



OPEN ACCESS

EDITED BY
Xukun Yin,
Xidian University, China

REVIEWED BY
Wei Ren,
Xi'an University of Posts and
Telecommunications, China
Jingjing Yang,
Hebei North University, China

*CORRESPONDENCE
Xiaojun Jin,
✉ xiaojun.jin@pku-iaas.edu.cn

RECEIVED 15 September 2024
ACCEPTED 10 January 2025
PUBLISHED 27 January 2025

CITATION
Jin H, Han K, Xia H, Xu B and Jin X (2025)
Detection of weeds in vegetables using image
classification neural networks and image
processing.
Front. Phys. 13:1496778.
doi: 10.3389/fphy.2025.1496778

COPYRIGHT
© 2025 Jin, Han, Xia, Xu and Jin. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Detection of weeds in vegetables using image classification neural networks and image processing

Huiping Jin¹, Kang Han^{2,3}, Hongting Xia³, Bo Xu¹ and
Xiaojun Jin^{2*}

¹Engineering Training Center, Nanjing Forestry University, Nanjing, Jiangsu, China, ²Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, Shandong, China, ³Jurong Institute of Smart Agriculture, Zhenjiang, Jiangsu, China

Weed management presents a major challenge to vegetable growth. Accurate identification of weeds is essential for automated weeding. However, the wide variety of weed types and their complex distribution creates difficulties in rapid and accurate weed detection. In this study, instead of directly applying deep learning to identify weeds, we first created grid cells on the input images. Image classification neural networks were utilized to identify the grid cells containing vegetables and exclude them from further analysis. Finally, image processing technology was employed to segment the non-vegetable grid images based on their color features. The background grid cells, which contained no green pixels, were identified, while the remaining cells were labeled as weed cells. EfficientNet, GoogLeNet, and ResNet models achieved overall accuracies of over 0.956 in identifying vegetables in the testing dataset, demonstrating exceptional identification performance. Among these models, the ResNet model exhibited the highest computational efficiency, with a classification time of 12.76 ms per image and a corresponding frame rate of 80.31 fps, satisfying the requirement for real-time weed detection. Effectively identifying vegetables and differentiating weeds from soil significantly reduces the complexity of weed detection and improves its accuracy.

KEYWORDS

weed detection, deep learning, image classification neural networks, image processing, weed management

1 Introduction

Vegetables constitute one of the largest categories of crops in Chinese agriculture. Weeds compete with vegetables for growth resources, resulting in reduced yields. Statistics reveal that there are over 1,400 weed species in vegetable fields, with more than 130 species causing significant harm to vegetables [1]. For example, the yield of tomato (*Solanum lycopersicum* L.) was reduced by 48%–71% due to competition with weeds, including common cocklebur (*Xanthium strumarium* L.), crabgrass (*Digitaria sanguinalis* (L.) Scop.), jimsonweed (*Datura stramonium* L.), and tall morning glory (*Ipomoea purpurea* (L.) Roth) [2]. With a density of 65 broadleaf weeds per square meter, the yield of lettuce (*Lactuca sativa* L.) declined by more than 90% [3]. Effective weed management is critical for enhancing crop yield and quality [4, 5].

At present, the control of weeds in vegetable fields is heavily dependent on manual weeding, a method that is both labor-intensive and inefficient [6]. As labor costs

continue to escalate, the overall expense of vegetable cultivation rises accordingly [2]. This increase significantly impacts the sustainable development of the vegetable industry [7]. Consequently, there is an urgent need for the research and promotion of innovative and environmentally friendly weeding techniques, such as precision mechanical weeding or electric weeding [1, 2, 8]. With the advancement of precision agriculture technology, a growing number of researchers have initiated studies on intelligent weeding equipment [8–10]. However, the ability to rapidly and accurately identify weeds is critical for achieving effective precision weeding.

Numerous researchers have conducted extensive studies on weed identification [11–14]. Herrmann et al. utilized hyperspectral imaging technology to acquire images of wheat fields and distinguish wheat from weeds using the least squares classifier [15]. Nieuwenhuizen et al. combined color and texture features to classify sugar beets under fixed and changing lighting conditions using an adaptive Bayesian classifier, achieving recognition rates of 89.8% and 67.7%, respectively [16]. Deng et al. extracted 48 features, including color, shape, and texture, from maize leaves and designed a model using Support Vector Machines (SVMs) for weed identification [17]. Traditional weed identification methods typically use features such as color, shape, texture, and spatial distribution, or combinations thereof, and employ techniques such as wavelet analysis, Bayesian discriminant models, and SVMs to distinguish crops from weeds [4, 18]. These methods require manual feature selection, rely heavily on designer expertise, and are susceptible to small sample sizes and human subjectivity [19]. Due to factors such as lighting changes, background noise, and diverse target shapes, it is difficult to design feature extraction models with high adaptability and stability. With the development of deep learning technology, deep convolutional neural networks have been widely applied in the field of machine vision and have achieved excellent results [12, 20].

With the advancement of deep learning technologies, Convolutional Neural Networks (CNNs) have seen widespread application in various research fields, including Natural Language Processing (NLP) [21], computer vision [22], and speech recognition [23]. In recent years, research on weed identification based on deep learning has also been extensively conducted [24]. For example, Adel et al. used an SVM and an Artificial Neural Network to identify four common weeds associated with sugar beet crops, achieving an accuracy of 95% [25]. Wang et al. proposed a recognition method based on the Shift Window Transformer network, which can be used to identify corn and weeds during the seedling stage. They used four models for real-time recognition, with an average recognition rate of 95.04% [26]. Mu et al. used the Faster R-CNN model to identify weeds in images of cropping areas, with a recognition accuracy of >95% [27].

Detecting weeds using deep convolutional neural networks requires a substantial amount of images of diverse weed species for training the model, resulting in a high cost of constructing a training dataset [28]. Prior studies indicate that deep learning methods can only accurately identify a limited number of weed species [29]. To simplify weed detection and enhance its accuracy, this research presented a method for identifying weeds growing in vegetables by combining convolutional neural networks and image processing technology. The proposed method first uniformly divided the original images captured by the camera into various grid images. Afterward, the deep convolutional neural network models were

utilized to identify the grid images containing vegetables, while the remaining images were further divided into weed and soil background grids via image segmentation technology. During the neural network training, images with vegetables were labeled as true positives, while those without were considered true negatives. In other words, plant images without vegetables were marked as containing weeds, allowing the neural network model to focus specifically on identifying vegetable targets. This approach effectively reduced the cost of building the training dataset covering various weed species and enhanced the model's recognition robustness. Additionally, the size of the grid images corresponds to the operational range of intelligent weeding equipment, allowing for the seamless integration of intelligent weed recognition algorithms with practical weeding operations. The objectives of this study were to 1) assess and compare the performance of various deep learning models in detecting vegetables and 2) explore the feasibility of integrating image classification neural networks with image processing techniques for vegetable and weed detection.

2 Materials and methods

2.1 Image classification neural networks

Three image classification deep convolutional neural networks, including EfficientNet [30], GoogLeNet [31], and ResNet [32], were selected for evaluating the feasibility of using deep convolutional neural networks for detecting weeds growing in vegetables. EfficientNet, GoogLeNet, and ResNet are three of the most influential architectures in the field of deep learning, particularly in the context of image recognition and classification tasks. Their unique approaches to network design and scalability have contributed significantly to advancements in computer vision.

EfficientNet is a family of convolutional neural networks (CNNs) that sets new standards for efficiency, achieving much higher accuracy with significantly fewer parameters compared to previous models. Introduced by Mingxing Tan and Quoc V. Le in 2019, EfficientNet utilizes a novel scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient, allowing for a balanced network expansion [30]. This methodical scaling up results in models that are both lightweight and high-performing, making EfficientNet ideal for applications where computational resources are limited.

GoogLeNet, also known as Inception-v1, was introduced by researchers at Google in 2014 [31]. It is notable for its inception module, which allows the network to choose from filters of various sizes in each layer. This architecture enables the model to capture information at various scales effectively. GoogLeNet significantly increased the depth of networks while maintaining computational efficiency, winning the 2014 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) with a record-low error rate. Its design minimizes the use of computational resources by incorporating dimensionality reduction techniques, allowing for efficient processing and less overfitting.

ResNet, short for Residual Networks, introduced by He et al. [32], revolutionized deep learning by enabling the training of extremely deep neural networks. Through the introduction of residual blocks, ResNet allows for the easy flow

of gradients throughout the network, solving the vanishing gradient problem that plagued earlier architectures. This innovation enables the construction of networks with hundreds, or even thousands, of layers, achieving unprecedented performance on various benchmarks. ResNet's design is simple yet effective, significantly reducing training time without compromising on accuracy.

EfficientNet, GoogLeNet, and ResNet each offer unique advantages in the realm of image recognition. EfficientNet provides a highly efficient model scalable across different resource constraints while maintaining accuracy. GoogLeNet introduces a versatile architecture capable of capturing complex patterns through its inception modules. ResNet enables the training of very deep networks by addressing the vanishing gradient problem, leading to remarkable improvements in accuracy. These architectures complement each other, covering a broad spectrum of needs in image processing tasks, from efficiency and versatility to depth and complexity. Consequently, this study selected these three models to investigate their effectiveness in recognizing vegetables and weeds, leveraging their distinct strengths to address the challenges of this specific application.

2.2 Image acquisition

A total of 600 original images of bok choy (*brassica rapa* spp. *chinensis*) and weeds were acquired in September 2020 from multiple vegetable fields in Baguazhou, Nanjing, Jiangsu Province, China (32.12°N, 118.48°E). To ensure dataset diversity, the collected images encompassed variations in environment, lighting, and growth stages of vegetables and weeds. A digital camera (HV1300FC, DaHeng Image, Inc., Beijing, China) was used to capture vertical shots from approximately 60 cm above the ground. The images had a resolution of $1,792 \times 1,344$ pixels and were in JPEG format. The collected images were uniformly divided into 48 grid images in a 6-row and 8-column arrangement, with each grid image having a resolution of 224×224 pixels. These grid images can be categorized into four types: those containing only soil (Figure 1A), those containing only vegetables (Figure 1B), those containing vegetables and weeds (Figure 1C), and those containing only weeds (Figures 1D–F). After being manually categorized, grid images were used as the image data for training and testing the neural network models.

2.3 Training and testing

In the process of data preparation, the acquired original images were divided into 48 grid images arranged in a 6-row and 8-column layout, with each grid image having a resolution of 224×224 pixels. This grid division was implemented using a custom Python script, which provided flexibility and ensured compatibility with the experimental framework. During the manual inspection and classification phase, grid images were categorized into two classes: those featuring only vegetables or both vegetables and weeds were grouped into the “vegetable” class, while those without vegetables were placed in the “background” class. This classification resulted in a total of 6,200 samples for each class. To ensure robust model training and evaluation, these samples were then randomly allocated

into training, validation, and test sets in proportional amounts. Specifically, the training set comprised 5,000 samples from each class, the validation or testing set included 600 samples from each class.

To rigorously evaluate the models' performance and ensure their ability to generalize across diverse conditions, three subsets (Datasets 1, 2, and 3) were created using a standard three-fold cross-validation approach. Each subset was randomly sampled from a diverse dataset that includes variations in lighting conditions, weed distributions, and vegetable growth stages. By dividing the original dataset into three subsets, this approach allows each model to be trained and tested on different combinations of data, thereby enhancing the reliability of the evaluation. This diversity ensures that the datasets reflect a wide range of real-world scenarios, further improving the robustness and generalizability of the models. Each dataset contains 5,000 training samples and 600 validation or testing samples per class.

In this study, image classification neural networks underwent training and testing within the PyTorch (version 1.8.1) deep learning framework (available at <https://pytorch.org/>; developed by Facebook, San Jose, California, United States) on an NVIDIA GeForce RTX 2080 Ti graphics processing unit (GPU). The initialization of weights and biases utilized transfer learning, employing pre-trained CNNs from the ImageNet dataset [33, 34]. To ensure a fair comparison across the evaluated deep learning models, we used consistent hyperparameter settings, as outlined in Table 1. Specifically, we employed the cross-entropy loss function, which is well-suited for classification tasks, and optimized the networks using Stochastic Gradient Descent (SGD) with a momentum of 0.9. All models were trained with a base learning rate of 0.001, adjusted dynamically during training using a LambdaLR scheduler. To mitigate overfitting, regularization techniques, including weight decay (L2 regularization) with a factor of $1e-4$, were applied. Each model was trained for 24 epochs with a batch size of 16.

The performance of the neural networks for image classification was assessed using a binary classification confusion matrix, which delineated four outcomes: true positives (tp), true negatives (tn), false positives (fp), and false negatives (fn). To evaluate the effectiveness of the image classification neural networks, we employed four classification metrics, including precision, recall, overall accuracy, and the F_1 score.

Precision measures the accuracy of positive predictions made by the neural network. It is the ratio of correctly identified positive samples to all samples predicted as positive (Equation 1) [35]. High precision indicates fewer false positives.

$$\text{precision} = \frac{tp}{tp + fp} \quad (1)$$

Recall, also known as sensitivity or true positive rate, measures the neural network's ability to identify all positive instances correctly. It is the ratio of correctly identified positive samples to all actual positive samples (Equation 2) [35]. High recall indicates fewer false negatives.

$$\text{recall} = \frac{tp}{tp + fn} \quad (2)$$

Overall accuracy represents the proportion of correctly classified samples (both true positives and true negatives) among all samples

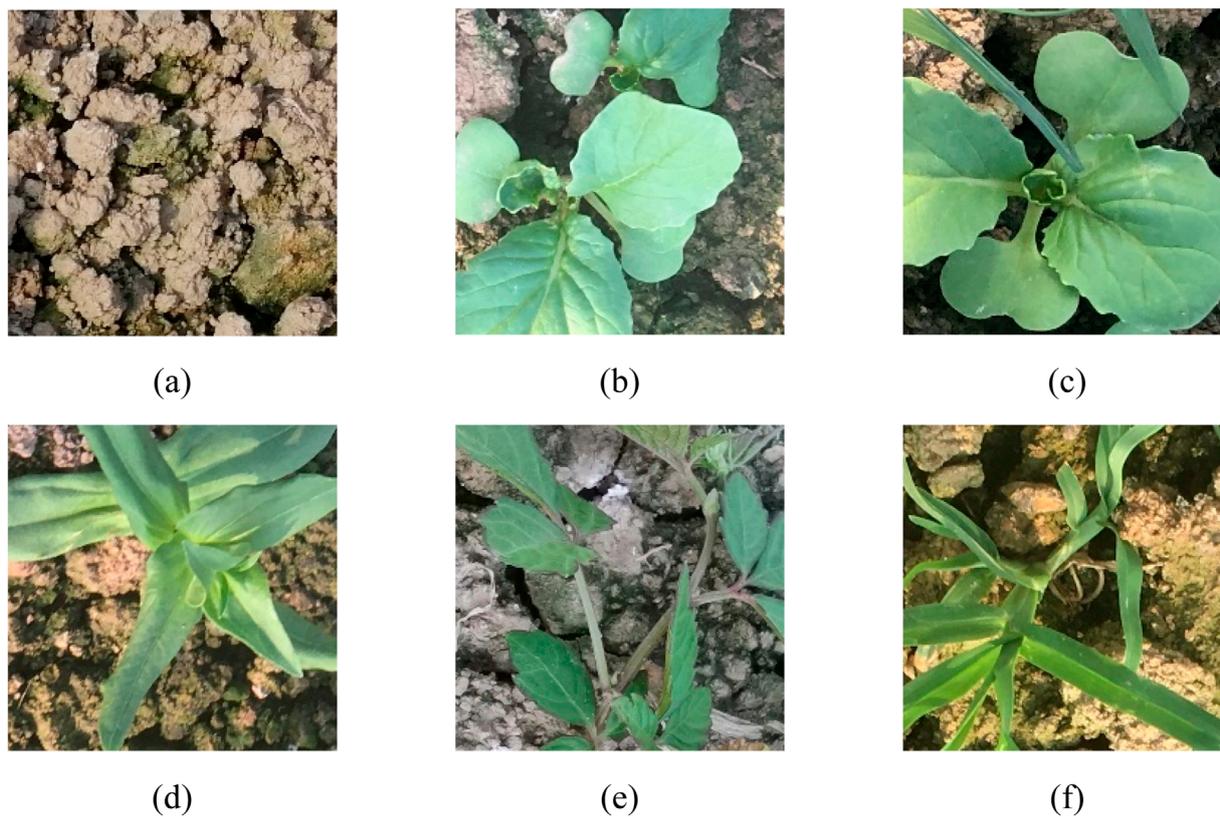


FIGURE 1 Sample grid images. (A) Soil only (B) bok choy only (C) bok choy and weed (D–F) weed only.

TABLE 1 Hyper-parameters used for training the neural networks.

Neural networks	Optimizer	Base learning rate	Learning rate policy	Batch size	Training epochs
EfficientNet	SGD	0.001	LambdaLR	16	24
GoogLeNet	SGD	0.001	LambdaLR	16	24
ResNet	SGD	0.001	LambdaLR	16	24

(Equation 3) [35]. It is a general measure of the model’s correctness.

$$\text{Overall accuracy} = \frac{tp + tn}{tp + fp + fn + tn} \tag{3}$$

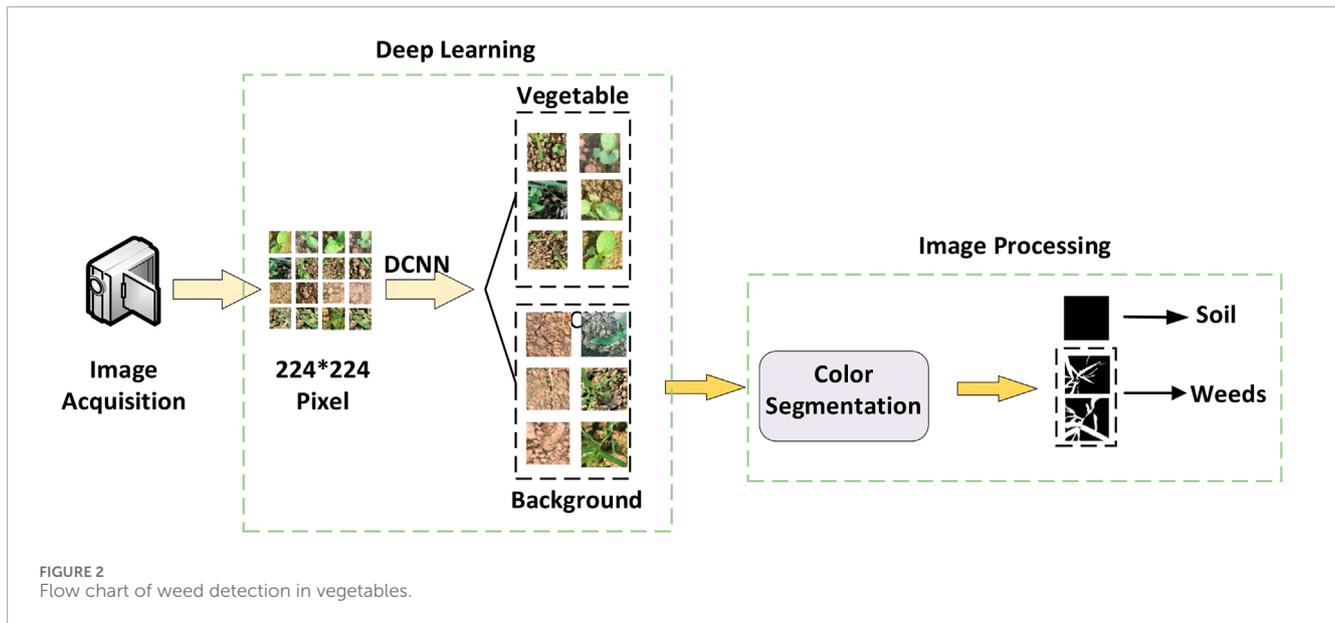
The F_1 score is the harmonic mean of precision and recall (Equation 4) [35]. It provides a balance between precision and recall, especially when there is an imbalance between positive and negative samples. It is a useful metric for binary classification tasks.

$$F_1 \text{ score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{4}$$

In the context of image processing speed, frames per second (FPS) measures the number of frames (images) processed by the neural network per second. It indicates the neural network’s real-time processing capability, which is crucial for applications requiring fast decision-making based on images or video streams.

2.4 Weed detection

In this study, vegetables were identified to indirectly detect weeds. Initially, image classification neural networks were used to identify grid images containing vegetables. The remaining grid images, which included either weeds or soil, were processed using image processing techniques to identify green pixels within the grid; those containing green pixels were classified as weeds, while those without were classified as soil. Specifically, during the training phase of the classification neural network, images containing vegetables were categorized as true positives, and all others as true negatives. Images of green plants that did not include vegetables were identified as containing weeds. The process for weed detection is illustrated in Figure 2, which can be divided into three stages:



- 1) Grid division. The original images collected from the vegetable fields were uniformly divided into 48 grid images in a 6-row and 8-column layout, with each grid image having a resolution of 224×224 pixels.
- 2) Vegetable identification. Image classification neural networks were trained to identify all grid images containing vegetables.
- 3) Weed detection. Color features were used to identify soil background images that do not contain green pixels, with the remaining grid images being classified as weeds.

Weeds and vegetables have similar colors but differ significantly from the soil background. After the neural networks identified grid images containing vegetables, color segmentation technology was used to distinguish images that contained only soil from those that contained weeds, thereby achieving final weed detection. Both vegetables and weeds appear green in images, while soil appears yellow-brown. This study employed a color index proposed by Jin et al. [7] to segment green plants, with conditional modifications made to further improve image segmentation results. As shown in Equation 5, in the RGB color space, the G (green) component of vegetables and weeds is greater than the R (red) or B (blue) components. By iterating over each pixel in the grid image, it first checks if the G component is less than either the R or B component. If so, the pixel value is set to 0 (background). Otherwise, the pixel value is calculated according to the color index.

$$index = \begin{cases} 0, & \text{if } (G < R \text{ or } G < B) \\ -19R + 24G - 2B > 862, & \text{otherwise} \end{cases} \quad (5)$$

During the image processing, the segmented images may contain several noise points due to the influence of background color values. Area filtering was employed on the segmented images to eliminate noise points and enhance the segmentation results. By calculating and labeling the pixel connected regions, areas below a certain area threshold were marked as noise and removed from the image.

3 Results

3.1 Performances of the neural networks

Table 2 presents the performances of various image classification neural network architectures, including EfficientNet, GoogLeNet, and ResNet, on detecting and classifying sub-images containing vegetables in validation datasets. EfficientNet showcased remarkable performance with a precision of 0.969 and recall of 0.982 for bok choy recognition, achieving an overall accuracy and F_1 score of 0.975. This suggested that EfficientNet not only accurately identified bok choy but also minimized false positives, effectively distinguishing bok choy from the background. GoogLeNet demonstrated slightly lower metrics compared to EfficientNet, with precision and recall for bok choy at 0.952 and 0.965, respectively, and an overall accuracy and F_1 score of 0.958. Although GoogLeNet was highly capable, it showed a marginal reduction in performance metrics relative to EfficientNet. ResNet exhibited a high recall of 0.988 for background identification but a slightly lower recall of 0.950 for bok choy. Its precision for bok choy was the highest at 0.988, with overall accuracy and F_1 score both at 0.969. This indicated ResNet's strong ability to correctly identify bok choy images, albeit with a slight increase in false negatives compared to EfficientNet.

The performance metrics of various image classification neural networks in the testing datasets exhibited a slight decline compared to the validation datasets. However, they still maintained commendable recognition capabilities (Table 3). EfficientNet demonstrated balanced performance with precision scores of 0.979 for the background and 0.955 for bok choy, and recall scores of 0.953 for the background and 0.980 for bok choy. The overall accuracy and F_1 scores never fell below 0.966, indicating consistent and reliable recognition capability across both classes. GoogLeNet showed slightly lower precision and recall scores compared to EfficientNet, with precision scores of 0.966 for the background and 0.946 for bok choy, and recall scores of 0.945 for the background and 0.967 for bok choy. The overall

TABLE 2 Evaluation matrix of image classification neural networks in validation datasets.

Neural networks	Labels	Precision	Recall	Overall accuracy	F ₁ score
EfficientNet	Background	0.981	0.968	0.975	0.974
	Bok choy	0.969	0.982	0.975	0.975
GoogLeNet	Background	0.965	0.952	0.958	0.958
	Bok choy	0.952	0.965	0.958	0.958
ResNet	Background	0.952	0.988	0.969	0.970
	Bok choy	0.988	0.950	0.969	0.969

TABLE 3 Evaluation matrix of image classification neural networks in testing datasets.

Neural networks	Labels	Precision	Recall	Overall accuracy	F ₁ score
EfficientNet	Background	0.979	0.953	0.967	0.966
	Bok choy	0.955	0.980	0.967	0.967
GoogLeNet	Background	0.966	0.945	0.956	0.955
	Bok choy	0.946	0.967	0.956	0.956
ResNet	Background	0.952	0.982	0.966	0.967
	Bok choy	0.981	0.950	0.966	0.965

accuracy and F₁ scores were approximately 0.956, suggesting a slight decrease in model performance for distinguishing between background and bok choy compared to EfficientNet. ResNet, in this comparison, demonstrated a well-rounded performance with an overall accuracy and F₁ score of 0.966. Notably, its precision for bok choy, the highest among the three neural networks at 0.981, suggested an exceptional ability to accurately identify bok choy with minimal false positives. Furthermore, its recall rate for the background was 0.982, underscoring its effectiveness in background classification, albeit with a slightly increased rate of false negatives for bok choy. Generally, EfficientNet and ResNet both exhibited exceptional capabilities in recognizing bok choy compared to GoogLeNet.

Figure 3 presents the confusion matrices for the EfficientNet, GoogLeNet, and ResNet models in the testing datasets. It was observed that the GoogLeNet model incorrectly identified 33 background images as vegetable images, while mistaking 20 vegetable images for background images. The lower recognition rate of the EfficientNet model was primarily due to 28 background images being erroneously classified as vegetables. Compared to vegetable recognition, the ResNet model was better at identifying non-vegetables, with only 11 background images misidentified as vegetables, and 30 vegetable images misclassified as background.

To further validate the performance of the models, we conducted three-fold cross-validation experiments. As shown in Table 4, all three neural networks, including EfficientNet, GoogLeNet, and ResNet, demonstrated consistently high performance across the

different datasets. EfficientNet achieved average overall accuracies of 96.6%, 96.9%, and 95.6% on Datasets 1, 2, and 3, respectively, with corresponding F₁ scores reflecting similar trends. GoogLeNet attained accuracies ranging from 94.8% to 96.1%, while ResNet consistently performed well with accuracies between 95.2% and 97.1%. The precision and recall values for each model remained high across all datasets, indicating reliable classification capabilities. These results confirm the robustness of our findings and further support the effectiveness of the proposed method for weed detection in vegetables.

3.2 Detection speeds of the neural networks

Rapid detection capabilities are essential for neural networks to enable real-time precision weeding. The average detection speed, measured in terms of FPS and milliseconds, was calculated using images from the testing dataset (1,200 images). This large-scale testing approach ensures that the reported detection speeds are representative and reliable. Table 5 shows the detection speeds of the EfficientNet, GoogLeNet, and ResNet models for identifying vegetables. Since the grid images were obtained by uniformly dividing the original images taken by the camera into a 6-row and 8-column layout (each original image divided into 48 grid images), the batch size for calculating detection speed can be set to 48. This yielded the detection speed for the original images. According to

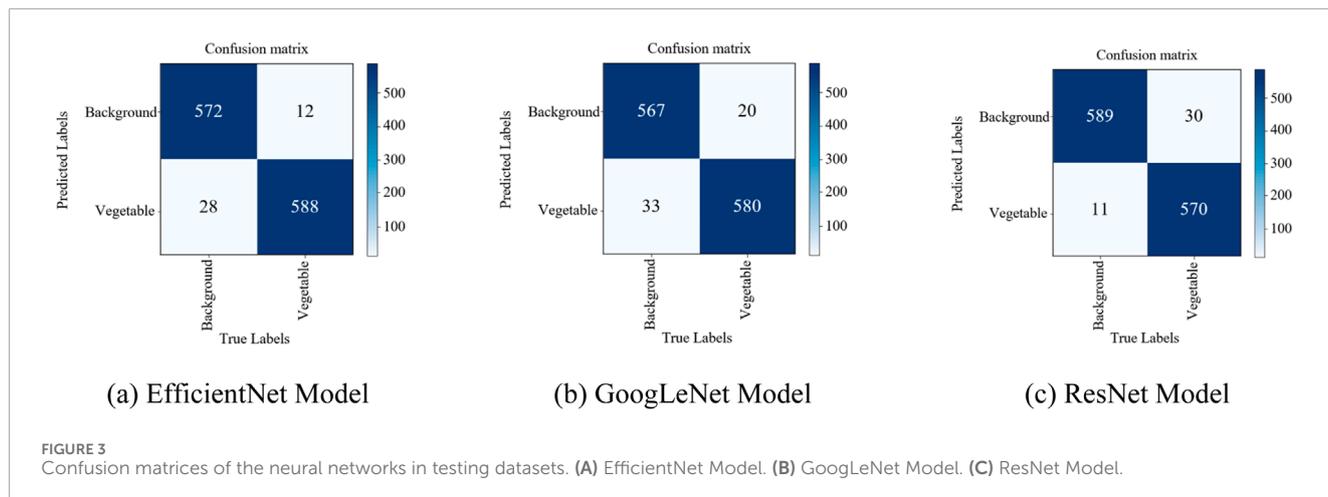


TABLE 4 Cross-validation results of image classification neural networks.

Neural networks	Cross-validation dataset	Average precision	Average recall	Average overall accuracy	Average F ₁ score
EfficientNet	Dataset 1	0.9665	0.9655	0.9660	0.9655
	Dataset 2	0.9700	0.9690	0.9690	0.9695
	Dataset 3	0.9565	0.9560	0.9560	0.9560
GoogLeNet	Dataset 1	0.9595	0.9575	0.9580	0.9580
	Dataset 2	0.9625	0.9605	0.9610	0.9610
	Dataset 3	0.9485	0.9480	0.9480	0.9480
ResNet	Dataset 1	0.9680	0.9665	0.9670	0.9665
	Dataset 2	0.9710	0.9710	0.9710	0.9705
	Dataset 3	0.9535	0.9525	0.9520	0.9525

TABLE 5 Detection speeds of the neural networks.

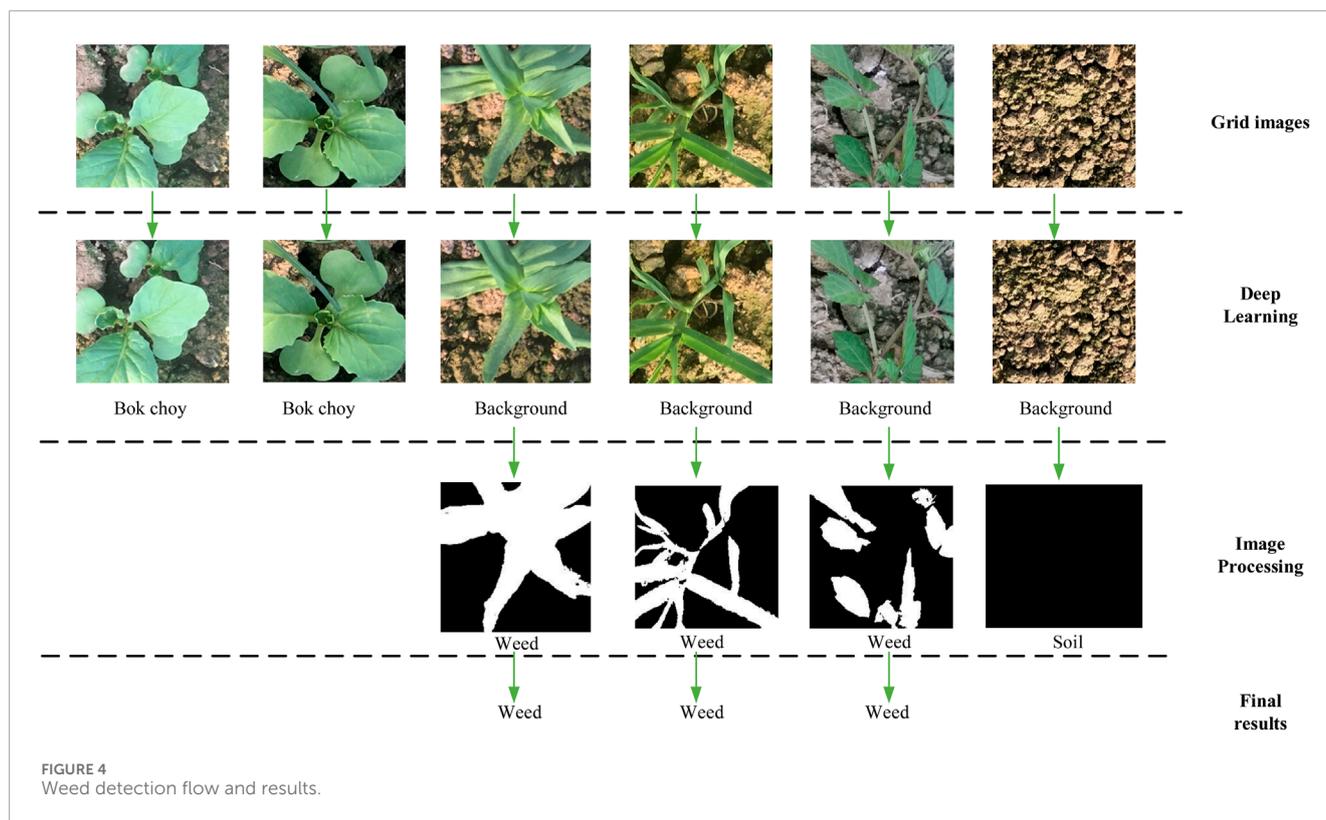
Neural networks	Batch size	Image calculations	Recognition speed	Frame per second
EfficientNet	48	1,200	19.44	30.35
GoogLeNet	48	1,200	12.38	77.96
ResNet	48	1,200	12.76	80.31

Table 5, the detection speeds of the ResNet and GoogLeNet models were similar, taking 12.76 milliseconds and 12.38 milliseconds, respectively, to recognize an original image. The EfficientNet model had the slowest recognition speed, at 19.44 milliseconds. Generally, a frame rate higher than 30 fps is considered to have the capability for real-time image processing [36]. Therefore, in this study, all neural networks possessed the ability for real-time vegetable recognition. Considering both recognition accuracy and speed, the ResNet model was the optimal choice for identifying vegetables, with a

recognition frame rate of 78.34 fps, meeting the requirements for real-time recognition applications.

3.3 Green plant segmentation

After recognizing all grids containing vegetables, the remaining grids were analyzed further. These grids were either occupied by weeds or consisted solely of soil. To differentiate between these two,



additional image processing techniques were employed, focusing on color characteristics to identify green pixels within the grid images. Grids containing green pixels were classified as containing weeds, while those lacking green pixels were identified as soil grids. This color-based segmentation method effectively distinguished between vegetable pixels and weed pixels, as demonstrated by the results (Figure 4). In grids displaying only a soil background, where the segmented image revealed an absence of plant pixels, such areas could be directly excluded from weeding operations. Conversely, grids that featured plant targets were classified as containing weeds, indicating the presence of weed infestations.

In natural settings, the similar colors of weeds and vegetables often lead to misidentification, and their intertwined growth further complicates the task of precisely localizing each plant. To address this challenge, our study introduced an innovative approach that involved cropping the original image into multiple grid images, each corresponding to the operational range of the weeding mechanism. This approach simplified the weed detection process, as depicted in Figure 4, by focusing on identifying the presence of weeds within these grid images. Each grid was assessed individually and classified into one of four scenarios: grids containing only vegetables, only weeds, a mix of both, or solely soil. Image classification neural networks were employed to identify and exclude grids that contained vegetables, as shown in the first and second rows of Figure 4. The remaining grids, which contained either weeds or soil, were then differentiated using color segmentation techniques. As demonstrated in the third row, the color index successfully isolated areas with green pixels - indicative of weed presence - while grids that displayed only soil showed no segmentation. Thus, grids in the first three columns were identified as containing weeds.

Precision mapping of weeds using custom software integrated with image classification neural networks developed in this study is illustrated in Figure 5. As previously mentioned, each original image was divided into 40 grid images, with 8 grid images highlighted in red to indicate the presence of weeds. The remaining 32 cells contained either vegetables or soil and were identified as areas not requiring weeding. This proposed method of weed mapping enables targeted weeding by guiding the weeding executor to the specific grid images that contain weeds.

4 Discussion

Compared to object detection neural networks, image classification neural networks generally offer higher accuracy and faster recognition speeds. However, image classification neural networks are limited to categorizing images and are unable to locate or identify the position of targets within the image. In the context of weed detection, while object detection networks are capable of pinpointing the exact location of weeds and delineating them with bounding boxes, integrating this detection method directly into a weeding execution system presents challenges. This complexity arises because weeds vary significantly in size, leading to bounding boxes of inconsistent dimensions. In contrast, the operational range of a weeding unit within a weeding mechanism is typically fixed, necessitating additional steps to align the detected location and size of the weed with the weeding executor's functional range.

In this study, image classification neural networks were utilized to distinguish between vegetables and weeds by recognizing grid images that contain vegetables, thus indirectly determining the



FIGURE 5
Weed detection and precision weed mapping.

positions of weeds. Each original image was divided into grid images, and once the dimensions of these grids were established, their precise locations within the original image could be calculated. Therefore, identifying all grids that contain weeds effectively locates the weed areas in the original image. In other words, this approach translates the task of identifying weed positions in the original image into locating grid images known to contain weeds. This enables the weeding machinery to directly navigate to the grids occupied by weeds for targeted weeding operations. For practical deployment, the size of the grid images can be adjusted based on the operational range of the weeding machinery actuators, ensuring that the machinery's working range corresponds to the physical size of one grid area.

It is worth noting that when vegetables or weeds occupy only a very small part of an image (as shown in Figure 5, the grid images in the third row, columns one, two, and five), the model faced challenges in accurately classifying them. While this situation may lower the recognition rate for vegetables, it does not significantly impact the weeding process. If vegetables were incorrectly identified as weeds, the error was likely to cause minimal damage due to the small size of the target area. Conversely, if weeds were misclassified as vegetables, this may result in overlooking the weedy part of that

area. However, the main clusters of weeds in nearby grids were likely to be correctly identified. Additionally, by adjusting the area filtering threshold during image processing, images where vegetables and weeds cover a smaller area can be excluded. Such images are treated as grid areas not requiring weeding operations, ensuring a more focused approach to weeding.

In this research, convolutional neural network models were employed for vegetable identification, classifying grid images into two distinct categories: vegetables and everything else as background. This simplification means the model is trained to discern the presence of vegetables, treating all other elements, including a range of weeds and background objects, as irrelevant to the classification task. Despite the potential variability in the appearance of weeds and background materials, the singular focus on a specific type of vegetable allows the model to hone in on accurately identifying vegetable presence. By shifting the identification emphasis from weeds to vegetables, we not only streamline the task of recognition but also significantly enhance the neural network model's robustness and reliability in distinguishing vegetables.

It should be noted that with the design of this research, grid images containing both vegetables and weeds will be classified as

vegetable images. This approach focuses on recognizing the presence of vegetables, treating all other elements, including weeds and soil, as background. While our experimental dataset consisted solely of bok choy, the proposed method can indeed be extended to fields with multiple vegetable species. By grouping all vegetable types under a single *vegetable* class label during training, the model can recognize any cultivated vegetable plant regardless of its specific type. However, a limitation of this method is the inability to precisely distinguish between grids that contain both vegetables and weeds. In organic farming, where herbicides are typically not used and mechanical weeding is preferred, applying mechanical weeding to such mixed areas could potentially damage the crops. To mitigate this risk and enhance the method's applicability, separation of vegetables and weeds into different grids can be achieved through strategies such as capturing images from different angles or taking additional images at varied positions. Additionally, employing smaller grid sizes allows for more precise identification. Despite this limitation, the contributions of this study highlight its practical and scientific value. The proposed framework offers an effective and simplified solution for weed detection, characterized by low complexity, enhanced robustness, and superior generalization capabilities, making it a promising approach for efficient weed management across diverse agricultural settings.

5 Conclusion

This study integrated deep convolutional neural networks with image processing techniques to indirectly distinguish weeds from vegetables by focusing on vegetable recognition. The experimental findings demonstrated that this approach not only simplified the process of weed recognition but also minimized the expenses associated with developing training sets for neural network models. Consequently, this strategy offers a viable solution for identifying vegetables and weeds, characterized by its low complexity, enhanced robustness, and superior generalization capabilities. The EfficientNet, GoogLeNet, and ResNet models all achieved overall accuracies exceeding 0.956 when classifying vegetables in the testing dataset, demonstrating outstanding performance in identification. Notably, the ResNet model stood out due to its superior computational efficiency, with an inference time of only 12.76 milliseconds per image and a corresponding frame rate of 80.31 frames per second, making it suitable for real-time weed detection. The proposed method for detecting weeds in vegetables can be integrated into a machine vision subsystem to facilitate precise weed control within a smart sensing system.

References

- Singh M, Kaul A, Pandey V, Bimbraw AS. Weed management in vegetable crops to reduce the yield losses. *Int J Curr Microbiol Appl Sci* (2019) 8(7):1241–58. doi:10.20546/ijcmas.2019.807.148
- Mennan H, Jabran K, Zandstra BH, Pala F. Non-chemical weed management in vegetables by using cover crops: a review. *Agronomy* (2020) 10(2):257. doi:10.3390/agronomy10020257
- Kapusta G, Krausz RF. Weed control and yield are equal in conventional, reduced- and No-tillage soybean (*Glycine max*) after 11 years. *Weed Technol* (1993) 7(2):443–51. doi:10.1017/s0890037x0002786x
- Slaughter DC, Giles DK, Downey D. Autonomous robotic weed control systems: a review. *Comput Electron Agric* (2008) 61(1):63–78. doi:10.1016/j.compag.2007.05.008
- Jin X, Liu T, Chen Y, Yu J. Deep learning-based weed detection in turf: a review. *Agronomy* (2022) 12(12):3051. doi:10.3390/agronomy12123051
- Lanini W, Strange M. Low-input management of weeds in vegetable fields. *Calif Agric* (1991) 45(1):11–3. doi:10.3733/ca.v045n01p11
- Jin X, Che J, Chen Y. Weed identification using deep learning and image processing in vegetable plantation. *IEEE Access* (2021) 9:10940–50. doi:10.1109/ACCESS.2021.3050296

Data availability statement

The datasets presented in this article are not readily available because they involve other unpublished research findings. Requests to access the datasets should be directed to XJ at xiaojun.jin@pku-iaas.edu.cn.

Author contributions

HJ: Conceptualization, Formal Analysis, Methodology, Writing—original draft. KH: Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft. HX: Investigation, Resources, Validation, Visualization, Writing—original draft. BX: Investigation, Resources, Writing—review and editing. XJ: Conceptualization, Formal Analysis, Funding acquisition, Methodology, Writing—review and editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Weifang Science and Technology Development Plan Project (Grant No. 2024ZJ1097).

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

8. Utstumo T, Urdal F, Brevik A, Dørum J, Netland J, Overskeid Ø, et al. Robotic in-row weed control in vegetables. *Comput Electron Agric* (2018) 154:36–45. doi:10.1016/j.compag.2018.08.043
9. Jin X, Liu T, Yang Z, Xie J, Bagavathiannan M, Hong X, et al. Precision weed control using a smart sprayer in dormant bermudagrass turf. *Crop Prot* (2023) 172:106302. doi:10.1016/j.cropro.2023.106302
10. Jin X, McCullough PE, Liu T, Yang D, Zhu W, Chen Y. A smart sprayer for weed control in bermudagrass turf based on the herbicide weed control spectrum. *Crop Prot* (2023) 170:106270. doi:10.1016/j.cropro.2023.106270
11. Asad MH, Bais A. Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network. *Inf Process Agric* (2019) 7:535–45. doi:10.1016/j.inpa.2019.12.002
12. Hasan AM, Sohel F, Diepeveen D, Laga H, Jones MG. A survey of deep learning techniques for weed detection from images. *Comput Electron Agric* (2021) 184:106067. doi:10.1016/j.compag.2021.106067
13. Zhuang J, Li X, Bagavathiannan M, Jin X, Yang J, Meng W, et al. Evaluation of different deep convolutional neural networks for detection of broadleaf weed seedlings in wheat. *Pest Manag Sci* (2022) 78(2):521–9. doi:10.1002/ps.6656
14. Jin X, Sun Y, Che J, Bagavathiannan M, Yu J, Chen Y. A novel deep learning-based method for detection of weeds in vegetables. *Pest Manag Sci* (2022) 78(5):1861–9. doi:10.1002/ps.6804
15. Herrmann I, Shapira U, Kinast S, Karnieli A, Bonfil D. Ground-level hyperspectral imagery for detecting weeds in wheat fields. *Precis Agric* (2013) 14:637–59. doi:10.1007/s11119-013-9321-x
16. Nieuwenhuizen A, Hofstee J, Van Henten E. Adaptive detection of volunteer potato plants in sugar beet fields. *Precis Agric* (2010) 11:433–47. doi:10.1007/s11119-009-9138-9
17. Deng W, Huang Y, Zhao C, Wang X. Discrimination of crop and weeds on visible and visible/near-infrared spectrums using support vector machine, artificial neural network and decision tree. *Sensors and Transducers* (2014) 26:26.
18. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine learning in agriculture: a review. *Sensors* (2018) 18(8):2674. doi:10.3390/s18082674
19. Wang A, Zhang W, Wei X. A review on weed detection using ground-based machine vision and image processing techniques. *Comput Electron Agric* (2019) 158:226–40. doi:10.1016/j.compag.2019.02.005
20. Schmidhuber J. Deep learning in neural networks: an overview. *Neural Netw* (2015) 61:85–117. doi:10.1016/j.neunet.2014.09.003
21. Collobert R, Weston J, Bottou L, Karlen M, Kavukcuoglu K, Kuksa P. Natural Language processing (almost) from scratch. *J Mach Learn Res* (2011) 12:2493–537.
22. Jin X, Bagavathiannan M, McCullough PE, Chen Y, Yu J. A deep learning-based method for classification, detection, and localization of weeds in turfgrass. *Pest Manag Sci* (2022) 78(11):4809–21. doi:10.1002/ps.7102
23. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* (2012) 29(6):82–97. doi:10.1109/MSP.2012.2205597
24. Jin X, Han K, Zhao H, Wang Y, Chen Y, Yu J. Detection and coverage estimation of purple nutsedge in turf with image classification neural networks. *Pest Manag Sci* (2024) 80:3504–15. doi:10.1002/ps.8055
25. Bakhshpour A, Jafari A. Evaluation of support vector machine and artificial neural networks in weed detection using shape features. *Comput Electron Agric* (2018) 145:153–60. doi:10.1016/j.compag.2017.12.032
26. Wang Y, Zhang S, Dai B, Yang S, Song H. Fine-grained weed recognition using swin transformer and two-stage transfer learning. *Front Plant Sci* (2023) 14:1134932. doi:10.3389/fpls.2023.1134932
27. Mu Y, Feng R, Ni R, Li J, Luo T, Liu T. A faster R-cnn-based model for the identification of weed seedling. *Agronomy* (2022) 12(11):2867. doi:10.3390/agronomy12112867
28. Chen X, Liu T, Han K, Jin X, Yu J. Semi-supervised learning for detection of sedges in sod farms. *Crop Prot* (2024) 179:106626. doi:10.1016/j.cropro.2024.106626
29. Jin X, Bagavathiannan M, Maity A, Chen Y, Yu J. Deep learning for detecting herbicide weed control spectrum in turfgrass. *Plant Methods* (2022) 18:94. doi:10.1186/s13007-022-00929-4
30. Tan M, Le Q, editors. Efficientnet: rethinking model scaling for convolutional neural networks. *International conference on machine learning*. Long Beach, CA: PMLR (2019).
31. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015).
32. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016).
33. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L, editors. Imagenet: a large-scale hierarchical image database. *2009 IEEE conference on computer vision and pattern recognition*. IEEE (2009).
34. Lu J, Behbood V, Hao P, Zuo H, Xue S, Zhang G. Transfer learning using computational intelligence: a survey. *Knowledge-Based Syst* (2015) 80:14–23. doi:10.1016/j.knosys.2015.01.010
35. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag* (2009) 45(4):427–37. doi:10.1016/j.ipm.2009.03.002
36. Lee RB, editor. Realtime mpeg video via software decompression on a Pa-risc processor. *Digest of papers COMPCON'95 technologies for the information superhighway*. IEEE (1995).