. Although the number of phenotypes in the dataset used here and described in Oellrich, Walls et al. (2015) is relatively small, the results of this work suggest utility of original text representations as a powerful means of calculating similarity between phenotypes, especially within a single species. Computationally generated EQ statements, which in the context of this study do not often meet the criteria for a fully logical curated EQ statement, were demonstrated to be more useful in any other approach for recovering biologically related genes across species.

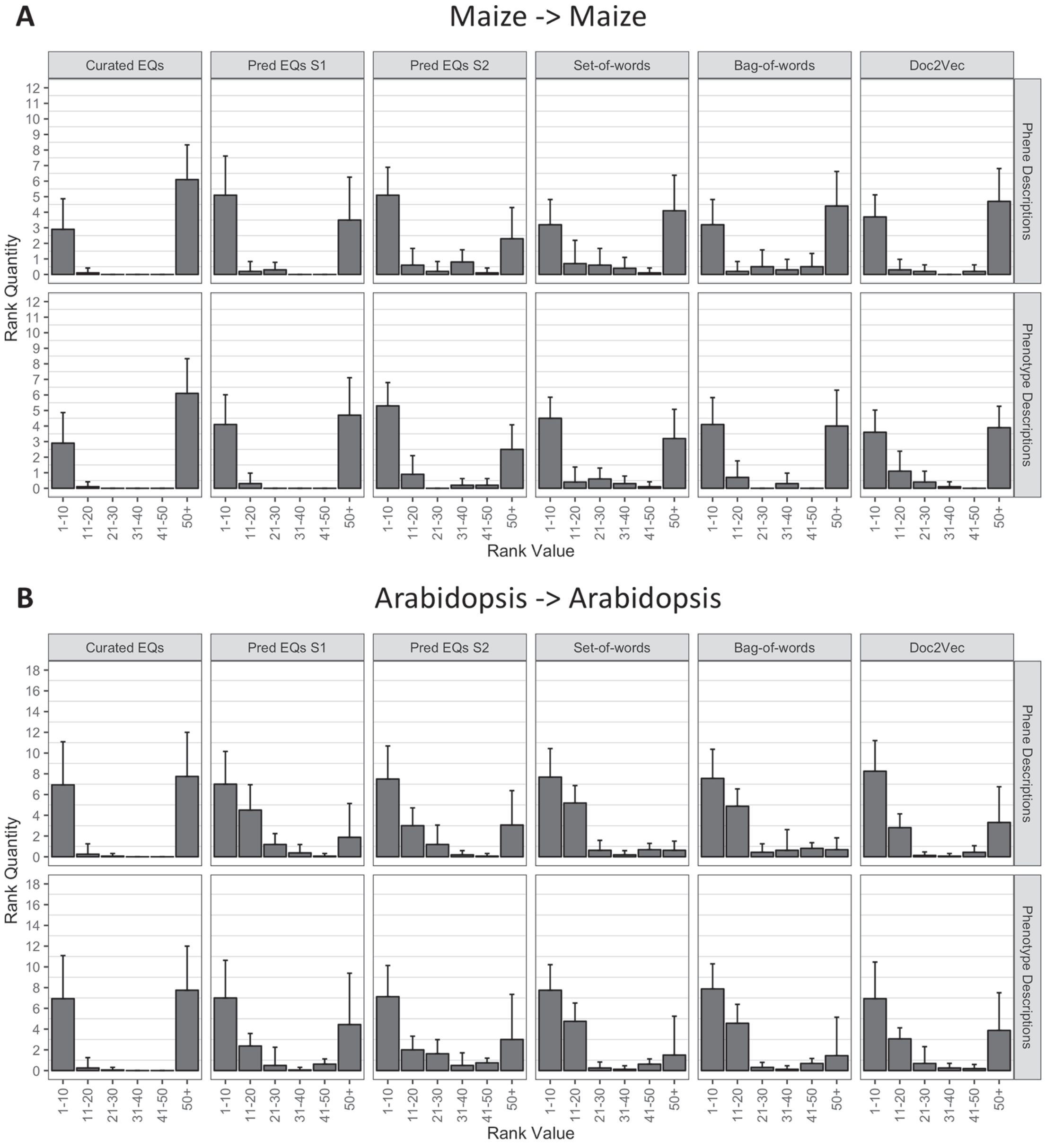


FIGURE 5 | Rankings of anthocyanin biosynthesis genes in either maize **(A)** or *Arabidopsis* **(B)** upon querying phenotype similarity networks generated with genes from the same species. Phenotype networks are organized by the method used to generate them (*columns*) and by whether those methods were applied to phenotype or phene descriptions (*rows*). Rank value specifies a range of rankings for each bar in the plots (1–10, 11–20, etc.) and rank quantity indicates the average number of anthocyanin biosynthesis genes that were ranked in a given range over all queries. *Error bars* indicate one standard deviation of the rank quantities in each range over all queries.

Ensemble methods are often applied in the field of machine learning, where multiple methods are used to solve a problem, with a higher-level model determining which method will be most useful in solving each new instance of the problem. It is possible that such an approach could be applied to measuring similarity between phenotypes to generate a single large-scale network, where similarity values are based on the best possible method to assess the text representations of each pair of particular phenotypes.

Additional Challenges With EQ Statement Representation

Although ontology terms and EQ statements composed of ontology terms are an information-rich representation of phenes and phenotypes, flexibility in which terms and statements can represent a particular phenotype can limit the ability to computationally recognize true biological similarity. The graph structures of the ontologies themselves, the metrics used to assess semantic similarity, and the ambiguity inherent

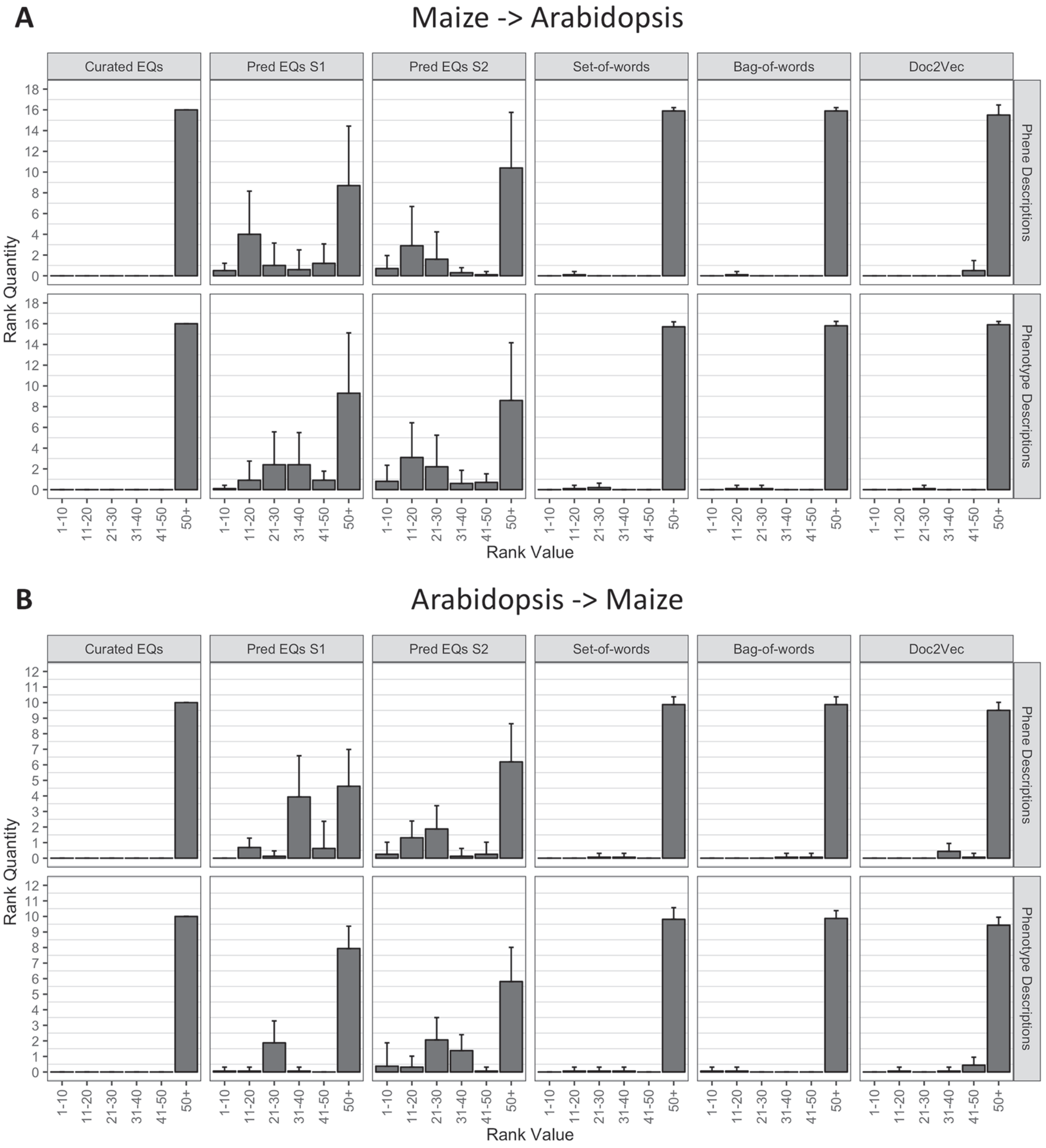


FIGURE 6 | Rankings of anthocyanin biosynthesis genes in either maize **(A)** or *Arabidopsis* **(B)** upon querying phenotype similarity networks generated with genes from the other species. Phenotype networks are organized by the method used to generate them (*columns*) and by whether those methods were applied to phenotype or phene descriptions (*rows*). Rank value specifies a range of rankings for each bar in the plots (1–10, 11–20, etc.) and rank quantity indicates the average number of anthocyanin biosynthesis genes that were ranked in a given range over all queries. *Error bars* indicate one standard deviation of the rank quantities in each range over all queries.

in both natural language and EQ statement representations of phenes and phenotypes can all potentially contribute to this problem.

As one example in the Oellrich, Walls et al. (2015) dataset used here, the phene description "complete loss of flower formation" was annotated with an EQ statement whose entity is *flower development*, whereas the computationally identified entity using the methods described in this work was *flower formation*. In this instance, the Jaccard similarity between these two ontology terms was 0.286, which by comparison is less than the Jaccard similarity between *flower formation* and *leaf formation* in the context of the ontology graph. This selected example illustrates the possible discrepancies between true biological similarity and semantic similarity as measured using graph-based metrics. Although each semantic similarity metric calculates this value differently, those that use the hierarchical nature of the ontology are all constrained by the structure of the graph itself.

Variation in how humans and computational methods interpret how a phenotype as a whole should be conceptualized also has the potential to produce representations that obscure true similarity, as measured by graph-based metrics. In another example from the Oellrich, Walls et al. (2015) dataset, the phene description "stamens transformed to pistils" was annotated with two different EQ statements. The first EQ statement uses the relational quality *has fewer parts of type* to indicate the absence of stamen in this phenotype, and the second uses the relational quality *has extra parts of type* to indicate the presence of pistils in this phenotype. This representation of the phenotype makes logical sense, but is not easy to generate computationally because it abstractly describes the outcome of the transformation that is explicitly present in the natural language description and is dissimilar from computationally generated representations that focus on the explicit content (i.e., those which use the relational quality *transformed to*).

Finally, this study looked at a dataset consisting entirely of phenotypic descriptions in English, and the generalizability of these methods to other languages is not discussed. It is certainly likely that structural differences between languages would result in differences in how certain methods of computing over descriptions in those languages perform, but such analysis is outside the scope of this work.

Extending This Work to the Wealth of Text Data Available in Databases and the Literature

We plan to apply the methods of semantic annotation, ontology-based semantic similarity calculation, and natural language-based semantic similarity calculation to the wealth of text data available in existing plant model organism databases and biological literature. For the latter, doing so will involve the additional challenge of extracting phenotype descriptions as well as the genes causative to those phenotypes as a separate identification and processing step. We plan to leverage existing work in the areas of named entity recognition specific to genes (Wei et al., 2015) and relation extraction, as well as existing methods for extracting information related to phenotypes such as those developed using vector-based representations of phenotype descriptions (Xing et al., 2018) and grammar-tree representations of phenotype descriptions (Collier et al., 2015). As the size of the applicable dataset is increased by these means, we will continue to analyze the performance of methods from the domains of machine learning and NLP towards constructing biologically meaningful networks from this phenotypic data, including additional techniques that were not included in the results presented here. For example, Sent2Vec (Pagliardini et al., 2018) is another technique for assessing text similarity that takes a different approach from Doc2Vec for embedding text as numerical vectors and has been shown to perform well when trained on life science corpora (Chen et al., 2018). These next steps are anticipated to enable researchers to begin to compute on phenotype descriptions directly and will drive a promising future for forward genetics research approaches where phenotypes can be used for novel candidate gene prediction as easily as sequence similarity searches can be used to identify putative homologs from sequence data.

DATA AVAILABILITY STATEMENT

The dataset of phenotype and phene descriptions and the corresponding hand-curated EQ statements used in this work are available as supplemental data of Oellrich, Walls et al. (2015). The hierarchical functional categorization of the set of *Arabidopsis* genes used in this work is available as supplemental data of Lloyd and Meinke (2012). The code used to produce the results of this work is available at github.com/irbraun/phenologs. Files necessary to reproduce the discussed results, datasets used to generate figures presented in this work, and other supplemental files are available at doi.org/10.5281/zenodo.3255020. This data repository also includes versions of the previously described datasets available as supplemental data of Oellrich, Walls et al. (2015) and Lloyd and Meinke (2012), for the purpose of making this study reproducible without any additional external files.

FUNDING

The authors were supported to carry out this work by an Iowa State University Presidential Interdisciplinary Research Seed Grant, the Iowa State University Plant Sciences Institute Faculty Scholars Program, and the Predictive Plant Phenomics NSF Research Traineeship (#DGE-1545453).

AUTHORS CONTRIBUTIONS

IB and CL-D together contributed to the conception and design of the study. IB organized the data, performed the analyses, and wrote the manuscript. IB and CL-D contributed to manuscript revision and read and approved the final version.

ACKNOWLEDGMENTS

This manuscript has been released as a preprint at doi.org/10.1101/689976. We thank Lisa Harper, Sowmya Vajjala, and Ramona Walls for helpful discussions and suggestions. We are grateful to the NSF Phenotype Ontology RCN (#DBI-0956049) for creating foundations for this work by bringing plant and computational biologists together to develop a common vocabulary and for their support to the Plant Phenotype Pilot Project participants who developed the Oellrich, Walls et al. (2015) datasets that our analyses relied upon. We thank the reviewers for their valuable guidance. Based upon their suggestions, the manuscript was improved significantly.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2019.01629/full#supplementary-material

REFERENCES

Altenhoff, A. M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D. A., DeLuca, T., Forslund, K., et al. (2016). Standardized benchmarking in the quest for orthologs. *Nat. Methods* 13, 425. doi: 10.1038/nmeth3830

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Appelhagen, I., Thiedig, K., Nordholt, N., Schmidt, N., Huep, G., Sagasser, M., et al. (2014). Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta* 240 (5), 955–970. doi: 10.1007/s00425-014-2088-0

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556.Gene. arXiv: 10614036

Bada, M., Eckert, M., Evans, D., Garcia, K., Shipley, K., Sitnikov, D., et al. (2012). Concept annotation in the CRAFT corpus. *BMC Bioinf.* 13, 161.doi: 10.1186/1471-2105-13-161

Braun, I., Balhoff, J. P., Berardini, T. Z., Cooper, L., Gkoutos, G., Harper, L., et al. (2018). Computable' phenotypes enable comparative and predictive phenomics among plant species and across domains of life. *Appl. Semant. Technol. Biodivers. Sci.* 187–206.

Chen, Q., Yifan, P., and Zhiyong, L. (2018). *BioSentVec: creating sentence embeddings for biomedical texts.* arXiv: 1810.09302 [cs.CL].

Collier, N., Groza, T., Smedley, D., Robinson, P. N., Oellrich, A., and Rebholz-Schuhmann, D. (2015). PhenoMiner: from text to a database of phenotypes associated with OMIM diseases. *Database* 2015 (1), 1–12. doi: 10.1093/database/bav104

Cooper, L., Walls, R. L., Elser, J., Gandolfo, M. A., Stevenson, D. W., Smith, B., et al. (2013). The plant ontology as a tool for comparative plant anatomy and genomic analyses. *Plant Cell Physiol.* 54 (2), 1–23. doi: 10.1093/pcp/pcs163

Cui, H. (2012). CharaParser for fine-grained semantic annotation of organism morphological descriptions. *J. Am. Soc. Inf. Sci. Technol.* 63 (4), 738–754. doi: 10.1002/asi. arXiv: 08031716

Cui, H., Dahdul, W., Dececchi, A. T., Ibrahim, N., Mabee, P., Balhoff, J. P., et al. (2015). CharaParser+EQ: Performance evaluation without gold standard". *Proc. Assoc. Inf. Sci. Technol.* 52 (1), 1–10. doi: 10.1002/pra2.2015.145052010020

Dahdul, W., Manda, P., Cui, H., Balhoff, J. P., Dececchi, T. A., Ibrahim, N., et al. (2018). Annotation of phenotypes using ontologies: a gold standard for the training and evaluation of natural language processing systems. *Database (Oxford).* 2018, 1–17. doi: 10.1093/database/bay110

Fahlgren, N., Malia, A. G., and Ivan, B. (2015). Lights, camera, action: high-throughput plant 32 phenotyping is ready for a close-up. *Curr. Opin. Plant Biol.* 24, 93–99. doi: 10.1016/j.pbi.2015.02.006

Gehan, M. A., Fahlgren, N., Abbasi, A., Berry, J. C., Callen, S. T., Chavez, L., et al. (2017). PlantCV v2: Image analysis software for high-throughput plant phenotyping. *PeerJ.* 5, e4088. doi: 10.7717/peerj.4088

Gkoutos, G. V., Green, E. C. J., Mallon, A.-m., Hancock, J. M., and Davidson, D. (2005). Using ontologies to describe mouse phenotypes. *Genome Biol.* 6 (1), R8. doi: 10.1186/gb-2004-6-1-r8. R8

Green, J. M., Appel, H., Rehrig, E. M., Harnsomburana, J., Chang, J.-F., Balint-Kurti, P., et al. (2012). PhenoPhyte: a flexible affordable method to quantify 2D phenotypes from imagery. *Plant Methods* 8, 45. doi: 10.1186/1746-4811-8-45

Hailu, N. D., Bada, M., Hadgu, A. T., and Hunter, L. E (2019). Biomedical concept recognition using deep neural sequence models. *bioRxiv*. 1–21. doi: 10.1101/530337

Hastings, J., De Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., et al. (2013). The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Res.* 41 (D1), 456–463. doi: 10.1093/nar/gks1146

Hoehndorf, R., Paul, N. S., and Georgios, V. G. (2011). PhenomeNET: a whole-phenome approach to disease gene discovery. *Nucleic Acids Res.* 39 (18), e119–e119. doi: 10.1093/nar/gkr538

Homma, N., Takei, Y., Tanaka, Y., Nakata, T., Terada, S., Kikkawa, M., et al. (2003). Kinesin superfamily protein 2A (KIF2A) functions in suppression of collateral branch extension. *Cell* 114, 229–239. doi: 10.1016/S0092-8674(03)00522-1

Lau, J. H., and Baldwin., T. (2016). *An Empirical Evaluation of doc2vec with Practical Insights into Document Embedding Generation.* arXiv: 1607.05368 [cs.CL].

Le, Q. V., and Mikolov., T. (2014). *Distributed Representations of Sentences and Documents.* arXiv: 1405.4053 [cs.CL].

Li, T., Zhang, W., Yang, H., Dong, Q., Ren, J., Fan, H., et al. (2019). Comparative transcriptome analysis reveals differentially expressed genes related to the tissue-specific accumulation of anthocyanins in pericarp and aleurone layer for maize. *Sci. Rep.* 9, 1–12. doi: 10.1038/s41598-018-37697-y

Lloyd, J., and Meinke, D. (2012). A comprehensive dataset of genes with a loss-of-function mutant phenotype in *arabidopsis*. *Plant Physiol.* 158 (3), 1115–1129. doi: 10.1104/pp.111.192393

Lu, L., Lee, Y.-R. J., Pan, R., Maloof, J. N., and Liu, B. (2004). An internal motor kinesin is associated with the golgi apparatus and plays a role in trichome morphogenesis in *arabidopsis*. *Mol. Biol. Cell* 16, 811–823. doi: 10.1091/mbc.e04-05-0400

McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., and Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc. Natl. Acad. Sci.* 107 (14), 6544–6549. doi: 10.1073/pnas.0910200107. arXiv: 1408.1149

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space.* arXiv: 1301.3781 [cs.CL].

Miller, N. D., Haase, N. J., Lee, J., Kaeppler, S. M., de Leon, N., and Spalding, E. P. (2017). A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. *Plant J.* 89, 169–178.doi: 10.1111/tpj.13320

Mungall, C. J., Gkoutos, G. V., Smith, C. L., Haendel, M. A., Lewis, S. E., and Ashburner, M. (2010). Integrating phenotype ontologies across multiple species. *Genome Biol.* 11 (1), 1–16. doi: 10.1186/gb-2010-11-1-r2

Musen, M. A., Noy, N. F., Shah, N. H., Whetzel, P. L., Chute, C. G., Story, M. A., et al. (2012). The National center for biomedical ontology. *J. Am. Med. Informatics Assoc.* 19, 190–195. doi: 10.1136/amiajnl-2011-000523

Oellrich, A., Walls, R. L., Cannon, E. K., Cannon, S. B., Cooper, L., Gardiner, J., et al. (2015). An ontology approach to comparative phenomics in plants. *Plant Methods* 11 (1), 1–15. doi: 10.1186/s13007-015-0053-y

Pagliardini, M., Prakhar, G., and Martin, J. (2018). Unsupervised learning of sentence embeddings using compositional n-Gram features, In: Proceedings of the 2018 Conference of the North American Chapter 77 of the Association for Computational Linguistics: Human Language Technologies Volume 1 (Long Papers) (Association for Computational Linguistics), 528–540. doi: 10.18653/v1/n18-1049

Resnik, P. (1999). Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. *J. Artif. Intell. Res.* 11, 95–130.doi: 10.1613/jair.514. arXiv: 1105.5444

Schlap̈fer, P., Zhang, P., Wang, C., Kim, T., Banf, M., Chae, L., et al. (2017). Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* 173, 2041–2059. doi: 10.1104/pp.16.01942.15

Sharma, M., Cortes-Cruz, M., Ahern, K. R., McMullen, M., Brutnell, T. P., and Chopra, S. (2011). Identification of the pr1 gene product completes the anthocyanin biosynthesis pathway of maize. *Genetics* 188, 69–79. doi: 10.1534/genetics.110.126136

Slimani, T. (2013). Description and evaluation of semantic similarity measures approaches. *Int. J. Comput. Appl.* 80, 25–33. doi: 10.5120/13897-1851

Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6, e21800. doi: 10.1371/journal.pone.0021800

Thessen, A. E., Hong, C., and Dmitry, M. (2012). Applications of natural language processing in biodiversity science. *Adv. Bioinf.* 2012, 1–17. doi: 10.1155/2012/391574

Tseytlin, E., Mitchell, K., Legowski, E., Corrigan, J., Chavan, G., and Jacobson, R. S. (2016). NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. *BMC Bioinf.* 17 (1), 32. doi: 10.1186/s12859-015-0871-y

Wei, C.-H., Hung-Yu, K., and Zhiyong, L. (2015). GNormPlus: an integrative approach for tagging genes, gene families, and protein domains. *Biomed. Res. Int.* 2015, 1–7. doi: 10.1155/2015/918710

Whetzel, P. L., Noy, N. F., Shah, N. H., Alexander, P. R., Nyulas, C., Tudorache, T., et al. (2011). BioPortal: enhanced functionality *via* new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Res*. 39, W541–W545. doi: 10.1093/nar/gkr469

Woods, J. O., Singh-Blom, U. M., Laurent, J. M., McGary, K. L., and Marcotte, E. M. (2013). Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. *BMC Bioinf.* 14 (1), 203.

Xing, W., Qi, J., Yuan, X., Li, L., Zhang, X., Fu, Y., et al. (2018). A gene-phenotype relationship extraction pipeline from the biomedical literature using a representation learning approach. *Bioinformatics* 34 (13), i386–i394. doi: 10.1093/bioinformatics/bty263

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ORIGINAL RESEARCH
published: 26 February 2020
doi: 10.3389/fpls.2020.00213

A Computation Method Based on the Combination of Chlorophyll Fluorescence Parameters to Improve the Discrimination of Visually Similar Phenotypes Induced by Bacterial Virulence Factors

Valérian Méline [1,2†], **Chrystelle Brin** [1†], **Guillaume Lebreton** [3], **Lydie Ledroit** [3], **Daniel Sochard** [3], **Gilles Hunault** [2,4], **Tristan Boureau** [1,2,3†] **and Etienne Belin** [2,3,5*†]

[1] Emersys, SFR 4207 QUASAV, IRHS, UMR1345, Université d'Angers, Angers, France, [2] ImHorPhen, SFR 4207 QUASAV, IRHS, UMR1345, Université d'Angers, Angers, France, [3] Phenotic Platform, SFR 4207 QUASAV, IRHS, UMR1345, Université d'Angers, Angers, France, [4] Laboratoire HIFIH, UPRES EA 3859, SFR 4208, Université d'Angers, Angers, France, [5] Laboratoire Angevin de Recherche en Ingénierie des Systèmes, Université d'Angers, Angers, France

Phenotyping biotic stresses in plant-pathogen interactions studies is often hindered by phenotypes that can hardly be discriminated by visual assessment. Particularly, single gene mutants in virulence factors could lack visible phenotypes. Chlorophyll fluorescence (CF) imaging is a valuable tool to monitor plant-pathogen interactions. However, while numerous CF parameters can be measured, studies on plant-pathogen interactions often focus on a restricted number of parameters. It could result in limited abilities to discriminate visually similar phenotypes. In this study, we assess the ability of the combination of multiple CF parameters to improve the discrimination of such phenotypes. Such an approach could be of interest for screening and discriminating the impact of bacterial virulence factors without prior knowledge. A computation method was developed, based on the combination of multiple CF parameters, without any parameter selection. It involves histogram Bhattacharyya distance calculations and hierarchical clustering, with a normalization approach to take into account the inter-leaves and intra-phenotypes heterogeneities. To assess the efficiency of the method, two datasets were analyzed the same way. The first dataset featured single gene mutants of a *Xanthomonas* strain which differed only by their abilities to secrete bacterial virulence proteins. This dataset displayed expected phenotypes at 6 days post-inoculation and was used as ground truth dataset to setup the method. The efficiency of the computation method was demonstrated by the relevant discrimination of phenotypes at 3 days post-inoculation. A second dataset was composed of transient expression (agrotransformation) of Type 3 Effectors. This second dataset displayed phenotypes that cannot be discriminated by visual assessment and no prior knowledge can be made on the respective impact of each Type 3 Effectors on leaf tissues. Using the computation method resulted in clustering the leaf samples according to the Type 3 Effectors, thereby demonstrating

Edited by:
Carmen R. Beuzón,
University of Málaga, Spain

Reviewed by:
Andrés Gárriz,
CONICET Institute of Biotechnological Research, National University of General San Martín (IIB-INTECH), Argentina
Simona Masiero,
University of Milan, Italy

***Correspondence:**
Etienne Belin
etienne.belin@univ-angers.fr

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: 21 June 2019
Accepted: 11 February 2020
Published: 26 February 2020

Citation:
Méline V, Brin C, Lebreton G, Ledroit L, Sochard D, Hunault G, Boureau T and Belin E (2020) A Computation Method Based on the Combination of Chlorophyll Fluorescence Parameters to Improve the Discrimination of Visually Similar Phenotypes Induced by Bacterial Virulence Factors. Front. Plant Sci. 11:213. doi: 10.3389/fpls.2020.00213

an improvement of the discrimination of the visually similar phenotypes. The relevant discrimination of visually similar phenotypes induced by bacterial strains differing only by one virulence factor illustrated the importance of using a combination of CF parameters to monitor plant-pathogen interactions. It opens a perspective for the identification of specific signatures of biotic stresses.

Keywords: imaging analysis, chlorophyll fluorescence parameters, Bhattacharyya distance, hierarchical clustering, biotic stress

INTRODUCTION

In recent years, plantx phenotyping has been significantly evolving. High-throughput plant phenotyping platforms have been developed to answer to the rapid improvement of plant genomic technologies. Time consuming expert-based approaches of traditional phenotyping is moving toward a technology-based approaches providing automatic and quantitative measurements of biotic or abiotic stresses.

Imaging analysis applied to plant phenotyping is a component of this evolution. Measurements based on automatic image analysis could provide higher throughput, accuracy, and reproducibility than human visual inspections (Bock et al., 2008). Imaging analysis can be applied in various parts of plant phenotyping domain, such as the characterization of plant structure at a given instant, the quantification of plant growth over time or the monitoring of plants interactions with the environment or with pathogens. Plant structure and growth are now accessible with various 3D imaging techniques (Fang et al., 2009; Jahnke et al., 2009; Dhondt et al., 2010; Alenya et al., 2011; Zhu et al., 2011; Bellasio et al., 2012; Paproki et al., 2012), while imaging of plant health is accessible with various functional imaging techniques (see Li et al., 2014; Mahlein, 2016 for recent reviews). Thermal, near infrared reflectance, hyperspectral reflectance and chlorophyll fluorescence imaging (CF imaging) are among the most popular imaging techniques for monitoring plant health.

CF imaging is of special interest as it can be considered as a non-invasive and non-destructive method to efficiently phenotype the impact of biotic (Baron et al., 2012; Rousseau et al., 2013; De Torres Zabala et al., 2015; Zhou et al., 2015; Montero et al., 2016; Perez-Bueno et al., 2016; Pineda et al., 2018) or abiotic stresses (Honsdorf et al., 2014; Mishra et al., 2014; Bresson et al., 2015; Sebela et al., 2018) on the photosystem II of plants. Based on an active imaging technique with illumination flashes, sequences of many images are acquired and exploited to then evaluate CF parameters. These CF parameters are studied both for basic research on photosynthetic processes (Genty et al., 1989, 1990; Lichtenthaler et al., 2005; Baker, 2008), or alternatively for applied purposes, such as screening for phenotypes of resistance to abiotic and biotic stresses. Contrary to the study of abiotic stresses, only few CF parameters have been exploited when studying biotic stresses (Gorbe and Calatayud, 2012). Among all CF parameters available, F_v/F_m and NPQ are commonly used for studying biotic stress (Baron et al., 2016; Kalaji et al., 2017). These parameters could give an efficient pre-symptomatic measure of the impact of several pathogen (Csefalvay et al., 2009; Pineda et al., 2011). However, contrasts obtained may differ among the numerous CF parameters used. Therefore, the use of only a subset of CF parameters may limit the ability to discriminate visually similar phenotypes (Berger et al., 2007; Pineda et al., 2008).

Bacteria belonging to the genus *Xanthomonas* are associated to plants, and numerous strains are responsible for disease on many important crops, such as rice, bean, soybean, tomato, sugarcane, wheat, oilseed rape, as well as on model plants, such as *Arabidopsis thaliana*. Even though more than 400 plant species may be infected by strains belonging to the genus *Xanthomonas*, a single strain only displays a narrow host range, restricted to one or several plant species. Even though most strains of *Xanthomonas* spp isolated from plants were described as pathogenic, non-pathogenic strains of *Xanthomonas* spp have been also been isolated and described, receiving an increasing attention in the recent years (Cesbron et al., 2015; Essakhi et al., 2015; Garita-Cambronero et al., 2016; Merda et al., 2016, 2017).

Among the multiple virulence factors that have been described for *Xanthomonas* spp strains, the type 3 secretion system (T3SS) encoded by the *hrp* gene cluster, is known to play a central role in the interactions with plants. Indeed, the inactivation of the T3SS usually abolishes the virulence of pathogenic *Xanthomonas* spp strains on their host plants. On the other hand, the acquisition of a T3SS by non-pathogenic bacteria may constitute an evolutionary step toward the emergence of novel plant pathogenic bacterial strains (Manulis and Barash, 2003). For example, strain 7698R of *Xanthomonas cannabis* is non-pathogenic on bean, its host of isolation. When inoculated on the non-host plant *Nicotiana benthamiana*, strain 7698R induces a rapid necrosis similar to the hypersensitive reaction (HR) observed in non-host resistance. The complementation of this non-pathogenic strain with a plasmid harboring genes encoding a T3SS, suppressed the onset of this HR-like necrosis, and only a mild chlorosis could be observed on inoculated tissues at 6 days post-inoculation (6 dpi) (Meline et al., 2019). Such a suppression of a defense layer could constitute a first step toward the emergence of novel pathogenic strains.

This secretion system enables the injection of numerous bacterial effector proteins called Type 3 effectors (T3Es) directly into the plant cells, that collectively suppress the plant defenses and subvert the plant's physiology to the benefit of the bacteria (Li et al., 2002; Büttner, 2016; Jacques et al., 2016; Wei and Collmer, 2018). Repertoires of T3Es vary importantly among strains of *Xanthomonas* belonging to a same species, and were reported to correlate to some extend with the host specificity of strains (Hajri et al., 2009, 2012). Knowledge on the functions and cellular targets in plant cells of some single T3Es has considerably increased over the last decade. Numerous targets in various compartments of the plant cell were described (for review Büttner, 2016). For example, once injected inside the plant cells, the T3E *XopAC* of *Xanthomonas campestris* uridylylates the cytoplasmic kinase BIK1, which blocks the transduction pathways leading to plant immunity and promote bacterial virulence on *Arabidopsis thaliana* (Feng et al., 2012; Guy et al., 2013a,b). Another example is that of TAL effectors of *Xanthomonas* spp, enter the plant nucleus, recognize and bind specific DNA sequences and induce the expression of plant genes (Boch et al., 2014). Plant immunity and more globally the physiology of plant cells may also be altered by T3Es targeting other cell organelles, such as cytoskeleton and stromules, or chloroplasts (Büttner, 2016; Erickson et al., 2018).

However, in most cases, functions of T3Es and their targets in plant cells still remain elusive. Importantly, mutants in single T3E genes often hardly display any phenotype visible to the eye, which may hinder deciphering their role in the interaction between plants and bacteria (Mutka and Bart, 2015). Therefore, innovative tools are needed to better phenotype the impact of single T3Es on plant tissues.

In the present paper, we propose an approach based on CF imaging aiming at the discrimination of the behavior of such mutants on leaves. To this purpose, we developed a computation method based on the combination of multiple images of CF parameters. This approach aims at using the information contained in all the images of CF parameters, without prior selection nor focusing on the physiological processes involved. As such, the approach developed in the present study is mainly intended to setup screening methods for phenotyping closely related biotic stresses on plants.

MATERIALS AND METHODS

Xanthomonas Mutant Strains

Mutant strains of *Xanthomonas* were already described in Meline et al. (2019). Strain 7698R is an environmental strain belonging to the species *X. cannabis* isolated from bean seeds, but not pathogenic on bean. As largely documented in Meline et al. (2019), when inoculated on *N. benthamiana*, this strain induces a necrosis of leaf tissues at 6 dpi. Strain 7698R was complemented with the plasmid pIJ3225 that carries a 20 kb *hrp* cluster encoding a major bacterial virulence determinant: the Type 3 Secretion System and four Type 3 Secreted Proteins (Arlat et al., 1991). On *N. benthamiana* at 6 dpi, the complemented strain 7698R pIJ3225 does not induce any necrosis, the inoculated tissues only display a chlorotic phenotype. Derivatives of pIJ3225 harboring a Tn5 insertion that inactivates one gene in the *hrp* cluster were characterized in Meline et al. (2019). These derivatives were transformed into strain 7698R to obtain strains 7698R pIJ3225::Tn5 (G9, G2, F2, F15, C3, or G1). After inoculation on *N. benthamiana* leaves, Tn5 insertions G9 and G2 restore the onset of the necrosis, the Tn5 insertions F2, F15, and C3 partially restore the necrosis, whereas Tn5 insertion G1 does not alter the phenotype conferred by pIJ3225. Strain 7698R and complemented strains 7698R pIJ3225::Tn5 were used in the present paper for inoculation on *N. benthamiana* and subsequent CF imaging at 3 dpi. The complete list of Tn5 derivatives used is reported in **Supplementary Table 1**.

Cloning of Bacterial Virulence Factors

Six T3E genes (*xopAF*, *xopL*, *xopG*, *xopV*, *xopT*, and *xopAK*) of the sequenced strain *Xanthomonas citri* pv. *fuscans* CFBP 4834 were chosen for cloning into Gateway vectors, for subsequent transient expression in *N. benthamiana*. These T3E genes were chosen as previous knowledge suggests they are involved in various biological processes (Darrasse et al., 2013). Hence they constituted a good set of candidates for setup of an approach aiming at discriminating the impact of various virulence factors on plants. The *xopG* and *xopT* genes were chosen because they often are found in highly aggressive strains of *X. citri* on bean. In the genome of the model strain *X. citri* pv. *fuscans* CFBP 4834, these genes are flanked by insertion sequences suggesting that they may be horizontally transferred to other bacteria (Rousseau et al., unpublished data). The *xopV* and *xopAK* genes were chosen as they were proposed by Moreira et al. (2010) to be involved in the specificity of symptoms caused by *X. citri*. The *xopL* gene was described to impact stromule formation in plant cells, and is widely distributed among sequenced model strains of *Xanthomonas*. Conversely, the *xopAF* gene is poorly distributed among the sequenced model strains of *Xanthomonas*, and its distribution suggests this gene may be involved in tissue specificity of *Xanthomonas* strains (Bogdanove et al., 2011). Sequences of these T3Es were amplified by PCR using adequate primers summarized in **Supplementary Table 2**, and *AccuPrime*^{TM} Taq DNA Polymerase High Fidelity. Amplified sequences were cloned in pENTR/D-TOPO vector using pENTR Directional TOPO cloning kit (Invitrogen). Then, cloned sequences were transferred to the pB7WG2 binary vector, obtained from VIB (Belgium) (Karimi et al., 2002) using Gateway LR clonase II (Invitrogen). The nucleotide sequences of all constructs were confirmed by sequencing.

Agrobacterium tumefaciens-Mediated Transient Expressions

Strain EHA105 of *A. tumefaciens* was transformed with the binary vector pB7WG2 containing the bacterial genes of interest fused to *gfp*. The subsequent derivatives of strain EHA105 were used to prepare an inoculum (see subsection Preparation of Inocula and Inoculation procedures) infiltrated into fully expanded leaves to perform transient expressions of the cloned genes in *N. benthamiana*. The transient expression of six T3Es and β-glucuronidase (GUS) were performed. To control the efficiency of the transient expression process, we revealed the GUS activity of samples of leaf tissues inoculated with the EHA105 pB7FWG2-gus at 24 h post-inoculation (hpi), 36 and 48 hpi (data not shown). The GUS activity of samples was revealed using a buffer containing X-Gluc (1 mM), $K_3Fe(CN)_6$ (4 mM), $K_4[Fe(CN)_6]\cdot 3H_2O$ (0.05 mM), EDTA (10 mM), Na_2HPO_4 (50 mM) and NaH_2PO_4 (50 mM) in a buffer phosphate. The efficiency of the transient expression of each T3E gene was controlled by observation at 24 and 48 hpi of a GFP signal resulting from the expression of the *gfp* fusion obtained by the cloning the T3E gene of interest into pB7FWG2.

Plant Material

Nicotiana benthamiana plants were grown and inoculated in environmentally controlled growth room under a 16 h photoperiod and 8 h dark period at 22°C and 80% of relative humidity under a light intensity of 100 μE$\cdot$m$^{-2}\cdot$s^{-1} throughout the whole experiment. At the optimal developmental stage (6 weeks old) the plants had at least five fully developed true leaves.

Preparation of Inocula and Inoculation

Xanthomonas mutant strains were cultured on classical media TSA (Trypticase Soy Agar: Tryptone 17 g$\cdot$L^{-1}; peptone soja 3 g$\cdot$L^{-1}; glucose 2.5 g$\cdot$L^{-1}; NaCl 5 g$\cdot$L^{-1}; KH_2PO_4 5 g$\cdot$L^{-1}; agar 15 g$\cdot$L^{-1}; pH 7.2, supplemented with the adequate antibiotics) and incubated at 28°C. For the inoculation, bacterial suspensions were calibrated at 10^8 cfu.ml^{-1}.

A. tumefaciens was cultured into 20 mL of Luria Bertani medium (Tryptone 10 g$\cdot$L^{-1}; NaCl 10 g$\cdot$L^{-1}; Yeast extract 5 g$\cdot$L^{-1}) in the presence of selective antibiotics (50 mg L^{-1} spectinomycin and 50 mg$\cdot$L^{-1} gentamycin), and grown overnight in a rotary shaker at 150 rpm and 28°C. Cells were harvested by centrifugation and resuspended to a final concentration of 0.3 OD600 in a solution containing 10 mM $MgCl_2$, 10 mM MES, 10 g$\cdot$L^{-1} sucrose, pH 5.6 and 150 μM acetosyringone and incubated at 28°C for 3 h before agroinfiltration.

For *Xanthomonas* or *Agrobacterium* inoculation, three fully expanded leaves per plant were inoculated by pressing the blunt end of a 1 ml needleless syringe to the lower side of the leaf while supporting the leaf with a gloved finger.

Datasets

In a validation approach, two datasets were used and processed the same way to setup and assess the computation method.

The first dataset was used as ground truth dataset to validate the computation method. It featured 85 leaves of *N. benthamiana* at 3 dpi. Each leaf was inoculated with three controls and one tested-strain as described in Meline et al. (2019). The three controls were, (i) strain 7698R used as necrosis control, (ii) strain 7698R pIJ3225 used as chlorosis control, and (iii) mock-inoculated control used as no symptom control. The tested-strains were strains 7698R pIJ3225::Tn5 (G9, G2, F2, F15, C3, or G1). Each inoculated area was cropped and considered as an independent sample. Thereby, 340 samples (85 × 4) were generated. For each sample, 70 images of CF parameters were obtained. The total size of this first dataset was of 23800 images. As expected, Tn5 insertions G9 and G2 restored the onset of the necrosis, the Tn5 insertions F2, F15, and C3 partially restored the necrosis, whereas Tn5 insertion G1 led to chlorosis phenotype at 6 dpi. However, at 3 dpi, all the controls and tested-strains displayed phenotypes that could hardly be discriminated by visual assessment.

The second dataset was used to assess the relevance of the computation method to discriminate leaf tissues impacted by biotic stresses without any prior knowledge. It featured 108 leaves of *N. benthamiana* each inoculated with three controls and one tested-strain. The three controls were, (i) *A. tumefaciens* strain EHA105 not expressing any exogenous protein, (ii) transient expression of β-glucuronidase as non-deleterious exogenous protein control, and (iii) mock-inoculated control used as no symptom control. The tested-strains were *A. tumefaciens* EHA105 transiently expressing one of the *X. citri* pv. *fuscans* virulence genes (T3Es): *xopL*: L, *xopT*: T, *xopV*: V, *xopAK*: AK, *xopAF*: AF, *xopG*: G. Each inoculated area was cropped and considered as an independent sample. Thereby, 432 samples (108 × 4) were generated. For each sample, 70 images of CF parameters were obtained. The total size of this second dataset was of 30240 images. This second dataset is of interest as phenotypes cannot be discriminated by visual assessment: no necrosis nor chlorosis can be observed either at 3 or 6 dpi. Therefore, no prior knowledge can be made on the respective impact of the transient expression of each T3Es on leaf tissues.

Chlorophyll Fluorescence Imaging System

Acquisition of fluorescence images is performed with a PSI Open FluorCam FC 800-O. The system sensor is a CCD camera with a pixel resolution of 512 by 512 and a 12-bit dynamic. The system includes 4 LED panels divided in 2 pairs. One pair provides an orange actinic light with a wavelength around 618 nm, with an intensity up to 400 μmol$\cdot$m$^{-2}\cdot$s^{-1}. The other pair provides a saturating pulse in blue wavelength, typically 455 nm, with an intensity up to $3,000$ μmol$\cdot$m$^{-2}\cdot$s^{-1}. The acquisition protocol is a quenching analysis protocol (Kolber et al., 1998), producing a raw file containing 70 images of CF parameters. A schematic description of the quenching protocol is proposed in **Figure S1**. To measure the parameter F_0, a modulated light of 0.1 μmol$\cdot$m$^{-2}\cdot$s^{-1} is used. Then orange actinic light with intensities of 20% of the 400 μmol$\cdot$m$^{-2}\cdot$s^{-1} is used during the light-adapted period of 60 s. The protocol also provides 6 pulses of 0.8 s duration of blue saturating light with an intensity of 50% of the 3,000 μmol$\cdot$m$^{-2}\cdot$s^{-1}: 5 pulses during the light-adapted period and 1 pulse during the dark-relaxation period. The 50% of the saturating light pulse was considered as a good intensity as it provides a ratio $(F_m - F_0)/F_m$ of 0.82 measured on non-inoculated area, and being closed to optimal value of 0.83 (Bjorkman and Demmig, 1987). This was measured after a dark adaptation of 30 min. The whole duration of the illumination protocol is about 95.8 s. For evaluating the performance of our computation method on datasets obtained with such protocol, the 70 images given in gray-level intensity are processed in batch to provide histograms of pixels that statistically represent regions of interest, i.e., the four areas of infection on the leaf. Distances between the histograms of each leaf are then calculated according to Bhattacharyya distance.

Bhattacharyya Distance

The Bhattacharyya distance measures the similarity of two discrete probability distributions or histograms (Bhattacharyya, 1943). This measure, regularly used in classification problems in the field of computer vision (Kailath, 1943), determines the relative closeness of two histograms being considered. The

Bhattacharrya distance is known to be particularly useful to give a contrast scalar directly connected to detection performance in noisy images (Goudail et al., 2004). The Bhattacharyya distance B_d is defined as:

$$B_d = -ln \sum \sqrt{(h_A h_B)}$$

where h_A and h_B are the normalized histograms for two different areas A and B.

3D Euclidean Distance

According to Deza and Deza (2009), the distance between two points (p,q) in a three-dimensional Euclidean space (x, y, z) is defined as

$$E_d = \sqrt{(x_p - x_q)^2 + (y_p - y_q)^2 + (z_p - z_q)^2}$$

In a three dimensional Euclidean space, we also defined intra-modality and inter-modality distances. A modality is composed of several images of the same kind of inoculation. For each modality, the intra-modality distance is the mean Euclidean distance between each image and the centroid of the modality, constituting the dispersion of the modality. The inter-modality distance is the Euclidean distance between the centroid of each modality. Such distances were useful to evaluate contribution of the addition of CF parameters for the clustering ability. The addition of parameters was performed using a sequential forward sequence method (SFS) (Agrawal and Srikant, 1995).

Clustering Based on Dendrogram

We used hierarchical clustering algorithm based on a Ward linkage method (Ward, 1963). This method is the only one among the agglomerative clustering methods that is based on a classical sum of squares criterion, producing groups that minimize within group dispersion at each merging step (Murtagh and Legendre, 2014). In R, the Ward.D2 algorithm of the function hclust is the one attributed to Ward. This function requires Euclidean distances as input dissimilarities. Several studies point out that this method outperforms other hierarchical clustering methods (Blashfield, 1976; Hands and Everitt, 1987). In our case this method allows to cluster modalities with a better accuracy.

To evaluate the clustering abilities of dendrograms, we computed the sensitivity and the specificity of the clustering. According to Parikh et al. (2008), the sensitivity and the specificity can be expressed as:

$$\text{sensitivity} = \frac{\text{true positive}}{\text{true positive + false negative}}$$

$$\text{specificity} = \frac{\text{true negative}}{\text{true negative + false positive}}$$

where, for the sensitivity, a true positive represented an image of one modality (for instance necrosis control) classified in the correct corresponding cluster (necrosis phenotype cluster) and false negative represented an image of the same modality (necrosis control) classified in the other uncorrect clusters (chlorosis and mock-inoculated clusters). For the specificity, true negative represented an image of the two other modalities (chlorosis and mock-inoculated) classified in one of these two corresponding correct cluster (chlorosis phenotype and mock-inoculated phenotype cluster) and false positive represented an images of the two other modalities (chlorosis and mock-inoculated controls) classified in a uncorrect cluster (necrosis phenotype cluster).

RESULTS

Distance Calculation Between Tested-Strain and Controls for Each Image of CF Parameter

In this study, phenotyping of interactions between *N. bethamiana* leaves and *Xanthomonas* strains was confronted to an inter-leaf heterogeneity. As illustrated in **Figure 1**, the intensity of the necrosis development, after inoculation with the same strain, could vary between leaves. To circumvent this inter-leaves heterogeneity, for both dataset, each tested modality was inoculated with their respective controls on each leaf as illustrated in **Figure 1**. In a normalization approach, each tested-strain was then characterized according to its distance with its respective controls on each leaf. This image processing procedure occurring in 4 steps was described for the computation of the first dataset there after.

Firstly, as illustrated in **Figures 2A,B**, from a gray-level intensity image, the normalized histograms of the cropped four areas were extracted (**Figures 2C,D**). Secondly, for each image of CF parameter, Bhattacharyya distances were calculated between the histogram of the tested-strain and each histogram of the three controls. As shown in **Figure 2E**, the three Bhattacharyya distances defined the axes of a three-dimensional plot, where each tested-strain modality can be represented. This representation had the advantage of taking into account the inter-leaves heterogeneity as each tested-strain was compared to its three controls on each leaf. Thirdly, our CF imaging protocol generated 70 images of CF parameters for each tested-strain. Therefore, each tested-strain can be characterized by an array of 70 three-dimensional Euclidean distances. Finally, tested-strains were then clustered according to their array of 70 three-dimensional Euclidean distances by hierarchical ascendant classification based on Ward's agglomeration method and represented with a dendrogram (**Figure 3**).

The Discrimination of the Contrasted Phenotypes of Controls at 3 Days Post-inoculation Validated the Computation Method

To validate the computation method, the relevance of the 70 CF images combination method was evaluated through the discrimination efficiency of the three controls which induced contrasted phenotypes at 3 dpi. The first dataset was computed according to the method described in the previous paragraph. On the dendrogram of **Figure 3**, the

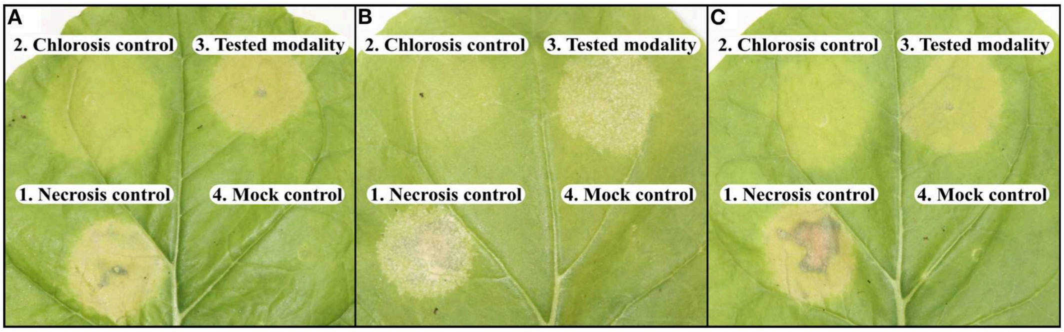


FIGURE 1 | Visual observation of the inter-leaf heterogeneity at 3 dpi. **(A–C)** Were representative of the diversity of results obtained with the inoculations of the same strains on three different leaves. 1. necrosis: strain 7698R. 2. chlorosis: strain 7698R pIJ3225. 3. tested-strain: strain G9. 4. mock-inoculated.

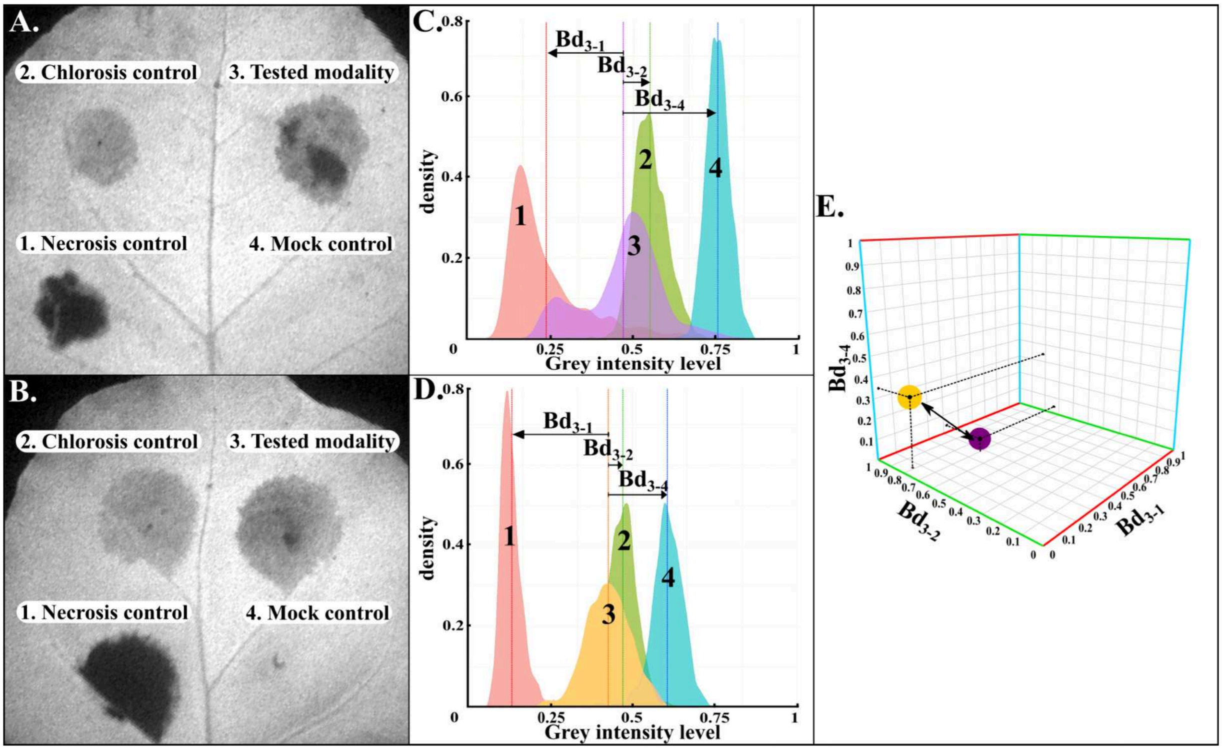


FIGURE 2 | Method for the calculation of the distance between two different tested strains. **(A,B)** Illustrations of gray-level intensity images for two different leaves corresponding to F_m parameter. Gray-level intensity was coded between 0 and 1. The four areas corresponded to necrosis, chlorosis, tested strain, and mock-inoculated, respectively. **(C,D)** Extracted normalized histograms from areas 1 to 4. From these histograms, three Bhattacharyya distances (Bd_{3-1}, Bd_{3-2}, Bd_{3-4}) were computed between tested strains and their three respective controls. Plant tissues could also display a heterogeneity intra the inoculated area. Bhattacharyya distances allowed to take into account each pixel value to consider this heterogeneity across each inoculated area. These three distances defined axes for a three-dimensional space where tested strains could be represented and compared to each others using Euclidean distance calculation. **(E)** Illustrated the three-dimensional representation and the three-dimensional Euclidean distance calculation between two tested strains (yellow and violet spots) for one image.

three controls were clearly grouped in three distinct clusters referred to as necrosis phenotype, chlorosis phenotype and no symptom phenotype clusters. A first level of analysis revealed an efficient clustering of mock-inoculated control. Practically, all samples (except two samples) of this control were grouped into a single cluster (No symptom phenotype cluster). All the other clusters of the dendrogram are clearly discriminated from this latter cluster, which highlighted the impact of inoculation. At a second level of analysis, all samples of necrosis control and all samples of chlorosis control were respectively grouped into distinct clusters (Necrosis phenotype cluster and Chlorosis phenotype cluster, respectively). Results obtained using this method were in complete accordance with the observed phenotypes at 3 dpi, and with the

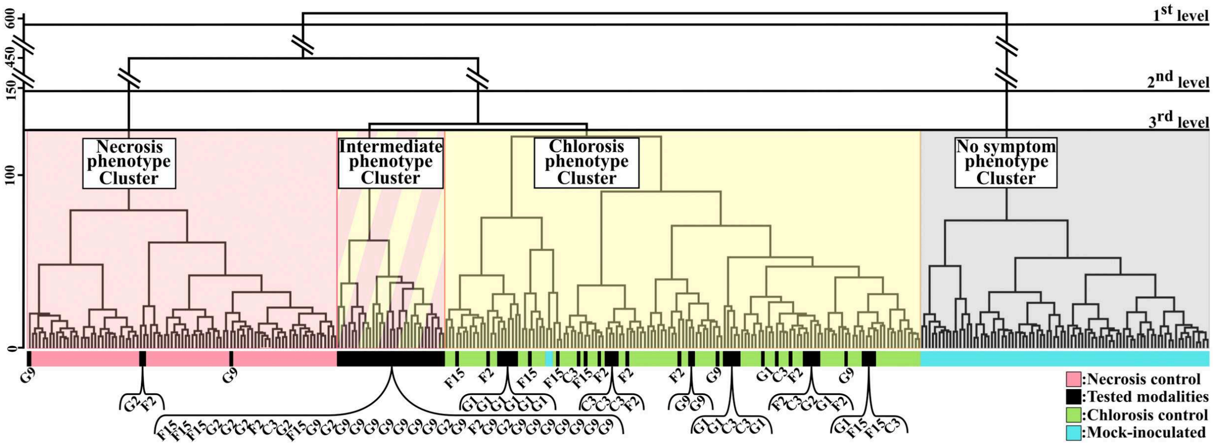


FIGURE 3 | Clustering of samples of the first dataset according to the combination of the 70 images of CF parameters. Dendrogram based on three-dimensional Euclidean distance of 70 images combination and Ward agglomeration method. Black horizontal bars corresponded to the different levels of analysis of the dendrogram. Four clusters corresponding to the four phenotypes visually observed (necrosis, intermediate, chlorosis, and no symptom phenotypes) were obtained. Necrosis (pink) corresponds to leaf tissues inoculated with strain 7698R. Chlorosis (green) corresponds to leaf tissues inoculated with strain 7698R pIJ3225. Mock-inoculated (blue) corresponds to leaf tissues inoculated with water. Tested-strains (black) correspond to leaf tissues inoculated with strains G9, G2, F2, F15, C3, or G1.

expected phenotypes at 6 dpi for the controls used as ground truth (Meline et al., 2019).

To better assess the relevance of the approach, we compared the sensitivity and the specificity of the clustering using 70 images of CF parameters, or using a single parameter (**Table 1**). To perform clustering using single one, we chose F_v/F_m or *NPQ*, as these parameters are commonly used in CF phenotyping (Baron et al., 2016; Kalaji et al., 2017). Related to F_v/F_m and *NPQ*, two supplementary dendrograms were then built and shown in **Figure S2**. Using the combination of 70 images of parameters, the mock-inoculated control have been efficiently discriminated from other controls (sensitivity almost of 0.98 and specificity of 1). When using clustering based only on single parameter, contrasting results were obtained according to which parameter was used. Using the sole F_v/F_m parameter, mock-inoculated control was also well-discriminated from other controls (sensitivity of 0.99 and a specificity of 1). Inversely, the sole *NPQ* parameter did not allow such an efficient discrimination as only 60% of the mock-inoculated control images were correctly classified into the no symptom phenotype cluster (sensitivity of 0.6). Moreover, by observing the associated dendrogram (dendrogram C. in **Figure S2**), we noted that the remaining 40% images of this control were grouped into a cluster where all the necrosis control images were also grouped. The necrosis and chlorosis controls have been efficiently discriminated using the combination of 70 images of CF parameters (sensitivities of 1 and specificities of 1 and 0.99, respectively). Using the sole F_v/F_m parameter, the sensitivities of discrimination between necrosis and chlorosis phenotypes decreased down to 0.68 and 0.56, respectively, although its specificities remained high as 1 and 0.99, respectively. Such low sensitivities values could be explained with some mild necrosis or chlorosis phenotypes that were grouped in the intermediate phenotype cluster (dendrogram B. in **Figure S2**). Such results highlighted the interest of using a combination of 70 images of CF parameters to improve the discrimination of the three controls.

TABLE 1 | Sensitivity (sens.) and specificity (spec.) of the clustering for the controls according to CF parameters for the two datasets.

	First dataset					
	Mock-inoc.		Chlorosis		Necrosis	
Parameter	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
Combined 70	0.98	1	1	0.99	1	1
F_v/F_m	0.99	1	0.56	0.99	0.68	1
NPQ	0.60	1	0.75	1	1	0.57
	Second dataset					
	Mock-inoc.		GUS		EHA105	
Parameter	Sens.	Spec.	Sens.	Spec.	Sens.	Spec.
Combined 70	1	0.87	1	0.65	0.87	0.67

In the first dataset, controls are mock-inoculated, chlorosis and necrosis. In the second dataset, controls are mock-inoculated, GUS and EHA105.

Assessment of the Contribution of Each Image of CF Parameters in the Combination

According to the results exposed in last section, the combined 70 images of CF parameters provided a better clustering ability of the three controls than only one classical parameter, e.g., F_v/F_m or *NPQ* in this study. To further confirm the relevance of the combination of images of CF parameters, it was therefore suitable to evaluate how the clustering ability is impacted according to

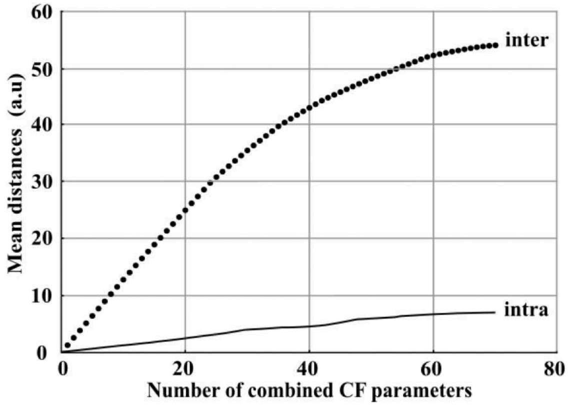


FIGURE 4 | Evolution of the mean inter-modality distance (black spot) and the mean intra-modality distance (black line) between the three controls according to the number of images of CF parameters combined from 1 to 70.

the number of images of CF parameters which are combined. For these purpose, images were sequentially combined from 1 to 70 using sequential forward sequence method (SFS) based on the maximal mean inter-modality distance between the three controls. For each combination, we measured the evolution of the inter-modality and intra-modality distance between the three controls. For our application, the inter-modality distance can be considered as the useful information to discriminate the three controls and the intra modality distance represented the dispersion of all the images of one modality and can therefore be considered like degrading information as the dispersion increased. These mean distances are plotted in **Figure 4**.

The mean inter-modality distance increased quickly with the combination of images between 1 and 40. It then slowly reached a saturation level from 40 to 70. At the same time, the mean intra-modality distance remained at least five times lower than inter-modality distance values. As a consequence, while the combination of a large number of images increased the mean inter-modality distances, it was not a source of degrading information regarding the positive ratio between inter and intra-modality distance.

Application of the Computation Method to Successfully Discriminate the Tested-Strains From the First Dataset

At 3 dpi, among the six tested-strains of the first dataset, only two different phenotypes were visually annotated as shown in **Figure 5**. Tested-strains G9 and G2 were identified as intermediate phenotypes between necrosis and chlorosis phenotypes whereas the phenotypes of the other tested-strains (F2, F15, C3, G1) were identified as a chlorosis phenotype.

To assess the relevance of the computation method, we analyzed the clustering of these tested-strains (**Figure 3**). Samples from the tested-strains G9 and G2 were mainly grouped into a fourth cluster distinct from the three previously described clusters. This fourth cluster corresponded to samples displaying intermediate phenotypes. Such results were expected, as G9 and G2 corresponded to mutant strains for which Tn5 insertions fully inactivated the secretion system encoded on pIJ3225. In contrast Tn5 insertions of F2, F15, C3, G1 tested-strains only inactivated one secreted protein that transit through the secretion system encoded on pIJ3225. For the tested-strains (F2, F15, C3, G1), the inoculation resulted in phenotypes visually identified as chlorosis. Regarding these latter strains, the clustering ability of the method showed more differences than the visual observation. Tested-strains C3 and G1 were grouped almost exclusively into the chlorosis cluster. Whereas, F15 was grouped into intermediate and chlorosis clusters considering almost equal proportions, and F2 was mainly grouped into chlorosis cluster but with a small proportion also grouped in necrosis cluster. The clustering of these single gene mutants of *Xanthomonas* were in accordance with the expected impact of each inactivated gene in the ability of the plasmid pIJ3225 to suppress the necrosis development at 6 dpi, as described in Meline et al. (2019). Therefore, the combination of images of CF parameters allowed at 3 dpi a predictive discrimination between samples that could previously be discriminated only at 6 dpi. Proportions (in percentage) of samples of each tested-strain grouped into necrosis, chlorosis, intermediate or no symptom phenotype clusters were summarized in **Table 2**.

Application of the Computation Method on a Second Dataset to Discriminate Phenotypes for Which No Prior Knowledge Is Available

To test the interest of the computation method on situations that cannot be discriminated by visual assessment and for which no prior knowledge is available, we used a second dataset composed of bacterial virulence factors (T3Es) transiently expressed on leaves (*xopAF*: AF, *xopL*: L, *xopG*: G, *xopV*: V, *xopT*: T or *xopAK*: AK). As shown in **Figure 6**, 3 dpi, no difference was observed visually between mock-inoculated control and the transient expressions of β-glucuronidase protein (GUS) or five T3Es among the six tested. Only transient expression of *xopV* could in rare case induce a mild chlorosis. The computation method was applied to the second dataset (**Figure 7**) and proportions (in percentage) of images of each tested-strain grouped into four clusters were summarized in **Table 2**.

At a first level of analysis, a relevant clustering of mock-inoculated control vs. leaf tissues inoculated with bacterial strains was obtained. All samples corresponding to mock-inoculated control were grouped in cluster 1 with a sensitivity equal to 1. Few samples corresponding to other modalities were misclustered, resulting in a specificity of 0.87 in cluster 1. Here again, plant tissues inoculated with bacterial strains could be discriminated from mock-inoculated strains in the absence of any phenotype visible to the eye (**Figure 6**). At a second level of analysis, the transient expressions of the GUS reporter protein could be discriminated from the transient expression of T3Es. Indeed, all samples of leaf tissues expressing GUS were clustered in cluster 2.1 whereas most samples of leaf

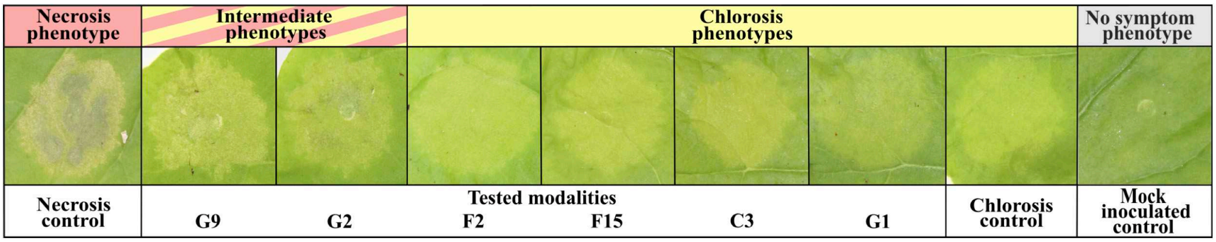


FIGURE 5 | Visual observation of controls and tested-strains of the first data set at 3 dpi. Two phenotypes for the tested-strains (strains G9, G2, F2, F15, C3, G1) were identified as intermediate and chlorosis phenotypes.

TABLE 2 | Proportions of clustering of the tested strains for the two datasets, based on the combination of the 70 images of CF parameters.

		First dataset		
Tested strains	**No symptom phenotype cluster (%)**	**Chlorosis phenotype cluster (%)**	**Intermediate phenotype cluster (%)**	**Necrosis phenotype cluster (%)**
G9	0	17	74	9
G2	0	11	78	11
F2	0	73	18	9
F15	0	60	40	0
C3	0	90	10	0
G1	0	100	0	0

		Second dataset		
Tested strains	**Cluster 1 (%)**	**Cluster 2.1 (%)**	**Cluster 2.2.1 (%)**	**Cluster 2.2.2 (%)**
AF	0	11	39	50
L	0	0	11	89
G	6	11	44	39
V	11	11	17	61
T	0	6	0	94
AK	22	0	39	39

Proportions are expressed in percentage. In the first dataset, tested strains are G9, G2, F2, F15, C3, G1. In the second dataset, tested strains are transient expression of xopAF: AF, xopL: L, xopG: G, xopV: V, xopT: T, xopAK: AK.

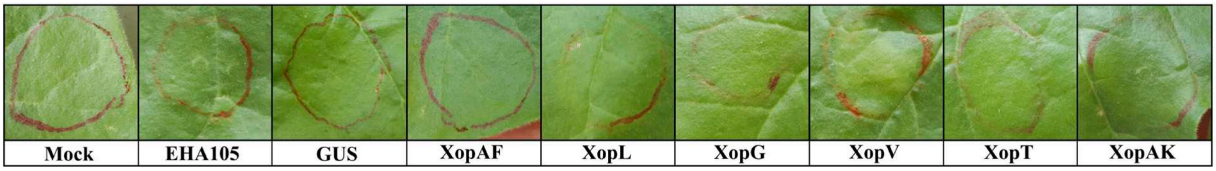


FIGURE 6 | Visual observation of controls and tested-strains of the second dataset at 3 dpi. Mock-inoculated (Mock), *A. tumefaciens* strain no protein expressed (EHA105), the transient expression of β-glucuronidase protein (GUS) were considered as the three controls on each leaf. The six transient expression of T3E genes were considered as the tested-strains (transient expression of *xopAF*: AF, *xopL*: L, *xopG*: G, *xopV*: V, *xopT*: T, *xopAK*: AK, respectively). Constructs for transient expression are described in **Supplementary Table 1**.

tissues expressing T3Es were clustered in cluster 2.2. At this level, the method allowed to discriminate the impact on leaf tissues of presumably deleterious bacterial virulence proteins (T3Es) vs. non-deleterious exogenous proteins (GUS). The sensitivity of GUS clustering was of 1. Its specificity decreased down to 0.65 because of the presumably deleterious effect on leaf tissues of the inoculation with *A. tumefaciens* (EHA105). The sensitivity of EHA105 clustering in cluster 2.2 was of 0.87 with a specificity 0.67. The clustering ability in terms of sensitivity and specificity for these three controls (mock-inoculated, GUS, and EHA105) is summarized in **Table 1**. When further focusing, cluster 2.2.2 grouped 94, 88, and 61% images of leaf tissues inoculated with *xopT*, *xopL*, and *xopV*, respectively. Moreover, the computation method discriminated the transient

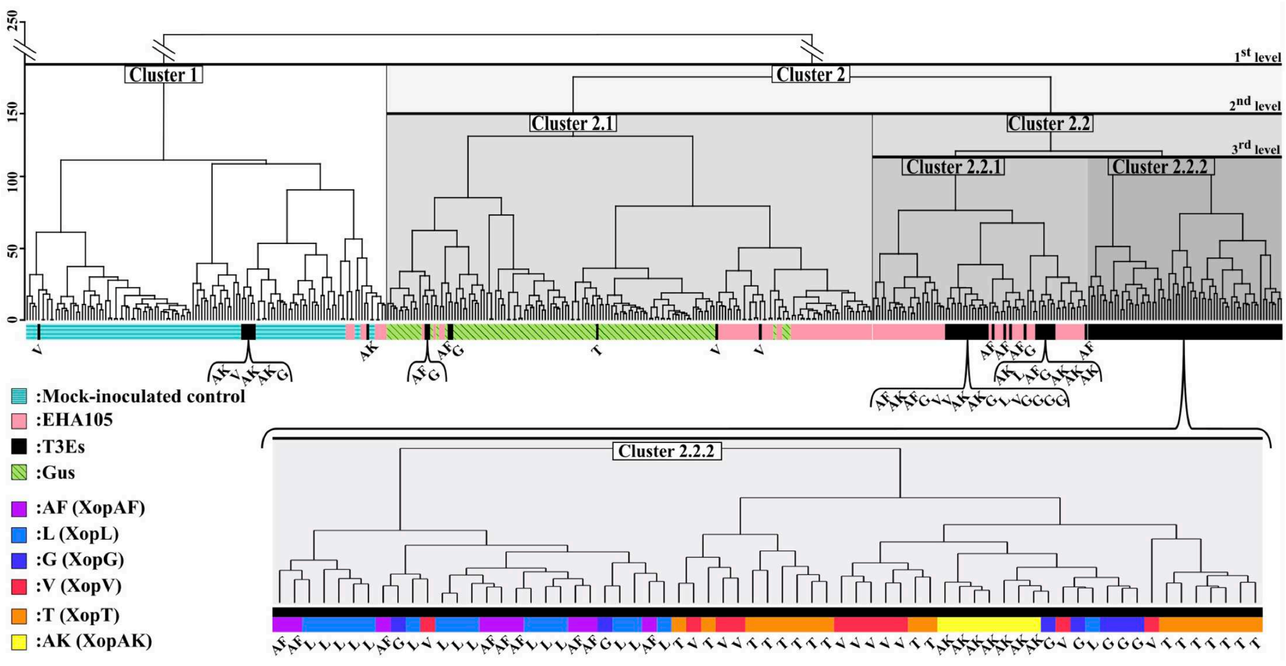


FIGURE 7 | Clustering of samples of the second dataset according to the combination of the 70 images of CF parameters. Black horizontal bars corresponded to the different levels of analysis of the dendrogram. The lower part of the figure presents a focus on the cluster 2.2.2 regrouping the majority of the transient expression of T3Es considered as tested-strains.

expression of *xopL* vs. the transient expression of *xopT* and *xopV* (**Figure 7**, cluster 2.2.2).

DISCUSSION

CF imaging is a powerful technique to study quantitatively plant-pathogen interactions (Perez-Bueno et al., 2016; Pineda et al., 2018) and has been used to highlight potential physiological mechanisms underlying disease symptoms (De Torres Zabala et al., 2015; Zhou et al., 2015). However, some CF parameters are still very contentious (Kalaji et al., 2014) and caution must be exercised when attempting to interpret their significance (Baker, 2008; Murchie and Lawson, 2013). Moreover, careful setups of illumination protocols have to be considered for the assessment of physiological studies. In the context of our study, we developed a computational method to improve the discrimination between close phenotypes based on the combination of images of multiple CF parameters. In the present study, these multiple CF parameters are generated from non-optimized illumination protocols which preclude any physiological interpretation. However, this approach would be fully compatible with optimized illumination protocols. Such an optimization of illumination protocols would probably result in improved performances of the discrimination. The use of optimized illumination protocols in combination with the method described in the present paper may also allow the discrimination of distinct physiological status of plants. However, this latter is beyond the scope of this paper.

Although a large variety of CF parameters is available, only a subset of parameters have been used in the literature (Kalaji et al., 2014). Parameters empirically selected do not provide necessarily the best contrast between studied phenotypes. For instance, F_v/F_m parameter is a relatively stable ratio as impacts of stress could be detected rather late (Lichtenthaler et al., 2005). *NPQ* parameter could be one of the most appropriate one to distinguish plant-pathogen interactions (Rodriguez-Moreno et al., 2008; Perez-Bueno et al., 2015). But it has also been shown to be an unstable temporal parameter (Bonfig et al., 2006; De Torres Zabala et al., 2015). Inversely to subjective selections of these parameters, the depicted computation method provided an objective method to exploit all the information from all the available images of CF parameters. However, although the present paper presents a computation method to gather all the information provided by multiple images of CF parameters, the illumination protocol remains to be optimized to provide more physiologically relevant information.

It had been considered that CF imaging and thermography techniques could have a lack of potential to identify specific diseases in contrary to RGB and hyperspectral imaging techniques (Martinelli et al., 2015; Mahlein, 2016). Indeed, taken individually, F_v/F_m parameter could be impacted either by any abiotic or biotic stress factor like water deficit (Bresson et al., 2015) or bacterial disease development (Perez-Bueno et al., 2016). However, it has been shown that the inoculation of an avirulent *Pseudomonas syringae* pv. *glycinea* on soybean leaves was associated with a decrease in F_v/F_m and an increase in *NPQ* parameters (Zou et al., 2005) whereas an inoculation of an avirulent *Pseudomonas syringae* pv. *tomato* on *arabidopsis thaliana* was associated with a decrease in both F_v/F_m and *NPQ* parameters (Bonfig et al., 2006). Thereby, combination of several CF parameters could constitute a specific signature of visually similar stress, as reviewed in Baron et al. (2012).

During infections, plant-pathogen interactions could have effects on several mechanisms of the plant. Regulation of stomatal aperture could impact temperature dynamic which could be study using thermography imaging (Sankaran et al., 2013; Maes et al., 2014). Production of phenolic compounds involved in plant defense could be measured with blue green fluorescence (BGF) which can be study using Multi Color Fluorescence Imaging (MCFI) (Perez-Bueno et al., 2015; Montero et al., 2016; Ortiz-Bustos et al., 2017). CF imaging could measure performance of the photosystem II which could be altered by the development of a pathogen (Rousseau et al., 2015). Using several imaging techniques in parallel could improve the performance of detection and allowed pre-symptomatic identification of a stress. It could also be used to improve the discrimination abilities to identify disease signatures for a specific pathogen (Mahlein et al., 2012; Baron et al., 2016). The method presented in this study could be generically applied on a combination of multi-modal images.

Adding a temporal monitoring may improve the phenotyping for the characterization of the plant-pathogen interactions. Indeed, monitoring the temporal dynamic revealed useful to discriminate temporal spectral signatures of three foliar pathogens of barley leaves (Wahabzada et al., 2015). As well, Berger et al. (2007) showed that infections by virulent and avirulent strains of *Pseudomonas syringae* result in distinct temporal dynamics of CF parameters, although the same CF parameters were involved. For example, the response to *Pseudomonas syringae* pv. *tomato* DC3000 could be discriminated from the response to its mutant inactivated in the type 3 secretion system by a transient increase of *NPQ* parameter between 6 and 12 hpi (De Torres Zabala et al., 2015). Furthermore, spatio-temporal phenotyping of the response to virulent and avirulent strains of *Pseudomonas syringae* provided non-redundant information (Perez-Bueno et al., 2015). Therefore, combining our computation method with a temporal monitoring could improve significantly the discrimination of biotic stresses on leaves.

The aim of our paper is essentially to present and discuss a computation method combining for each sample 70 images of CF parameters, without any selection, and computing a normalization in order to take into account the inter-leaves and intra-phenotypes heterogeneities. We show its implementation and its usability to provide significant results, in the study-case of the discrimination of visually similar phenotypes induced by bacterial virulence factors. We showed our method to be useful for phenotyping of the impact of single T3E on plant tissues following transient expression. Any T3E could be tested this way and our approach could reveal of interest as well for phenotyping the impact on plant tissues of T3Es of other plant pathogenic bacteria, such as *Pseudomonas syringae* or *Ralstonia solanacearum*. The genericity of the present computation method also resides in the possibility to be applied systematically to any dataset where controls can be defined. In that way, controls can be defined within each leaf (as described in the present paper) or within each experiment if it is not possible to apply different stresses to the same leaf. Furthermore, the use of CF imaging for the phenotyping is widely documented for the study of both biotic (Chaerle et al., 2004) and abiotic stresses (Yusuf et al., 2010).

Although we showed in this study that the combination of a large number of CF parameters was not degrading information, we could not rule out that some parameters could carry redundant information. Moreover, it had been demonstrated that in some case, the combination of eleven or less CF parameters involved the most important fluorescence signatures and could be sufficient to classify tissues inoculated with different strains (Mishra et al., 2014; Cen et al., 2017). It could be interesting to process the datasets using selective and reductive methods (as decision tree for instance) to select and identify parameters which could discriminate tissues inoculated with different strains and to reduce inherent redundancy and overfeeding. However, we could consider that the present method remained an efficient alternative to such learning methods which could require large annotated datasets that can be tough to obtained in plant phenotyping domain. In that direction, further studies aiming at selecting the most informative CF parameters in each situation may further increase the discriminative power of the method described in the present paper. Indeed different biological situations may impact differentially the various CF parameters, therefore the most informative CF parameters may differ among the biological situations. The selection of the most informative CF parameters in each biological situation would thus provide a first step toward specific signatures of particular stresses. This constitutes the scope of further studies in order to complete and improve the present method.

In plant phenotyping, heterogeneity linked to each plant constitutes an important and widespread limitation. In this paper, the computation method based on multiple Bhattacharyya distance calculation between tested-strain and controls on each leaf allowed to circumvent the inter-leaves heterogeneity and to take into account the heterogeneity of each phenotype. Plant-pathogen interactions studies often face phenotypes that could be hardly discriminated by visual assessment. The computation method based on the combination of images of multiple CF parameters provided an efficient discrimination of visually similar phenotypes which differed only by one protein secretion or protein expression. The results obtained in the present paper support the idea that combination of images of CF parameters improve the discrimination between distinct biotic stresses compared to single CF approach. Furthermore, the combination of several imaging techniques using this computation method could constitute an advance in the identification of specific signature of these biotic stresses.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

VM, TB, and EB conceived the study, drafted the manuscript, designed the experimental protocols, and developed the computational method applied in this study. VM, LL, DS, and CB realized the experimentations. GL contributed to the management and the analysis of the chlorophyll fluorescence images. GH worked on the Bhattacharyya distance applications to data of the study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work received support from the RFI vegetal, région Pays de la Loire in the framework of the project Phenoscreen (2015–2018) for the thesis of Valérian Méline. This work was also supported by INRA-SPE (Phénotypage de l'impact sur les tissus végétaux d'Effecteurs de Type 3 par imagerie de fluorescence de chlorophylle) and Le Ministère de l'Agriculture (project CASDAR Phénotypage des interactions entre les espèces végétales potagères et les *Xanthomonas*). In the framework of the CASDAR project, we thank Perrine David and Clément Poisblaud of HM-Clause for fruitful discussions on CF imaging of *Xanthomonas* plant interactions. This work benefited from the scientific support of the EPPN^{2020} network.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.00213/full#supplementary-material

REFERENCES

Agrawal, R., and Srikant, R. (1995). "Mining sequential patterns," in *Data Engineering, Proceedings of the Eleventh International Conference on Data Engineering* (Taipei: IEEE), 3–14.

Alenya, G., Dellen, B., and Torras, C. (2011). "3D modelling of leaves from color and ToF data for robotized plant measuring," in *IEEE International Conference on Robotics and Automation* (Shanghai), 3408–3414.

Arlat, M., Gough, C. L., Barber, C. E., Boucher, C., and Daniels, M. J. (1991). *Xanthomonas campestris* cantains a cluster of hrp genes related to the larger hrp cluster of *Pseudomonas solanacearum*. *Mol. Plant Microbe Interact.* 4, 593–601.

Baker, N. R. (2008). Chlorophyll fluorescence: a probe of photosynthesis *in vivo*. *Annu. Rev. Plant Biol.* 59, 89–113. doi: 10.1146/annurev.arplant.59.032607.092759

Baron, M., Flexas, J., and Delucia, E. H. (2012). *Photosynthetic Responses to Biotic Stress*, Chapter 22. Cambridge: Cambridge University Press.

Baron, M., Pineda, M., and Perez-Bueno, M. L. (2016). Picturing pathogen infection in plants. *Z. Naturforsch. C* 71, 355–368. doi: 10.1515/znc-2016-0134

Bellasio, C., Olejnickova, J., Tesar, R., Sebela, D., and Nedbal, L. (2012). Computer reconstruction of plant growth and chlorophyll fluorescence emission in three spatial dimensions. *Sensors* 12, 1052–1071. doi: 10.3390/s120101052

Berger, S., Benediktyova, Z., Matous, K., Bonfig, K. B., Mueller, M. J., Nedbal, L., et al. (2007). Visualization of dynamics of plant-pathogen interaction by novel combination of chlorophyll fluorescence imaging and statistical analysis: differential effects of virulent and avirulent strains of *P. syringae* and of oxylipins on *A. thaliana*. *J. Exp. Bot.* 58, 797–806. doi: 10.1093/jxb/erl208

Bhattacharyya, A. (1943). The divergence and Bhattacharyya distance measures in signal selection. *IEEE Trans. Commun. Technol.* 15, 52–60.

Bjorkman, O., and Demmig, B. (1987). Photon yield of O_2 evolution and chlorophyll fluorescence characteristics at 77 K among vascular plants of diverse origins. *Planta* 170, 489–504.

Blashfield, R. (1976). Mixture model tests of cluster analysis: accuracy of four agglomerative hierarchical methods. *Psychol. Bull.* 83, 377–388.

Boch, J., Bonas, U., and Lahaye, T. (2014). TAL effectors–pathogen strategies and plant resistance engineering. *New Phytol.* 204, 823–832. doi: 10.1111/nph.13015

Bock, C. H., Parker, P. E., Cook, A. Z., and Gottwald, T. R. (2008). Visual rating and the use of image analysis for assessing different symptoms of citrus canker on grapefruit leaves. *Plant Dis.* 92, 530–541. doi: 10.1094/PDIS-92-4-0530

Bogdanove, A. J., Koebnik, R., Lu, H., Furutani, A., Angiuoli, S. V., Patil, P. B., et al. (2011). Two new complete genome sequences offer insight into host and tissue specificity of plant pathogenic *Xanthomonas* spp. *J. Bacteriol.* 193, 5450–5464. doi: 10.1128/JB.05262-11

Bonfig, K. B., Schreiber, U., Gabler, A., Roitsch, T., and Berger, S. (2006). Infection with virulent and avirulent *P. syringae* strains differentially affects photosynthesis and sink metabolism in Arabidopsis leaves. *Planta* 225, 1–12. doi: 10.1007/s00425-006-0303-3

Bresson, J., Vasseur, F., Dauzat, M., Koch, G., Granier, C., and Vile, D. (2015). Quantifying spatial heterogeneity of chlorophyll fluorescence during plant growth and in response to water stress. *Plant Methods* 11:23. doi: 10.1186/s13007-015-0067-5

Büttner, D. (2016). Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol. Rev.* 40, 894–937. doi: 10.1093/femsre/fuw026

Cen, H., Weng, H., Yao, J., He, M., Lv, J., Hua, S., et al. (2017). Chlorophyll fluorescence imaging uncovers photosynthetic fingerprint of citrus Huanglongbing. *Front. Plant Sci.* 8:1509. doi: 10.3389/fpls.2017.01509

Cesbron, S., Briand, M., Essakhi, S., Gironde, S., Boureau, T., Manceau, C., et al. (2015). Comparative genomics of pathogenic and nonpathogenic strains of *Xanthomonas arboricola* unveil molecular and evolutionary events linked to pathoadaptation. *Front. Plant Sci.* 6:1126. doi: 10.3389/fpls.2015.01126

Chaerle, L., Hagenbeek, D., De Bruyne, E., Valcke, R., and Van Der Straeten, D. (2004). Thermal and chlorophyll-fluorescence imaging distinguish plant-pathogen interactions at an early stage. *Plant Cell Physiol.* 45, 887–896. doi: 10.1093/pcp/pch097

Csefalvay, L., Di Gaspero, G., Matous, K., Bellin, D., Ruperti, B., and Olejnickova, J. (2009). Pre-symptomatic detection of *Plasmopara viticola* infection in grapevine leaves using chlorophyll fluorescence imaging. *Eur. J. Plant Pathol.* 125, 291–302. doi: 10.1007/s10658-009-9482-7

Darrasse, A., Carrère, S., Barbe, V., Boureau, T., Arrieta-Ortiz, M. L., Bonneau, S., et al. (2013). Genome sequence of *Xanthomonas fuscans* subsp. fuscans strain 4834-R reveals that flagellar motility is not a general feature of xanthomonads. *BMC Genomics* 14:761. doi: 10.1186/1471-2164-14-761

De Torres Zabala, M., Littlejohn, G., Jayaraman, S., Studholme, D., Bailey, T., Lawson, T., et al. (2015). Chloroplasts play a central role in plant defence and are targeted by pathogen effectors. *Nat. Plants* 1:15074. doi: 10.1038/nplants.2015.74

Deza, M., and Deza, E. (2009). *Encyclopedia of Distances*. Berlin: Springer.

Dhondt, S., Vanhaeren, H., Van Loo, D., Cnudde, V., and Inze, D. (2010). Plant structure visualization by high-resolution X-ray computed tomography. *Trends Plant Sci.* 15, 419–422. doi: 10.1016/j.tplants.2010.05.002

Erickson, J. L., Adlung, N., Lampe, C., Bonas, U., and Schattat, M. H. (2018). The *Xanthomonas* effector XopL uncovers the role of microtubules in stromule extension and dynamics in *Nicotiana benthamiana*. *Plant J.* 93, 856–870. doi: 10.1111/tpj.13813

Essakhi, S., Cesbron, S., Fischer-Le Saux, M., Bonneau, S., Jacques, M. A., and Manceau, C. (2015). Phylogenetic and variable-number tandem-repeat analyses identify nonpathogenic *Xanthomonas arboricola* lineages lacking the canonical type III secretion system. *Appl. Environ. Microbiol.* 81, 5395–5410. doi: 10.1128/AEM.00835-15

Fang, S., Yan, X., and Liao, L. H. (2009). 3D reconstruction and dynamic modeling of root architecture *in situ* and its application to crop phosphorus research. *Plant J.* 60, 1096–1108. doi: 10.1111/j.1365-313X.2009.04009.x

Feng, F., Yang, F., Rong, W., Wu, X., Zhang, J., Chen, S., et al. (2012). A *Xanthomonas* uridine 5′-monophosphate transferase inhibits plant immune kinases. *Nature* 485, 114–118. doi: 10.1038/nature10962

Garita-Cambronero, J., Palacio-Bielsa, A., López, M. M., and Cubero, J. (2016). Comparative genomic and phenotypic characterization of pathogenic and non-pathogenic strains of *Xanthomonas arboricola* reveals insights into the infection process of bacterial spot disease of stone fruits. *PLoS ONE* 11:e0161977. doi: 10.1371/journal.pone.0161977

Genty, B., Briantais, J. M., and Baker, N. R. (1989). The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. Acta* 990, 87–92.

Genty, B., Harbinson, J., Briantais, J. M., and Baker, N. R. (1990). The relationship between non-photochemical quenching of chlorophyll fluorescence and the rate of photosystem 2 photochemistry in leaves. *Photosynth. Res.* 25, 249–257.

Gorbe, E., and Calatayud, A. (2012). Applications of chlorophyll fluorescence imaging technique in horticultural research: a review. *Sci. Horticult.* 138, 24–35. doi: 10.1016/j.scienta.2012.02.002

Goudail, F., Refregier, P., and Delyon, G. (2004). Bhattacharyya distance as a contrast parameter for statistical processing of noisy optical images. *Data Eng.* 21, 1231–1240. doi: 10.1364/JOSAA.21.001231

Guy, E., Genissel, A., Hajri, A., Chabannes, M., David, P., Carrere, S., et al. (2013b). Natural genetic variation of *Xanthomonas campestris* pv. campestris pathogenicity on arabidopsis revealed by association and reverse genetics. *MBio* 4:e00538-12. doi: 10.1128/mBio.00978-13

Guy, E., Lautier, M., Chabannes, M., Roux, B., Lauber, E., Arlat, M., et al. (2013a). xopAC-triggered immunity against *Xanthomonas* depends on Arabidopsis receptor-like cytoplasmic kinase genes PBL2 and RIPK. *PLoS ONE* 8:e73469. doi: 10.1371/journal.pone.0073469

Hajri, A., Brin, C., Hunault, G., Lardeux, F., Lemaire, C., Manceau, C., et al. (2009). A "repertoire for repertoire" hypothesis: repertoires of type three effectors are candidate determinants of host specificity in *Xanthomonas*. *PLoS ONE* 4:e6632. doi: 10.1371/journal.pone.0006632

Hajri, A., Pothier, J. F., Fischer-Le Saux, M., Bonneau, S., Poussier, S., Boureau, T., et al. (2012). Type three effector gene distribution and sequence analysis provide new insights into the pathogenicity of plant-pathogenic *Xanthomonas arboricola*. *Appl. Environ. Microbiol.* 78, 371–384. doi: 10.1128/AEM.06119-11

Hands, S., and Everitt, B. (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivar. Behav. Res.* 22, 235–243.

Honsdorf, N., March, T. J., Berger, B., Tester, M., and Pillen, K. (2014). High-throughput phenotyping to detect drought tolerance QTL in wild barley introgression lines. *PLoS ONE* 9:e97047. doi: 10.1371/journal.pone.0097047

Jacques, M. A., Arlat, M., Boulanger, A., Boureau, T., Carrère, S., Cesbron, S., et al. (2016). Ecology, physiology, and genomics to understand host specificity in *Xanthomonas*. *Annu. Rev. Phytopathol.* 54, 163–187. doi: 10.1146/annurev-phyto-080615-100147

Jahnke, S., Menzel, M. I., van Dusschoten, D., Roeb, G. W., Bühler, J., Minwuyelet, S., et al. (2009). Combined MRI-PET dissects dynamic changes in plant structures and functions. *Plant J.* 59, 634–644. doi: 10.1111/j.1365-313X.2009.03888.x

Kailath, T. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.* 35, 99–109.

Kalaji, H. M., Schansker, G., Brestic, M., Bussotti, F., Calatayud, A., Ferroni, L., et al. (2017). Frequently asked questions about chlorophyll fluorescence, the sequel. *Photosynth. Res.* 132, 13–66. doi: 10.1007/s11120-016-0318-y

Kalaji, H. M., Schansker, G., Ladle, R. J., Goltsev, V., Bosa, K., Allakhverdiev, S. I., et al. (2014). Frequently asked questions about *in vivo* chlorophyll fluorescence: practical issues. *Photosynth. Res.* 122, 121–158. doi: 10.1007/s11120-014-0024-6

Karimi, M., Inze, D., and Depicker, A. (2002). GATEWAY^{TM} vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Kolber, Z., S., Prasil, O., and Falkowski, P. G. (1998). Measurements of variable chlorophyll fluorescence using fast repetition rate techniques: defining methodology and experimental protocoles. *Biochim. Biophys. Acta* 1367, 88–106.

Li, C. M., Brown, I., Mansfield, J., Stevens, C., Boureau, T., Romantschuk, M., et al. (2002). The Hrp pilus of Pseudomonas syringae elongates from its tip and acts as a conduit for translocation of the effector protein HrpZ. *EMBO J.* 21, 1909–1915. doi: 10.1093/emboj/21.8.1909

Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078

Lichtenthaler, H. K., Buschmann, C., and Knapp, M. (2005). How to correctly determine the different chlorophyll fluorescence parameters and the chlorophyll fluorescence decrease ratio R_{Fd} of leaves with the PAM fluorometer. *Photosynthetica* 25, 379–393. doi: 10.1007/s11099-005-0062-6

Maes, W. H., P.Minchin, E. H., Snelgar, W. P., and Steppe, K. (2014). Early detection of Psa infection in kiwifruit by means of infrared thermography at leaf and orchard scale. *Funct. Plant Biol.* 41:1207. doi: 10.1071/FP14021

Mahlein, A. K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Dis.* 100, 241–251. doi: 10.1094/PDIS-03-15-0340-FE

Mahlein, A. K., Steiner, U., Hillnhutter, C., Dehne, H. W., and Oerke, E. C. (2012). Hyperspectral imaging for small scale analysis of symtoms caused by different sugar beet diseases. *Plant Methods* 8:3. doi: 10.1186/1746-4811-8-3

Manulis, S., and Barash, I. (2003). *Pantoea agglomerans* pvs. *gypsophilae* and *betae*, recently evolved pathogens? *Mol. Plant Pathol.* 4, 307–314. doi: 10.1046/j.1364-3703.2003.00178.x

Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi P., et al. (2015). Advanced methods of plant disease detection. A review. *Agron. Sustain. Dev.* 35, 1–25. doi: 10.1007/s13593-014-0246-1

Meline, V., Delage, W., Brin, C., Belin, E., Boureau, T., Arlat, M., et al. (2019). Role of the acquisition of a type 3 secretion system in the emergence of novel pathogenic strains of *Xanthomonas*. *Mol. Plant Pathol.* 20, 33–50. doi: 10.1111/mpp.12737

Merda, D., Bonneau, S., Guimbaud, J. F., Durand, K., Brin, C., Boureau, T., et al. (2016). Recombination-prone bacterial strains form a reservoir from which epidemic clones emerge in agroecosystems: recombinant strains as a reservoir for epidemics. *Environ. Microbiol. Rep.* 8, 572–581. doi: 10.1111/1758-2229.12397

Merda, D., Briand, M., Bosis, E., Rousseau, C., Portier, P., Barret, M., et al. (2017). Ancestral acquisitions, gene flow and multiple evolutionary trajectories of the type three secretion system and effectors in *Xanthomonas* plant pathogens. *Mol. Ecol.* 26, 5939–5952. doi: 10.1111/mec.14343

Mishra, A., Heyer, E. K., and Mishra, K. (2014). Chlorophyll fluorescence emission can screen cold tolerance of cold acclimated *Arabidopsis thaliana* accessions. *Plant Methods* 10:38. doi: 10.1186/1746-4811-10-38

Montero, R., Perez-Bueno, M. L., Baron, M., Florez-Sarasa, I., Tohge, T., Fernie, A. R., et al. (2016). Alterations in primary and secondary metabolism in Vitis vinifera Malvasia de Banyalbufar upon infection with Grapevine leafroll–associated virus 3. *Physiol. Plant.* 157, 442–452. doi: 10.1111/ppl.12440

Moreira, L. M., Almeida, N. F., Potnis, N., Digiampietri, L., Adi, S. S., Bortolossi, J. C., et al. (2010). Novel insights into the genomic basis of citrus canker based on the genome sequences of two strains of *Xanthomonas fuscans* subsp. aurantifolii. *BMC Genomics* 11:238. doi: 10.1186/1471-2164-11-238

Murchie, E. H., and Lawson, T. (2013). Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *J. Exp. Bot.* 64, 3983–3998. doi: 10.1093/jxb/ert208

Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31, 274–295. doi: 10.1007/s00357-014-9161-z

Mutka, A. M., and Bart, R. S. (2015). Image-based phenotyping of plant disease symptoms. *Front. Plant Sci.* 5:734. doi: 10.3389/fpls.2014.00734

Ortiz-Bustos, C. M., Perez-Bueno, M. L., Baron, M., and Molinero-Ruiz, L. (2017). Use of blue-green fluorescence and thermal imaging in the early detection of sunflower infection by the root parasitic weed orobanche cumana Wall. *Front. Plant Sci.* 8:833. doi: 10.3389/fpls.2017.00833

Paproki, A., Sirault, X., Berry, S., Furbank, R., and Fripp, J. (2012). A novel mesh processing based technique for 3D plant analysis. *BMC Plant Biol.* 12:63. doi: 10.1186/1471-2229-12-63

Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., and Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian J. Ophthalmol.* 56, 45–50. doi: 10.4103/0301-4738.37595

Perez-Bueno, M. L., Granum, E., Pineda, M., Flors, V., Rodriguez-Palenzuela, P., Lopez-Solanilla, E., et al. (2016). Temporal and spatial resolution of activated plant defense responses in leaves of *Nicotiana benthamiana* infected with *Dickeya dadantii*. *Front. Plant Sci.* 6:1209. doi: 10.3389/fpls.2015.01209

Perez-Bueno, M. L., Pineda, M., Diaz-Casado, E., and Baron, M. (2015). Spatial and temporal dynamics of primary and secondary metabolism in *Phaseolus vulgaris* challenged by *Pseudomonas syringae*. *Physiol. Plant.* 153, 161–174. doi: 10.1111/ppl.12237

Pineda, M., Olejnickova, J., Csefalvay, L., and Baron, M. (2011). Tracking viral movement in plants by means of chlorophyll fluorescence imaging. *J. Plant Physiol.* 168, 2035–2040. doi: 10.1016/j.jplph.2011.06.013

Pineda, M., Perez-Bueno, M. L., and Baron, M. (2018). Detection of bacterial infection in melon plants by classification methods based on imaging data. *Front. Plant Sci.* 9:164. doi: 10.3389/fpls.2018.00164

Pineda, M., Soukupova, J., Matous, K., Nedbal, L., and Baron, M. (2008). Conventional and combinatorial chlorophyll fluorescence imaging of tobamovirus-infected plants. *Photosynthetica* 46, 441–451. doi: 10.1007/s11099-008-0076-y

Rodriguez-Moreno, L., Pineda, M., Soukupova, J., Macho, A., Beuzon, C., Baron, M., et al. (2008). Early detection of bean infection by *Pseudomonas syringae* in asymptomatic leaf areas using chlorophyll fluorescence imaging. *Photosynth. Res.* 96, 27–35. doi: 10.1007/s11120-007-9278-6

Rousseau, C., Belin, E., Bove, E., Rousseau, D., Fabre, F., Berruyer, R., et al. (2013). High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. *Plant Methods* 9:17. doi: 10.1186/1746-4811-9-17

Rousseau, C., Hunault, G., Gaillard, S., Bourbeillon, J., Montiel, G., Simier, P., et al. (2015). Phenoplant: a web resource for the exploration of large chlorophyll fluorescence image datasets. *Plant Methods* 11:24. doi: 10.1186/s13007-015-0068-4

Sankaran, S., Maja, J., Buchanon, S., and Ehsani, R. (2013). Huanglongbing citrus greening detection using visible, near infrared and thermal imaging techniques. *Sensors* 13, 2117–2130. doi: 10.3390/s130202117

Sebela, D., Quinones, C., Cruz, C. V., Ona, I., Olejnickova, J., and Jagadish, K. S. (2018). Chlorophyll fluorescence and reflectance-based non-nvasive uantification of blast, bacterial blight and drought stresses in rice. *Plant Cell Physiol.* 59, 23–43. doi: 10.1093/pcp/pcx144

Wahabzada, M., Mahlein, A. K., Bauckhage, C., Steiner, U., Oerke, E. C., and Kersting, K. (2015). Metro maps of plant disease dynamics-automated mining of differences using hyperspectral images. *PLoS ONE* 10:e0116902. doi: 10.1371/journal.pone.0116902

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244.

Wei, H. L., and Collmer, A. (2018). Defining essential processes in plant pathogenesis with *Pseudomonas syringae* pv. tomato DC3000 disarmed polymutants and a subset of key type III effectors. *Mol. Plant Pathol.* 19, 1779–1794. doi: 10.1111/mpp.12655

Yusuf, M., A., Kumar, D., Rajwanshi, R., Strasser, R., J., Tsimilli-Michael, M., and Sarin, S. N. (2010). Overexpression of γ-tocopherol methyl transferase gene in transgenic *Brassica juncea* plants alleviates abiotic stress: physiological and chlorophyll a fluorescence measurements. *Biochim. Biophys. Acta* 1797, 1428–1438. doi: 10.1016/j.bbabio.2010.02.002

Zhou, J., Zneg, L., Liu, J., and Xing, D. (2015). Manipulation of the xanthophyll cycle increases plant susceptibility to *Sclerotinia sclerotiorum*. *PLoS Pathog.* 11:e1004878. doi: 10.1371/journal.ppat.1004878

Zhu, J., Ingram, P. A., Benfey, P. N., and Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. *Curr. Opin. Plant Biol.* 14, 310–317. doi: 10.1016/j.pbi.2011.03.020

Zou, J., Rodriguez-Zas, S., Aldea, M., Li, M., Zhu, J., Gonzales, D. O., et al. (2005). Expression profiling soybean response to *Pseudomonas syringae* reveals new defense-related genes and rapid HR-specific downregulation of photosynthesis. *Mol. Plant Microbe Interact.* 18, 1161–1174. doi: 10.1094/MPMI-18-1161

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 27 February 2020
doi: 10.3389/fpls.2020.00159

Assessment of Mixed Sward Using Context Sensitive Convolutional Neural Networks

Christopher J. Bateman[1]*, *Jaco Fourie*[1], *Jeffrey Hsiao*[1], *Kenji Irie*[2], *Angus Heslop*[3], *Anthony Hilditch*[3], *Michael Hagedorn*[2], *Bruce Jessep*[3], *Steve Gebbie*[3] and *Kioumars Ghamkhar*[4]

[1] *Lincoln Agritech Limited, Lincoln University, Lincoln, New Zealand,* [2] *Red Fern, Solutions Limited, Christchurch, New Zealand,* [3] *Development Engineering, Lincoln Research Centre, AgResearch, Lincoln, New Zealand,* [4] *Forage, Science, Grasslands Research Centre, AgResearch, Palmerston North, New Zealand*

Edited by:
Cyril Pommier,
INRA Centre Versailles-Grignon,
France

Reviewed by:
Giuseppe Jurman,
Fondazione Bruno Kessler,
Italy
Susanne Barth,
The Irish Agriculture and Food
Development Authority,
Ireland

***Correspondence:**
Christopher J. Bateman
christopher.bateman@lincolnagritech.co.nz

Specialty section:
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *02 July 2019*
Accepted: *03 February 2020*
Published: *27 February 2020*

Citation:
Bateman CJ, Fourie J, Hsiao J,
Irie K, Heslop A, Hilditch A,
Hagedorn M, Jessep B, Gebbie S and
Ghamkhar K (2020) Assessment of
Mixed Sward Using Context Sensitive
Convolutional Neural Networks.
Front. Plant Sci. 11:159.
doi: 10.3389/fpls.2020.00159

Breeding higher yielding forage species is limited by current manual harvesting and visual scoring techniques used for measuring or estimation of biomass. Automation and remote sensing for high throughput phenotyping has been used in recent years as a viable solution to this bottleneck. Here, we focus on using RGB imaging and deep learning for white clover (*Trifolium repens* L.) and perennial ryegrass (*Lolium perenne* L.) yield estimation in a mixed sward. We present a new convolutional neural network (CNN) architecture designed for semantic segmentation of dense pasture and canopies with high occlusion to which we have named the local context network (LC-Net). On our testing data set we obtain a mean accuracy of 95.4% and a mean intersection over union of 81.3%, outperforming other methods we have found in the literature for segmenting clover from ryegrass. Comparing the clover/vegetation fraction for visual coverage and harvested dry-matter however showed little improvement from the segmentation accuracy gains. Further gains in biomass estimation accuracy may be achievable through combining RGB with complimentary information such as volumetric data from other sensors, which will form the basis of our future work.

Keywords: forage yield, clover, ryegrass, biomass, semantic segmentation, deep learning

INTRODUCTION

The increase in demand for meat and dairy over the last few decades has led to an intensification of forage based farming. Breeding for the improvement of forage yield and nutrient composition of grassland forage species adds value to these industries (Smith and Spangenberg, 2014; Capstaff and Miller, 2018; Gebremedhin et al., 2019). The length of time required to develop stable new forage cultivars can however take up to 10–15 years (Lee et al., 2012). One of the bottlenecks in this process

---

Abbreviations: CNN, convolutional neural network; DM, dry matter; FCN, fully convolutional network for semantic segmentation (Shelhamer et al., 2016); HSV, hue, saturation, value; IoU, intersection over union; LC-Net, local context network; LCPP, local context pyramid pooling; LED, light emitting diode; LiDAR, light detection and ranging; MSICS, multiple scanning imaging capacity system; RGB, red, green, and blue; RTK-GPS, real-time kinematic global positioning system; SIG, synthetic image generation.

is that growth rates and yield measurements for these forage species are generally done by visual scoring and/or manual harvesting. Both of these practices require a considerable amount of time and labour to perform, which in turn limits the size of breeding trials (Araus et al., 2018; Gebremedhin et al., 2019). With advances in sensors, computing technologies, and more recently in artificial intelligence tools and methods, there has been a surge of interest around automated and high throughput phenotyping techniques for overcoming this bottleneck (Walter et al., 2012; Araus et al., 2018).

Numerous techniques have been developed over the last 15 years for assessing clover and ryegrass forage yield using image data. (Bonesmo et al., 2004) developed a method for pixel-level segmentation of clover and ryegrass using thresholding and morphological filtering on RGB images. This method was applied to predicting the visible area of clover within a specified region. They showed that their automated method could achieve correlation of $R^2 = 0.81$ compared to marking out the same areas manually. This method was extended by (Himstedt et al., 2012) through performing segmentation in the HSV colour space. They then combined this information with a linear model for different legumes to estimate the legume biomass. Although high correlations between measured and predicted legume content were observed (R^2 = 0.90–0.94), the validation set only consisted of 44 data points spread between their different legumes and monoculture examples. They also limited their validation method to swards with less than 2,800 kgDM/ha. (Mortensen et al., 2017) also investigated improvements to this method by looking at different ways of combining the colour information in RGB space. They were also able to obtain a biomass correlation of $R^2 = 0.93$ with a multivariate linear model trained separately for mixed sward grown from different seed mixtures. This was done using a similar sized validation set, however spanning a larger forage dry matter range (up to 3,084 kgDM/ha). An issue all of these researchers found with the morphological approach is that the erosion operations have a tendency to eliminate small clover leaves producing a bias — especially in younger pasture. There are also significant differences between the camera setup, lighting, and image resolutions in each of their setups—requiring tuning of the parameters provided in (Bonesmo et al., 2004) and derivative methods to use them. (Skovsen et al., 2017) investigated use of deep learning for segmentation of clover and ryegrass. They trained the fully convolutional network for semantic segmentation (FCN) model (Shelhamer et al., 2014) using synthetic pasture data constructed from clover, ryegrass, and weed leaves cropped from photographs. The network was shown to have a significantly higher segmentation accuracy than the method in (Bonesmo et al., 2004)—with a pixel accuracy of 83.4% and mean intersection over union (IoU) of 65.5%. (Skovsen et al., 2017) did not regress their segmentation results directly to the clover dry matter as done by others mentioned. Instead their analysis was limited to the correlation of clover-vegetation fraction between dry matter measurements and estimated coverage area.

The use of deep learning in place of traditional machine vision techniques has become increasingly common since (Krizhevsky et al., 2012) demonstrated that CNNs are unreasonably effective for solving image classification type problems. Over the last five years, significant efforts have been made toward adapting CNNs to image segmentation. The majority of these networks follow the same general procedure: the image is first down-sampled while extracting semantic information, followed by up-samping and extrapolation of the semantic information back to the image's original size. The simplest segmentation network (FCN) makes coarse predictions from down-sampled features, then uses learned deconvolutions, skip-connections, and bilinear interpolation to upscale the predictions back to the original image size Shelhamer et al., 2016). SegNet built on this by following a more explicit encoder-decoder architecture, replacing deconvolutions with inverse-pooling operations that are subsequently followed by convolution layers (Badrinarayanan et al., 2016). The latest iteration of the DeepLab architecture (DeepLabV3+) is currently one of the more popular networks being used for high accuracy image segmentation. This replaces some of the down-sampling steps with separable atrous convolutions—which achieve the same effect without losing spatial resolution. They also reinforce the predictions using an additional decoder and global context information for the image (Chen et al., 2018). For a more general review on deep learning based image segmentation, we direct the reader to (Garcia-Garcia et al., 2017).

FCN was the first deep learning network architecture developed for pixel-level segmentation. In the two years between since its introduction and subsequent application to pasture segmentation there have been dozens of networks developed that can achieve significantly higher segmentation accuracy (Garcia-Garcia et al., 2017). State of the art networks such as DeepLabV3+, PSPNet, and EncNet all achieve an mean IoU above 80% on standard computer vision benchmark datasets (Zhao et al., 2017; Chen et al., 2018; Zhang et al., 2018). It is worth noting however that the benchmarks these networks are developed generally focus on urban settings—and therefore are not guaranteed to perform as well for agricultural applications. These benchmarks often contain images with objects sparsely distributed with relatively little occlusion. Pasture images however have objects densely clustered with a high degree of occlusion.

We reported on a mobile multisensor platform for high throughput phenotyping of ryegrass to augment selective breeding (Ghamkhar et al., 2018). Here, we build upon the work of (Skovsen et al., 2017) and (Ghamkhar et al., 2018), with the aim of improving the accuracy of clover/ryegrass segmentation. The focus of this paper is on the measurement of percentage white clover (*Trifolium repens* L.) and perennial ryegrass (*Lolium perenne* L.) yields in mixed sward using top-down view RGB images and deep learning. We introduce a new CNN architecture which we have named local context network (LC-Net) which applies design principles from the aforementioned networks to segmentation of complex agricultural images. We show that LC-Net can differentiate clover and ryegrass with a significantly higher accuracy than previously applied deep learning based methods reported by (Skovsen et al., 2017).

MATERIALS AND METHODS

Data Collection

All data used in the manuscript was collected at the AgResearch Lincoln farm in Canterbury, New Zealand (−43.667799, 172.470908). Data used for training our neural network was gathered from location's A and B in **Figure 1**. A mixed sward trial at location C was used for independently comparing the RGB segmentation results to dry matter harvest measurements. All three of these locations contained varying mixtures of perennial ryegrass and white clover. Data was collected using our mobile data collection platform for forage assessment—the *Multiple Scanning Imaging Capacity System* (MSICS) described by (Barrett et al., 2018).

Mobile Data-Capture Platform

The MSICS contains a range of different sensors—including RGB, LiDAR, as well as VNIR and SWIR hyperspectral cameras, RTK-GPS, and a wheel encoder (**Figure 2**). The MSICS contains two hoods, each with controlled lighting specifically designed for the sensors mounted within the respective hood. The Teledyne Dalsa Genie Nano C1920 RGB camera used for this work is mounted in the second hood along with the LiDAR (**Figure 3A**). The underside of the hood is lined with a custom rig of focused high-powered RGB LED lights that illuminate the ground directly below the camera (**Figure 3B**). During these experiments the RGB camera was setup to acquire 14 frames per second. During data-capture the MSICS was driven at speeds ranging from 0.14 ms^{-1} to 0.37 ms^{-1}.

Training Data

The data used for training our neural network was collected at location's A and B in **Figure 1**. RGB data was acquired while the MSICS was driven at full speed (approx. 0.4 ms^{-1}) randomly across the mixed sward in each location. Note that the MSICS only had lighting installed on one side of the hood when data was collected at location B. The result of this is that approximately half of the data has deeper shadows than the other half. We retained this data in the final training set to improve the network's robustness to variable lighting conditions.

Validation Trial

A mixed sward trial at site C (**Figure 1**) containing ryegrass and clover was used for comparing visual clover fraction estimates to harvest measurements. The clover used in this trial are half white sib white clover (*Trifolium repens L.*) breeding lines along with two commercial checks, planted in October 2017. The grass cultivar is Rohan perennial ryegrass, which was sown in April 2017. Efforts have been made to minimise weed presence in the trial. Irrigation has not been used. The trial is row-column layout, with two replicates of 240 plots (a total of 480 plots). Each plot is approx 1.5 m by 1.5 m in size. The trial is scored when it reaches 2,800 kg/ha– 3,200 kg/ha (approximately 8–10 times per year), and then cattle grazed. The whole trial is scored for growth (total clover biomass) then a selection of plots, which are a fair representative of the whole trial are selected for harvest measurements. This equates to 20%–30% of the trial plots harvested, or approximately three representatives of each score per replicate. The trial was between 12 and 18 months old when the data for this paper was collected.

RGB data was collected for a subset of the harvested plots during two separate scoring events. First, a total of 30 measurements were taken on the 22nd January 2019, and then another 40 measurements were conducted on the 15th April 2019. Of these, five were taken at the edge of the trial to increase the amount of data collected with low clover fraction. Measured

FIGURE 1 | All data was collected at the AgResearch farm in Lincoln, New Zealand (Lat:-43.627799, Long:172.470908). Data used for training the neural network was collected from sites A and B. An independent mixed sward trial at site C was used for validating against harvest measurements. Satellite image retrieved from Google Maps (16 April 2019) and follows attribution guidelines for redistribution.

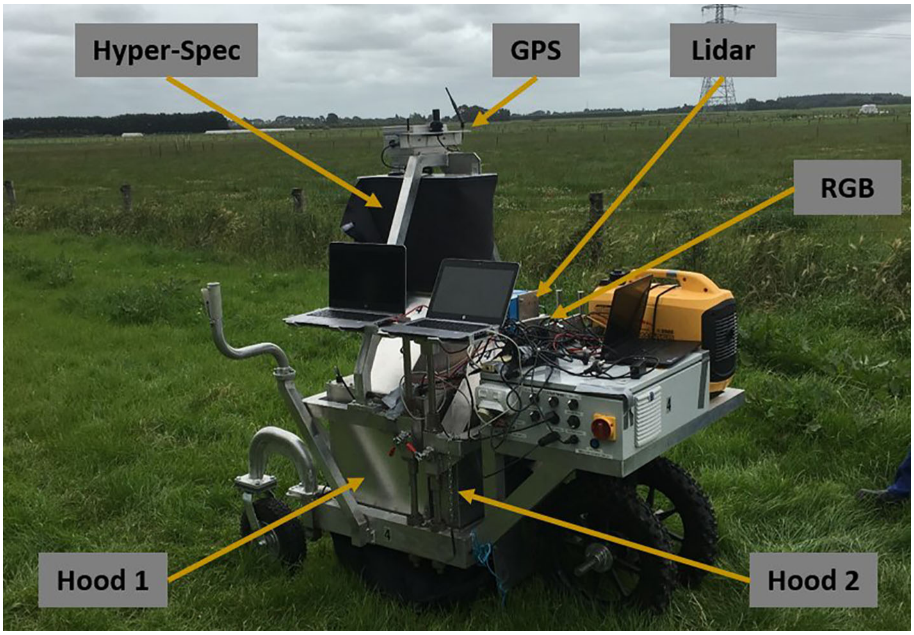


FIGURE 2 | The multiple scanning imaging capacity system (MSICS) used for high throughput phenomic screening of forage. Hood 1 contains SWIR and VNIR hyperspectral cameras. Hood 2 contains light detection and ranging (LiDAR) and red, green, and blue (RGB) sensors. Each hood contains its own custom lighting setup. A black skirt is installed around the base of the hoods to block out ambient light. An encoder is installed on the back wheel for measuring distance. The MSICS also incorporates real-time kinematic global positioning system (RTK-GPS) technology to enable geo-referencing.

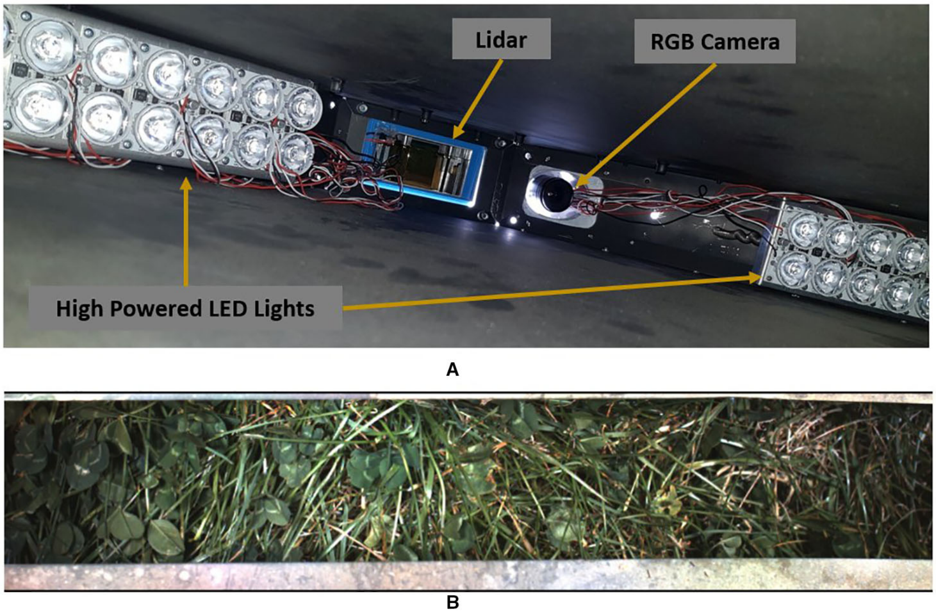


FIGURE 3 | (A) Interior of Hood 2 on the multiple scanning imaging capacity system (MSICS) showing the custom lighting setup alongside the LiDAR an red, green, and blue (RGB) camera. **(B)** A typical image captured from the RGB camera during operation.

plots were chosen specifically to have a wide range of biomass and proportion of clover/ryegrass.

For each harvested plot, a 25 cm × 25 cm quadrat was placed on the plot in a position that visually corresponded to the plot's growth score (**Figure 6**). After RGB data was collected, the mixed material inside each quadrat was harvested using electric shears and packaged. Following this, clover and grass were separated from each sample before being dried and weighed.

Training Dataset

To train convolutional neural networks, thousands of labelled images are typically needed. This problem is exacerbated for image segmentation where a label is required for every pixel. Although techniques for semisupervised and unsupervised learning do exist, their success varies between different problems. Therefore, we decided to construct our data set using two different labelling approaches. The majority of the data set is built from synthetic data that has been constructed using a variant of the synthetic image generation (SIG) method used by (Skovsen et al., 2017). Here samples of individual plants and leaves are cropped from RGB images captured by the MSICS, then randomly overlapped to create new images where the label is already known. Augmentations such as scaling, flips, and gamma transformations are also applied to these samples to further increase diversity of images in the dataset. This process enables the synthesis of thousands of unique labelled images from a relatively few number of samples. We have also included a number of partially labelled images to provide some examples that are more realistic and target configurations that are not easily simulated—such as clusters of small gaps in the pasture canopy, object boundaries, and complex shadowing effects. Each partially labelled image was augmented with vertical and horizontal flips to quadruple the amount of partially labelled data.

Our final training data set is made of 4,500 training images and 600 testing images as shown in **Table 1**. Samples used for generating synthetic data for the training and testing data sets were kept separate to avoid cross-contamination. Each image in this dataset have both a height and width of 100 pixels. We also experimented with larger image sizes, however the 100×100 size provided a good balance between training speed and image complexity.

Synthetic Image Generation

The SIG process we followed was similar to that used by (Skovsen et al., 2017). Each synthetic image contains a random background (soil image), and between 2 and 20 random samples that are augmented and overlaid in random positions. The selection probability for clover and grass samples was subjectively adjusted to account for the average sizes of each to ensure a balanced distribution of ratio and sparseness. Sample augmentations included: horizontal and vertical flips; scale ±25%; gamma ±10%; and saturation ±25%. The ground sample distance for our data is approximately 2–3 pixels per mm. As such we have omitted rotation augmentations of our samples as it disrupted the texture information at this resolution. We also apply Gaussian drop shadows to approximately half of our synthetic images to darken underlying soil and leaves. Instead of simulating shadows that are realistic for our data-capture system, we vary the extent and intensity of the Gaussian drop shadows in order to improve our networks robustness to different lighting conditions. Examples of our synthetic data are shown in **Figure 4**. Unlike Skovsen et al. who focused on making photo-realistic synthetic images, we have embraced the inelegance of simple stitching as a further form of data lighting augmentation. We also identify the boundary pixels in each sample so that they can be ignored during network training to reduce influence of edge artefacts.

TABLE 1 | Composition of the training and testing data sets.

	Clover	Ryegrass	Background	Synthetic	Partial	Total
Training Set	308	230	54	2700	450 (x4)	4500
Testing Set	50	50	50	400	50 (x4)	600

Each set contained independent clover, ryegrass, and background samples. The amount of partially labelled data was quadrupled using flip augmentations.

Local Context Network

LC-Net follows an encoder-decoder structure (**Figure 5**), taking design inspiration from a number of recent segmentation networks. The input image is encoded using the first five blocks of VGG16 (Simonyan and Zisserman, 2015). We have also followed standard practice of removing the max-pooling layer from final block. The decoder part of the network has two branches. The primary branch contains an FC block (equivalent role to the Fully Connected layers of VGG16) and three Decoder blocks which have skip connections to the corresponding encoding blocks. The second branch takes the features from the last VGG16 block and feeds them through a custom pyramid pooling module we have designed, which we have named local context pyramid pooling (LCPP). The output of the LCPP module is then resized to be the same size as the output of the primary branch (both in spatial and feature dimensions) and concatenated to it. This is then put through mixing convolution layers for combining the information from each branch before the logits are calculated. At this point the output has a stride of 2 and the up-sampling is completed with bilinear interpolation. Final predictions are then made by applying softmax activation to this output.

FC Block

This consists of three 3×3 convolutions then a 1x1 convolution. Each convolution is followed by ReLU activation then batch normalisation. We choose the number of output channels for these convolutions to be 1,024 as we generally find VGG16's original 4,096 feature length to be excessive for problems targeting a small number of classes.

Decoder Block

The output from the previous layer is bilinearly resized to match the output size of the corresponding VGG16 block output. These are then concatenated and passed through two mixing convolution layers. There are different numbers of channels for each input into the concatenation—this causes the less developed features in the earlier VGG16 blocks to have less influence on the semantic information in the final output.

Mixing Convolution

These follow the same design principle shown by the DeepLabV3 + decoder; i.e. it is beneficial to follow concatenation operations with several convolutions to properly combine the features. We use a 3×3 convolution followed by ReLU activation then batch normalisation for this. The number of output channels is fixed to 128 for all mixing convolutions in this network.

LCPP Module

The input to this module is sent separately through two average pooling layers with kernel sizes 3×3 and 5×5. "Same" padding is used to maintain spatial shape. The outputs of these pooling layers are then concatenated before being sent through two 1×1

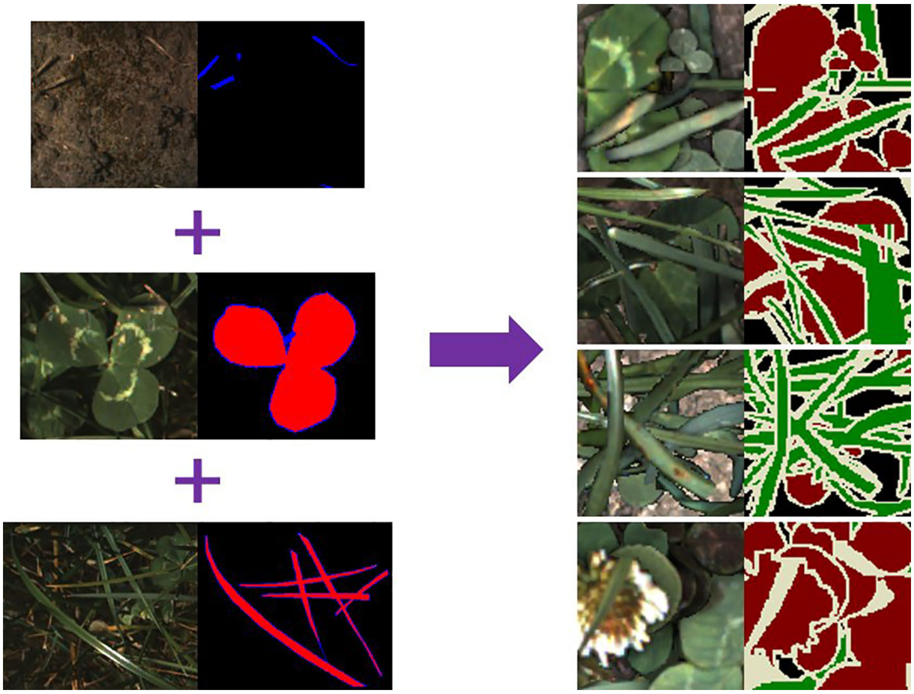


FIGURE 4 | Example of synthetic image generation. Plant samples use the labelling convention: red = object of interest; blue = pixels to include but ignored during training—such as edges or ambiguous pixels; black = excluded from sample. The synthetically generated labels (right) use a convention based on the (Everingham et al., 2012) colour scheme: red = clover; green = grass; black = background; white = pixels marked to be ignored during training.

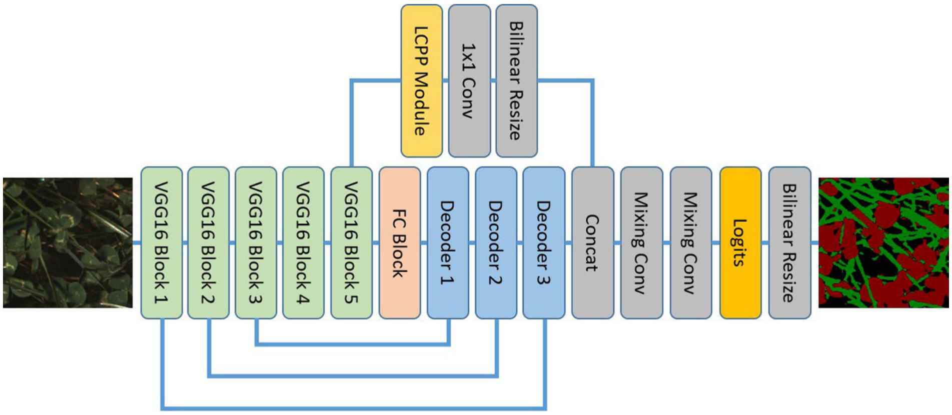


FIGURE 5 | The local context network architecture using a VGG16 backbone and local context pyramid pooling.

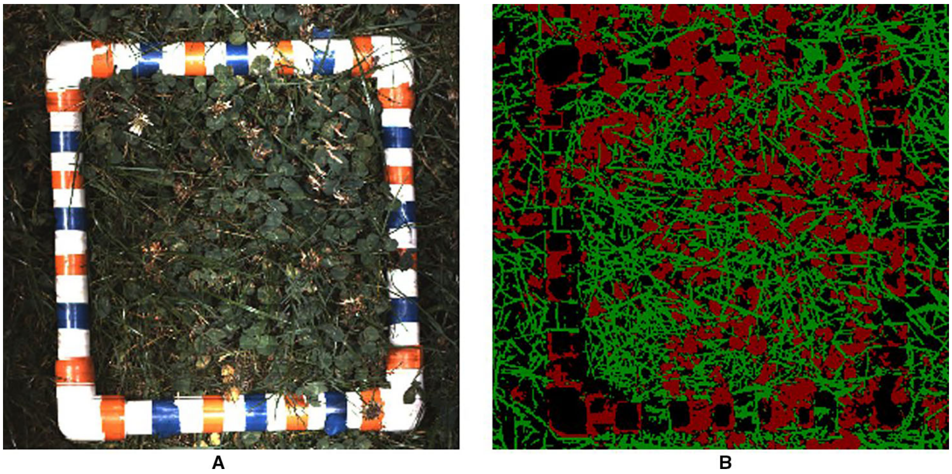


FIGURE 6 | **(A)** visual and **(B)** segmentation images for one of the validation trial plots after rasterising. Each image has a pixel size of 1.2 × 1.2 mm^2. The harvest frame has ambiguous segmentation as the convolutional neural network (CNN) has not been trained to recognise it.

convolution + ReLU activation + batch normalisation blocks. The length of the features returned by these convolutions are the same as the module input.

It has been demonstrated by other networks (Chen et al., 2018) that predictions can be improved by incorporating the global context vector (the image described by single rich feature vector with zero spatial extent) into the final decision process. However, it would be ambiguous to apply this with forage images of different sizes and scales (i.e., visually forage is relatively densely packed and continuous across the extent of the image). The LCPP module is designed to play the role of the global context vector in this situation. It does this by incorporating multiple context vectors that are local to large regions of the image at different scales. The dependencies on image size and scale are limited as a result.

Network Training

LC-Net was trained to classify pixels to three classes—Clover, Ryegrass, and Background. The background class includes regions containing soil and exceptionally dark shadows. Pixels of other pastoral species are explicitly ignored during training—therefore will have indeterminate classification during inference. ImageNet initialisation was used for the VGG16 layers (Tensorflow, 2016). The other weights were initialised using standard Xavier initialisation. The network was trained for 400 epochs with a batch size of 24, using Adam optimisation, partial categorical cross-entropy loss and a learning rate of 5e-5. Training LC-Net with the above data set took 3.5 h on a desktop PC using a GTX 1080Ti GPU and an Intel i7-7700K CPU.

Data Postprocessing

Segmentation networks provide a prediction score for each class in every pixel which is used to determine the classification for that pixel. After processing each image with the network, we defined any pixel with a prediction confidence less than 80% to be background (i.e., uncertain predictions). All other pixels were defined as as either clover or ryegrass depending on which the had the highest prediction confidence for the respective pixel.

When processing the data for the validation trial the harvest frame did not fit in the camera field of view in the driving direction. Furthermore, the height of the hood was adjustable—meaning that the camera footprint (i.e., pixels/mm) needed to be calculated dynamically. A rasterising-like approach to stitching the RGB data was taken as per below.

First cross-correlation between adjacent images was performed to determine the number of overlapped pixels. The overlap strip was then divided into 22 strips, taking the central row of pixels as the representative for each strip. This was then followed by using the wheel encoder information to determine the distance travelled in this overlap region, and therefore the relative position of each strip. The pixels/mm along each strip is determined from information obtained by the LiDAR unit on the MSICS. These samples were then averaged in a raster grid with a pixel size of 1.2 × 1.2 mm. An example of this rasterising process applied to both an RGB image and its associated segmentation mask is shown in **Figure 6**. The rasterised RGB image was not processed with LC-Net. Instead, LC-Net is used to create segmentation masks for the raw RGB images. The same rasterising parameters calculated for stitching the raw RGB images is also used to stitch the associated segmentation masks.

RESULTS

Network Training

In addition to LC-Net, two more networks were trained for comparison (**Table 2** and **Figure 7**). We used two standard metrics to assess the quality of each network—mean pixel

TABLE 2 | Comparison of networks trained for clover-ryegrass segmentation.

Method	mAcc	mIoU	mAcc (CV)	mIoU (CV)
FCN (Skovsen et al., 2017)	83.4	65.5	–	–
FCN	92.7	74.8	92.0	73.4
LC-Net (without LCCP)	93.7	79.1	93.4	77.3
LC-Net	**95.4**	**81.3**	**95.0**	**79.0**

The mean pixel accuracy (mAcc) and mean intersection over union (mIoU) metrics are class-weighted averages. The first two columns are results for our original dataset configuration. The last two columns are averaged results from a stratified fivefold cross-validation (CV).

accuracy and mean IoU. Both of these metrics are class weighted averages. Due to the non-linear behaviour of these metrics the mean IoU was the more informative one during training after the pixel accuracy exceeded 90%.

FCN was trained for one of the comparison networks. Although data from (Skovsen et al., 2017) consists of higher resolution images, the FCN model we trained in our study significantly higher accuracy —with an IoU close to that reported for benchmark datasets (Shelhamer et al., 2016). As illustrated in **Figure 7**, FCN lacks the ability to properly delineate fine filament like structures or corners, which is fundamental limitation of this architecture.

To show the benefits provided by the LCPP module, the LC-Net architecture both with and without the LCPP branch. Despite the LCPP module only improving the mean IoU by 2.2%, there is a significant boost in performance as presented in **Figure 7**. Clover in the segmentation are more filled in; and prominent gaps in the grass from the network without LCPP are filled in when LCPP is included. Overall, the full LC-Net outperformed all other networks tested for this application.

To test the robustness of our training process and dataset we applied a stratified 5-fold cross validation to each of the tested networks (**Figure 8**). We split each sub dataset (17/07/2018 samples, 05/12/2018 samples, and partially labelled) independently and aggregated the respective folds. This was so the samples with different lighting conditions would not be mixed in the synthetic data generation process. The accuracy and mIoU for these cross validation runs are on average a few percent lower than obtained from the original dataset configuration (**Table 2**). However, the overall network ranking remains unchanged. Note

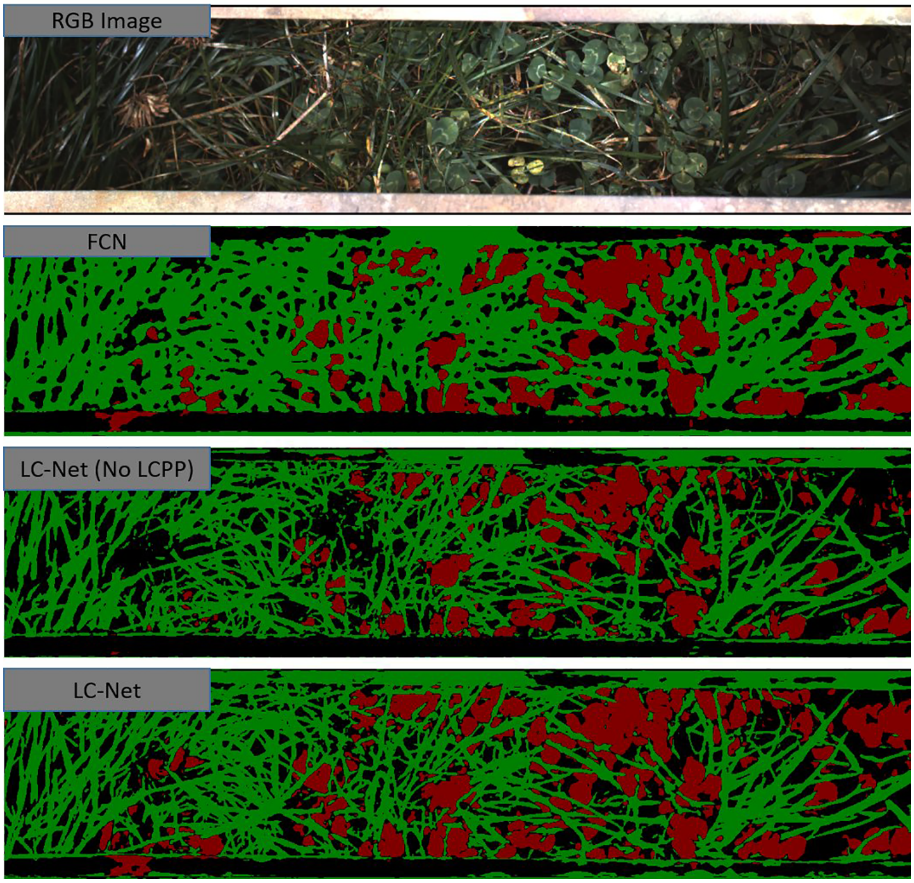


FIGURE 7 | A red, green, and blue (RGB) image that has been captured by the multiple scanning imaging capacity system (MSICS) at site C, and processed using each of the trained networks—fully convolutional network for semantic segmentation (FCN), local context network (LC-Net) [without the local context pyramid pooling (LCPP) module], and the full LC-Net. The segmentation masks show clover (red), grass (green), and background (black). The background class includes soil, very dark shadows, and pixels that have less than an 80% confidence score in their classification.

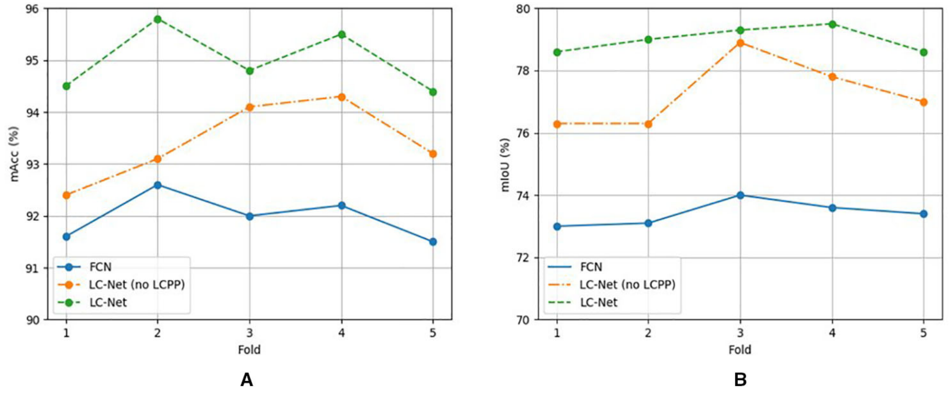


FIGURE 8 | Mean accuracy **(A)** and mean intersection over union **(B)** results for each fold for models trained using stratified fivefold cross-validation. Results are relatively consistent between folds for each model.

the original dataset configuration used approximately 11% more samples in the training set than the cross validation. This indicates that further improvements may still be gained through increasing the number of samples in the dataset.

Sward Composition

To assess the improvement in estimating sward composition using LC-Net over FCN we used 70 data points from our validation trial. A small number of data points were excluded from this analysis due to insufficient data for to completing the rasterising process. The fraction of clover estimated from the RGB images were taken as the ratio of clover pixels to clover and ryegrass pixels as identified by the respective networks. The clover fraction for the harvest measurements is obtained from the equivalent calculation using dry weights instead of pixels.

As shown in **Figure 9**, the clover fraction correlation obtained from LC-Net (R^2 = 0.825) is only marginally better than what was obtained for FCN (R^2 = 0.793). Overall this similarity is not surprising since they provide approximations to the same coverage areas. Data points with similar forage density are also spread relatively evenly throughout the scatter cloud, which suggests that a significant component of the variation is due to occlusion of underlying vegetation by the top layer of pasture.

DISCUSSION

Deep Learning methods have a tendency to produce results that appear much better compared to traditional methods, especially on classification type problems. However, since the field of deep

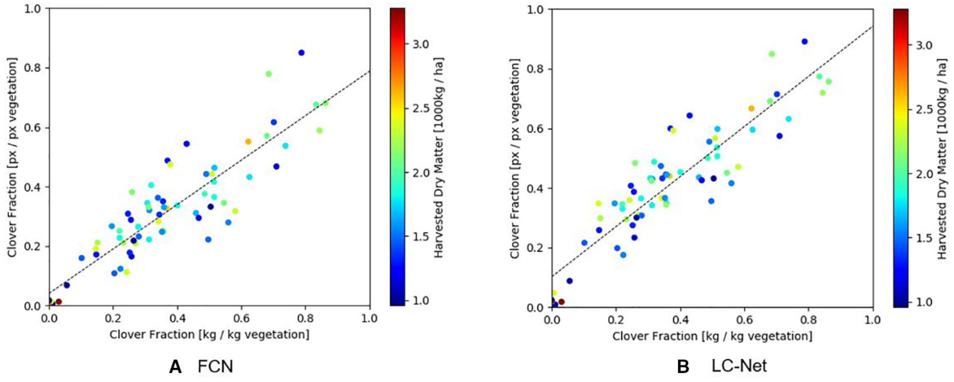


FIGURE 9 | Comparison of the clover-vegetation ratio obtained from RGB images and harvested dry matter measurements. **(A)** Red, green, and blue (RGB) results from the fully convolutional network for semantic segmentation (FCN) network has a linear correlation of R^2 = 0.793 with harvested dry matter. **(B)** RGB results from the local context network (LC-Net) and has a linear correlation of R^2 = 0.825 with harvested dry matter.

learning is advancing at a very high rate it is challenging for other fields to uptake state-of-the-art techniques into new applications. The first pre-print of the FCN network was published two years before it was applied to segmenting ryegrass from clover (Shelhamer et al., 2014; Skovsen et al., 2017). During this time dozens of new segmentation networks were published that significantly outperformed FCN (Garcia-Garcia et al., 2017). Also, most CNN architectures have been developed on standard computer vision benchmark data sets, many of which are focused on urban settings. Agricultural applications can offer challenges that are either not present or less prominent than those found in these benchmarks. For example, the object detection networks such as Faster R-CNN and YOLO are now commonly applied to fruit and weed detection. However, these networks have severe limitations around how many objects they can detect which can cause problems when looking at a tree with hundreds of targets—requiring them to be applied in innovative ways. Based on our experience in this research, global context information for proximal pasture images contain sufficient ambiguity to make it challenging to robustly incorporate it into networks for this application. During the development of LC-Net we trialled DeepLabV3+ (a state-of-the-art network which utilises global context) on this problem. We found that it required both a fixed ground sample distance and a consistent input image size between both the training and deployment versions of the model. This restriction is not practical as it is too time consuming and costly to train and maintain a different model for every potential hardware configuration we may need to use. The LCPP module is a pragmatic alternative to global context—providing similar benefit without being restrictive on the development of our data-collection platform.

For this work we did not restricted ourselves to using predefined CNN architectures, instead we developed our own specifically for the application of canopy segmentation in an agricultural context. In this paper we have applied used this new architecture (LC-Net) for analysing pasture. The design of LC-Net is influenced by a number of different networks. Although we don't use separable convolutions, the decoder modules and mixing convolutions are derived from the structures used for DeepLabV3+. The network also trained more robustly when we used skip-connections during up-sampling rather than atrous convolutions—although it is not clear whether this observation is specific to our application. We also used a pyramid pooling module of our own design (LCPP) to incorporate local context information at different scales instead using a single global context. By including a context layer in our network, we effectively force it to make a high level visual score and incorporate that into the decision process. Another advantage of LCPP is that it uses pre-defined pooling operations, therefore requires significantly fewer trainable parameters than other pyramid pooling techniques. The description for LCPP we have given is very specific, however it follows a generic formula. It can be adapted by changing the number of pooling layers along with different kernel sizes and/or changing the number of convolutions and feature sizes following the concatenation. The design of this module will almost certainly need tuning for specific applications/data depending on factors such as relative object size. This aspect is being investigated further. Our experience working with CNNs is that those built upon VGG16 are easier to train for canopy assessment compared to deeper networks such as Xception (Chollet, 2017) and ResNet (Kaiming et al., 2016). This is the primary reason for choosing VGG16 as the backbone of the first iteration of our LC-Net. In addition to pasture, we are also currently investigating extensions of LC-Net for segmentation of orchard and vineyard canopy applications, in addition to how well it performs on the standard benchmark data sets.

We showed that LC-Net significantly out-performs FCN for segmenting clover from ryegrass in mixed swards. Comparing our FCN results and those of (Skovsen et al., 2017) two major differences are noticeable. First, our images have half the ground sample distance of theirs. Given that FCN has an output stride of 8, the edges of objects in its predictions should be more uncertain with our data. Second, (Skovsen et al., 2017) included a class for predicting weeds whereas we have not. They used very few examples of weeds in their training data which negatively impacted their reported accuracy. It is likely their FCN model would have an accuracy similar to ours if either more weed examples were added to their data set—or if the weed was class removed entirely. Over the following season we are planning to incorporate other plants into our dataset so we can also adapt our network to be effective in weedy conditions.

We are planning to publish the dataset associated with this work separately. Before doing so however, there are several additions and improvements that we believe would add significant value—which include labelling data from different seasons, and inclusion of additional plants/weeds. It is worth noting that the dataset we have compiled for this study is highly specific to the camera and lighting setup we have used. For example, a few minor tests we performed demonstrated that our network is not generalised for natural lighting conditions (a drop in mIoU of roughly 30 – 40%). As such, networks trained by our dataset will likely perform poorly on data collected from setups substantially different from our own.

Comparing LC-Net and FCN models by the correlation between visual and harvested clover ratios, we see little difference in the results. Work by (McRoberts et al., 2016) and (Rayburn, 2014) using more traditional methods showed that it is also possible to obtain reasonable estimation through classification with coarse super-pixels and sparse subsampling. This indicates that significant boosts in pixel-level segmentation accuracy only translates to small (sometimes negligible) improvements in clover-vegetation fraction estimation with RGB imaging. There are two situations we identify that could benefit from higher segmentation accuracy: (1) when combining visual information with that of other sensors such as LiDAR to which can potentially compensate for the lack of plant density, volume, and occlusion information; and (2) when assessing pasture for more species than just clover and ryegrass. We are currently investigating both of these areas. Another motivation for using deep learning models is that they are in many cases simpler to work with, maintain, and extend when compared to more traditional feature crafting methods.

An observation of significance we made during training each of these networks was that training and testing losses could not be used to monitor for overfitting. After approximately 20–30 epochs in every training run the testing loss starts to increase (a typical sign of the network memorising the training set), however the accuracy and IoU metrics continued to improve regularly. This increase in validation loss seemed to be focused around the edges of the leaves in the images. Due to the mesh-like structure of pasture there is an unusually high proportion of class boundary pixels to interior pixels (%56 in our training data) that therefore provide a significant contribution to the loss. The synthetic images in our data set has been constructed in a manner that provides little information to the network about what true edges look like. This peculiarity may be resolved through improvements to how the training set is labelled, however the cost is likely higher than the benefit from doing so. The networks trained appear to work well in practice despite the elevated loss, and the accuracy metrics did not indicate overfitting nor appear to be significantly affected by this.

Overall, our results for assessing clover fraction using convolutional neural networks are comparable to those obtained by (Skovsen et al., 2017). We have demonstrated that relationship between visual and harvested clover fraction is reproducible using different networks, lighting setup, camera, image resolution, less photogenic synthetic data, and different postprocessing procedures.

CONCLUSIONS

A new CNN architecture (LC-Net) designed for segmentation of agricultural canopies is showing promise for component identification in mixed sward. This architecture can segment clover and ryegrass in RGB images with higher accuracy than any other methods publicly available for the same application. We have also achieved this with half of the image resolution (pixels/mm) used by the next best method. Our comparisons between visual and harvested dry matter clover-vegetation ratios indicate that these improvements in segmentation accuracy do not yield similar improvements to biomass estimation. However, we predict that refined segmentation is necessary for improving biomass predictions when it is combined with information from other sensors.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

CB was the primary developer of LC-Net along with writing the associated code, performing the analysis, and writing the paper. JF and JH aided in the development of LC-Net through discussions around critical design choices and software debugging. JH also contributed to the literature review. Data collection and preprocessing was performed by CB, KI, AHe, and AHi. KI and MH developed the RGB image capture software used, the LED rig, and configured the setup. BJ installed the RGB hardware into the MSICS and was involved in the integration design decisions. SG oversaw and managed MSICS hardware changes and developments. KG guided the style of the manuscript writing and managed and directed the overall programme of research. All authors contributed to review and editing of the paper.

FUNDING

This work was supported by Pastoral Genomics, a joint venture co-funded by DairyNZ, Beef+Lamb New Zealand, Dairy Australia, AgResearch Ltd, New Zealand Agriseeds Ltd, Grasslands Innovation Ltd, and the Ministry of Business, Innovation and Employment (New Zealand).

ACKNOWLEDGMENTS

The authors would like to thank Sophie Rebbeck for managing the Lincoln Agritech coordination and contracts with AgResearch colleagues. Sophie Rebbeck and Ian Boddy, for discussions about simple description of LC-Net and AgResearch engineering team involved in the construction of the MSICS machine.

REFERENCES

Araus, J., Kefauver, S., Zaman-Allah, M., Olsen, M., and Cairns, J. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2016). SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *CoRR* abs/1511.00561.

Barrett, B., Favelle, M., Ghamkhar, K., and Carena, M. (2018). Developing new tools for pasture plant. *J. New Z. Grasslands* 80, 225–262.

Bonesmo, H., Kaspersen, K., and Kjersti Bakken, A. (2004). Evaluating an image analysis system for mapping white clover pastures. *Plant Sci.* 54, 76–82. doi: 10.1080/09064710410024462

Capstaff, N., and Miller, A. (2018). Improving the yield and nutritional quality of forage crops. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.00535

Chen, L., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation, in: *Computer Vision - ECCV 2018 - 15th European Conference*, Munich, Germany, September 8-14, 2018. pp. 833–851, *Proceedings, Part VII*. doi: 10.1007/978-3-030-01234-2_49

Chollet, F. (2017).Xception: deep learning with depthwise separable convolutions, in: *IEEE Conference on Computer Vision and Pattern Recognition* doi: 10.1109/CVPR.2017.195

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2012). The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/ind.html.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., and Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.

Gebremedhin, A., Badenhorst, P., Wang, J., Spangenberg, G., and Smith, K. (2019). Prospects for measurement of dry matter yield in forage breeding programs using sensor technologies. *Agronomy* 9. doi: 10.3390/agronomy9020065

Ghamkhar, K., Irie, K., Hagedorn, M., Hsiao, W.-H., Fourie, J., Gebbie, S., et al. (2018). Using LIDAR for Forage Yield Measurement of Perennial Ryegrass (Lolium perenne L.) Field Plots. *Breed. Grasses Protein Crops In Era Genomics*, 203–208. doi: 10.1007/978-3-319-89578-9_37

Himstedt, M., Fricke, T., and Wachendorf, M. (2012). The benefit of color information in digital image analysis for the estimation of legume contribution in legume-grass mixtures. *Crop Sci.* 52, 943–950. doi: 10.2135/cropsci2011.04.0189

Kaiming, H., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, . doi: 10.1109/CVPR.2016.90

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet Classification with Deep Convolutional Neural Networks." in *Advances in Neural Information Processing Systems 25*. Eds. F. Pereira, C. J. C. Burges, L. Bottou and K. Q. Weinberger (Curran Associates, Inc.), 1097–1105.

Lee, J., Matthew, C., Thom, E., and Chapman, D. (2012). Perennial ryegrass breeding in New Zealand: a dairy industry perspective. *Crop Pasture Sci.* 63, 107–127. doi: 10.1071/CP11282

McRoberts, K., Benson, B., Mudrak, E., Parsons, D., and Cherney, D. (2016). Application of local binary patterns in digital images to estimate botanical composition in mixed alfalfa-grass fields. *Comput. Electron. In Agric.* 123, 95–103. doi: 10.1016/j.compag.2016.02.015

Mortensen, A., Karstoft, H., Søegaard, K., Gislum, R., and Jørgensen, R. (2017). Preliminary results of clover and grass coverage and total dry matter estimation in clover-grass crops using image analysis. *J. Imaging* 3. doi: 10.3390/jimaging3040059

Rayburn, E. (2014). Measuring legume content in pastures using digital photographs. *Forage Grazinglands* 12, 1–6. doi: 10.1016/j.compag.2016.02.015

Shelhamer, E., Long, J., and Darrell, T. (2014). Fully convolutional networks for semantic segmentation Https://arxiv.org/abs/1411.4038v1, (pre-print).

Shelhamer, E., Long, J., and Darrell, T. (2016). Fully convolutional networks for semantic segmentation. *CoRR* abs/1605.06211.

Simonyan, K., and Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition, in: *3rd International Conference on Learning Representations, ICLR 2015*, San Diego, CA, USA, May 7-9, 2015, *Conference Track Proceedings*.

Skovsen, S., Dyrmann, M., Mortensen, A., Steen, K., Green, O., Eriksen, J., et al. (2017). Estimation of the botanical composition of clover-grass leys from RGB images using data simulation and fully convolutional neural networks. *Sensors (Switzerland)* 17. doi: 10.3390/s17122930

Smith, K., and Spangenberg, G. (2014). Forage breeding for changing environments and production systems: an overview. *Crop Pasture Sci.* 65, i–ii. doi: 10.1071/CPv65n11_FO

Tensorflow (2016). http://download.tensorflow.org/models/vgg_16_2016_08_28.tar.gz.

Walter, A., Studer, B., and Kölliker, R. (2012). Advanced phenotyping offers opportunities for improved breeding of forage and turf species. *Ann. Bot.* 110, 1271–1279. doi: 10.1093/aob/mcs026

Zhang, H., Dana, K., Shi, J., Zhang, Z., Wang, X., Tyagi, A., et al. (2018). Context Encoding for Semantic Segmentation 7151–7160. doi: 10.1109/CVPR.2018.00747

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network 6230–6239. doi: 10.1109/CVPR.2017.660

Conflict of Interest: The authors declare that this study received funding from Pastoral Genomics—a joint venture co-funded by DairyNZ, Beef+Lamb New Zealand, Dairy Australia, AgResearch Ltd, New Zealand Agriseeds Ltd, Grasslands Innovation Ltd, and the Ministry of Business, Innovation and Employment (New Zealand). Pastoral Genomics has interests in developing technology that has potential to be commercialised for adding value to the forage breeding industry, and have involvement in setting the overall goals and vision of the Pastoral Genomics research programme. This relationship did not influence study design, data collection and analysis, decision to publish, or preparation of the manuscript. Lincoln Agritech Ltd. and Red Fern Solutions are independent research organisations that were subcontracted to conduct part of this research. Development of the LC-Net architecture was internally funded by Lincoln Agritech Limited. All other authors declare no competing interests.

The handling editor is currently organizing a Research Topic with one of the authors KG and confirms the absence of any other collaboration.

frontiers
in Plant Science

METHODS
published: 28 February 2020
doi: 10.3389/fpls.2020.00141

Doing More With Less: A Multitask Deep Learning Approach in Plant Phenotyping

Andrei Dobrescu[1]*, *Mario Valerio Giuffrida*[2] and *Sotirios A. Tsaftaris*[1]

[1] IDCOM, University of Edinburgh, Edinburgh, United Kingdom, [2] School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

Image-based plant phenotyping has been steadily growing and this has steeply increased the need for more efficient image analysis techniques capable of evaluating multiple plant traits. Deep learning has shown its potential in a multitude of visual tasks in plant phenotyping, such as segmentation and counting. Here, we show how different phenotyping traits can be extracted simultaneously from plant images, using multitask learning (MTL). MTL leverages information contained in the training images of related tasks to improve overall generalization and learns models with fewer labels. We present a multitask deep learning framework for plant phenotyping, able to infer three traits simultaneously: (i) leaf count, (ii) projected leaf area (PLA), and (iii) genotype classification. We adopted a modified pretrained ResNet50 as a feature extractor, trained end-to-end to predict multiple traits. We also leverage MTL to show that through learning from more easily obtainable annotations (such as PLA and genotype) we can predict a better leaf count (harder to obtain annotation). We evaluate our findings on several publicly available datasets of top-view images of *Arabidopsis thaliana*. Experimental results show that the proposed MTL method improves the leaf count mean squared error (MSE) by more than 40%, compared to a single task network on the same dataset. We also show that our MTL framework can be trained with up to 75% fewer leaf count annotations without significantly impacting performance, whereas a single task model shows a steady decline when fewer annotations are available. Code available at https://github.com/andobrescu/Multi_task_plant_phenotyping.

Keywords: plant phenotyping, deep learning, multitask, leaf count, PLA, genotype

Edited by:
Tony Pridmore,
University of Nottingham,
United Kingdom

Reviewed by:
Asa Ben-Hur,
Colorado State University,
United States
Elizabeth Van Volkenburgh,
University of Washington,
United States

****Correspondence:***
Andrei Dobrescu
a.dobrescu@ed.ac.uk

Specialty section:
This article was submitted to Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: *06 June 2019*
Accepted: *29 January 2020*
Published: *28 February 2020*

Citation:
Dobrescu A, Giuffrida MV and Tsaftaris SA (2020) Doing More With Less: A Multitask Deep Learning Approach in Plant Phenotyping.
Front. Plant Sci. 11:141.
doi: 10.3389/fpls.2020.00141

INTRODUCTION

Nondestructive, image-based plant phenotyping is a growing trend in how scientists and breeders engage in plant characterization. Due to the advances in image acquisition systems (Qiu et al., 2018) and development of affordable hardware and software framework (Dobrescu et al., 2017b; Minervini et al., 2017), high throughput plant image capture is becoming widespread. In particular, machine learning has shown that it can be applied effectively in processing vasts amounts of data, including in plant phenotyping problems (Scharr et al., 2016). For example, segmenting whole plants (Minervini et al., 2014; Aich and Stavness, 2017), or each individual leaf (Romera-Paredes and

Torr, 2016; Ren and Zemel, 2017; Ward et al., 2018), synthetic image synthesis (Giuffrida et al., 2017; Zhu et al., 2018), and leaf counting (Aich and Stavness, 2017; Dobrescu et al., 2017a; Giuffrida et al., 2015; Pape and Klukas, 2015; Giuffrida et al., 2018b; Itzhaky et al., 2018) are all phenotyping tasks that have been recently addressed using machine learning and deep learning, technologies that are becoming more common in the plant-research community. In fact, the fourth edition of the *Computer Vision Problems in Plant Phenotyping*[1] workshop (CVPPP 2019) shows an increasing interest from people inside and outside the plant phenotyping community to invest efforts to develop newer machine learning based approaches.

Leaf count has been an area of interest for plant phenotyping, as it is related to developmental stages (Boyes et al., 2001) and can be an indicator for yield potential (Ngouajio et al., 1999) and plant health (Rahnemoonfar and Sheppard, 2017). Two have been proposed to infer leaf count: (i) determining the leaf count as a subproduct of per-leaf segmentation; or (ii) tackling the problem as a holistic regression task. Several different algorithms have been proposed for a per-leaf segmentation approach. Scharr et al. (2016) discusses four methods to achieve per-leaf segmentation, where machine learning was used for the first time for this task. Romera-Paredes and Torr (2016); Ren and Zemel (2017); Ward et al. (2018), and Zhu et al. (2018) have proposed several deep learning approaches for per-leaf segmentation, obtaining remarkable results in terms of segmentation accuracy. However, the main issue with such methods is that they require per-leaf segmentations to train the algorithms that are often time-consuming, laborious, and expensive to acquire. Although Minervini et al. (2015; 2017) have proposed semiautomatic graphical tools, they still require experienced users to obtain an adequate per-leaf segmentation. Another type of annotation used for leaf counting is to mark each leaf with a dot on the center, rather than the whole leaf segmentation. Although it is an easier way to provide topological and localisation information, it still requires a human to click on the center of each leaf. Itzhaky et al. (2018) use such annotation to train a leaf detector which is used in conjunction with a leaf regressor (named D+R) to achieve state-of-the-art leaf count.

Alternatively, leaf counting can be addressed as a holistic regression task, where an algorithm predicts the total leaf count in an image. In this context, the machine learning algorithm requires just the total number of leaves, which is an easier annotation to obtain, compared to the per-leaf segmentations (Minervini et al., 2015; Giuffrida et al., 2018a). The first studies to use machine learning techniques reported encouraging results (Giuffrida et al., 2015; Pape and Klukas, 2015), although more recently approaches based on deep neural networks have become the state of the art. Dobrescu et al. (2017a) proposed a deep neural network based on a ResNet50 (He et al., 2016), where leaf counting was learned by agglomerating data from multiple sources. Further to this, Giuffrida et al. (2018b) proposed a versatile network that demonstrated that leaf counting could be better learned using data from multiple imaging modalities using a single architecture. Itzhaky et al. (2018) also describe another approach (named MSR) which uses a feature pyramid network architecture (Lin et al., 2017) to learn a direct regressor at multiple scale levels of a plant and then fuse them to output a single leaf count prediction. Ubbens and Stavness (2017) proposed several specialised deep network architectures to count leaves in different datasets, as well as to infer other tasks such as projected leaf area (PLA) and genotype prediction.

The success of machine learning, and especially deep learning, is attributed to the ability to relate images to a given task. Deep neural networks extract meaningful information from images (typically referred to as *image features*), even when they contain complex structures like plants. In the current paradigm, many machine learning models are specialised to perform a single task (i.e., learn one plant trait at a time).

However, plant phenotyping traits, such as the total leaf count, can often be related to other traits, such as the total leaf area, age, and genotype. Incorporating such related traits in the deep learning framework would help the deep neural network better learn all the traits (Caruana, 1997).

Multitask learning (MTL) has been shown to improve the accuracy and the generalization performance of each task (Caruana, 1997). The benefits of MTL are multifold, especially when tasks are related to each other. Firstly, one one network is trained to perform multiple tasks at the same time, in contrast with Ubbens and Stavness (2017), where several networks with different architectures were trained separately to extract phenotyping traits. The learning of multiple tasks enforces the network to learn good representations, thus increasing the generalization capability of the model. Since information sharing is the core of MTL, learning multiple tasks simultaneously reduces overfitting, even in presence of reduced datasets (Baxter, 1997). Additionally, from an implementation perspective, MTL allows having just one shared model instead of independent models per task. This helps reduce storage space, decreases training times and is easier to deploy and maintain. MTL is a special case of transfer learning (Pan and Yang, 2010), where (i) there is no distinction between tasks; and (ii) the objective is to increase performance for all the involved tasks.

Surprisingly, despite the benefits of MTL and its application in several other areas of computer vision (Ramsundar et al., 2015; Kokkinos, 2017; Ranjan et al., 2019), it has been under-explored in addressing problems in plant phenotyping. Pound et al. (2017) proposed the earliest application in MTL for plant phenotyping, where a deep neural network that can both detect and count wheat spikes, as well as classify the presence of awns.

In this paper, we propose an MTL architecture aimed to infer leaf counting, together with the PLA and genotype classification (**Figure 1**). We use the dataset *Ara2013* (Minervini et al., 2017) and show that multiple tasks help to achieve more precise predictions of these three plant traits. The tasks were chosen, as they are relevant and well known plant phenotyping objectives as well as being correlated to each other, which helps the training process. The PLA and genotype annotations are less tedious and time-consuming to gather. The PLA can be obtained with a plant segmentation algorithm (Aich and Stavness, 2017; Dobrescu et al., 2017a;

---

[1] More information available at https://www.plant-phenotyping.org/CVPPP2019

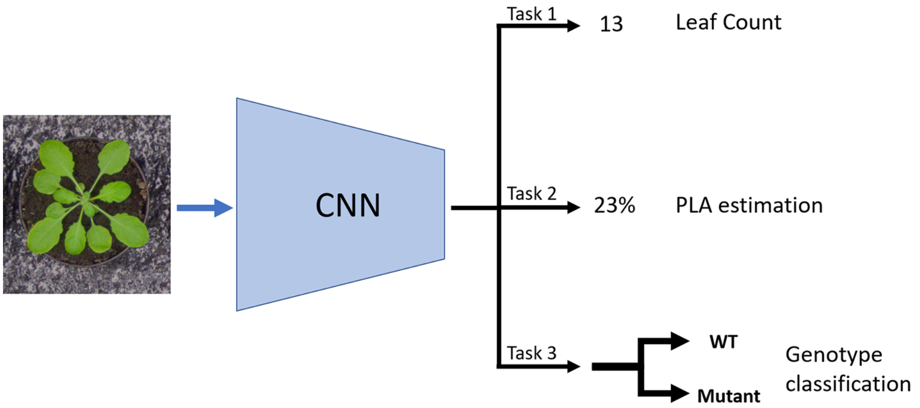


FIGURE 1 | Schematic of the multitask learning (MTL) model: The model takes in an image as input and it uses a CNN to learn three tasks concurrently: Leaf count outputted as a scalar (Task 1); projected leaf area (PLA) estimation as a percentage of area the plant occupies in the image (Task 2); and genotype classification between mutant and wild type (Task 3).

Minervini et al., 2015), whereas the genotype is generally known *a priori* to the scientists. The leaf counting task and PLA estimation are treated as direct regression problems having only the total leaf count and total PLA as respective annotations. The genotyping task is addressed as a binary classification between wild-type and mutant.

The contributions of this paper are multifold:

- Our end-to-end MTL architecture predicts several traits at the same time, in particular leaf counting, PLA, and genotype. Having one unified model for multiple tasks improves performance in leaf count compared to a single task model. Amongt the other tasks, leaf counting is the hardest to predict from a computer vision perspective, due to huge variability between leaves as well as occlusions in the images.
- We show that our proposed method can be trained with fewer leaf count annotations without significantly impacting leaf count performance. Our results show that when annotations for one task are available, performance can be improved by using correlated tasks for the same images.
- We show which count annotations have the most impact on the model performance. Understanding this key aspect would help guide the annotation of a new dataset, highlighting which images should be first annotated in a new dataset.

METHODS

For this study we developed an MTL deep learning model that takes in as input a top-down color (RGB) image of a rosette plant (e.g. *Arabidopsis thaliana*) to infer the total number of leaves, PLA, and binary genotype classification.

The Feature Extractor

The first part of the model (**Figure 2** Top) is a ResNet50 (He et al., 2016) neural network and works as a feature extractor. We used a ResNet50 pretrained on ImageNet (Krizhevsky et al., 2012), as it has been demonstrated to perform well on plant phenotyping tasks (Dobrescu et al., 2017a; Giuffrida et al., 2018b). The architecture of the model is composed of 16 convolutional blocks, each consisting of three convolutional layers of increasing filter sizes to maintain complexity per layer (He et al., 2016). This model is a residual neural network, which means that the convolutional layers are not just stacked on top of each other, but also additional connections between the convolutional blocks (residual connections) are present between neighboring blocks. These skip connections help propagate the error signal faster across these very deep networks layers, yielding improved results over other network designs. We modified the reference ResNet50, by removing the last layer intended for classification and replaced it with a fully connected layer containing 1536 nodes, which acts as a shared representation for the three training tasks. Up to and including the shared representation, we leverage hard parameter sharing, meaning the network layers are shared between all the tasks. This approach reduces the risk of overfitting which is important when training deep learning models.

The Task Branches

The second part of the model (**Figure 2** Bottom) consists of the three task-specific branches that are each responsible for computing one of the tasks. The branches receive information from the shared representation above and specialise on one task. The first one computes the leaf count and it consists of a fully connected layer of 512 nodes and a 1 node layer which outputs the count prediction. The second for estimating the PLA, has the same design as the leaf count branch. The PLA output is normalised as the percentage that the plant occupies in relation to the whole image. Genotype classification is determined by the third branch and contains 3 fully connected layers of 512, 256 and 2 nodes respectively. The activation functions of the fully connected layers in the branches are

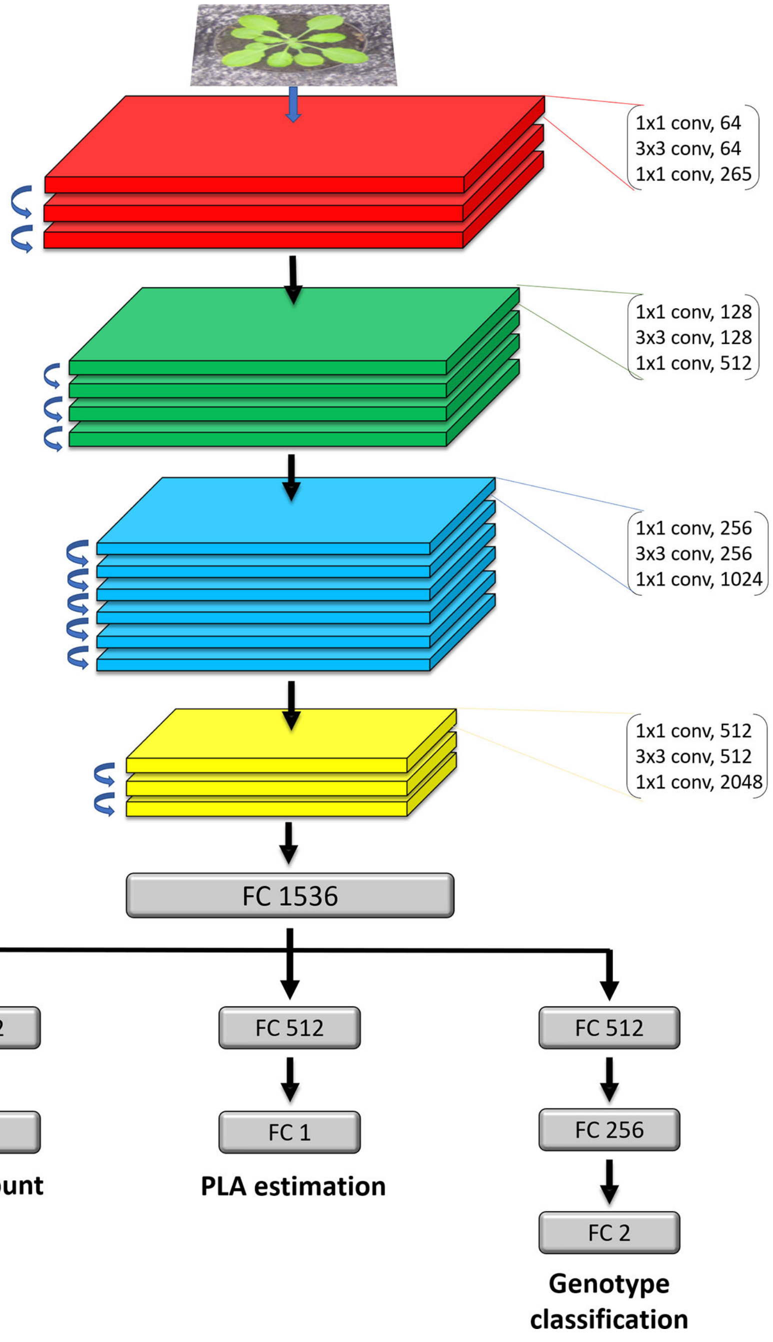


FIGURE 2 | Detailed architecture of the model. The network takes in as input an RGB image of a rosette plant. The main feature extractor is a ResNet50 deep residual network, which is composed of 16 residual blocks which consist of three stacked layers with residual connections between the input and the output of each block. FC layers represent fully connected layers of a given size. The FC1536 is a shared dense representation layer from which each task branch off into their respective output. Each branch is then is specialised for a specific task.

rectified linear units (ReLU), except for the final genotype and PLA prediction layers, which are sigmoid and LeakyReLU respectively. On layers before the final prediction layers for all three tasks we apply an L2 regularization of 0.04 to penalize layer activity during training and prevent overfitting.

Losses

All tasks are learned at the same time in the MTL model. Each task has a specific loss tailored to the specifications. For the leaf counting and PLA estimation tasks the loss is mean squared error (MSE). However, when comparing to the other tasks the values were very low. To balance it out, we multiplied the error values by 10 to maintain comparable values. For the task of genotype prediction the loss is binary cross entropy using a sigmoid final layer activation to get the output between 0 and 1.

Datasets

Three different datasets were used in this study that contain top-down RGB images of *Arabidopsis thaliana* plants. The *Ara2013* (Minervini et al., 2017) dataset consists of 24 separate plants of 5 different genotypes: Col-0 (wild-type), ein2 (Guzman, 1990), ctr (Kieber et al., 1993), adh1 (Perata and Alpi, 1993), pgm (Caspar et al., 1985). Images were captured of each plant twice a day for 26 days. Example images from the dataset can be seen in **Figure 3**. The different genotypes represent a wide range of visual phenotypes when compared to the wild type (Col-0). Ein2 and adh1 are visually similar to the wild-type while the ctr and pgm are more distinct. Two additional datasets part of the CVPPP leaf counting challenge (LCC), hereafter denoted as A1 (Minervini et al., 2016) and A4 (Bell and Dee, 2016) were also used in evaluating the model. The total number of images in the datasets are 1248, 128, 624 of resolutions 317×309, 500×530, and 441×441 in *Ara2013*, A1, and A4, respectively. The datasets were captured with different experimental setups, so the quality of the images as well as the background appearance varies.

Data Augmentation

Data augmentation is a method widely used in deep learning to increase the size of available datasets and to give more diverse examples to the neural network during training. The aim is to instill in the model a level of invariance to nuisance factors meaning that the network should give the same results if the same image of a plant is just rotated or shifted. It also helps the network to ignore background variability such as different planting trays, camera setups and soil appearance. For this study, data augmentation was performed when training all models, in the form of assigning random affine transformations from a pool of random rotations between 0 and 180 degrees, shifting the image between 0% and 10% of its size as well as flipping the image on the horizontal or vertical axis.

Data Preprocessing and Model Training

Before training the neural network, all images were resized to 320 × 320 as a preprocessing step to optimise training times while retaining important features, such as distinct small leaves. Out of the 24 plants in the *Ara2013* dataset, 19 were used for training and the remaining five plants were used for testing. As the five genotypes present in the dataset can be visually distinct, it is important to present the network with an adequate sample of each so that it can learn each genotype particularities. As a result, the five testing plants were chosen as to contain one plant of each genotype. We perform fourfold cross validation where the dataset was randomly divided into four nonoverlapping subsets so that all images are present in the test set once. There is an imbalance between mutants and wild type in the training datasets so a training class weight was added to the genotype classification branch to increase training importance of the wild-type images. The class weight was chosen to be proportional to the class imbalance in each training scenario.

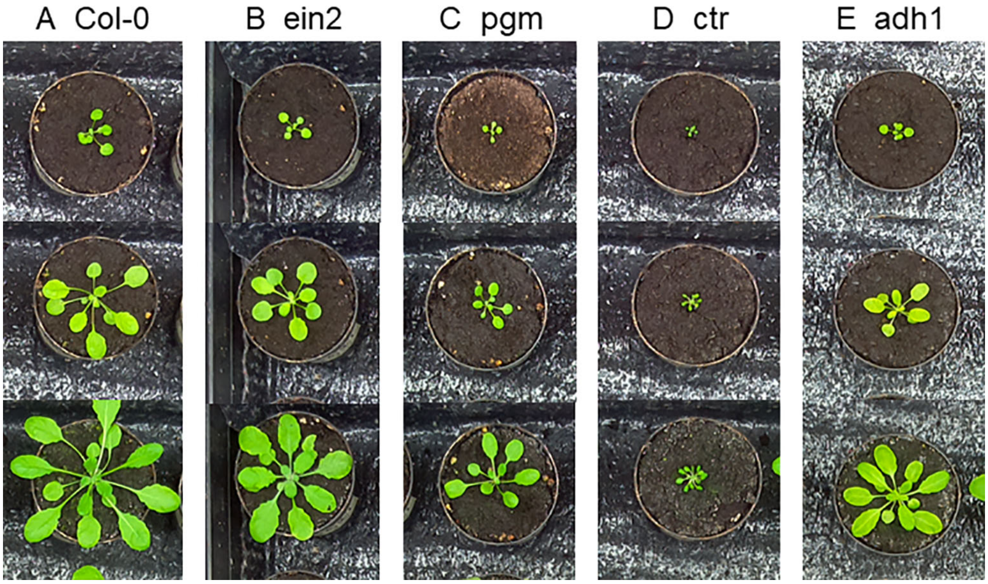


FIGURE 3 | Example images from the *Ara2013* dataset. The dataset is composed of time series images of 24 plants of 5 *Arabidopsis thaliana* accessions. The different genotypes vary in size, shape and color hue, making it a challenging dataset.

The annotations used during training for each image were the total leaf count as an integer, the PLA, and whether the plant was a mutant or wild type. We normalised the PLA values between 0 and 1 by computing the total area covered by the plant as a percentage of the whole image. In the experiment testing how the model performs with less training annotations, leaf count labels were removed in incremental steps leaving 75%, 50%, and 25% from the total number in the training set. The labels were removed to maintain an even distribution of plant ages and genotypes in the training set (i.e., every 4th label removed for the 75% step). Next, in the experiment analyzing the different strategies of annotating a dataset, three methods of removing labels were employed: we either removed count annotations corresponding to the most juvenile plants, or we removed labels corresponding to the most mature plants, or lastly we removed labels randomly. The same 25% increments were used. During all experiments where we trained the model with fewer leaf count annotations, the PLA and genotype annotations were still provided for all images.

The model was trained on an Nvidia TitanX GPU using the Adam optimizer with a learning rate of 0.0001. All the tasks were concurrently learned end-to-end, with an early stopping criterion based on the validation loss, in order to avoid overfitting. Model selection was according to the overall validation loss for all tasks in the cross-validation.

RESULTS

In this section, we offer experimental evidence of the effectiveness of our model. To evaluate our model in the leaf counting task, we use CVPPP evaluation metrics (Scharr et al., 2016; Giuffrida et al., 2018b).

They are the difference in count (DiC), absolute DiC (|DiC|), MSE, percentage agreement and coefficient of determination (R^2). The agreement metric represents the percentage of instances where the network prediction corresponds exactly with the ground truth.

Evaluation of the MTL Model

We first trained our MTL model on the *Ara2013* dataset, as it is the only dataset that contains plants of different genotypes. We then added the A1 and the A4 datasets in order to gauge impact of visually diverse datasets to our model. The results are displayed in **Table 1** for all three tasks. The results show that the network predictions display a strong correlation with the ground truth in the leaf count task with a R^2 of 0.95 and an overall test MSE of 0.93. The PLA estimation task shows a small MSE equating to an average difference of 2.1% between the ground truth and the predicted PLA. The genotype classification task shows a promising 91.1% test accuracy. As illustrated in the confusion matrix in **Figure 4**, wrong predictions occur rarely. Moreover, the model shows resistance to nuisance variability (i.e., different backgrounds and soil), as we evaluated different datasets grown in different growth scenarios.

Next, we assessed if the addition of MTL increases performance for the leaf counting task compared to a single task model. To make the single task leaf count variant of our model we removed the other branches. We maintained the same training procedure for both models and the dataset used was the *Ara2013* dataset. The results are shown in **Table 2**. Overall, the results of the MTL model are improved for all metrics analyzed, demonstrating that MTL reduces prediction errors when multiple related tasks are learned jointly. To test whether there is a statistically significant difference between the multitask and single-task models, we performed a bootstrapped paired t-test (Rodriguez, 2011) between the results of the |DiC| for the two approaches at 100% leaf count labels with a null hypothesis that they are equal. We perform the bootstrapped t-test because the

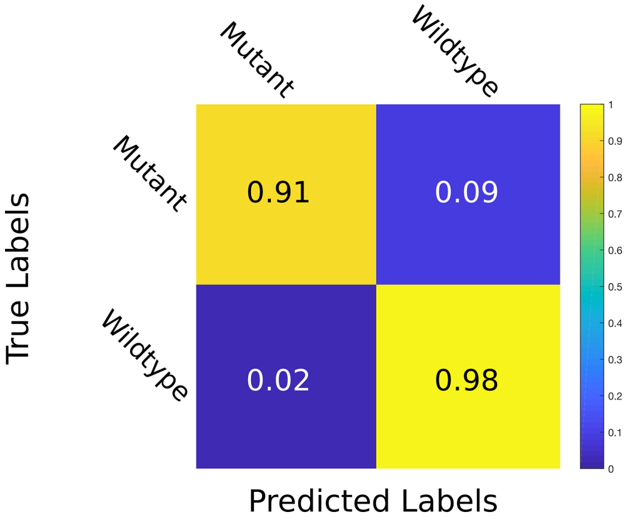


FIGURE 4 | Confusion matrix for genotype prediction at test time. The multitask learning (MTL) model learns to classify whether a plant is a mutant or wild type with an accuracy of 98% correct wild type and 91% correct mutant classification. The values are given at test time in the *Ara2013* dataset.

TABLE 1 | Results for the multitask learning (MTL) network for leaf count and projected leaf area (PLA) and genotype classification.

Dataset	DiC	Count				PLA	Genotype
		|DiC|	Agreement	MSE	R^2	MSE	Accuracy
Ara2013	−0.22 (0.93)	0.67 (0.69)	45	0.93	0.95	0.021	91.1
Ara2013 + A1A4	−0.21 (1.09)	0.77 (0.79)	44	1.23	0.96	0.025	95.6

The values are computed at test time for the model trained on first just the Ara2013 *dataset and then the extended dataset of* Ara2013+*A1+A4. The small drop in performance in the extended dataset can be attributed to the increase in dataset difficulty by adding more challenging examples.*

TABLE 2 | Results for the multitask learning (MTL) network vs. the single task network for leaf counting task trained on the *Ara2013* dataset.

Model	DiC	|DiC|	Agreement	MSE	R^2
Single task	0.40 (1.09)	0.80 (0.84)	41	1.35	0.92
Multitask	0.22 (0.93)	0.67 (0.69)	45	0.93	0.95

All parameters are improved in the MTL model, with the mean squared error (MSE) showing an improvement of 40%.

output of our model for the |DiC| does not follow a Gaussian distribution required for a valid t-test. The result is a two tailed p-value of 0.0093 after 10^5 bootstrapped samples. The p-value of <0.05 demonstrates that there is a significant difference between the MTL and single task models, confirming the superiority of MTL for the leaf counting task.

We then compare our MTL framework to current state of the art specialised leaf counting models. The results can be seen in **Table 3**. We trained our MTL model on the A1 dataset but with just the leaf count and PLA tasks as there are only wild-type plants present. Our model outperforms the results of Dobrescu et al. (2017a) in all categories. We achieve similar results to the best reported values in Itzhaky et al. (2018), D+R method that utilise the leaf center as additional training annotation, while we are using only a direct regression method.

Substituting Hard to Get Annotations With MTL

In this experiment, we assess whether we could compensate for the lack of expensive training annotations in the leaf counting task by using an MTL approach and providing other, easier to acquire, annotations. When training the network, we removed parts of the leaf count labels, but we retained all the PLA and genotype labels. Leaf count labels were removed in incremental steps leaving 75%, 50%, and 25% from the total number in the training set to check how the models perform when increasingly fewer count annotations are available.

Experimental results are shown in **Table 4**. It can be noted that the MTL model remain consistent even when only 25% of the original count labels are used in training. Furthermore, the standard deviation of the DiC in the MTL model remains nearly constant for all the label steps, indicating that the predictions are consistently close to the reported mean. On the other hand, the single-task model sees a significant decline in performance when less training annotations are present. The MSE increases from 1.45 when 100% of the labels are present to 5.49 and then to 17.2 at the 50% and 25% count label steps respectively.

TABLE 3 | Comparison of our proposed multitask learning (MTL) model with state-of-the-art results in leaf counting on the *Computer Vision Problems in Plant Phenotyping* workshop (CVPPP) A1 test set.

Method	DiC	|DiC|	Agreement	MSE
Romera-Paredes and Torr (2016)**	0.20(1.40)	1.1(0.9)	–	–
Aich and Stavness (2017)†	–0.33(1.38)	1.00(1.00)	30.3	1.97
Dobrescu et al. (2017a)†	–0.39(1.17)	0.88(0.86)	33.3	1.48
Itzhaky et al. (2018) MSR†	–0.27(1.21)	0.70(1.02)	57.0	1.48
Itzhaky et al. (2018) D+R**	–0.12(1.11)	0.73(0.84)	45.5	1.21
Proposed Multi-Task Model†	–0.09(1.10)	0.78(0.77)	39.0	1.22

*The results show an improvement in mean squared error (MSE) on previous works that use just the total leaf count as annotation. The results are similar to the current state-of-the-art specialized leaf counting networks. The table only shows results of the leaf counting task as there is no benchmark for the other tasks. †Method uses just the total leaf count as annotation. **Method uses stronger annotations.*

The same trend is visible in the R^2 values as well declining from 0.92 at 100% to near 0 when only 25% of the count labels are available. To test whether there is a significant difference between the results of the different count label thresholds in **Table 4**, we computed the same type of bootstrapped paired sample t-tests mentioned in Section 3.1 between the results of the |DiC| for the multitask and single-task models trained with 100% and 25% of the labels respectively, using the standard threshold of 0.05 as a significance level to indicate whether there is a true mean difference between the two samples. The performance drop is more noticeable in single task-model at all levels of omitted labels and the bootstrapped two tailed p-value well below <0.001 reflects the results. On the other hand, in the MTL model, the results remain stable and do not differ significantly as the number of training labels decrease (bootstrapped two tailed p-value of 0.097, above significance threshold). This means that the model successfully compensates from the lack of leaf counting data by learning from the other tasks.

The distribution of count predictions at the 25% count label step can be seen in **Figure 5**. The MTL model maintains a more leptokurtic distribution, with 91% of the predictions fall within ±1 of the ground truth, compared to the single task model where only 50% of predictions are within ±1 of the ground truth.

Which Labels Are Most Important

Given that it is possible to obtain reliable leaf count predictions with only the 25% of the training count labels in the MTL model, an important question is: *Which 25% of labels are most important for the model to successfully train?* Understanding this key aspect would help guide the first annotation of a new dataset, highlighting which 25% of images should be first (and potentially only) annotated in a new dataset.

TABLE 4 | Effect of incrementally decreasing leaf count annotations in the multitask learning (MTL) (multi) and single-task (single) models during training.

	Count Labels	100%	75%	50%	25%
DiC	Single	0.40 (1.09)	0.82 (1.68)	1.16 (2.04)	1.18 (3.98)
	Multi	–0.22 (0.93)	–0.14 (0.94)	–0.23 (0.95)	–0.46 (0.94)
|DiC|	Single	0.80 (0.84)	1.28 (1.36)	1.62 (1.69)	2.68 (3.16)
	Multi	0.67 (*0.69)	0.62 (0.71)	0.75 (0.72)	0.75 (0.75)
Agreement	Single	41	33	23	21
	Multi	45	48	40	42
MSE	Single	1.35	3.48	5.50	17.2
	Multi	0.93	0.91	1.08	1.13
R^2	Single	0.92	0.80	0.68	0.02
	Multi	0.95	0.95	0.94	0.94

All the projected leaf area (PLA) and genotype labels are still present during training of the MTL model. The MTL model maintains steady performance in all label steps while the single task model shows significant decline. We show the results on the leaf count task because it is the most challenging task.

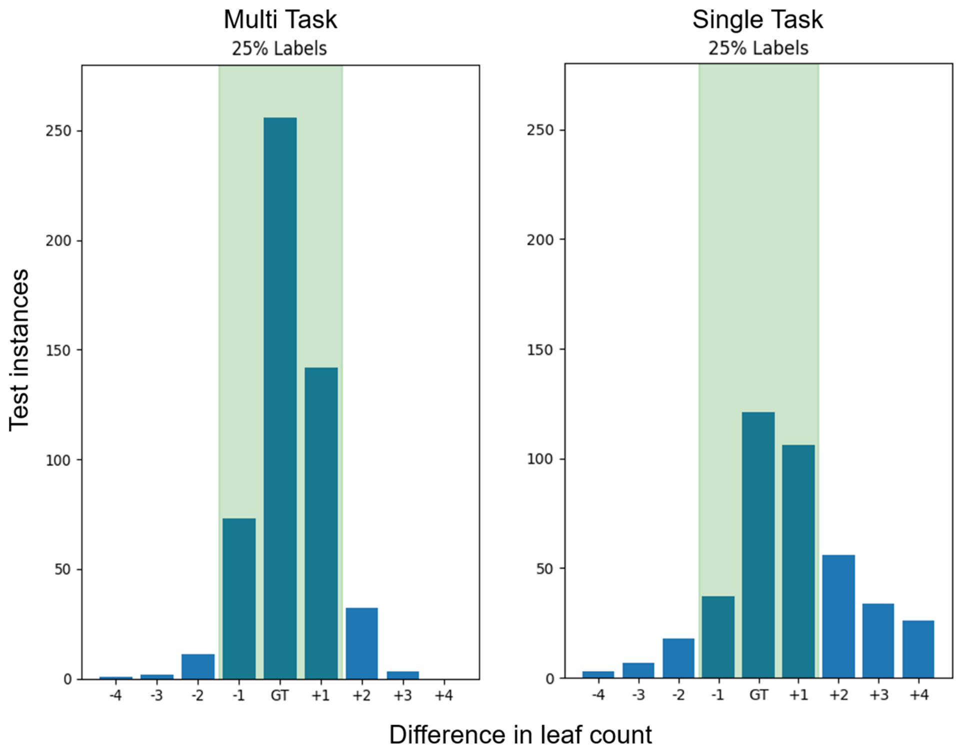


FIGURE 5 | Leaf count test accuracy when training with 25% of the count labels. The bars represent the sum of predictions in the *Ara2013* test set which are equal the ground truth as well as the differences in count on either side. The green shaded region represents the region of ground truth ±1 leaves which is similar to human accuracy. The multitask learning (MTL) variants have a ±1 accuracy of 91% while the single task models only have a 50% ± 1 accuracy and a much wider spread of difference in count errors.

We evaluated three different annotating strategies and results are shown in **Table 5**. Firstly, we removed the count labels starting with the youngest plants up to the designated thresholds of 75%, 50%, and 25%. For example, at 50% labels there were no count labels present for the first half of the plant's life. Using this method, the results show similar results between 100% and 75% count labels so we conclude that the youngest plants have little impact on the training of the model. The results then start declining until there is only a R^2 value of 0.07 when only the oldest 25% of the plants were present.

The next method is the reverse of the previous one, meaning we removed the count labels starting with the oldest plants. We observed a decline in results, even at 75% labels. At the next step threshold, the model failed to learn any of the tasks. Lastly, we excluded annotations from the dataset selecting at random plants across the time span. This method, perhaps as expected, gave

TABLE 5 | The impact on the multitask learning (MTL) model different strategies for annotating a dataset by determining the impact on the MTL model count labels and their impact on the MTL model.

		Count				PLA	Genotype
Selection Method	Count Labels	|DiC|	Agreement	MSE	R^2	MSE	Accuracy
All count labels	100%	0.67 (0.69)	45	0.93	0.95	0.021	90
Removed juvenile plants	75%	0.65 (0.68)	46	0.89	0.95	0.025	91
	50%	1.66 (1.73)	28	5.76	0.68	0.032	88
	25%	3.45 (2.26)	6	17.08	0.07	0.030	81
Removed mature plants	75%	1.36 (1.84)	34	5.27	0.71	0.015	63
	50%	4.83 (5.80)	21	53.03	N/A	0.013	73
	25%	6.91 (6.20)	17	86.16	N/A	0.019	67
Random Selection	75%	0.70 (0.68)	42	0.97	0.94	0.010	91
	50%	0.67 (0.71)	44	0.96	0.94	0.024	88
	25%	1.39 (1.24)	27	3.49	0.81	0.045	81

This can also be seen as what count labels are most important when annotating a new image based plant growth dataset. The values shown were obtained training on the Ara2013 *dataset. The count labels were removed in increments of 25%. First the labels of the most juvenile plants were removed. Then the labels of the oldest plants were removed. The third category removes count labels in a random fashion at the designated percentage steps.*

results which are comparable to having an equal distribution of labels as in **Table 4**. At the 25% step the results worsened, but this could be explained due to random chance of how the count label distribution was selected.

Determining Important Image Regions

Training a deep neural network model with less annotations generally makes it more difficult for the model to learn. To assess this impact in our model, we investigated what parts of the image the network considers important. We aim to see if the most important regions correspond to the plant or the network is influenced by information found in the background (e.g. the soil or plant pot). We performed the test by imposing a black sliding window on a sample of test images and predicted the leaf count, genotype, and PLA using our model on the images as the sliding window was traversing it. The method developed in Zeiler and Fergus (2014) is similarly used in Dobrescu et al. (2017a). The aim was to understand what are the important parts of the image from the trained network';s perspective, as obstructing such a part would give rise to errors in the predictions.

For the leaf counting task, we carried out this test on models trained using 25% count annotations in MTL and single task models to gauge if there is a difference in how the errors are distributed in the two approaches when less annotations were available. In the PLA estimation and genotype classification we compare MTL models trained with 100% and 25% count labels to determine if they were learning properly, and if they were still focusing on relevant image parts at the two annotation increments. The results are shown in **Figure 6**, showing that the network does actually focus mostly on the image areas corresponding to the plants. Additionally in the MTL model the errors generated and the regions impacted are similar between the models trained with 100% and 25% leaf count labels.

DISCUSSION

We show that an MTL deep learning approach is superior to just single task models for the purposes of characterizing visually challenging plant traits, such as leaf counting. We treat the leaf counting problem as a holistic regression task. One of the

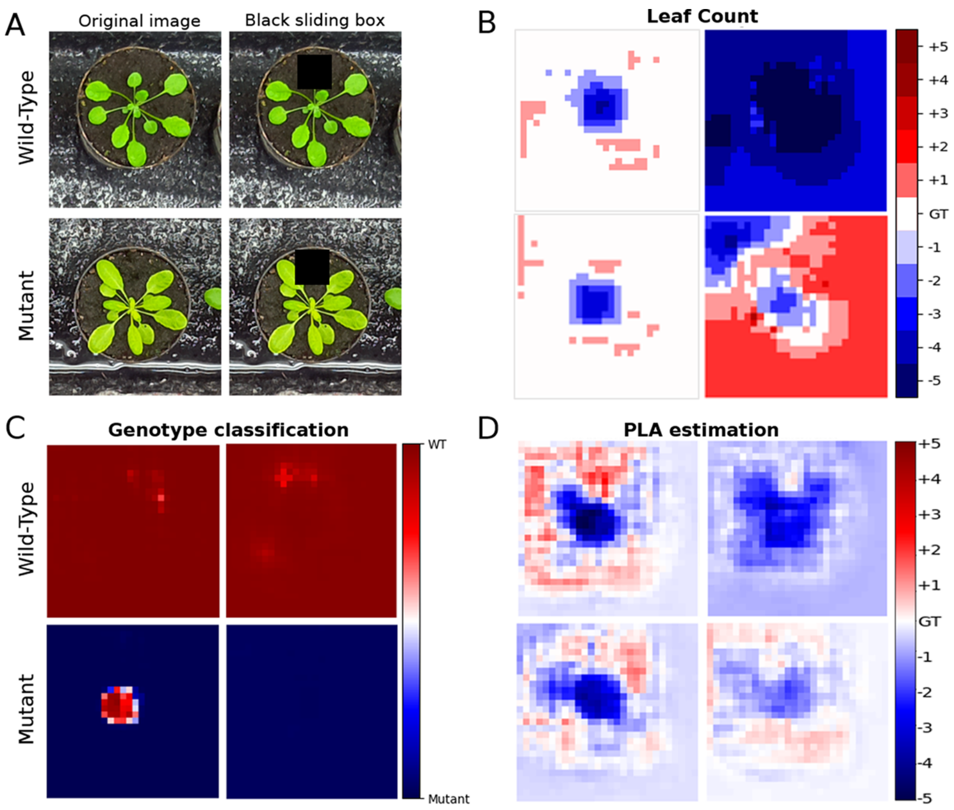


FIGURE 6 | Test showing the focus of the network using a sliding window. **(A)** is the original test image, along with the count ground truth. Overlaid is a black sliding window (60 × 60 pixels), which traverses the original images. The top row is an example of a wild type plant and the bottom row is an example of a mutant. **(B)** represents the prediction accuracy as the sliding window is traversing the image of the multitask learning (MTL) model (left column) and the single task model (right column) trained with just 25% count labels. The errors are expected to be confined to the area where the plant is located as the box obscures whole or parts of leaves the overall count prediction should decrease. **(C, D)** represent the sliding box test only in the MTL model for Genotype classification accuracy and projected leaf area (PLA) estimation. These were only performed on the MTL model comparing between models trained with 100% (right column) and 25% (left column) labels. The color bar for the PLA task shows increments in percentage points. The rows correspond to the images in part A. GT signifies the ground truth.

limitations of such approaches is that the network needs to learn good image representation from each image, based only on the total leaf count number. Employing an MTL model offers extra information to the model easing the training process.

In agreement with Caruana (1997), an MTL model can learn also from the other tasks leading to better generalization performance and more robust extraction of features. The benefits can be seen in **Tables 2** and **4** where the MTL model outperforms the single-task model. The performance of deep neural networks is known to be strongly influenced by the quantity of annotated data used during training (Sun et al., 2017). By omitting leaf count labels in our approach, the model is essentially training the leaf counting task with fewer annotated examples and, therefore, it would be expected to have an important negative impact on the results. However that is not the case for our MTL model, which can overcome the extra difficulty of training from less annotations without having a statistically significant drop in performance. Furthermore, during training, the MTL model was more stable when fewer count annotations were available compared to the single task models (see Additional **Figure 1** for more details).

Getting a sense of what regions the network considers important, provides an insight if the model was successfully trained to get information from the appropriate image areas (i.e., the plant not the background). There is a clear difference between the MTL and single task models in the leaf counting task when trained with just 25% count annotations **Figure 6**. As the sliding box moves over the image, the errors that produce a lower count prediction are very specific to regions containing the plant suggesting that the model learned well the area of interest. On the other hand, the single task model yields more pronounced prediction errors in all regions of the image so it does not focus on the plant region as well as the MTL model. For the genotype classification and PLA estimation tasks we investigated if there are differences between MTL models trained with 100% and 25% leaf count annotations. There is no visibly significant difference between them meaning that both models learned to focus mostly on the plant areas.

Two of the mutants present in the *Ara2013* dataset seen in **Figure 3**, (ein2 and adh1) are visually similar to the Col-0 wild type, making genotype classification a challenging task. The errors we observed occur mainly when the model misclassifies these mutants as wild-type in the early and middle part of the growth cycle. However, the overall classification accuracy remains >90%, demonstrating that the network is not biased towards a specific class.

When assessing the best strategy to select labeled data for the leaf counting task in the MTL model, we can find what are the most important time points in the plant growth stage for the network to learn in **Table 5**. The network performance is directly affected when the count labels are missing from mature plants, while minor changes are seen when the juvenile 25% are removed, showing similar behavior as a random selection of up to 50%. This means that most important information for these tasks is learned from the mature plants, while the juvenile plants contribute less in the learning process. The other tasks reflect this trend as well. We can conclude that the best strategy is to provide the most balanced dataset, that provides the widest-ranging examples to the neural network during training. Next, in order from best to worst would be to just randomly choose which labels to provide, then omitting the juvenile plants and lastly is to omit the mature ones.

CONCLUSIONS

In this paper we have proposed a framework for multitask deep learning (MTL) for plant phenotyping. We showed that MTL architecture outperforms the single-task models trained on the same datasets. We have achieved an improvement on the state-of-the-art for leaf counting compared to direct regression approaches for the datasets tested. We achieve a similar performance to state-of-the-art methods which use additional annotations for training. To the best of our knowledge, this is the first work that studies and compares the benefits of MTL versus single task in plant phenotyping. We show that the proposed MTL model can be used to compensate for missing labels in plant phenotyping, leveraging other related traits. We have also explored different leaf count annotation strategies and showed which segments of the plant images are most important to be labeled. Lastly we have shown that the MTL model correctly focuses on the parts of the image that correspond to the plant and largely disregards the background when computing prediction for all three tasks.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.plant-phenotyping.org/CVPPP2017-challenge.

AUTHOR CONTRIBUTIONS

AD, MG, and ST conceived the study. AD performed the experiments. AD prepared the manuscript with feedback from MG and ST. All authors read and approved the final manuscript.

FUNDING

AD is currently supported by an EPSRC DTP PhD fellowship (EP/N509644/1). MG is supported by the BBSRC grant BB/P023487/1. ST is partly supported by the UK Biotechnology and Biological Sciences Research Council (BB/P023487/1) and The Alan Turing Institute under the EPSRC grant EP/N510129/1.

ACKNOWLEDGMENTS

We thank NVIDIA for providing a TitanX GPU essential for this work.

REFERENCES

Aich, S., and Stavness, I. (2017).Leaf counting with deep convolutional and deconvolutional networks, in: *IEEE International Conference on Computer Vision CVPPP Workshop* (IEEE Computer Society), IEEE. 2080–2089. doi: 10.1109/ICCVW.2017.244

Baxter, J. (1997). A bayesian/information theoretic model of learning to learn *via* multiple task sampling. *Mach. Learn.* 28, 7–39. doi: 10.1023/A:1007327622663

Bell, J., and Dee, H. M. (2016). Aberystwyth leaf evaluation dataset. Zenodo. doi: 10.5281/zenodo.168158

Boyes, D. C., Zayed, A. M., Ascenzi, R., McCaskill, A. J., Hoffman, N. E., Davis, K. R., et al. (2001). Growth stage-based phenotypic analysis of Arabidopsis: a model for high throughput functional genomics in plants. *Plant Cell* 13, 1499–1510. doi: 10.1105/tpc.13.7.1499

Caruana, R. (1997). Multitask learning. *Mach. Learn.* 28, 41–75. doi: 10.1023/A:1007379606734

Caspar, T., Huber, S. C., and Somerville, C. (1985). Alterations in growth, photosynthesis, and respiration in a starchless mutant of arabidopsis thaliana (l.) deficient in chloroplast phosphoglucomutase activity. *Plant Physiol.* 79, 11–17. doi: 10.1104/pp.79.1.11

Dobrescu, A., Giuffrida, M. V., and Tsaftaris, S. A. (2017a), Leveraging multiple datasets for deep leaf counting, in: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. IEEE. 2072–2079. doi: 10.1109/ICCVW.2017.243

Dobrescu, A., Scorza, L. C. T., Tsaftaris, S. A., and McCormick, A. J. (2017b). A “Do-It-Yourself” phenotyping system: measuring growth and morphology throughout the diel cycle in rosette shaped plants. *Plant Methods* 13, 95–107. doi: 10.1186/s13007-017-0247-6

Giuffrida, M. V., Minervini, M., and Tsaftaris, S. (2015). “Learning to count leaves in rosette plants,” in *CVPPP workshop - BMVC* (British Machine Vision Association), The British Machine Vision Association (BMVA). 13. doi: 10.5244/C.29.CVPPP.1

Giuffrida, M. V., Scharr, H., and Tsaftaris, S. A. (2017). “ARIGAN: synthetic arabidopsis plants using generative adversarial network,” in *ICCVW - Workshop in Computer Vision Problems in Plant Phenotyping*. IEEE. doi: 10.1109/ICCVW.2017.242

Giuffrida, M. V., Chen, F., Scharr, H., and Tsaftaris, S. A. (2018a). Citizen crowds and experts: observer variability in image-based plant phenotyping. *Plant Methods* 14, 12–26. doi: 10.1186/s13007-018-0278-7

Giuffrida, M. V., Doerner, P., and Tsaftaris, S. A. (2018b). Pheno-deep counter: a unified and versatile deep learning architecture for leaf counting. *Plant J.* 96, 880–890. doi: 10.1111/tpj.14064

Guzman, P. E. J. (1990). Exploiting the triple response of arabidopsis to identify ethylene-related mutants.´. *Plant Cell* 2, 513–523. doi: 10.1105/tpc.2.6.513

He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Identity mappings in deep residual networks,” in *Computer Vision – ECCV 2016*. Eds. B. Leibe, J. Matas, N. Sebe and M. Welling (Cham: Springer International Publishing), 630–645.

Itzhaky, Y., Farjon, G., Khoroshevsky, F., Shpigler, A., and Bar-Hillel, A. (2018).Leaf counting: multiple scale regression and detection using deep cnns, in: *British Machine Vision Conference 2018, BMVC 2018*, Northumbria University, Newcastle, UK, September 3-6, 2018. The British Machine Vision Association (BMVA). 328.

Kieber, J. J., Rothenberg, M., Roman, G., Feldmann, K. A., and Ecker, J. R. (1993). Ctr1, a negative regulator of the ethylene response pathway in arabidopsis, encodes a member of the raf family of protein kinases. *Cell* 72, 427–441. doi: 10.1016/0092-8674(93)90119-B

Kokkinos, I. (2017).Ubernet: training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 5454–5463. doi: 10.1109/CVPR.2017.579

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*. Neural Information Processing Systems Foundation. 1097–1105. doi: 10.1109/CVPR.2009.5206848

Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017).Feature pyramid networks´ for object detection, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE. 936–944. doi: 10.1109/CVPR.2017.106

Minervini, M., Abdelsamea, M. M., and Tsaftaris, S. (2014). Image-based plant phenotyping with incremental learning and active contours. *Ecol. Inf.* 23, 35–48. doi: 10.1016/j.ecoinf.2013.07.004

Minervini, M., Giuffrida, M. V., and Tsaftaris, S. (2015).An interactive tool for semi-automated leaf annotation, in: *CVPPP workshop - BMVC* (British Machine Vision Association), 13, 204–216. doi: 10.5244/C.29.CVPPP.6

Minervini, M., Fischbach, A., Scharr, H., and Tsaftaris, S. A. (2016). Finely-grained annotated datasets for image-based plant phenotyping. *Pattern Recognit. Lett.* 81, 80–89. doi: 10.1016/j.patrec.2015.10.013

Minervini, M., Giuffrida, M., Perata, P., and Tsaftaris, S. (2017). Phenotiki: an open software and hardware platform for affordable and easy image-based phenotyping of rosette-shaped plants. *Plant J.* 90, 204–216. doi: 10.1111/tpj.13472

Ngouajio, M., Lemieux, C., and Leroux, G. D. (1999). Prediction of corn (zea mays) yield loss from early observations of the relative leaf area and the relative leaf cover of weeds. *Weed Sci.* 1, 297–304. doi: 10.1017/S0043174500091803

Pan, S. J., and Yang, Q. (2010). A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 22, 1345–1359. doi: 10.1109/TKDE.2009.191

Pape, J.-m., and Klukas, C. (2015). Conference proceedings of British Machine Vision Conference (BMVC). *Proc. Comput. Vision Prob. In Plant Phenotyping (CVPPP)*. The British Machine Vision Association (BMVA) 1–12. doi: 10.5244/C.29.CVPPP.3

Perata, P., and Alpi, A. (1993). Plant responses to anaerobiosis. *Plant Sci.* 93, 1–17. doi: 10.1016/0168-9452(93)90029-Y

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017).Deep learning for multi-task plant phenotyping, in: *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. IEEE. 2055–2063. doi: 10.1109/ICCVW.2017.241

Qiu, R., Wei, S., Zhang, M., Li, H., Sun, H., Liu, G., et al. (2018). Sensors for measuring plant phenotyping: a review. *Int. J. Agric. Biol. Eng.* 11, 1–17. doi: 10.25165/j.ijabe.20181102.2696

Rahnemoonfar, M., and Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. *Sensors* 17, 905–916. doi: 10.3390/s17040905

Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., and Pande, V. (2015). Massively multitask networks for drug discovery. *arXiv Prepr. arXiv:1502.02072.*

Ranjan, R., Patel, V. M., and Chellappa, R. (2019). Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 121–135. doi: 10.1109/TPAMI.2017.2781233

Ren, M., and Zemel, R. S. (2017).End-to-end instance segmentation with recurrent attention, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 293–301. IEEE. doi: 10.1109/CVPR.2017.39

Rodriguez, R. N. (2011). Sas. *Wiley Interdiscip. Rev.: Comput. Stat* 3, 1–11.

Romera-Paredes, B., and Torr, P. H. S. (2016). “Recurrent Instance Segmentation,” in *ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI* (Cham: Springer International Publishing), 312–329. doi: 10.1007/978-3-319-46466-419

Scharr, H., Minervini, M., French, A. P., Klukas, C., Kramer, D. M., Liu, X., et al. (2016). Leaf segmentation in plant phenotyping: a collation study. *Mach. Vision Appl.* 27, 585–606. doi: 10.1007/s00138-015-0737-3

Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017).Revisiting unreasonable effectiveness of data in deep learning era, in: *Computer Vision (ICCV), 2017 IEEE International Conference on (IEEE)*. IEEE. 843–852.

Ubbens, J. R., and Stavness, I. (2017). Deep plant phenomics: a deep learning platform for complex plant phenotyping tasks. *Front. Plant Sci.* 8, 1190–1201. doi: 10.3389/fpls.2017.01190

Ward, D., Moghadam, P., and Hudson, N. (2018). Deep leaf segmentation using synthetic data. *Comput. Vision Prob. In Plant Phenotyping (CVPPP)*. The British Machine Vision Association (BMVA). 327–337. doi: 10.5244/C32.CVPPP

Zeiler, M. D., and Fergus, R. (2014). “Visualizing and understanding convolutional networks,” (Springer International Publishing), 818–833. doi: 10.1007/978-3-319-10590-1_53

Zhu, Y., Aoun, M., Krijn, M., Vanschoren, J., and Campus, H. T. (2018). Data augmentation using conditional generative adversarial networks for leaf counting in arabidopsis plants. *Comput. Vision Prob. In Plant Phenotyping (CVPPP)*. The British Machine Vision Association (BMVA). 327–337. doi: 10.5244/C32.CVPPP.2

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 10 June 2020
doi: 10.3389/fpls.2020.00617

Rapeseed Stand Count Estimation at Leaf Development Stages With UAV Imagery and Convolutional Neural Networks

Jian Zhang*[1,2†], *Biquan Zhao*[1,2†], *Chenghai Yang*[3], *Yeyin Shi*[4], *Qingxi Liao*[5], *Guangsheng Zhou*[6], *Chufeng Wang*[1,2], *Tianjin Xie*[1,2], *Zhao Jiang*[1,2], *Dongyan Zhang*[7], *Wanneng Yang*[6], *Chenglong Huang*[5]* and *Jing Xie*[8]

[1] Macro Agriculture Research Institute, College of Resource and Environment, Huazhong Agricultural University, Wuhan, China, [2] Key Laboratory of Arable Land Conservation (Middle and Lower Reaches of Yangtze River), Ministry of Agriculture, Wuhan, China, [3] Aerial Application Technology Research Unit, USDA-Agricultural Research Service, College Station, TX, United States, [4] Department of Biological Systems Engineering, University of Nebraska–Lincoln, Lincoln, NE, United States, [5] College of Engineering, Huazhong Agricultural University, Wuhan, China, [6] College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, [7] Anhui Engineering Laboratory of Agro-Ecological Big Data, Anhui University, Hefei, China, [8] College of Science, Huazhong Agricultural University, Wuhan, China

Edited by:
Tony Pridmore,
University of Nottingham,
United Kingdom

Reviewed by:
Dong Xu,
University of Missouri, United States
Alexei E. Solovchenko,
Lomonosov Moscow State University,
Russia

***Correspondence:**
Chenglong Huang
hcl@mail.hzau.edu.cn
Jing Xie
xiejing625@mail.hzau.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Technical Advances in Plant Science, a section of the journal Frontiers in Plant Science

Received: *07 August 2019*
Accepted: *22 April 2020*
Published: *10 June 2020*

Citation:
Zhang J, Zhao B, Yang C, Shi Y, Liao Q, Zhou G, Wang C, Xie T, Jiang Z, Zhang D, Yang W, Huang C and Xie J (2020) Rapeseed Stand Count Estimation at Leaf Development Stages With UAV Imagery and Convolutional Neural Networks. Front. Plant Sci. 11:617. doi: 10.3389/fpls.2020.00617

Rapeseed is an important oil crop in China. Timely estimation of rapeseed stand count at early growth stages provides useful information for precision fertilization, irrigation, and yield prediction. Based on the nature of rapeseed, the number of tillering leaves is strongly related to its growth stages. However, no field study has been reported on estimating rapeseed stand count by the number of leaves recognized with convolutional neural networks (CNNs) in unmanned aerial vehicle (UAV) imagery. The objectives of this study were to provide a case for rapeseed stand counting with reference to the existing knowledge of the number of leaves per plant and to determine the optimal timing for counting after rapeseed emergence at leaf development stages with one to seven leaves. A CNN model was developed to recognize leaves in UAV-based imagery, and rapeseed stand count was estimated with the number of recognized leaves. The performance of leaf detection was compared using sample sizes of 16, 24, 32, 40, and 48 pixels. Leaf overcounting occurred when a leaf was much bigger than others as this bigger leaf was recognized as several smaller leaves. Results showed CNN-based leaf count achieved the best performance at the four- to six-leaf stage with F-scores greater than 90% after calibration with overcounting rate. On average, 806 out of 812 plants were correctly estimated on 53 days after planting (DAP) at the four- to six-leaf stage, which was considered as the optimal observation timing. For the 32-pixel patch size, root mean square error (RMSE) was 9 plants with relative RMSE (rRMSE) of 2.22% on 53 DAP, while the mean RMSE was 12 with mean rRMSE of 2.89% for all patch sizes. A sample size of 32 pixels was suggested to be optimal accounting for balancing performance and efficiency. The results of this study confirmed that it was feasible to estimate rapeseed stand count in field automatically, rapidly, and accurately. This study provided a special perspective in phenotyping and cultivation management for estimating seedling count for crops that have recognizable leaves at their early growth stage, such as soybean and potato.

Keywords: stand counting, field-based phenotyping, optimal observation timing, convolutional neural network, precision agriculture

INTRODUCTION

Next to soybean and oil palm, rapeseed (*Brassica napus* L.) is the third largest oil crop worldwide (Wang et al., 2010; Berrocoso et al., 2015; Bouchet et al., 2016). Statistical data from the Food and Agriculture Organization of the United Nations have shown that the world production of rapeseed in 2016 was more than 68 million tons, mainly from Canada (19.5 million tons), China (13.1 million tons), and India (6.8 million tons)[1]. Increasing rapeseed yield is a major focus for rapeseed researchers and cultivators (Godfray et al., 2010). The crop stand count at early growth stages is one of the most important parameters for the prediction of yield, density, and growth status (Jin et al., 2017; Liu S. et al., 2017; Zhao et al., 2018). Rapeseed leaf development in early growth stages includes cotyledons completely unfolded, first leaf unfolded, two leaves unfolded, three leaves unfolded, until nine, or more leaves unfolded (Weber and Bleiholder, 1990; Lancashire et al., 1991). Overlapping is intense throughout the entire leaf development stage. Moreover, small and irregular spacing makes rapeseed seedlings clustered. As a result, it is hard to detect and count each individual rapeseed seedling (Zhao et al., 2018). The traditional way of counting rapeseed seedlings at early growth stages is based on ground-level investigation which is labor-intensive and time-consuming (Jin et al., 2017; Liu T. et al., 2017; Naito et al., 2017). Since rapeseed seedlings are in small plant size, irregular spacing and complex overlapping at their early growth stages, timing of rapeseed seedling counting by field investigation depends on empiricism. Accordingly, the obtained records and data are subjective (Bucksch et al., 2014; Deng et al., 2018). The most serious problem is that ground-level investigation is destructive to the field crops (Jin et al., 2017; Liu S. et al., 2017; Deng et al., 2018). In addition, manual investigation brings more external factors into the plant growth environment, resulting in artificial error (Zhao et al., 2018). Therefore, an objective, precise, and automated rapeseed stand counting method will benefit researchers and producers (Araus and Cairns, 2014).

In recent years, plant scientists worldwide have shown great interest in phenotyping since this technology will bring a brand new perspective for agricultural planting and breeding (Furbank and Tester, 2011; Yang et al., 2014; Naito et al., 2017). Phenotyping provides a new tool to reveal phenotype traits determined by environmental and genetic factors (White et al., 2012; McCouch et al., 2013; Ghanem et al., 2015) and to estimate the growth status of plants and crops (Sadras et al., 2013; Maimaitijiang et al., 2017; Yang et al., 2017). Remote sensing technology provides an efficient means for crop phenotype data collection (Verger et al., 2014; Shi et al., 2016; Yu et al., 2016; Wendel and Underwood, 2017), which can record phenotyping traits, such as plant height, canopy temperature, architecture, stress, and color (Walter et al., 2012; Rahaman et al., 2015; Mir et al., 2019). In particular, UAVs draw much attention due to their unique advantages, such as noninvasive observation at low altitude, high resolution, frequent data collection, and deployment flexibility (Zhang and Kovacs, 2012; Ballesteros et al., 2014; Huang et al., 2016). Accordingly, UAVs are used as a platform to collect data and estimate vegetation growth parameters including biomass (Bendig et al., 2014), leaf area index (Córcoles et al., 2013), height (Van Iersel et al., 2018), yield (Geipel et al., 2014), canopy cover, and structure (Cunliffe et al., 2016). Overall, as a powerful and reliable platform, UAVs have shown their advantages to be used to collect crop data for phenotyping.

In studies of crop stand counting, Gnädinger and Schmidhalter (2017) found maize plant numbers had a strong correlation (R^2 = 0.89) with the enhanced color digital counts using UAV imagery. Jin et al. (2017) extracted 13 object features containing color and texture from UAV images and further employed SVM for wheat classification, counting, and density estimation. Their results indicated that wheat density can be estimated when wheat plants had one to two leaves (Jin et al., 2017). Besides, field imagery for crop stand counting has been applied in some other crops such as corn (Shi et al., 2013; Varela et al., 2018), potato (Zheng et al., 2016; Sankaran et al., 2017), and cotton (Chen et al., 2018) with various remote sensing platforms. However, in most previous studies of crop stand counting, data were derived only from one observation at a certain day during the growth stages, and to our knowledge, there has been little research with data from multiple observations for rapeseed.

Object identification, classification, and counting are major tasks in image analysis (Davies, 2009; Blaschke, 2010; Ma et al., 2017; Zanotta et al., 2018). Several approaches have been developed for fast image processing and classification (Schowengerdt, 2012), including SVM, RF. These approaches are developed from using individual spectral, spatial, or textural information to integrating all the information (Linker et al., 2012). Gnädinger and Schmidhalter (2017) used regression analysis to find the correlation between maize plant number and green digital pixel counts based on spectral information. Jin et al. (2017) employed SVM model to classify wheat and to

---

Abbreviations: CNN, convolutional neural network; DAP, day after planting; GCP, ground control point; GPS, global positioning system; LOOCV, leave-one-out crossing validation; MAE, mean absolute error; RF, random forest; RMSE, root mean square error; rRMSE, relative root mean square error; SVM, support vector machine; UAV, unmanned aerial vehicle; UGV, unmanned ground vehicle.

[1] www.fao.org/home/en/

estimate seedling count and density by using spectral and textural features. These approaches can be used for counting trees (Gomes et al., 2018), fruits (Qureshi et al., 2017), and flowers based on satellite remote sensing imagery or UGV/UAV remote sensing imagery. Nevertheless, these regular non-automatic approaches required manual extraction of distinct features, which was time-consuming for a multi-observation study.

Convolutional neural networks have drawn wide attention due to their automated processing and good performance in image analysis (Lee et al., 2017; Pound et al., 2017; Sindagi and Patel, 2017). The CNN application for object detection and counting has been reported in previous studies, such as crowd detection and counting in public events (Sheng et al., 2016; Zhang et al., 2015, 2016) and animal detection and counting in the wild (Arteta et al., 2016). CNN approach is also reported to be applied for classification, detection, and counting for seeds, fruits, flowers, crops, and leaves in agriculture (Grinblat et al., 2016; Ramos et al., 2017). Pound et al. (2017) showed that CNNs were effective to distinguish, identify, and count wheat plants and their ears in glasshouse condition. Madec et al. (2019) used faster region-based CNN (Faster-RCNN) to detect and count wheat ears in a field and estimated the density of ears *via* high-resolution RGB imagery captured by a camera fixed on a boom. Besides, there are some studies using CNN for tree detection and counting. Li et al. (2016) employed a CNN algorithm to detect and estimate oil palm trees from four-band satellite images with 0.6 m spatial resolution, and they reported that more than 96% of trees were correctly detected. Csillik et al. (2018) used a customized CNN model for citrus tree detection and counting with overall accuracy greater than 95%. These studies demonstrated that CNN could be applied to remote sensing imagery captured from both satellites and UAVs. Crop stand counting using CNN has not been widely reported, especially for rapeseed stand counting. Ribera et al. (2017) reported a CNN model to estimate the number of sorghum plants by UAV imagery with a best mean absolute percentage error of 6.7%, which provided a solution for large field plant counting research using CNN. Using CNN in a field or in the wild is of more practical significance (Simonyan and Zisserman, 2014).

However, the situation of crops grown in the field was complex. Specifically, rapeseed seedlings grown in the field were in irregular spacing, complex overlapping, and different plant sizes, which made its counting difficult. Nevertheless, one of rapeseed phenotyping traits that UAV imagery collected was its leaf canopy. During the early growth stages of leaf development, rapeseed leaves play an important role reflecting their growth status (Weber and Bleiholder, 1990; Lancashire et al., 1991). Therefore, it is feasible to recognize individual rapeseed leaves with CNN and perform stand counting with reference to the existing knowledge of the number of rapeseed leaves per plant. The objectives of this study were to (1) recognize and count the individual rapeseed canopy leaves through UAV imagery with CNN, (2) establish and examine the models identifying the number of leaves per rapeseed seedling, and (3) propose an optimal timing to estimate rapeseed stand count.

MATERIALS AND METHODS

Study Area and Experimental Design

Figure 1 shows the study area and GCPs. The study area with center coordinates (30°28′57.11″N, 114°18′39.45″E) located near Huazhong Agricultural University in Wuhan, China. It covered an area about 50 m × 30 m with an average elevation of 27 m. In the field, a rapeseed cultivar named Huayouza 62 (*B. napus* L.) was sown with two planters including a valve-branch distributor-based centrifugal precision metering device and a rotating disk-type seeding device on November 4, 2017 (**Figure 1**). According to the experimental design, rapeseed was seeded by the two devices in eight rows simultaneously with 20 cm row spacing at a seeding rate of 5.5 kg/ha. No weed control management was implemented after sowing. There were 12 GCPs permanently arranged surrounding the study area. Reach RS+ (Emlid, United States) was used to collect GPS information.

Image Acquisition System

The image acquisition system was composed of a DJI rotocopter Matrice 600 UAV (DJI, Shenzhen, China) and a Nikon D800 camera (Nikon, Japan) in this study. The UAV could hover for 35 min without a payload. In this study, each mission took about 13 min with a payload of 2 kg. A GPS module was integrated into the UAV, which was tested with a horizontal accuracy of 0.5 m and a vertical accuracy of 1.5 m. The Nikon D800 camera fitted with a Nikon 50.0 mm f/1.4D lens was mounted on the UAV to collect nadir RGB images of the rapeseed field during flights. The complementary metal oxide semiconductor (CMOS) sensor of the camera had a size of 35.9 mm × 24.0 mm, capturing images with 7,360 × 4,912 pixels. The camera was also integrated with a GPS device to geotag the images and a wireless control to trigger the camera capturing imagery every 1.0 s automatically. An SD memory card was used to store JPEG images with a 24-bit format.

Each flight mission followed the same camera configuration during the whole rapeseed leaf growth stage. The frontal and side overlaps of the flight path were 80.0 and 70.0%, respectively. Two perpendicular flight paths were conducted to cover and image the rapeseed canopy. In this study, the UAV was flown at a ground speed of 3 m/s and at a height of approximately 30 m above ground level, and spatial resolution of image was about 0.18 cm.

Unmanned Aerial Vehicle Data Collection and Preprocessing

Image collection was scheduled from November 17, 2017, the 14th DAP, to January 12, 2018, with an interval of 7 days. This period covered the whole rapeseed leaf development stage (Weber and Bleiholder, 1990; Lancashire et al., 1991). Some adjustments were made because of rainy or heavy windy weather. As a result, data collection started on November 17, 2017, and ended on January 10, 2018, with specific dates on 14, 23, 32, 39, 46, 53, 58, and 68 DAP from 12:30 p.m. to 2:00 p.m. The one- to three-leaf stage was before 40 DAP and the three- to four-leaf stage lasted from 40 to 50 DAP. The four- to six-leaf stage

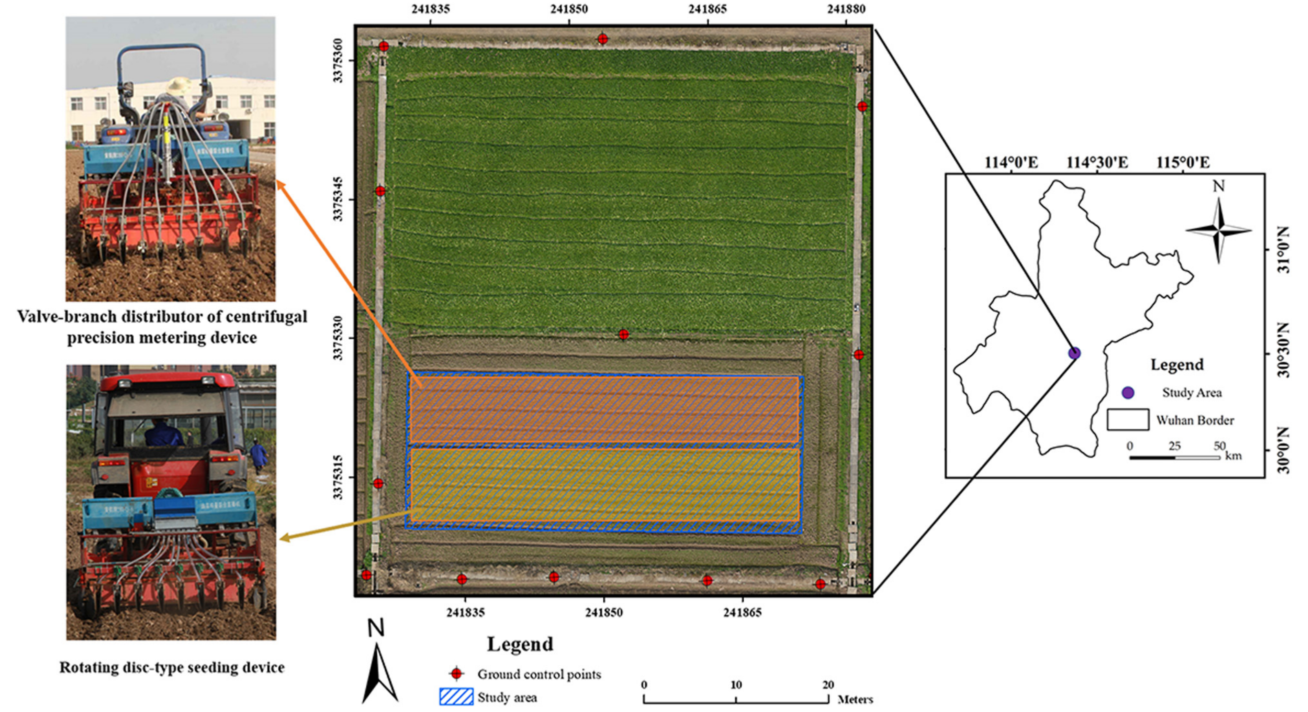


FIGURE 1 | Study area and ground control point (GCP) distribution.

ranged from 50 to 60 DAP, and the seven-leaf stage and beyond was after 60 DAP.

Free software Capture NX-D 1.2.1 from Nikon (Nikon, Japan) was utilized to correct geometric distortion for the UAV captured individual images. Secondly, an ortho-mosaic image for each collection was generated using Pix4Dmapper software (Pix4D, Switzerland). In this step, the 12 GCPs were added for image mosaicking. This study used eight sample plots (four plots per seeding device) for analysis, which were subsets from each ortho-mosaic image by ArcMap 10.3 (ESRI, United States). Each plot was in the size of 9.5 m × 2 m (**Figure 2B**).

These eight sample plots were divided into an image training dataset containing six sample plots and a test dataset consisting of two sample plots during the whole processing and analysis. Since manual counting in the field could possibly change the real field condition, making the data unreliable for multiple observations, this study manually interpreted and annotated the rapeseed leaves over the eight sample plots for each observation date using ArcMap 10.3 (an illustration in **Figure 2**), which was mainly described in the next section. This study also used image-based manual rapeseed seedling count as the ground truth reference. Data of 14 and 23 DAP were not used because rapeseed seedlings were too small to distinguish in imagery. Therefore, there were six remaining observation dates over the eight study plots (six for training and two for testing) in this study.

All the processing and analysis were executed on a computer with an Intel (R) Core (TM) i7-6800K CPU and one NVIDIA GeForce GTX 1060 6GB GPU, and the memory of the computer was 32 GB.

Image Processing and Data Analysis

Rapeseed Leaf Recognition Based on a Convolutional Neural Network

Convolutional neural networks are the most popular machine learning algorithms applied to various computer vision tasks, such as numeral recognition, face recognition, and handwriting recognition. Some software packages such as Python and Matlab provide a convenient environment for CNN modeling, and there are some open-source CNN codes online. However, it is still hard for a nonprofessional machine learning researcher to implement the entire flowchart of CNN modeling, including CNN software operation environment configuration, codes modification, as well as parameter adjustment. In addition, such CNN software is not suitable for processing geospatial information and geoinformation analysis. Thus, this study employed the easy-to-use image analysis software eCognition Developer 9.3 (Trimble, United States), which contains a CNN module (Trimble, 2018). The module can be used to recognize objects in images based on the Google TensorFlow^{TM} library-create (Csillik et al., 2018; Trimble, 2018).

This approach was a patch-based CNN algorithm according to the categorization reviewed by Sindagi and Patel (2017). It was convenient and interactive when researchers use CNN for image analysis in this software. The operation in this software was mainly composed of four steps: (1) to generate labeled sample patches; (2) to create a CNN; (3) to train a CNN; and (4) to apply a CNN (Trimble, 2018). In this study, the labeled sample patches were cropped from the training sample plot images according to three classes including rapeseed leaves, weeds, and bare soil.

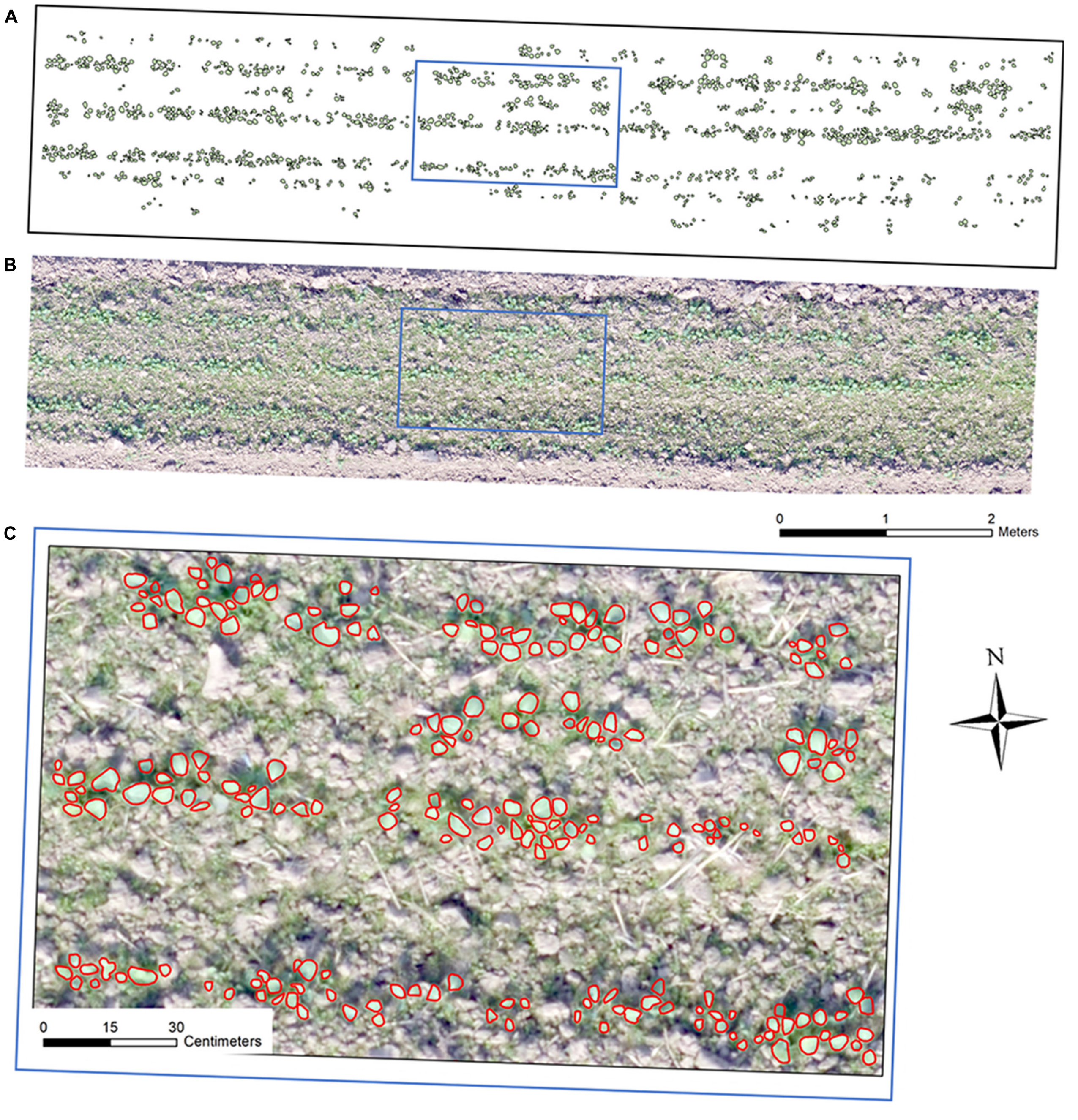


FIGURE 2 | Ortho-mosaic RGB image of a sample plot with a size of 9.5 m × 2 m: **(A)** manual interpreted and annotated rapeseed leaves, **(B)** ortho-mosaic RGB plot image, **(C)** RGB image overlapping with annotated rapeseed leaves.

To label the three classes, the rapeseed leaves from the eight sample plots for each observation date were manually interpreted and annotated using ArcMap 10.3, as shown in **Figure 2A**. The labeled leaves were used as a reference, representing the ground truth leaves. The overlapping of rapeseed leaves was unavoidable, and each recognized canopy rapeseed leaf was outlined. Furthermore, the weeds increased with time since there was no weed control management in the field to maintain the original ecological scene. As DAP increased, the rapeseed leaves and weeds were much distinguishable by size, color, texture, and pattern in high-resolution UAV images. Therefore, after interpreting and annotating the rapeseed leaves, we used ExG-ExR, a color vegetation index whose ability for green pixel identification was confirmed (Meyer and Neto, 2008; Zhao et al., 2018) to label the remaining green pixels as weeds by eCognition Developer. The pixels that were not in green were labeled as bare soil through normalized green minus red difference index (NGRDI). The pixels whose values of NGRDI less than 0 were classified as bare soil. In this study, the input data of eCognition Developer include RGB imagery in TIFF format exported from ArcMap, the manual-annotated rapeseed leaf polygon shapefile exported from ArcMap.

The parameters during the operation included sample patch size, number of hidden layers, kernel size, number of feature maps, max pooling, and learning rate. According to the rapeseed leaf development stages from one leaf to more than seven leaves during the investigation period, this study generated five sizes of sampling patches for CNN training, which included 16 × 16, 24 × 24, 32 × 32, 40 × 40, and 48 × 48. In addition, the default number of feature maps was 12, while the learning rate was at 0.0005 in the software. Based on the five patch sizes, the optimal kernel size was 5 × 5 after trial-and-error processing. Therefore, this study structured an initial CNN model as shown in **Figure 3**, containing two hidden layers, two max pooling layers, and one fully connected layer. A previous study using a similar CNN model showed great performance for oil palm tree detection and counting with satellite images (Li et al., 2016). We generated 20,000 sample patches of rapeseed leaves, 10,000 weed patches, and 10,000 bare soil patches from the training dataset in each observation date for five sample patches totally.

The number of recognized rapeseed leaves was counted after the dilate operation in the outputted heat map, according to the number of local maximal points. A heat map with each pixel value ranging from 0 to 1 was output by the CNN model, representing the possibility of the target class for rapeseed leaves. High pixel values close to 1 in the heat map indicated a high possibility of rapeseed leaves, while values close to 0 indicated a low possibility (Csillik et al., 2018; Trimble, 2018). Based on the pixel values, local maximal pixels whose threshold ranged from 0.5 to 1 (step on 0.01) were iteratively searched to locate a rapeseed leaf through the dilate operation in eCognition as the locations of rapeseed leaves were expected to coincide with local maximal and high values in the heat map (**Figure 4**).

In this study, the size of the dilation filter was defined as 5 × 5, which was the same as the kernel size in CNN. Afterward, a maximal layer was generated, reflecting the maximum value in the matrix of 25 pixels in the heat map (**Figure 4C**). Accordingly, when a pixel in the maximal layer has the same value with the pixel in the same location of the heat map, this pixel inherited the value. Otherwise, the pixel was valued 0 (**Figure 4D**). Each located point was considered to represent a CNN-detected rapeseed leaf. Thus, the number of local maximal points was considered as the number of recognized rapeseed leaves.

Overcounting occurred when the number of CNN-detected rapeseed leaves sometimes was larger than the actual number of rapeseed leaves, as a big leaf might correspond to several located maximum points (**Figure 5**). In **Figure 4**, if a pixel far away from the pixel of 0.98 also had a value of 0.98, two targets might be located in **Figure 4D**. Merging the adjacent points with a tolerance was a useful means for tree canopy recognition and counting (Li et al., 2016), but this method cannot be applied to rapeseed leaf counting in this study since the boundary of rapeseed leaves was less distinct than that of a tree canopy. Moreover, the sizes of rapeseed leaves were also different. Therefore, it was difficult to determine a precise distance for merging rapeseed leaf located points. Inspired by the method of using ground truth masks for assessing the accuracy of the estimated plant centers (Chen et al., 2017), we used the manual interpreted ground truth rapeseed leaf outlines as masks to record and analyze the overcounting rate (R_{oc}) of the local maximum points for each investigation (**Figure 5**).

The illustration in **Figure 5** was a subset of CNN recognition results on 53 DAP with the 32-pixel patch size and a local max value of 0.65. The blue lines showed rapeseed leaf outline masks, and each mask represented a ground truth rapeseed leaf. The green points illustrated the CNN-detected rapeseed leaves.

As mentioned, the sizes of leaves influenced the number of located maximum points representing the CNN-detected rapeseed leaves, and DAP was a significant factor of the leaf size. Thus, it could be assumed that in our study duration, the R_{oc} might be related to DAP. For CNN detection and counting, R_{oc}

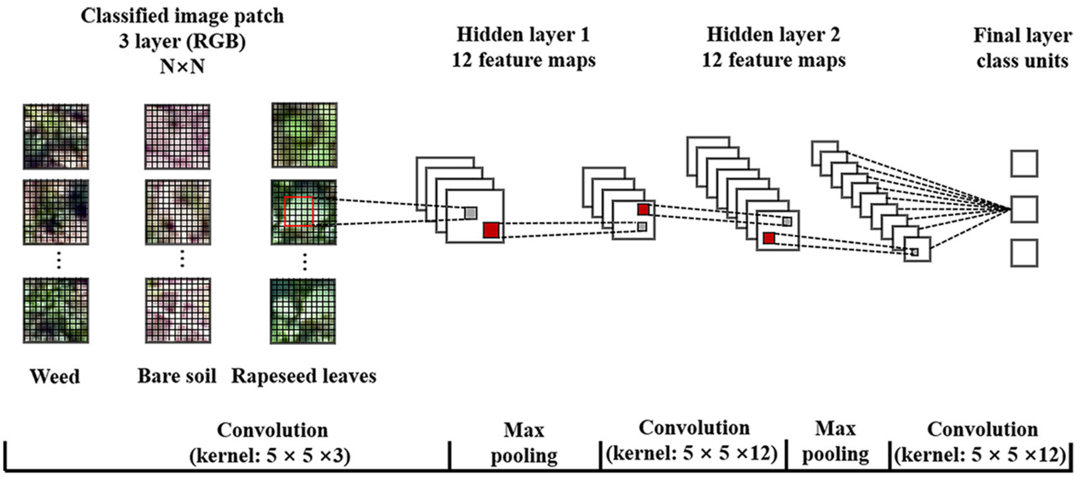
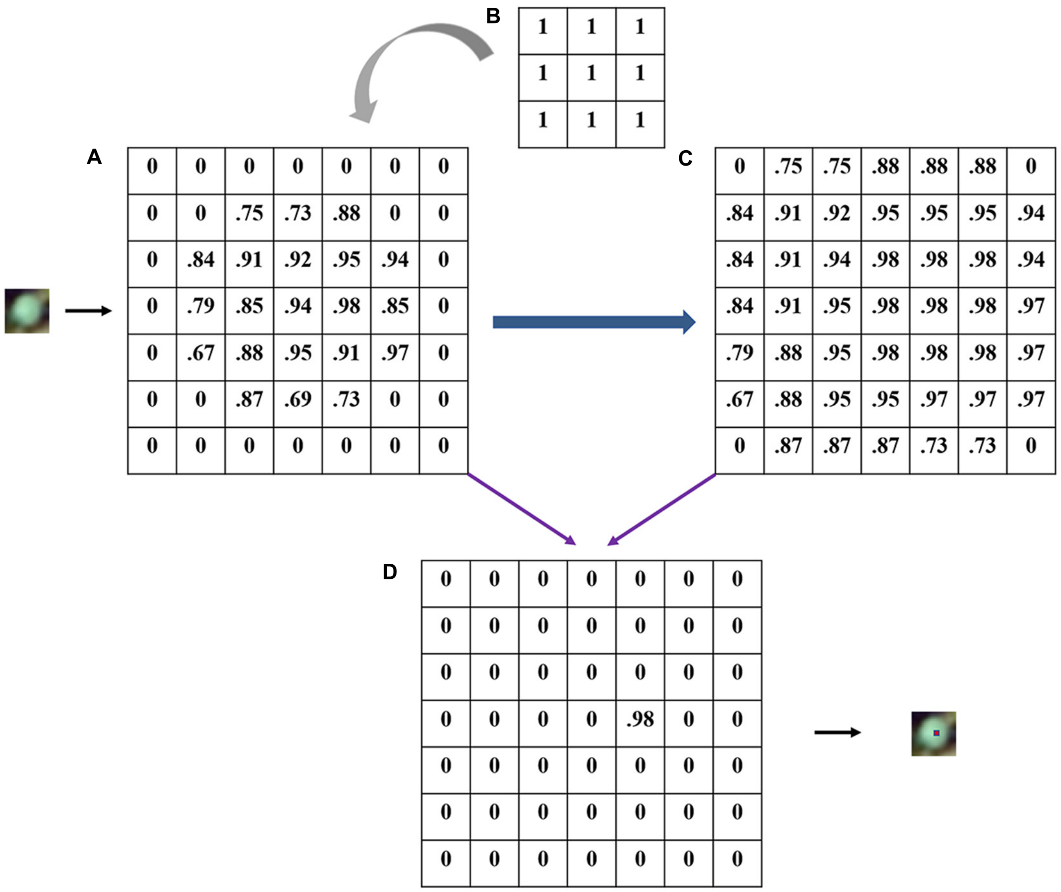


FIGURE 3 | Structure and parameters of the convolutional neural network (CNN) model used in this study. RGB, red-green-blue.



A

0	0	0	0	0	0	0
0	0	.75	.73	.88	0	0
0	.84	.91	.92	.95	.94	0
0	.79	.85	.94	.98	.85	0
0	.67	.88	.95	.91	.97	0
0	0	.87	.69	.73	0	0
0	0	0	0	0	0	0

B

1	1	1
1	1	1
1	1	1

C

0	.75	.75	.88	.88	.88	0
.84	.91	.92	.95	.95	.95	.94
.84	.91	.94	.98	.98	.98	.94
.84	.91	.95	.98	.98	.98	.97
.79	.88	.95	.98	.98	.98	.97
.67	.88	.95	.95	.97	.97	.97
0	.87	.87	.87	.73	.73	0

D

0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	.98	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0
0	0	0	0	0	0	0

FIGURE 4 | An illustration of local maximal possibility detection with 3 × 3 filter (actual 5 × 5 in the study): **(A)** possibility heat map; **(B)** a 3 × 3 dilation filter; **(C)** the maximal layer after dilating; **(D)** locating target.

was calculated by:

$$R_{oc} = \frac{\sum_{i=1}^{M}(C_i - 1)}{N} \times 100\%, \quad \left(\text{when}, \ C_i > 1\right) \tag{1}$$

where N is the number of CNN-detected rapeseed leaves, M is the number of ground truth rapeseed leaf masks, C_i is the number of located maximum points inside a ground truth rapeseed leaf mask for mask i, when mask i has more than one point.

Precision, *Recall,* and *F-score* were used in this study to evaluate leaf detection results (Xiong et al., 2017; Zhao et al., 2018). *Precision* and *Recall* are defined by true positive (TP), false positive (FP), and false negative (FN):

$$Precision = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Combined with the situation of overcounting, the following equations can be inferred by Eqs. (1), (2), and (3) in this study:

$$N = \text{TP} + \text{FP} + \sum_{i=1}^{M}(C_i - 1) \tag{4}$$

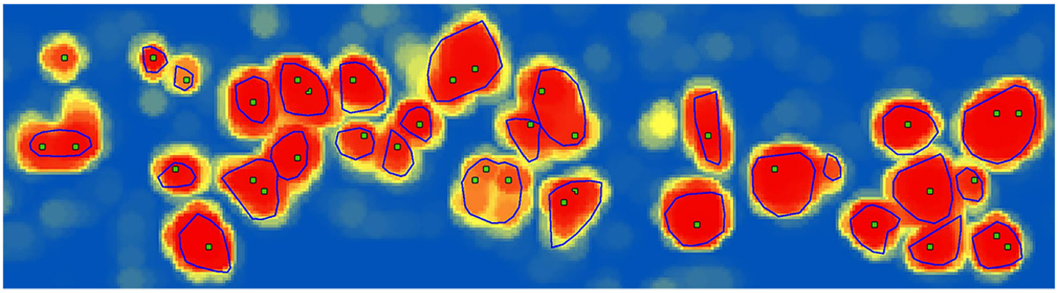
$$M = \text{TP} + \text{FN} \tag{5}$$

The unique recognized local maximum point that is inside a ground truth mask is considered as a TP. Accordingly, TP here is the accurate number of CNN-detected rapeseed leaves. If a local maximum point is outside a mask, then this maximum point is considered as a FP. A mask is identified as a FN if there is no point recognized inside (**Figure 5**). *F-score* was used as the final exponent to evaluate the CNN-detected rapeseed leaf recognition accuracy. *Precision*, *Recall*, and *F-score* in this study can be expressed as follows:

$$Precision = \frac{\text{TP}}{(1 - R_{oc}) \times N} \tag{6}$$

$$Recall = \frac{\text{TP}}{M} \tag{7}$$

$$F - score = \frac{2 \times Precision \times Recall}{\left(Precision + Recall\right)} \tag{8}$$



#	Mask	TP	FP	FN	Overcounting	*Precision*	*Recall*	*F-score*
A	29	38	1	1	-	97.44%	97.44%	97.44%
B	29	28	1	1	10	93.28%	96.55%	94.89%

FIGURE 5 | Illustration of overcounting statistics. CNN, convolutional neural network; FN, false negative; FP, false positive; TP, true positive. A was calculated without overcounting rate, B was calculated with overcounting rate.

Leave-One-Out Crossing Validation Regression Modeling for Rapeseed Seedling Counting

A strong relationship between the rapeseed seedling counting and the number of unfold leaves was expected because the identification of rapeseed at early growth stages was based on the number of unfolded leaves (for example, the one- to three-leaf stage and four- to six-leaf stage). In this study, the models that counting rapeseed seedlings through the number of manual-interpreted rapeseed leaves were first established and verified, which were considered as the reference describing the number of leaves per rapeseed seedling plant. Then the number of CNN-recognized rapeseed leaves was applied to these models to evaluate the accuracy of rapeseed stand count estimation at different growth stages.

The LOOCV regression modeling method was used to establish the models of seedling stand counting. LOOCV was effective in the case of small sample size. It was a special case of K-fold crossing validation, when K was equal to the number of samples. One sample was excluded for validation, and the rest samples were used for training. The same operation was repeated for K times so that each sample could be used for validation so that the results were unbiased. The LOOCV regression modeling was conducted by Python Spyder in Anaconda3 (64-bit) (Anaconda Software Distribution, 2016. Computer software).

In this study, there were six training sample plots for each investigation. Thus, K was 6, and LOOCV regression modeling was repeated for six iterations. The optimal rapeseed seedling model parameters of the corresponding DAP were further obtained by calculating an average value of the iteration results. Mean absolute error, RMSE, and coefficient of determination (R^2_{LOOC}) were used to verify these models. They were calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{9}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{10}$$

$$R^2_{LOOC} = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{11}$$

where n is the number of sample plots, y_{i} is the investigated ground truth rapeseed seedling stand count for sample i, $\hat{y}_{\mathrm{i}}$ is the model-predicted rapeseed seedling stand count for sample i, $\bar{y}$ is the average value of the investigated ground truth rapeseed seedling stand count for all samples in each observation date.

Performance Evaluation of Counting Seedlings

The number of CNN-recognized rapeseed leaves was used to evaluate the eventual performance of counting seedlings in this study. The number of CNN-recognized rapeseed leaves corresponding to the best value of *F-score* was applied to the models. Relative RMSE (rRMSE) was calculated as follows:

$$\mathrm{rRMSE} = \frac{\mathrm{RMSE}}{\bar{y}} \times 100\% \tag{12}$$

where $\bar{y}$ is the average value of the investigated ground truth rapeseed seedling stand count for all samples in each observation date.

RESULTS

Counting Rapeseed Leaves Recognized by Convolutional Neural Networks

The strong correlation ($R^2 = 0.831$) between R_{oc} and DAP conformed the impact of DAP on R_{oc} that R_{oc} increased as DAP advanced during the observed rapeseed leaf development stages (**Figure 6**). The bigger leaves caused the more detected local maximal points for a leaf, thus resulting in the overcounting situation. As shown in **Figure 6**, R_{oc} ranged from 7% on 39 DAP for the 32-pixel patch to 40% on 68 DAP for the 16-pixel patch. According to the correlation in **Figure 6**, R_{oc} was estimated to be 6.86% for 32 DAP, 12.25% for 39 DAP, 17.64% for 46 DAP, 23.03% for 53 DAP, 26.88% for 58 DAP,

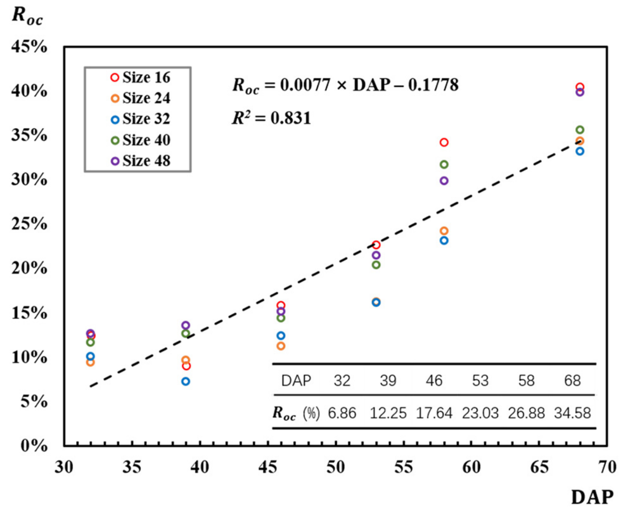


FIGURE 6 | Correlation between overcounting rate and days after planting (DAP).

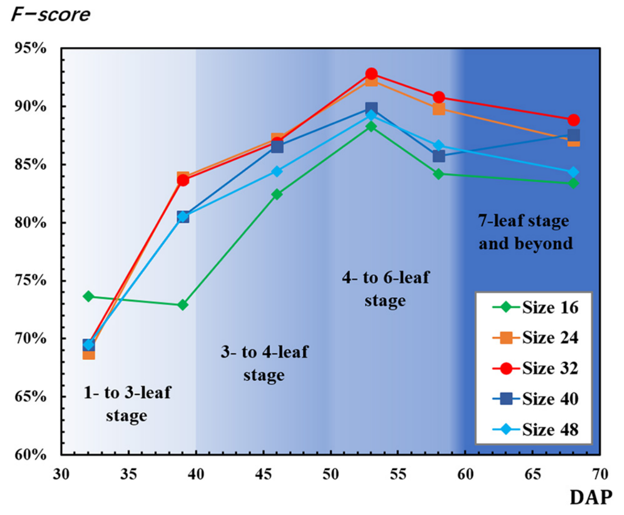


FIGURE 7 | Results of *F-score* for convolutional neural network (CNN)-detected rapeseed leaf recognition and counting.

and 34.58% for 68 DAP. In **Figure 6**, R_{oc} for the 32-pixel patch (blue legend) was lower than that for the other four patch sizes in most DAPs.

F-scores used for CNN-based recognition accuracy evaluation were calculated by equations (9), (10), and (11). The mean *F-scores* based on local maximal values ranging from 0.5 to 1 were calculated (**Figure 7**). For most of the patch sizes, *F-scores* increased from 32 to 53 DAP and decreased after 53 DAP. However, *F-scores* decreased from 32 to 39 DAP for the 16-pixel patch and increased from 58 to 68 DAP for the 40-pixel patch. For all patch sizes, *F-scores* had the highest values on 53 DAP. On 53 DAP, the ranking of *F-score* values from highest to lowest was 92.83% for the 32-pixel patch, 92.26% for the 24-pixel patch, 89.84% for the 40-pixel patch, 89.21% for the 48-pixel patch, and 88.26% for the 16-pixel patch.

Figure 8 shows the mean local max values achieving best *F-score* among the testing data during the whole observed leaf development stage. DAP also influenced the variation of local max values. The local max values increased with DAP and leveled off after 58 DAP. With these local max values, the number of rapeseed leaves recognized by CNN was counted for each observation.

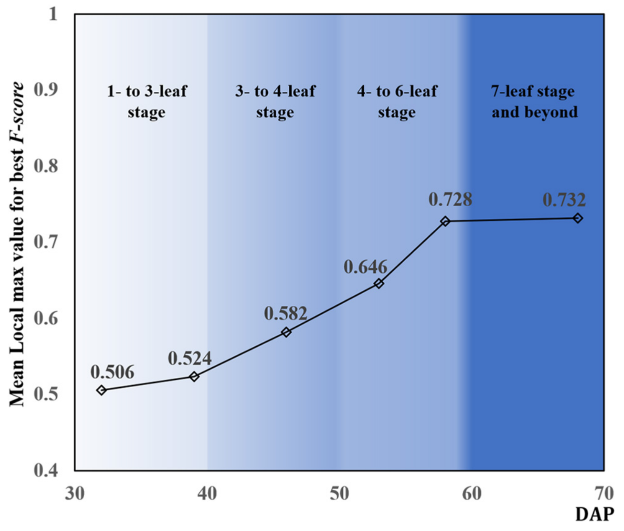


FIGURE 8 | Mean local max values for the best *F-score* according to days after planting (DAP).

Revealing the Number of Rapeseed Leaves per Seedling

A strong correlation between the seedling counting and the number of canopy rapeseed leaves was established by using the LOOCV method over data of training sample plots for each observation (**Table 1**). All sub-models were significant (p-value < 0.05) except for the sub-model of the first iteration on 32 DAP exhibiting a p-value of 0.058. Afterward, the relationship defining the number of rapeseed leaves per seedling was revealed by the formulas in **Table 1** for each observation.

These formulas were verified by the testing data. The ordering of R^2_{LOOC} from highest to lowest was 0.984 on 46 DAP, 0.926 on 53 DAP, 0.897 on 39 DAP, 0.886 on 58 DAP, 0.806 on 68 DAP, and 0.775 on 32 DAP. RMSE ranged from 14 plants to 54 plants, and MAE ranged from 13 plants to 47 plants. Results showed that the relationship revealing the number of rapeseed leaves per seedling obtained the best performance on 46 DAP and satisfactory performance on 53 DAP. As a result of the strong correlation revealing the number of rapeseed leaves per seedling, the approach using precise CNN-recognized leaves counting to estimate the seedling stand count was feasible and expected.

TABLE 1 | Results of optimal rapeseed seedling counting models for six observation periods.

Period	Formula	R^2_{LOOC}	MAE	RMSE
32 DAP	y = 0.354x + 14.340	0.775	47	54
39 DAP	y = 0.297x + 65.037	0.897	29	30
46 DAP	y = 0.322x + 19.490	0.984	13	14
53 DAP	y = 0.277x + 41.540	0.926	25	31
58 DAP	y = 0.220x + 72.334	0.886	31	35
68 DAP	y = 0.214x + 46.388	0.806	31	42

x represents the number of rapeseed canopy leaves. y stands for rapeseed seedling count. MAE and RMSE were rounded to integer. DAP, days after planting; MAE, mean absolute error; RMSE, root mean square error.

Performance of Estimating Seedling Stand Count With Convolutional Neural Network-Recognized Leaf Counting

Convolutional neural networks-recognized leaf counting (see section "Counting Rapeseed Leaves Recognized By Convolutional Neural Networks") was used to estimate seedling stand count according to the revealed relationship of the number of rapeseed leaves per seedling (**Table 1** in section "Revealing the Number of Rapeseed Leaves Per Seedling"). **Table 2** gives the results of estimating rapeseed seedling stand count. For all patch sizes, the best mean accuracy was achieved on 53 DAP with 99.26%. On average, 806 out of 812 plants were correctly estimated on 53 DAP at the four- to six-leaf stage. With the 32-pixel patch size, almost all the seedling stand counts were correctly estimated on 53 DAP. Some errors were counteracted mutually as a result of summing up the estimated seedling stand counts from two test sample plots. RMSE and rRMSE are presented in **Table 3** and **Figure 9**, respectively.

On 32 DAP, a maximum mean RMSE of 107 plants was observed, while on 53 DAP, a minimum mean RMSE of 12 plants showing the best performance was obtained. On 53 DAP, the rRMSE was 1.99% for the 40-pixel patch size (**Figure 9**). For the 32-pixel patch size, its RMSE was nine plants with an rRMSE of 2.22%. The ordering of mean RMSE from lowest to highest was 12 on 53 DAP, 48 on 46 DAP, 49 on 68 DAP, 57 on 58 DAP, 76 on 39 DAP, and 107 on 32 DAP, with the best performance on 53 DAP. Similarly, the ordering of rRMSE was 2.89% on 53 DAP, 11.71% on 68 DAP, 11.97% on 46 DAP, 11.97% on 58 DAP, 19.36% on 39 DAP, and 25.59% on 32 DAP.

DISCUSSION

Influence of Days After Planting Corresponding to Leaf Development Periods

Huayouza 62 (*B. napus* L.) used in the study is a member of the family Brassicaceae (Weber and Bleiholder, 1990; Lancashire et al., 1991). Based on the characteristics that its leaf development periods were highly related to DAP, the number of canopy leaves was used for seedling stand count modeling and estimation in this study. Chen et al. (2018) found that DAP had a strong correlation with the number of germinated seeds (R^2 = 0.938) and with average plant size (R^2 = 0.936). In our study, results also strongly illustrated that DAP played an important role on the leaf recognition and counting, revealing the relationship between seedling stand count and the number of leaves per seedling, and the eventual estimation of seedling stand count with CNN-recognized leaf counting. Finally, the best performance was achieved on 53 DAP for leaf recognition and counting as well as estimating seedling stand count with the number of CNN-recognized leaf counting (**Figures 7**, **9** and **Tables 2**, **3**).

This study not only used DAP but also tried to associate it with the leaf development period and to determine the influence of DAP corresponding to the leaf development periods on rapeseed leaf counting (**Figures 7**, **8**). DAP is an essential dimension unit to describe the growth process of crops, and it is easy to comprehend and quantify. Nevertheless, the growth situation based on DAP varies with different regions where crops and plants are cultivated with different treatments, even for the same cultivated variety. Sankaran et al. (2017) reported that it was hard to estimate the number of potato plants after 43 DAP because of the within-row canopy closure, even though the best correlation coefficient (r = 0.83) between manual plant counts and image-based counts was achieved on 32 DAP. This study presented a novel approach to understand the growth status of rapeseed seedlings and to count the seedlings in terms of leaf development periods.

In **Figure 6**, overcounting rate increased with DAP. From another perspective, the increasing overcounting rate was mainly caused by the development of leaves or the increasing size and number of rapeseed seedling canopy leaves. Rapeseed seedlings experience a process from new leaf unfolding to leaves gradually growing with DAP. In this study, rapeseed had more than seven leaves in the last observation period, which caused a serious overlapping. Moreover, the bigger leaf size influenced

TABLE 2 | Sum of rapeseed stand count estimation from two testing sample plots (unit: plants).

	Size 16	Size 24	Size 32	Size 40	Size 48	Average	Mean accuracy (%)	Ground-truth plants
32 DAP	593	672	657	640	590	630	75.05	840
39 DAP	567	734	668	640	657	653	83.42	783
46 DAP	686	723	727	712	714	712	89.05	800
53 DAP	801	806	812	804	807	806	99.26	812
58 DAP	835	833	823	854	823	834	88.12	946
68 DAP	803	802	805	805	800	803	96.51	832

Estimated results were rounded to integer. DAP, days after planting.

TABLE 3 | Results of RMSE for rapeseed stand count estimation (unit: plants).

	Size 16	Size 24	Size 32	Size 40	Size 48	Average
32 DAP	124	90	96	103	125	107
39 DAP	118	62	59	74	66	76
46 DAP	63	44	40	45	47	48
53 DAP	17	14	9	8	10	12
58 DAP	56	57	62	47	62	57
68 DAP	50	45	48	50	51	49

Means the optimal value in the column. Results were rounded to integer. DAP, days after planting; RMSE, root mean square error.

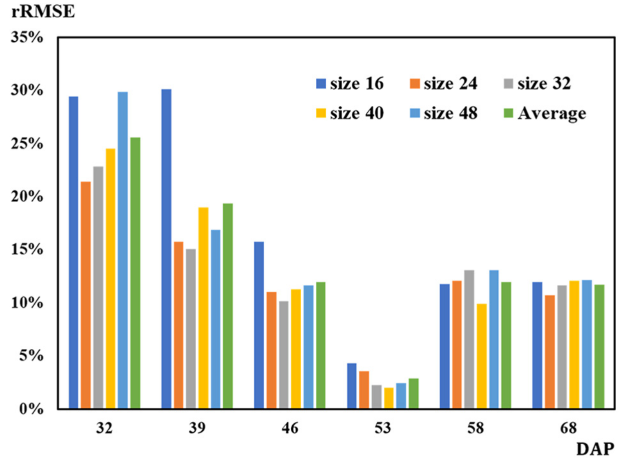


FIGURE 9 | Results of relative root mean square error (rRMSE) for rapeseed stand count estimation.

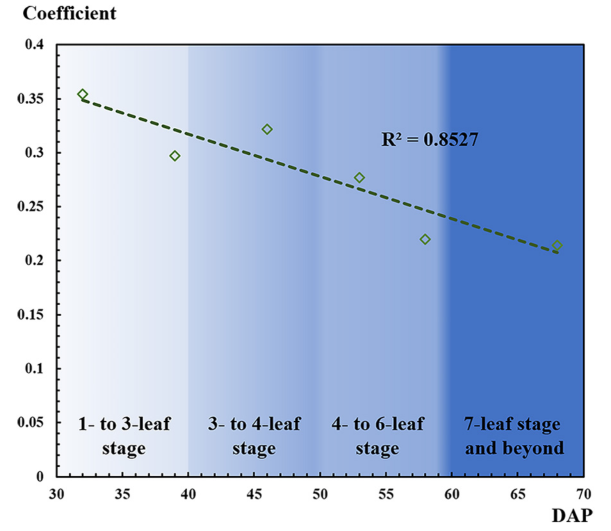


FIGURE 10 | Correlation between coefficient of the seedling counting models and days after planting (DAP).

the recognition. According to the trend and strong correlation between overcounting rate and DAP (R^2 = 0.831), when the number of leaves developed from one to more than seven leaves, overcounting rate might keep increasing.

As mentioned, as the number of leaves increased, leaf overlapping and saturation became more intensified. In **Figure 7**, F-score increased from the one- to six-leaf stage, and then decreased after that. This peak of F-score was mainly related to the leaf development. When rapeseed seedlings were in the one- to six-leaf stage, the number of unfolded leaves increased dramatically, and the characteristics of leaves became more and more obvious so that the recognition performance improved. After the six-leaf stage, leaf overlapping and saturation resulted in the decreasing recognition accuracy. On the other hand, in **Figure 8**, the local maximal value for CNN recognition did not increase after the six-leaf stage. Moreover, a strong correlation (R^2 = 0.835) between the coefficient of the seedling counting models (**Table 1**) and DAP was found (**Figure 10**).

When a DAP was replaced by its corresponding leaf development stage, the coefficients were approximately the inverse of the number of leaves at the given stage. For example, at the one- to three-leaf stage, the coefficients on 32 and 39 DAP were 0.354 and 0.297, which were approximately the inverse of 3 (1/3 or 0.333). Similarly, at the four- to six-leaf stage, the coefficients on 53 and 58 DAP were 0.277 and 0.220, which were approximately the inverse of 5 (1/5 or 0.200). To some extent, this correlation demonstrated another perspective to understand the counting models through the leaf development stage, but more data are needed to verify the result more scientifically.

Another interesting finding related to the leaf development stage of rapeseed seedlings was that the timing of 53 DAP corresponding to the four- to six-leaf stage was near winter solstice in the Chinese calendar. Winter solstice, a meaningful solar term, is known as a significant time for winter rapeseed planting as well as for other agricultural activities. Many Chinese agronomic researchers and cultivating specialists regard winter solstice as a critical timing for rapeseed characteristic measurement. It is also considered as a mid-growth phase of rapeseed leaf development with about five leaves. In fact, the day of winter solstice was 4 days before 53 DAP. Therefore, the DAP corresponding to the leaf development stage was confirmed to be reliable in this study. Our findings offered a scientific explanation for the agricultural practice in China and should be referential for relevant studies and practices.

Influence of Parameters in Convolutional Neural Network

Influence of Patch Size

It is necessary to define patch sizes according to the specific purposes and the image target for the application of CNN. In this study, sample patches were cropped from the entire labeled UAV image. Moreover, the trained CNN model was applied to a large-scale field image. This study employed five sample patch sizes with 16, 24, 32, 40, and 48 pixels.

These sizes were determined mainly by the image resolution and the growth situation of rapeseed. Madec et al. (2019) assumed that resolution around 0.3 mm could allow to detect and count wheat ears for high-throughput phenotyping based on UAV-captured RGB images using Faster-RCNN. However, their research used imagery captured by a camera fixed on a boom, and their assumption was not verified with UAV images in their work. The imagery used in this study was captured by UAV in the field 30 m above ground level. If the altitude was more than 30 m, the resolution was not enough for recognition. If the UAV imagery was captured below 30 m, it would reduce the efficiency of field plot data collection. Furthermore, the turbulence generated by the UAV made the rapeseed leaves shake, leading to the unfocused targets in the imagery.

Meanwhile, the leaf size of rapeseed was another factor determining the patch size. In this study, an overlarge patch size such as a 128-pixel patch could contain redundant and useless information. On the contrary, an undersized patch would cause underfitting recognition results because the patch would be so small that the features between rapeseed leaves and weeds could not be distinguished. Employing a patch-based CNN algorithm (Sindagi and Patel, 2017) brought about a problem that the patch could not cover an oversized rapeseed leaf, resulting in a larger number of detected objects than the actual number called overcounting in this study (**Figure 5**). Therefore, the patch size of UAV imagery was limited by its imaging mode and resolution, field status, and the characteristics of objects.

In this study, the areas of the five patches were calculated and matched with the leaf size of rapeseed for each observation. The area of the 48-pixel patch was calculated to be 74.6 cm^2, and the maximum area of rapeseed leaf obtained on 68 DAP was almost 63 cm^2. Based on this, a sampling patch cropped with 48 pixels could meet the minimum requirement for covering an entire rapeseed leaf. However, each patch size was used for all the observations, thus the same patch size had a dynamic influence on the counting results as DAP changed. In **Figure 7**, the 16-pixel patch exhibited the optimal performance on 32 DAP, which was attributed to the fact that the 16-pixel patch could match the leaf size better on 32 DAP than on the other DAPs. On the other DAPs, the 24-pixel patch and the 32-pixel patch matched the leaf size better than the other patch sizes achieving better *F-score* in **Figure 7**.

In addition, patch size significantly affected training time. **Figure 11** shows that training time increased exponentially with patch size. Training time was more than 5,200 s (about 90 min) when using the 48-pixel patch. Although using the 16-pixel patch could save time, its results were not desirable in this study. According to **Figure 11** and based on the results of leaf recognition and seedling estimation, the 32-pixel patch was selected as the optimal sampling size balancing the performance and efficiency in this study.

Influence of Learning Rate

Learning rates ranging from 0.0001 to 0.0014 (step on 0.0001) were used to analyze the learning rate function with the 32-pixel patch on 53 DAP based on the local max value of 0.65 (**Figure 12**). *F-scores were* higher than 90.00% for most of the results. However, it was difficult to determine the relationship between learning rate and *F-score* since there was an irregular fluctuation. In terms of the ranking of *F-scores*, the top four learning rates were 0.0002, 0.0006, 0.0005, and 0.0007 with all their corresponding *F-scores* higher than 92.00%.

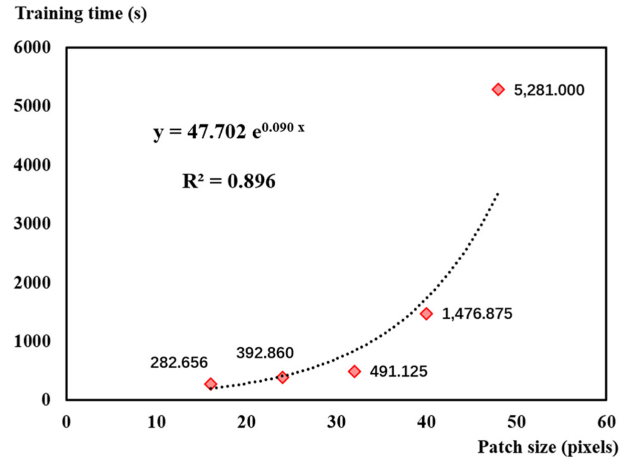


FIGURE 11 | Convolutional neural network training time consumed for five sampling patch sizes.

Learning rate plays an important role in CNN models (Trimble, 2018). It defines the weights used to adjust the gradient descent optimization. If learning rate is too small, the learning process will be slowed down and may not be close to the optimal settings. If learning rate is too large, the model may not reach the minimum boundary and produce results of null values (Trimble, 2017). Learning rate is considered as a hyper-parameter in machine learning and deep learning, which is mostly set up based on practices and empiricism (Senior et al., 2013). Finally, learning rates ranging from 0.0004 to 0.0006 were suggested in this study.

Influence of Overcounting Rate

Overcounting is common in object identification and counting tasks, especially in large-scale scenes (Li et al., 2016). It is challenging to use CNN algorithms for object detection and counting in a large-scale scene. Most studies detected flowers, leaves, and crops in the lab because the scenes were well controlled and the results were less influenced by external factors. This study detected rapeseed leaves in a field-based scene. The existing leaf overlapping was one of the main reasons for overcounting. With leaf size increasing, the overlapping became more complex, which caused serious overcounting and inaccurate counting results.

This study selected and employed different patch sizes, attempting to match the size of leaves and improve model performance. However, the overcounting was still difficult to avoid. Comparison was made between the *F-scores* with overcounting and those without overcounting based on the suggested learning rate of 0.0004 on 53 DAP (**Figure 13**). Even though two curves of *F-scores* exhibited a similar tendency, *F-scores* without overcounting calibration were higher than those with overcounting calibration. The former almost reached 100%, which was inaccurate.

A graphic example of *F-score* with and without overcounting calibration is shown in **Figure 5**. As shown in the figure, *precision* without overcounting calibration was greater

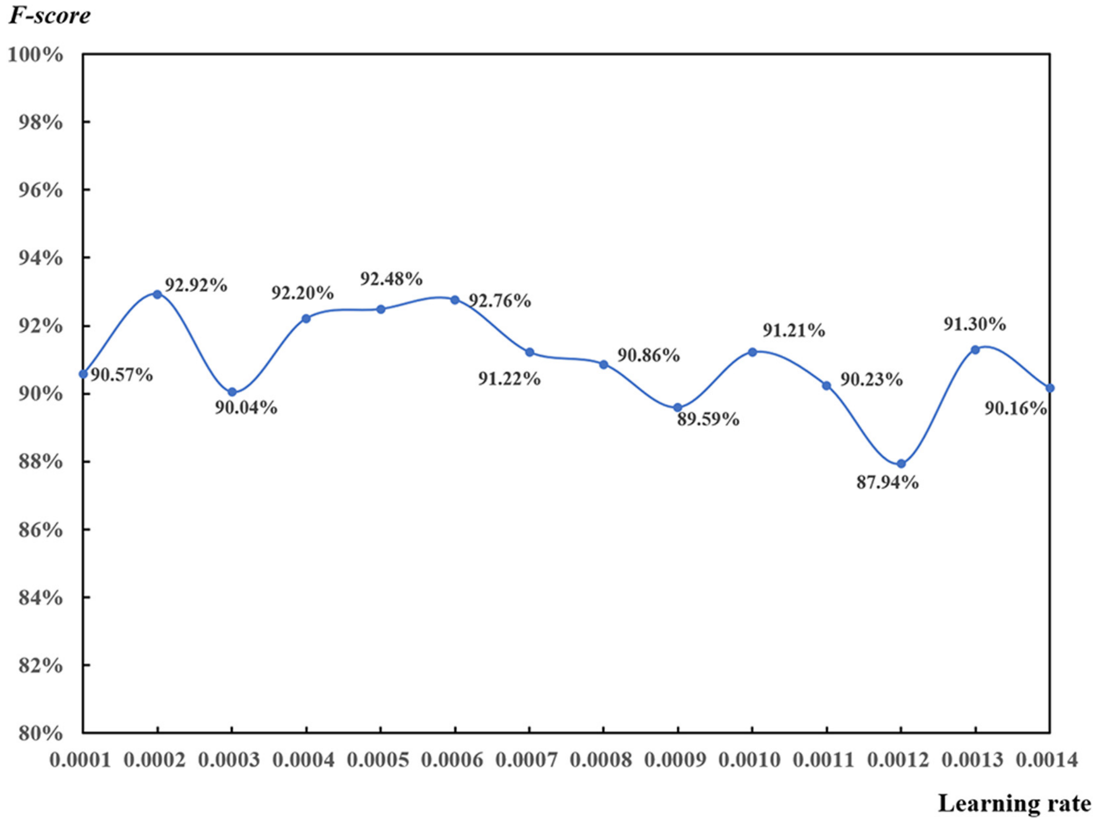


FIGURE 12 | Test of learning rate for the convolutional neural network-based rapeseed leaf recognition on 53 days after planting (32-pixel patch, local max value of 0.65).

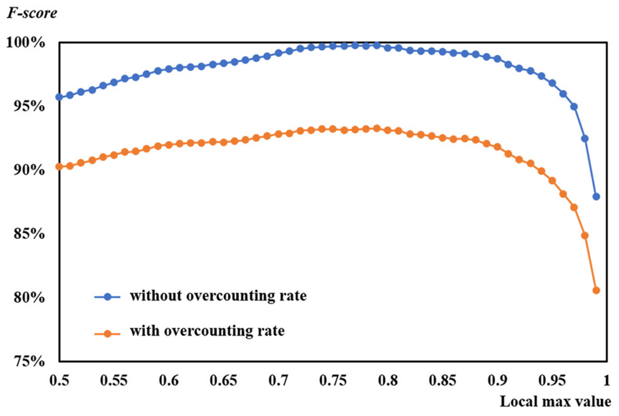


FIGURE 13 | Comparison of *F-scores* with and without overcounting rate at a learning rate of 0.0004 on 53 days after planting.

than that with overcounting calibration, resulting in an illusory high *F-score*. *F-scores* with overcounting calibration more accurately and objectively described the CNN-based leaf recognition performance. When the number of referencing masks was 29, the accurate number of detected rapeseed leaves was 28 using overcounting calibration. Without overcounting calibration, the number of detected rapeseed leaves was 38 (**Figure 5**). These results demonstrated that overcounting calibration was important to obtain precise rapeseed leaf counting. When overcounting calibration was considered, more reliable and objective rapeseed leaf counting results were achieved as demonstrated by the study.

Future Work

This study was a specific application of machine learning in agriculture for quantitative analysis of rapeseed seedling stand counting at a field level from UAV images. We aimed to estimate the rapeseed seedling count precisely and to offer a comprehensive study of field-based rapeseed seedling estimation throughout its early growth stages. Moreover, we tried to present a study showing general and user-friendly workflow for executing CNN methods. We also expected that our research offered another perspective in phenotyping and cultivation management for estimating seedling count for crops that have obvious tillering leaves at early growth stages such as soybean and potato. The results based on the particular case were desirable and promising. More detailed data and relevant technical information of our study were deposited to GitHub repository in "LARSC-Lab/Rapeseed_seedling_counting".

In the future, we will focus on the data accumulation including data from multiple growing seasons and multiple fields in different locations. Moreover, multiple approaches are expected to be employed, and their performances will be compared to identify optimal methods and improve accuracy. A more comprehensive dataset of rapeseed from UAV imagery is expected to be completed and published in the future based on our work, which will promote the research in phenotyping for rapeseed and other crops.

CONCLUSION

Utilizing a consumer-grade camera mounted on a UAV for crop phenotyping and vegetation investigation in the field is feasible and efficient. This study attempted to estimate rapeseed stand count in UAV-captured RGB imagery with machine learning. CNN algorithm was used for rapeseed leaf identification and counting. Regression modeling coupled with LOOCV method was used to establish and optimize the relationship between the seedling counting and the number of rapeseed leaves. When the number of CNN-detected rapeseed leaves was brought into seedling counting models, the results demonstrated that our proposed framework performed well and achieved great accuracy. In summary, the following conclusions can be drawn from this study:

(1) The effectiveness of our proposed CNN framework on rapeseed leaf recognition and counting was verified in this study. Overcounting is a common problem during leaf recognition and counting. The overcounting rate was related to the DAP reflecting of the rapeseed growth conditions during leaf development. CNN-recognized rapeseed leaf counting incorporated with overcounting calibration was reliable with an overall *F-score* of more than 90%. On average, 806 out of 812 plants were correctly estimated on 53 DAP at the four- to six- leaf stage. RMSE was nine plants with rRMSE of 2.22% on 53 DAP for the 32-pixel patch size, while the mean RMSE was 12 with mean rRMSE of 2.89% for all patch sizes.

(2) This study demonstrated that DAP influenced the overcounting rate, CNN-recognized leaf results, and seedling counting models. On 46 and 53 DAP, the counting models presented desirable performance. Moreover, a strong correlation ($R^2 = 0.835$) was also found between coefficients of counting models and DAP. The optimal observation period was on 53 DAP corresponding to the four- to six-leaf stage of rapeseed development.

(3) Based on machine learning approaches, this study proposed a framework for rapeseed stand counting in the field by high-resolution UAV imagery. Our future studies will focus on the collection and evaluation of multiple-year datasets to improve the robustness and reliability of the stand counting models.

DATA AVAILABILITY STATEMENT

The “Rapeseed_seedling_counting” data that support the findings of this study are available in “LARSC-Lab/Rapeseed_seedling_counting” in GitHub, which can be found at https://github.com/LARSC-Lab/Rapeseed_seedling_counting.

AUTHOR CONTRIBUTIONS

JZ, JX, and BZ designed and conducted the remote sensing part of the experiment. QL and GZ designed and conducted the agronomy part of the experiment. JZ, BZ, JX, and CH processed and analyzed the imagery as well as wrote the manuscript. CY guided the study design, advised in data analysis, and revised the manuscript. WY, YS, DZ, CW, TX, and ZJ were involved in the process of the experiment, ground data collection, or manuscript revision. All authors reviewed and approved the final manuscript.

FUNDING

This work was financially supported by the National Key Research and Development Program of China (2018YFD1000900) and the Fundamental Research Funds for the Central Universities (Grant Nos. 2662018PY101 and 2662018JC012).

ACKNOWLEDGMENTS

We acknowledge the field staff of Huazhong Agricultural University for daily management of the field experiments. We are grateful to the reviewers for their valuable comments and recommendations to improve the manuscript.

REFERENCES

Anaconda Software Distribution (2016). *Computer Software. Vers. 2-2.4.0*. Avaliable at: https://anaconda.com (accessed November, 2016).

Araus, J. L., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008

Arteta, C., Lempitsky, V., and Zisserman, A. (2016). *Counting in the Wild, in: European Conference on Computer Vision*. Berlin: Springer, 483–498.

Ballesteros, R., Ortega, J. F., Hernández, D., and Moreno, M. A. (2014). Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part I: description of image acquisition and processing. *Precis. Agric.* 15, 579–592. doi: 10.1007/s11119-014-9355-8

Bendig, J., Bolten, A., Bennertz, S., Broscheit, J., Eichfuss, S., and Bareth, G. (2014). Estimating biomass of barley using crop surface models (CSMs) derived from UAV-Based RGB imaging. *Remote Sens.* 6, 10395–10412. doi: 10.3390/rs61110395

Berrocoso, J. D., Rojas, O. J., Liu, Y., Shoulders, J., González-Vega, J. C., and Stein, H. H. (2015). Energy concentration and amino acid digestibility in high-protein canola meal, conventional canola meal, and soybean meal fed to growing pigs. *J. Anim. Sci.* 93, 2208–2217. doi: 10.2527/jas.2014-8528

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS J. Photogramm. Remote Sens.* 65, 2–16. doi: 10.1016/j.isprsjprs.2009.06.004

Bouchet, A.-S., Laperche, A., Bissuel-Belaygue, C., Snowdon, R., Nesi, N., and Stahl, A. (2016). Nitrogen use efficiency in rapeseed. A review. *Agron. Sustain. Dev.* 36:38. doi: 10.1007/s13593-016-0371-0

Bucksch, A., Burridge, J., York, L. M., Das, A., Nord, E., Weitz, J. S., et al. (2014). Image-based high-throughput field phenotyping of crop roots. *Plant Physiol.* 166, 470–486. doi: 10.1104/pp.114.243519

Chen, R., Chu, T., Landivar, J. A., Yang, C., and Maeda, M. M. (2018). Monitoring cotton (*Gossypium hirsutum* L.) germination using ultrahigh-resolution UAS images. *Precis. Agric.* 19, 161–177. doi: 10.1007/s11119-017-9508-7

Chen, Y., Ribera, J., Boomsma, C., and Delp, E. (2017). "Locating crop plant centers from UAV-based RGB imagery," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Venice, 2030–2037. doi: 10.1111/tpj.14799

Córcoles, J. I., Ortega, J. F., Hernández, D., and Moreno, M. A. (2013). Estimation of leaf area index in onion (*Allium cepa L.*) using an unmanned aerial vehicle. *Biosyst. Eng.* 115, 31–42. doi: 10.1016/j.biosystemseng.2013.02.002

Csillik, O., Cherbini, J., Johnson, R., Lyons, A., and Kelly, M. (2018). Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. *Drones* 2:39. doi: 10.3390/drones2040039

Cunliffe, A. M., Brazier, R. E., and Anderson, K. (2016). Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sens. Environ.* 183, 129–143. doi: 10.1016/j.rse.2016.05.019

Davies, E. R. (2009). The application of machine vision to food and agriculture: a review. *Imaging Sci. J.* 57, 197–217. doi: 10.1179/174313109X454756

Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., and Yan, Y. (2018). UAV-based multispectral remote sensing for precision agriculture: a comparison between different cameras. ISPRS-J. Photogramm. *Remote Sens.* 146, 124–136. doi: 10.1016/j.isprsjprs.2018.09.008

Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005

Geipel, J., Link, J., and Claupein, W. (2014). Combined spectral and spatial modeling of corn yield based on aerial images and crop surface models acquired with an unmanned aircraft system. *Remote Sens*. 6, 10335–10355. doi: 10.3390/rs61110335

Ghanem, M. E., Marrou, H., and Sinclair, T. R. (2015). Physiological phenotyping of plants for crop improvement. *Trends Plant Sci.* 20, 139–144. doi: 10.1016/j.tplants.2014.11.006

Gnädinger, F., and Schmidhalter, U. (2017). Digital counts of maize plants by unmanned aerial vehicles (UAVs). *Remote Sens.* 9:544. doi: 10.3390/rs9060544

Godfray, H. C. J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., et al. (2010). Food security: the challenge of feeding 9 billion people. *Science* 327, 812–818. doi: 10.1126/science.1185383

Gomes, M. F., Maillard, P., and Deng, H. (2018). Individual tree crown detection in sub-meter satellite imagery using marked point processes and a geometrical-optical model. *Remote Sens. Environ.* 211, 184–195. doi: 10.1016/j.rse.2018.04.002

Grinblat, G. L., Uzal, L. C., Larese, M. G., and Granitto, P. M. (2016). Deep learning for plant identification using vein morphological patterns. *Comput. Electron. Agric.* 127, 418–424. doi: 10.1016/j.compag.2016.07.003

Huang, Y., Thomson, S. J., Brand, H. J., and Reddy, K. N. (2016). Development and evaluation of low-altitude remote sensing systems for crop production management. *Int. J. Agric. Biol. Eng.* 9, 1–11.

Jin, X., Liu, S., Baret, F., Hemerlé, M., and Comar, A. (2017). Estimates of plant density of wheat crops at emergence from very low altitude UAV imagery. *Remote Sens. Environ.* 198, 105–114. doi: 10.1016/j.rse.2017.06.007

Lancashire, P. D., Bleiholder, H., Boom, T. V. D., Langelüddeke, P., Stauss, R., Weber, E., et al. (1991). A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* 119, 561–601.

Lee, S. H., Chan, C. S., Mayo, S. J., and Remagnino, P. (2017). How deep learning extracts and learns leaf features for plant classification. *Pattern Recognit.* 71, 1–13.

Li, W., Fu, H., Yu, L., and Cracknell, A. (2016). Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* 9:22. doi: 10.3390/rs9010022

Linker, R., Cohen, O., and Naor, A. (2012). Determination of the number of green apples in RGB images recorded in orchards. *Comput. Electron. Agric.* 81, 45–57. doi: 10.1016/j.compag.2011.11.007

Liu, S., Baret, F., Andrieu, B., Burger, P., and Hemmerlé, M. (2017). Estimation of wheat plant density at early stages using high resolution imagery. *Front. Plant Sci.* 8:739. doi: 10.3389/fpls.2017.00739

Liu, T., Li, R., Jin, X., Ding, J., Zhu, X., Sun, C., et al. (2017). Evaluation of seed emergence uniformity of mechanically sown wheat with UAV RGB imagery. *Remote Sens.* 9:1241. doi: 10.3390/rs9121241

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., and Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS J. Photogramm. Remote Sens.* 130, 277–293. doi: 10.1016/j.isprsjprs.2017.06.001

Madec, S., Jin, X., Lu, H., De Solan, B., Liu, S., Duyme, F., et al. (2019). Ear density estimation from high resolution RGB imagery using deep learning technique. *Agric. For. Meteorol.* 264, 225–234.

Maimaitijiang, M., Ghulam, A., Sidike, P., Hartling, S., Maimaitiyiming, M., Peterson, K., et al. (2017). (UAS)-based phenotyping of soybean using multi-sensor data fusion and extreme learning machine. *ISPRS J. Photogramm. Remote Sens.* 134, 43–58. doi: 10.1016/j.isprsjprs.2017.10.011

McCouch, S., Baute, G. J., and Bradeen, J. (2013). Agriculture: feeding the future. *Nature* 499, 23–24.

Meyer, G. E., and Neto, J. C. (2008). Verification of color vegetation indices for automated crop imaging applications. *Comput. Electron. Agric.* 63, 282–293. doi: 10.1016/j.compag.2008.03.009

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007

Naito, H., Ogawa, S., Valencia, M. O., Mohri, H., Urano, Y., Hosoi, F., et al. (2017). Estimating rice yield related traits and quantitative trait loci analysis under different nitrogen treatments using a simple tower-based field phenotyping system with modified single-lens reflex cameras. *ISPRS J. Photogramm. Remote Sens.* 125, 50–62. doi: 10.1016/j.isprsjprs.2017.01.010

Pound, M. P., Atkinson, J. A., Wells, D. M., Pridmore, T. P., and French, A. P. (2017). "Deep learning for multi-task plant phenotyping," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, 2055–2063.

Qureshi, W. S., Payne, A., Walsh, K. B., Linker, R., Cohen, O., and Dailey, M. N. (2017). Machine vision for counting fruit on mango tree canopies. *Precis. Agric.* 18, 224–244. doi: 10.1007/s11119-016-9458-5

Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the study of plant growth and development. *Front. Plant Sci* 6:619. doi: 10.3389/fpls.2015.00619

Ramos, P. J., Prieto, F. A., Montoya, E. C., and Oliveros, C. E. (2017). Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.* 137, 9–22. doi: 10.1016/j.compag.2017.03.010

Ribera, J., Chen, Y., Boomsma, C., and Delp, E. J. (2017). "Counting plants using deep learning," in *Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP). Presented at the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Montreal, QC: IEEE, 1344–1348. doi: 10.1109/GlobalSIP.2017.8309180

Sadras, V. O., Rebetzke, G. J., and Edmeades, G. O. (2013). The phenotype and the components of phenotypic variance of crop traits. *Field Crop. Res.* 154, 255–259. doi: 10.1016/j.fcr.2013.10.001

Sankaran, S., Quirós, J. J., Knowles, N. R., and Knowles, L. O. (2017). High-resolution aerial imaging based estimation of crop emergence in potatoes. *Am. J. Potato Res.* 94, 658–663. doi: 10.1007/s12230-017-9604-2

Schowengerdt, R. A. (2012). *Techniques for Image Processing and Classifications in Remote Sensing*. Cambridge, MA: Academic Press.

Senior, A., Heigold, G., Ranzato, M., and Yang, K. (2013). "An empirical study of learning rates in deep neural networks for speech recognition," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Presented at the ICASSP 2013 - 2013 IEEE International Conference* *on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC: IEEE, 6724–6728.

Sheng, B., Shen, C., Lin, G., Li, J., Yang, W., and Sun, C. (2016). Crowd counting via weighted vlad on dense attribute feature maps. *IEEE Trans. Circ. Syst. Video Technol.* 28, 1788–1797.

Shi, Y., Thomasson, J. A., Murray, S. C., Pugh, N. A., Rooney, W. L., Shafian, S., et al. (2016). Unmanned aerial vehicles for high-throughput phenotyping and agronomic research. *PLoS One* 11:e0159781. doi: 10.1371/journal.pone.0159781

Shi, Y., Wang, N., Taylor, R. K., Raun, W. R., and Hardin, J. A. (2013). Automatic corn plant location and spacing measurement using laser line-scan technique. *Precis. Agric.* 14, 478–494. doi: 10.1007/s11119-013-9311-z

Simonyan, K., and Zisserman, A. (2014). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. *arXiv preprint* arXiv:1409.1556. Available online at: http://arxiv.org/abs/1409.1556

Sindagi, V. A., and Patel, V. M. (2017). A survey of recent advances in cnn-based single image crowd counting and density estimation. *Pattern Recognit. Lett.* 107, 3–16.

Trimble (2017). *Tutorial 7 - Convolutional Neural Networks in eCognition.pdf*. Germany: Trimble Inc.

Trimble (2018). *eCognition Developer 9.3 Reference Book*. Germany: Trimble Inc.

Van Iersel, W., Straatsma, M., Addink, E., and Middelkoop, H. (2018). Monitoring height and greenness of non-woody floodplain vegetation with UAV time series. *ISPRS J. Photogramm. Remote Sens*. 141, 112–123. doi: 10.1016/j.isprsjprs.2018.04.01

Varela, S., Dhodda, P., Hsu, W., Prasad, P. V., Assefa, Y., Peralta, N., et al. (2018). UAS) and supervised learning techniques. *Remote Sens.* 10:343. doi: 10.3390/rs10020343

Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., and Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* 152, 654–664. doi: 10.1016/j.rse.2014.06.006

Walter, A., Studer, B., and Kölliker, R. (2012). Advanced phenotyping offers opportunities for improved breeding of forage and turf species. *Ann. Bot.* 110, 1271–1279. doi: 10.1093/aob/mcs026

Wang, X., Liu, G., Yang, Q., Hua, W., Liu, J., and Wang, H. (2010). Genetic analysis on oil content in rapeseed (*Brassica napus* L.). *Euphytica* 173, 17–24. doi: 10.1007/s10681-009-0062-x

Weber, E., and Bleiholder, H. (1990). Erläuterunger zu den BBCH–Dezimal-codes für die entwicklungsstadien von mais, raps, faba-bohne, sonnenblume und erbse-mit abbildungen. *Gesunde Pflanz* 42, 308–321.

Wendel, A., and Underwood, J. (2017). Illumination compensation in ground based hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* 129, 162–178. doi: 10.1016/j.isprsjprs.2017.04.010

White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crop. Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003

Xiong, X., Duan, L., Liu, L., Tu, H., Yang, P., Wu, D., et al. (2017). Panicle-SEG: a robust image segmentation method for rice panicles in the field based on deep learning and superpixel optimization. *Plant Methods* 13:104. doi: 10.1186/s13007-017-0254-7

Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8:1111. doi: 10.3389/fpls.2017.01111

Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087

Yu, N., Li, L., Schmitz, N., Tian, L., Greenberg, J. A., and Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. *Remote Sens. Environ.* 187, 91–101. doi: 10.1016/j.rse.2016.10.005

Zanotta, D. C., Zortea, M., and Ferreira, M. P. (2018). A supervised approach for simultaneous segmentation and classification of remote sensing images. *ISPRS J. Photogramm. Remote Sens.* 142, 162–173. doi: 10.1016/j.isprsjprs.2018.05.021

Zhang, C., and Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: a review. *Precis. Agric.* 13, 693–712. doi: 10.1007/s11119-012-9274-5

Zhang, C., Li, H., Wang, X., and Yang, X. (2015). "Cross-scene crowd counting via deep convolutional neural networks," in *Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference On*, Piscataway, NJ: IEEE, 833–841.

Zhang, Y., Zhou, D., Chen, S., Gao, S., and Ma, Y. (2016). "Single-image crowd counting via multi-column convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, 589–597.

Zhao, B., Zhang, J., Yang, C., Zhou, G., Ding, Y., Shi, Y., et al. (2018). Rapeseed seedling stand counting and seeding performance evaluation at two early growth stages based on unmanned aerial vehicle imagery. *Front. Plant Sci.* 9:1362. doi: 10.3389/fpls.2018.01362

Zheng, S., Wang, L., Wan, N., Zhong, L., Zhou, S., He, W., et al. (2016). Response of potato tuber number and spatial distribution to plant density in different growing seasons in Southwest China. *Front. Plant Sci.* 7:365. doi: 10.3389/fpls.2016.00365

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership