frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 28 August 2025
DOI 10.3389/fpls.2025.1526142

OPEN ACCESS

EDITED BY
Lei Shu,
Nanjing Agricultural University, China

REVIEWED BY
Sayantan Sarkar,
Texas A&M Research & Extension Center at Corpus Christi, United States
Shujin Qiu,
Shanxi Agricultural University, China

*CORRESPONDENCE
Yannan Shi
shiyannan@haafs.org
Peng Lv
pengl001@126.com

†These authors have contributed equally to this work

RECEIVED 11 November 2024
ACCEPTED 29 July 2025
PUBLISHED 28 August 2025

CITATION
Zhao J, Jiao Z, Wang J, Wang Z, Guo Y, Zhou Y, Chen S, Wu W, Shi Y and Lv P (2025) MOSSNet: multiscale and oriented sorghum spike detection and counting in UAV images. *Front. Plant Sci.* 16:1526142.
doi: 10.3389/fpls.2025.1526142

MOSSNet: multiscale and oriented sorghum spike detection and counting in UAV images

Jianqing Zhao[1†], Zhiyin Jiao[2,3†], Jinping Wang[2,3], Zhifang Wang[2,3], Yongchao Guo[2,3], Ying Zhou[1], Shiyi Chen[1], Wenjie Wu[1], Yannan Shi[2,3*] and Peng Lv[2,3*]

[1]Key Laboratory for Climate Risk and Urban-Rural Smart Governance, School of Geography, Jiangsu Second Normal University, Nanjing, China, [2]Institute of Millet Crops, Hebei Academy of Agriculture and Forestry Sciences, Hebei Branch of National Sorghum Improvement Center, Shijiazhuang, China, [3]Key Laboratory of Genetic Improvement and Utilization for Featured Coarse Cereals, Ministry of Agriculture and Rural Affairs, Key Laboratory of Minor Cereal Crops of Hebei Province, Shijiazhuang, China

Background: Accurate sorghum spike detection is critical for monitoring growth conditions, accurately predicting yield, and ensuring food security. Deep learning models have improved the accuracy of spike detection thanks to advances in artificial intelligence. However, the dense distribution of sorghum spikes, variable sizes and complex background information in UAV images make detection and counting difficult.

Methods: We propose a multiscale and oriented sorghum spike detection and counting model in UAV images (MOSSNet). The model creates a Deformable Convolution Spatial Attention (DCSA) module to improve the network's ability to capture small sorghum spike features. It also integrated Circular Smooth Labels (CSL) to effectively represent morphological features. The model also employs a Wise IoU-based localization loss function to improve network loss.

Results: Results show that MOSSNet accurately counts sorghum spike under field conditions, achieving mAP of 90.3%. MOSSNet shows excellent performance in predicting spike orientation, with RMSEa and MAEa of 14.6 and 12.5 respectively, outperforming other directional detection algorithms. Compared to general object detection algorithms which output horizonal detection boxes, MOSSNet also demonstrates high efficiency in counting sorghum spikes, with RMSE and MAE values of 9.3 and 8.1, respectively.

Discussion: Sorghum spikes have a slender morphology and their orientation angles tend to be highly variable in natural environments. MOSSNet 's ability has been proved to handle complex scenes with dense distribution, strong occlusion, and complicated background information. This highlights its robustness and generalizability, making it an effective tool for sorghum spike detection and counting. In the future, we plan to further explore the detection capabilities of MOSSNet at different stages of sorghum growth. This will involve implementing object model improvements tailored to each stage and developing a real-time workflow for accurate sorghum spike detection and counting.

KEYWORDS

sorghum spike, oriented detection boxes, angle feature, deep learning, UAV

1 Introduction

Sorghum [*Sorghum bicolor* (L.) Moench] is a C_4 drought tolerant cereal crop widely grown worldwide for food, feed, and biofuels (Bao et al., 2024). It plays a vital role in global food security, especially in arid regions, providing an important source of nutrition for humans and livestock (Liaqat et al., 2024; Hossain et al., 2022). Due to the reduction of arable land and the impact of extreme environments, humans have been committed to maximizing crop yields to feed the growing population (Baye et al., 2022). Sorghum is the fifth cereal crop, and increasing the yield per unit area of sorghum can provide a guarantee for food security. Therefore, accurate estimation of sorghum yield is crucial for farmers and breeders (Khalifa and Eltahir, 2023, Ostmeyer TJ et al., 2022, Wang et al., 2024). Traditional manual sampling methods are not only time consuming and labor intensive, but are susceptible to human error, especially in the case of large-scale planting (Wu et al., 2023). This makes it particularly urgent to achieve efficient and accurate yield estimation (Liang et al., 2024; Wang et al., 2023).

To address these challenges, many studies have used unmanned aerial vehicle (UAV) technology, which has excellent monitoring capabilities, as an effective tool for sorghum yield estimation (Ahmad et al., 2020, Jianqing et al., 2023). Equipped with high-precision sensors, UAVs can quickly capture images of large areas of farmland, greatly improving the speed and accuracy of data collection (Abderahman Rejeb et al., 2022). This aerial perspective provided by UAVs offers comprehensive crop coverage and helps reduce monitoring blind spots. It also enables real-time tracking of crop growth. These capabilities are crucial for identifying potential problems and making timely adjustments (Jin et al., 2021, Zhang et al., 2024a, Zhang et al., 2024b). The flexibility and cost-effectiveness of UAV operations enable efficient deployment in a variety of environments and terrains. These advantages help to significantly reduce labor costs while improving data reliability. As a result, they provide a strong basis for scientific evaluation of sorghum yields (Hasan MM et al., 2018, Perich G et al., 2020).

In recent years, deep learning have achieved groundbreaking success in solving complex problems in a variety of fields (Duan et al., 2024; Yan et al., 2022, Yang et al., 2020; Zhang et al., 2023). The rapid development of deep learning has provided a novel solution for the detection and counting of sorghum spikes (Huang et al., 2023; Sanaeifar et al., 2023). Deep neural networks show significant potential in tasks such as classification, detection and segmentation (Cai et al., 2021, James et al., 2024). While classification methods can effectively categorize images of sorghum spikes, it faces challenges in dealing with complex backgrounds and occlusions (Guo et al., 2018). Segmentation methods achieve precise pixel-level localization but requires significant computational resources, making it less suitable for real-time processing of large datasets (Chenyong et al., 2020; Genze et al., 2022; Malambo et al., 2019, Zarei et al., 2024). Detection methods combine the strengths of both classification and segmentation, efficiently identifying the position and shape of sorghum spikes while demonstrating strong adaptability to varying backgrounds (James et al., 2024; Gonzalo-Martín and García-Pedrero, 2021). Deep learning models improve the accuracy and robustness of sorghum spike detection, especially in complex environments. These approaches effectively meet the real-time processing requirements of large-scale agricultural automation.

Traditional object detection methods based on convolutional neural networks, such as Faster R-CNN, SSD and YOLO, typically use horizontal bounding boxes to localize objects in images. While effective for standard detection tasks, these methods have significant limitations when applied to sorghum panicle detection in UAV imagery. Sorghum panicles have a slender morphology and naturally occur in different orientations due to the influence of wind, gravity and plant structure. Horizontal bounding boxes cannot represent directional information, making it difficult for models to learn accurate spatial features (Gonzalo-Martín and García-Pedrero, 2021; Lin and Guo, 2020; Zhang et al., 2024). Sorghum panicles are small, densely packed and highly overlapping in UAV images. Horizontal boxes often lead to the merging of multiple panicles into a single detection, or the repeated detection of the same panicle in crowded scenes (Duan et al., 2024; Ming et al., 2021). Horizontal boxes tend to contain a large amount of irrelevant background, which interferes with the learning of target-specific features, reducing model accuracy and training efficiency. This problem is exacerbated by the complex field environment, where background elements such as leaves, soil and shadows introduce additional noise (Qiu et al., 2024, Wang et al., 2022; Xue et al., 2024). In recent years, oriented bounding boxes have been used with an additional angular parameter to better represent multi-oriented targets, making some progress in remote sensing and cluttered scene detection. However, simply incorporating orientation into the bounding box is insufficient to address the challenges of small target size, dense overlap, and complex spatial relationships. In UAV imagery, such approaches still struggle with inadequate feature extraction and poor modelling of object interactions, resulting in frequent false positives and missed detections (Sanaeifar et al., 2023). It is therefore critical to develop a detection model that is not only orientation-aware, but also capable of fine-grained feature extraction and robust spatial reasoning.

To address these challenges, this study integrates sorghum spike orientation information into convolutional neural networks and uses oriented bounding boxes to identify their spatial locations. We propose a multiscale and orientation-aware sorghum spike detection model for UAV images. This model is designed to effectively address the detection challenges posed by the complex morphology and significant occlusions commonly found in sorghum spikes.

2 Materials and methods

2.1 UAV images

The experiment was conducted in 2024 at the Qiema Experimental Station in Shijiazhuang, Hebei Province, China, with sorghum variety JN 4 planted in 3 plots of 900 m^2 each (Figure 1). A DJI Phantom 4 drone equipped with a visible light camera was used to capture images of sorghum spikes under sunny conditions between 10 am and 2 pm. The flight altitude was set at 10 meters, with a speed of 1 meters per second. The original drone images had a resolution of 4032×2268 pixels. To increase the efficiency of the model processing, the original images were cropped into 600×600 pixels sub-images. The sorghum spikes were manually annotated using the Rolabelimg tool. This process generated annotation files containing information on each spike's center point, dimensions, angle, and category.

2.2 Dataset preparation

Previous studies have shown that rotation and flipping can effectively improve the robustness of deep neural network models. In addition, brightness adjustment helps mitigate the effects of brightness variations caused by changes in ambient lighting and differences in sensors (Alzubaidi L et al., 2021, Zhang et al., 2023). Therefore, we applied data augmentation, including rotation, flipping, and brightness adjustment, to the annotated sorghum spike images and their associated annotation files (Figure 2). Specifically, rotation operations included rotating images by 90°, 180° and 270°, while flipping operations included both horizontal and vertical flipping. Brightness adjustment was achieved by increasing and decreasing the image brightness by 10% and 20% respectively. This resulted in a dataset of 6000 images of sorghum spikes. Since the YOLO model inherently extracts approximately 10% of the training set data as an internal validation subset during training. This internal validation procedure continuously can monitor model convergence, early stopping and hyperparameter tuning. The dataset was then randomly divided into training, validation and test sets in a ratio of 6:1:3.

2.3 Structure of MOSSNet

MOSSNet is constructed based on YOLOv8. MOSSNet performs feature extraction and fusion on the input image to generate multi-scale feature maps (Figure 3). A Deformable Convolution Spatial Attention (DCSA) module is proposed to improve the network's ability to capture features of small sorghum spikes. An angular feature extraction module is developed by incorporating circular smooth labels. This enables the network to capture the directional information of sorghum spikes, effectively representing their morphological characteristics in UAV images. A Wise-IoU based localization loss is applied to further optimize the network loss for bounding boxes and detection boxes of sorghum spikes in field environments. Finally, duplicate sorghum spike detection boxes are removed by calculating the overlap area of the oriented detection boxes. The size, position, angle, and category information of the remaining detection boxes are then output.

2.3.1 Deformable convolution spatial attention

This study proposes DCSA module to eliminate invalid features generated by traditional convolution operations when processing sorghum spikes of different shapes and sizes (Figure 4). The DCSA module has two main components. First, deformable convolution (Dai et al., 2017) is employed to overcome the limitations of conventional convolution in handling rectangular shapes. It allows the network to adaptively extract effective features based on the size of the sorghum spikes. (Figure 5). Second, the DCSA integrates deformable convolution with the Scaled Dot-Product Attention mechanism (Lovisotto et al., 2022). Notably, the scaled dot-product attention mechanism enhances the model's ability to focus on color-dimension features, thereby improving its adaptability to the color variations exhibited by sorghum spikes at different growth stages. By embedding this combination in the multi-scale feature extraction and fusion part of the network, the feature extraction capabilities are significantly enhanced. (Figure 5). The DCSA introduces a micro-scale detection layer branch designed to generate feature maps optimized for small sorghum spikes. These feature maps have a resolution of one quarter of the input image (Figure 4), ensuring better detection of multi-scale spikes in different stages.

2.3.2 Angle feature extraction

In this study, the oriented detection box of the sorghum spike is decoupled into a horizontal detection box. This box contains the center of mass coordinates (x, y), length and width information (w, h), angle information (θ), and category information (cls). θ is the angle between the long side of the detection box and the x-axis, and each angle is a category, with a total of 180 categories. Circular smooth labels address the issue of angular periodicity (Yang and Yan, 2020). Based on this, the study defines a periodicity labelling code to measure the angular distance between the oriented detection box and the bounding box using the following formula:

$$CSL(x) = \begin{cases} g(x), & \theta - r < x < \theta + r \\ 0, & otherwise \end{cases}$$

where θ is the angle of the sorghum spike orientation frame, $g(x)$ is the Gaussian function, and r is the window radius of the Gaussian function.

2.3.3 Optimized loss function

The IoU-based localization loss only considers the scenario where the detection frame and the bounding box intersect, without taking into account the complex spatial relationship between the two frames (Rezatofighi et al., 2019). This study is based on Wise IoU (Tong et al., 2021), which allows the evaluation of the spatial overlap between the bounding boxes. It focuses not only on the area of overlap, but also on the differences in shape and position between the bounding boxes. The formula is as follows:

$$L_{WIoU} = \alpha \cdot L_{IoU}$$

$$L_{IoU} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}(1 - IoU)$$

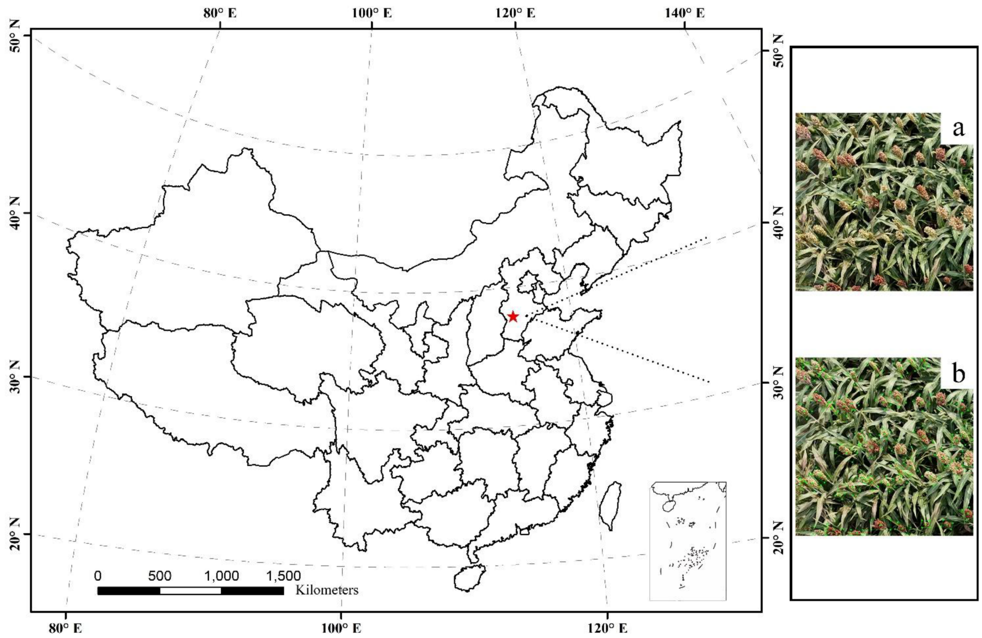

FIGURE 1
Research area. **(a)** UAV image, **(b)** Labeled image.

$$\alpha = \exp\left(\frac{(GT_x - B_x)^2 + (GT_y - B_y)^2}{(GT_w + GT_h)^2}\right)$$

where S^2 is the feature map size output from the detection layer and B is the detection frame. GT_x, GT_y, GT_h and GT_w are the center coordinates, length and width of the sorghum spike bounding box respectively. B_x and B_y are the center coordinates of the sorghum spike detection frame.

2.3.4 Duplicate detection boxes deleted

The traditional Non-Maximum Suppression (NMS) method is primarily used to compute the Intersection over Union (IoU) between horizontal bounding boxes. However, this approach presents challenges when applied to the oriented bounding boxes of sorghum spikes (Neubeck and Gool, 2006). To address this issue, this study divides the overlapping region between the sorghum detection box and the bounding box into multiple triangles that

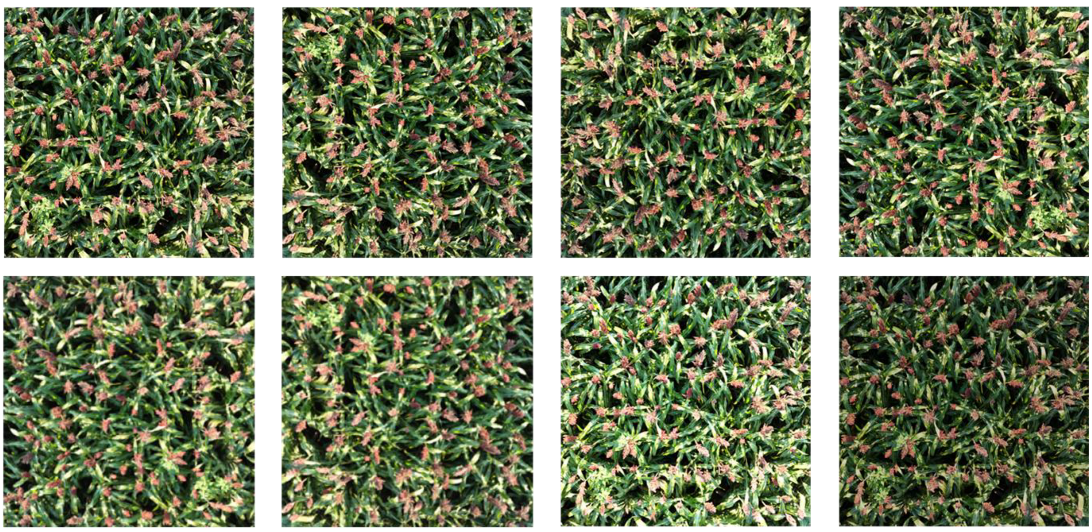

FIGURE 2
Data augmentation.

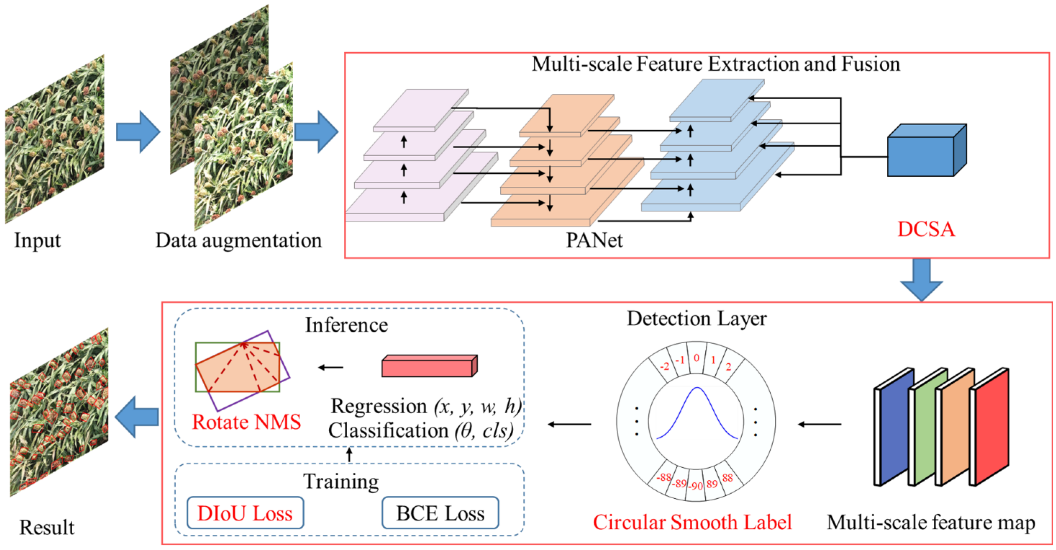

FIGURE 3
The overall framework of MOSSNet.

share the same vertices (Figure 6). The total overlapping area is obtained by calculating and summing the area of each triangle. In addition, the ratio of the overlapping area to the sum of the areas of the two bounding boxes is defined. The bounding box with the highest ratio is retained, while the others are discarded, resulting in the final detection result. Figure 6 illustrates the geometric principle for calculating the overlapping area between two oriented detection boxes.

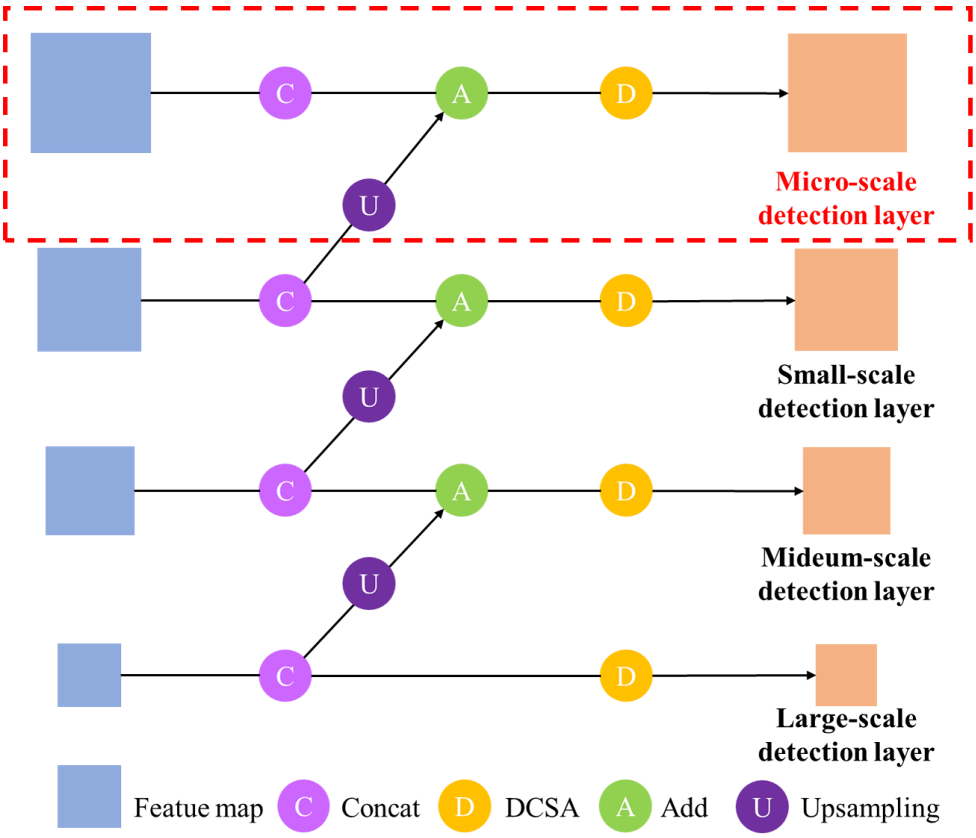

FIGURE 4
The structure of neck with DCSA.

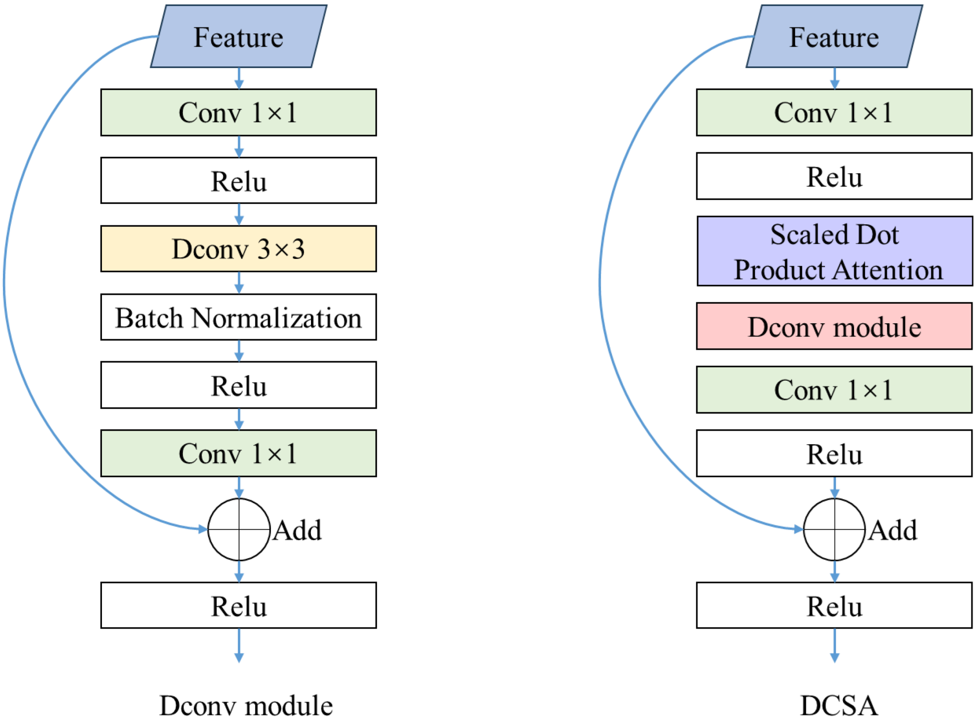

FIGURE 5
The structure of DCSA and Dconv module.

3 Experiment setup and evaluating indicators

3.1 Experiment settings

The deep network models were trained and tested on a server equipped with an Intel® Xeon Platinum 8268 CPU, NVIDIA TITAN V graphics processor (12 GB of graphics memory), and 500 GB of RAM, running Ubuntu 16. Considering the dataset size, model parameters, and computational resources, model training was performed with a learning rate of 0.001, a batch size of 32, a weight decay value of 1e-4, and a momentum of 0.9.

3.2 Evaluating indicators

This study evaluates the accuracy of MOSSNet and other deep learning network models in recognizing and counting sorghum spikes using five indicators grouped into three categories. Specifically, a mean Average Precision (mAP) is used to evaluate

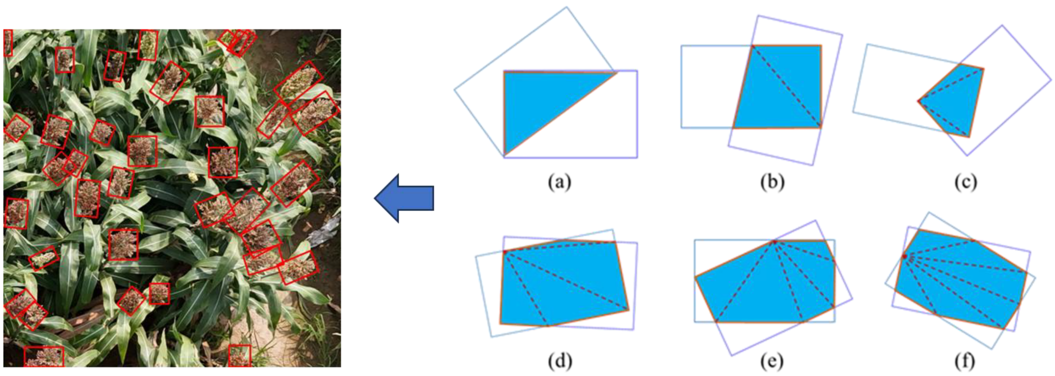

FIGURE 6
Area calculation for two overlapping oriented detection boxes. **(a)** intersecting graph is a triangle, **(b)** intersecting graph is a quadrilateral, **(c)** intersecting graph is a pentagon, **(d)** intersecting graph is a hexagon, **(e)** intersecting graph is a heptagon, **(f)** intersecting graph is an octagon.

the accuracy of MOSSNet and specialized detection models in detecting the number and location of sorghum spikes. $RMSE_a$ and MAE_a are used to evaluate the accuracy of the models in estimating the angle of sorghum spike. Meanwhile, *RMSE* and *MAE* are used to assess the performance of MOSSNet compared to general object detection methods for spike detection.

The mAP represents the mean accuracy across all categories within a recall range of 0 to 1, with higher values indicating better detection accuracy. In this study, the only object category is sorghum spikes and the formula for calculating mAP is as follows:

$$\mathrm{mAP} = \int_0^1 P(R)dR$$

where *P* and *R* are the precision and recall, respectively, defined as follows:

$$P = \frac{TP}{FP + TP}$$

$$R = \frac{TP}{FN + TP}$$

TP and *FP* are the number of correctly and incorrectly detected sorghum spikes, and *FN* is the number of undetected sorghum spikes.

$RMSE_a$ and MAE_a are used to evaluate the difference between the angle of the sorghum spike detection frame and the angle of the bounding box.

$$RMSE_a = \sqrt{\frac{1}{F}\sum_{i=1}^{F}(p_i^n - q_i^n)^2}$$

$$MAE_a = \frac{1}{F}\sum_{i=1}^{F}|p_i^n - q_i^n|$$

where *F* is the number of images for which the calculation was performed. p_i^n is the predicted angle of the nth sorghum spike in image *i* of the model calculation. q_i^n is the true angle of the corresponding sorghum spike.

As the output of the generalized one-stage and two-stage object detection methods is a horizontal detection box, it is not possible to use the mAP for comparison with MOSSNet. For the collected UAV images, we use manual annotation, where each sorghum spike was surrounded by a bounding box, and each predicted box from the model was correspondingly treated as a spike. To assess the accuracy of the model in predicting the total number of spikes and the associated variance, *RMSE* and *MAE* are used to compare the difference between the total number of spikes in the annotations and that predicted by the model.

$$RMSE = \sqrt{\frac{1}{F}\sum_{i=1}^{F}(a_i - t_i)^2}$$

$$MAE = \frac{1}{F}\sum_{i=1}^{F}|a_i - t_i|$$

where *t* is the number of sorghum spikes included in the model-calculated image *i*, and *a* is the number of manually labeled sorghum spikes in image *i*.

4 Results

This study evaluates the performance of MOSSNet in sorghum spike detection against R2CNN (Jiang et al., 2017), R3DET (Yang et al., 2019), RRPN (Liu JH et al., 2017), RIDET (Ming et al., 2021), SCRDET (Yang et al., 2019) and RSDET (Qian et al., 2021) using metrics such as mAP, $RMSE_a$ and MAE_a. These oriented object detection methods are primarily designed to detect rotated objects in remote sensing imagery, and their core mechanisms can be grouped into two categories. The first involves the incorporation of rotated bounding box regression (e.g., R2CNN, RRPN, R3Det), either at the proposal generation stage or at the regression stage, to model the angular information of objects. The second focuses on feature refinement and structure optimization (e.g. SCRDet, RIDet, RSDet) to improve detection performance for small objects or those with large angular variations in complex backgrounds. The results show that MOSSNet achieves a mean average precision (mAP) of 90.3% (Table 1), outperforming the other methods in terms of detection accuracy. MOSSNet also achieved the highest precision in predicting sorghum spike orientation, with $RMSE_a$ and MAE_a of 14.6 and 12.5 respectively. MOSSNet's frame per second (FPS) reached 35, demonstrating its ability to efficiently detect the orientation of sorghum spikes in UAV imagery.

Furthermore, this study compared MOSSNet with several mainstream horizontal bounding box object detection methods, including Faster RCNN (Ren et al., 2017), EfficientNet (Tan and Le, 2019), RetinaNet (Lin et al., 2017), SSD (Wei et al., 2016), YOLOv5, YOLOv8, and YOLOv11 (Figure 7). In Figure 7, the x-axis represents the actual number of spikes in a given image, while the y-axis represents the number of spikes predicted by the model. Each point on the scatterplot thus corresponds to an image that pairs its labelled spike count with the model's predicted spike

TABLE 1 Comparison between the proposed method and other oriented object detection approaches.

Method	mAP (%)	$RMSE_a$	MAE_a	FPS
Proposed	90.3	14.6	12.5	35
R2CNN	67.1	46.9	39.5	24
R3DET	76.5	31.3	25.3	32
RRPN	71.4	42.3	36.8	28
RIDET	84.7	19.5	18.3	20
SCRDET	78.3	26.6	21.2	27
RSDET	75.2	32.3	28.4	25

$RMSE_a$ and MAE_a are used to evaluate the difference between the angle of the sorghum spike detection frame and the angle of the bounding box. The subscript lowercase letter 'a' means that they are different from RMSE and MAE.

10.3389/fpls.2025.1526142

count. The distribution of these points around the diagonal visually illustrates the overall accuracy and error distribution of the model. The results show that each image contains on average about 30 sorghum spikes and MOSSNet still achieves the highest accuracy among all methods, with an *RMSE* of 9.3 and an *MAE* of 8.1. Comparing this benchmark with the outputs of different detection models, Faster R-CNN, EfficientNet, RetinaNet, SSD, YOLOv5, YOLOv8, YOLOv11, and the proposed MOSSNet detected approximately 81, 60, 43, 40, 34, 46, 33, and 31 spikes per image, respectively. These results indicate that some models significantly overestimate the number of spikes, whereas MOSSNet produces spike counts closest to the ground truth, demonstrating higher detection accuracy and better model alignment. As an improved model based on YOLOv8, MOSSNet reduced the *RMSE* by 29% and the *MAE* by 26% compared to YOLOv8. This study also tested the latest YOLOv11, released this year, and the results showed that MOSSNet reduced the *RMSE* by 13% and the *MAE* by 12% compared to YOLOv11. These results show that the proposed MOSSNet performs well in detecting and counting sorghum spikes in complex field environments.

5 Discussion

5.1 The influence of angle and size distribution for the detection

The number, length and width of crop spikes are critical parameters for crop growth monitoring and yield prediction, which rely heavily on spike angle information in images (Chang et al., 2021; Li et al., 2022). Traditional non-deep learning methods using features such as size, texture, color and morphology combined with operators such as Harris have successfully measured spike morphology. These methods highlight the importance of angular features for effective spike detection (AB et al., 2015, Guo et al., 2016). By manually annotating sorghum spike angles, MOSSNet introduces and extracts spike angle features, providing a more accurate description of sorghum spike morphology in UAV imagery. This study analyses UAV image data to quantify the number of sorghum spikes at different angles. Figure 8 shows the analysis of UAV images used to quantify the number of sorghum spikes at various angles. Specifically, the horizontal axis, which ranges from 0 to 180, represents the angles at which the sorghum spikes are rotated in the images. This allows their morphological features to be captured. The red and blue lines in the figure show the number of sorghum spikes detected at specific angles. At an angle of 0°, the corresponding number of detected sorghum spikes is 661. The numbers labelled in the figure represent the exact number of spikes at intervals based on a threshold of 100, which emphasizes the differences in spike quantities. These results show that sorghum spikes are unevenly distributed across different angles, appearing mainly between 0°-45°, 90° and 120°- 180°, with the highest number at 0°. This distribution may be influenced by visual preferences and operational habits during manual annotation (Zhao et al., 2022). Consequently, the uneven distribution of spike angles suggests variation in morphology and orientation in UAV imagery based on Figure 8. By incorporating directional features, MOSSNet can more accurately capture the true morphology of sorghum spike in UAV images. This improves the model's ability to provide reliable data for future UAV-based estimation of sorghum spike morphological parameters.

In this study, a detailed analysis of the bounding box length, width and area of sorghum spikes in the measured data was conducted (Figure 9). The bounding box length and width of sorghum spikes in drone images are generally less than 100 pixels. It indicates that the scale of sorghum spikes is relatively small and falls within the range of small object detection. On the other hand, the study used visible light images with a resolution of 4032×2268 taken by a drone flying at an altitude of 10 meters, resulting in a ground resolution of 0.4 cm/pixel. The combination of observation scale and representation scale determines that the acquired drone images have the characteristics of small size and high density of sorghum spikes. Under these conditions, the pixel area of small objects in the image is relatively small, resulting in blurred image representation, limited extractable features, and weaker feature representation ability (Zhao et al., 2021). Therefore, it is necessary to make improvements to the detection model based on the characteristics of small-sized objects to improve its performance in detecting and representing sorghum spikes.

5.2 Ablation study and generalization test

To further investigate the impact of DCSA, CSL and Wise IoU on the experimental results, this study evaluates the effectiveness of these improvements with mAP (Table 2). The results show that CSL is the most effective improvement, increasing mAP by 5.5%, while DCSA and Wise IoU contribute increases of 3.9% and 1.6% respectively. The baseline model only transforms sorghum spike angle information into standard categorical features, which is insufficient to capture the relative distances between different angles and the periodic nature of angle features. Comparing the results of oriented and horizontal bounding boxes for sorghum spike detection, it is clear that the inclusion of CSL for oriented detection offers significant advantages. The oriented bounding boxes are better able to capture the morphological features of the sorghum. In addition, they contain less background noise, which improves the overall detection accuracy (Figure 10). Without CSL, the horizontal bounding boxes tend to generate significant overlap, which interferes with sorghum spike detection and leads to missed detections. In addition, horizontal bounding boxes containing excessive background not only increase false detections, but also prevent further in-depth analysis of the detection boxes. Using our existing data resources, we conducted additional tests on JN3, JN6 and JN8 and their inbred lines (Figure 11). The results show that the proposed model has high generalization and robustness in different backgrounds and field conditions.

10.3389/fpls.2025.1526142

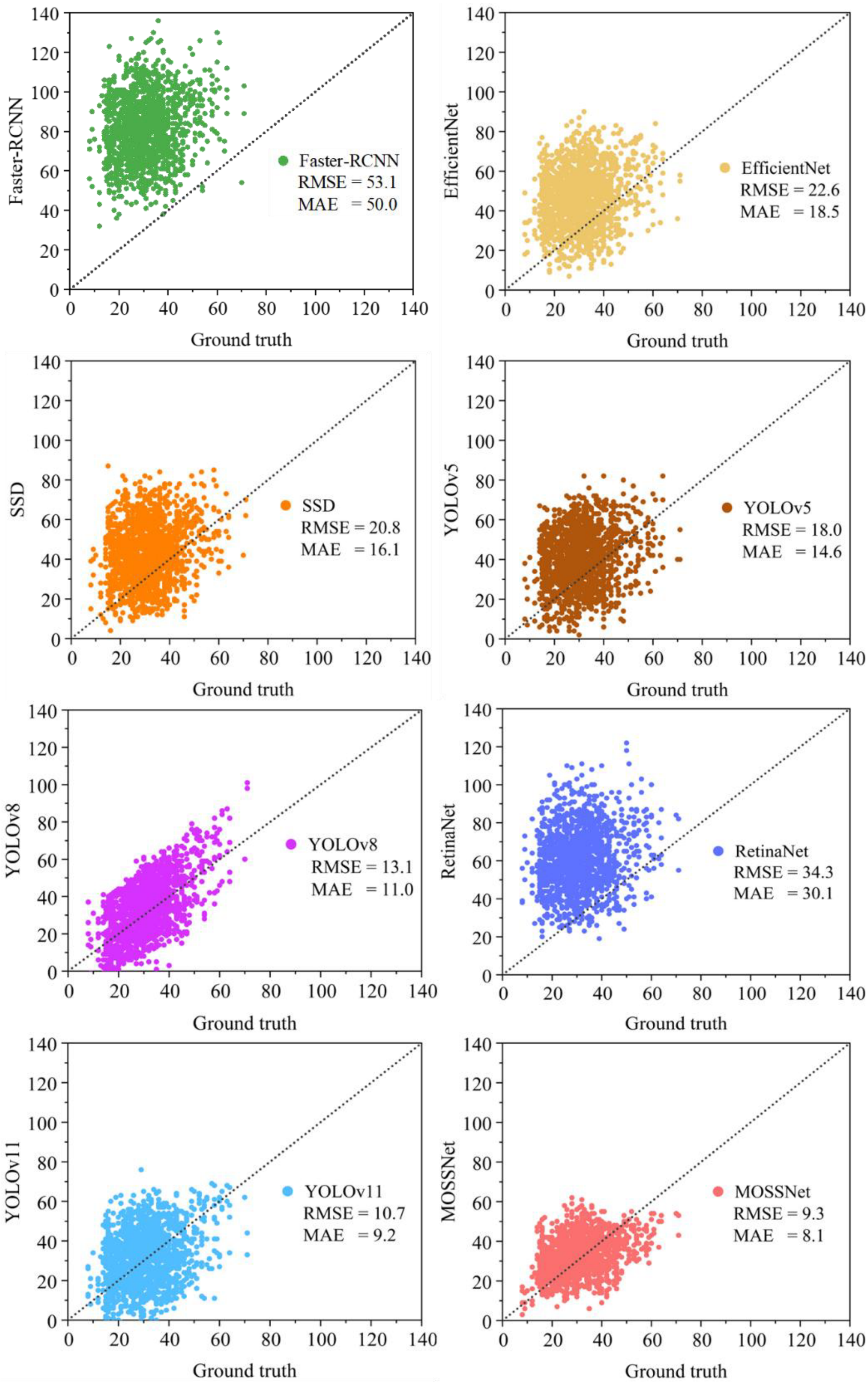

FIGURE 7
Comparison of the number of spikes labeled on the images by MOSSNet and other methods.

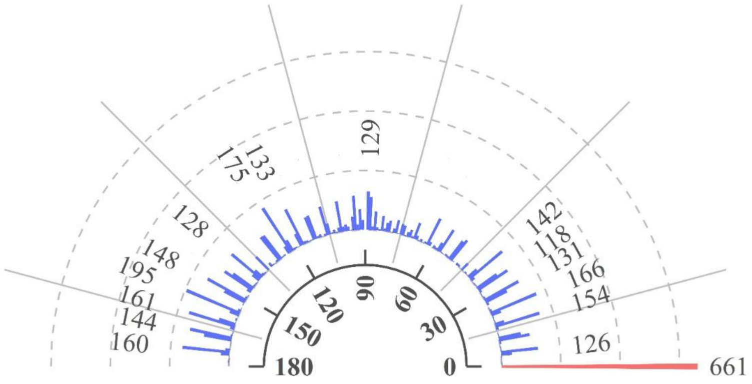

FIGURE 8
Classification for angles of sorghum spikes.

5.3 Optimized pathways for sustainable monitoring of sorghum spikes

Sorghum spikes have a slender morphology and their orientation angles tend to be highly variable in natural environments. Traditional horizontal detection boxes contain a significant number of non-target background pixels, reducing detection accuracy (Sanaeifar et al., 2023). The output does not capture the directional characteristics of rice spikes. The inclusion of complex background information in the horizontal bounding boxes can further degrade the model's training performance and make it difficult to visualize the detection results (Zhang et al., 2025). Sorghum plants have flexible structures, with dense, interlaced inflorescences that often overlap heavily, making false positives and false negatives more common than in standard object detection tasks. Under the influence of natural wind or UAV rotor wash, sorghum spikes can sway and intertwine, making it difficult to identify individual spikes. Our results show that sorghum spikes are unevenly distributed at different angles, occurring mainly between 0°-45°, 90° and 120°-180°, with the highest concentration at 0°. The model shows excellent adaptability, even in windy conditions. This angular distribution highlights a potential subjective bias in manual counting and serves as a reminder to farmers to consider the angular variations caused by wind, as failure to do so may compromise the accuracy of yield estimation. The model proposed in this study outputs oriented detection boxes with directional features that closely align with each sorghum ear, reducing background interference and improving counting accuracy.

On the other hand, the color of sorghum spikes in UAV images is influenced by the growth stage, which plays a dominant role in the performance of sorghum spike detection methods. This observation suggests that color-based detection approaches are generally used for specific growth stages (Chen H et al., 2024, Ghosal et al., 2019). In this study, the dataset covers both the heading and flowering stages of sorghum and includes several varieties (JN3, JN4, JN6 and JN8). Differences in cultivars and growth stages result in differences in spike color, such as green or brown. Some studies suggest that very early green spikes can exhibit significant differences in shape and size, requiring a larger amount of training data to maintain model generalizability and computational efficiency (Zhang et al., 2022). This study introduced the DCSA module, integrating Deformable Convolution and the Scaled Dot-Product Attention mechanism

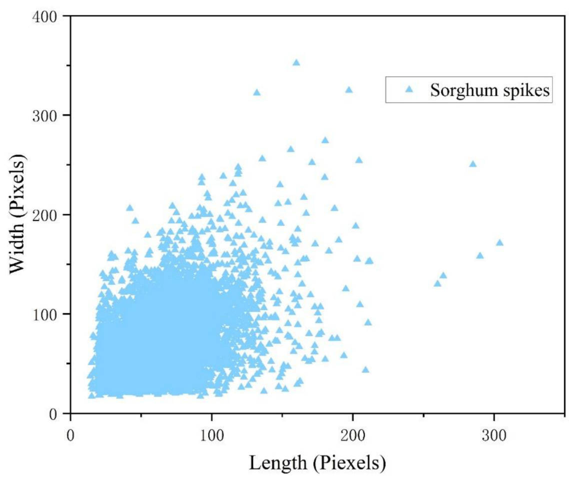

FIGURE 9
The size distribution of sorghum spikes in UAV images.

TABLE 2 Ablation study of components of MOSSNet.

CSL	DCSA		Wise IoU	mAP (%)
	Dconv	Scaled dot-product attention		
				79.3
√				84.8
√	√			85.9
√	√	√		88.7
√	√	√	√	90.3

Checkmark (√) indicates that the corresponding module is used. The absence of a checkmark indicates that the module is not used.

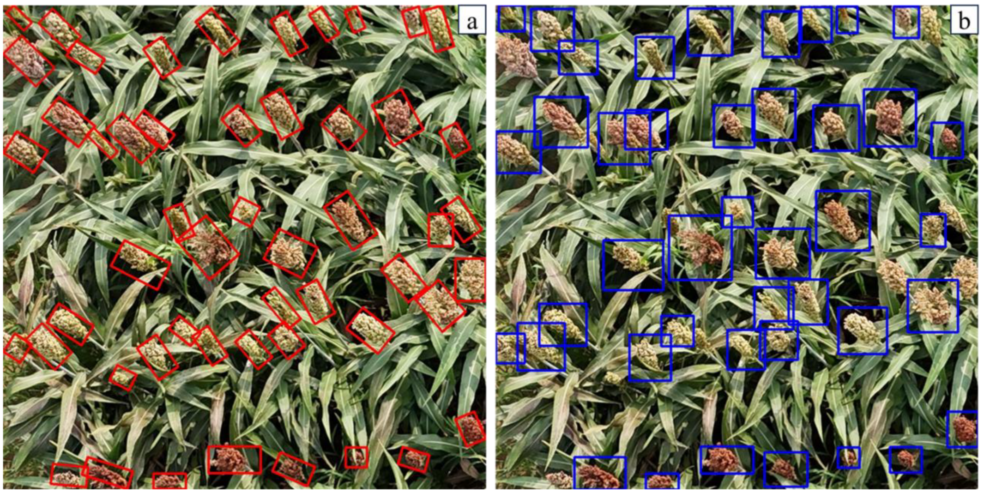

FIGURE 10
Detection results with MOSSNet **(a)** and the YOLOv8 **(b)**.

into the multi-scale feature extraction and fusion processes. The DCSA module combines deformable convolution with a scaled dot-product attention mechanism. The deformable convolution component introduces learnable 2D offsets to the regular sampling grid, enabling the sampling pattern to be deformed freely. This allows the model to focus adaptively on the key spatial features of the panicle regions in the image. Meanwhile, the scaled dot-product attention mechanism enables the model to focus on the most relevant parts of the input, thereby enhancing its ability to model information along the color dimension. Combining adaptive spatial sampling with color feature attention markedly improves the model's stability in recognizing sorghum panicles across images from different growth stages, effectively mitigating the drop in detection accuracy caused by variation in panicle color. we will also incorporate more rigorous statistical significance analyses in future studies to strengthen the reproducibility and persuasiveness of our results.

In addition, UAV flight parameters directly affect image quality, which has a significant impact on the accuracy of sorghum spike detection and yield estimation. Future research will focus on expanding and refining the sorghum spike dataset. It will also aim to identify the optimal UAV parameters and flight conditions. In addition, images from other growth stages will be incorporated to further improve the accuracy and applicability of the oriented sorghum spike detection model.

The proposed MOSSNet demonstrates significant effectiveness in detecting sorghum spikes of different sizes from UAV images, making it highly valuable for practical applications in field environments. Traditional detection methods struggle to capture the size and morphological characteristics of sorghum spikes in complex environments, especially given the dynamic changes in spike size and shape during the growth process (Guo et al., 2018; Salas Fernandez MG et al., 2017). The DCSA module in MOSSNet accurately captures the characteristics of small and densely spaced spikes. Horizontal detection boxes often contain excessive background information and are susceptible to interference from densely packed spikes (Chen H et al., 2024). By incorporating the CSL module, MOSSNet effectively discriminates the morphological and directional characteristics of sorghum spikes, thereby improving detection accuracy and reliability. In addition, MOSSNet, built with

FIGURE 11
Detection results on different varieties with MOSSNet. **(a)** JN3; **(b)** JN6; **(c)** JN8.

Pytorch, is a deep learning model that can be deployed on cloud servers or edge computing devices. It supports automated, real-time monitoring of sorghum. As a result, it provides farmers with efficient, accurate and informed planting and management strategies.

6 Conclusions

This paper proposes a multiscale and oriented sorghum spike detection method in UAV images. Experimental results show that MOSSNet accurately identifies and counts sorghum spike under field conditions, achieving mAP of 90.3%. MOSSNet shows excellent performance in predicting spike orientation, with $RMSE_a$ and MAE_a of 14.6 and 12.5 respectively, outperforming other directional detection algorithms. Com-pared to general object detection algorithms which output horizonal detection boxes, MOSSNet also demonstrates high efficiency in counting sorghum spikes, with *RMSE* and *MAE* values of 9.3 and 8.1, respectively. These results demonstrate the model's ability to handle complex scenes with dense distribution, strong occlusion, and complicated background information. This highlights its robustness and generalizability, making it an effective tool for sorghum spike detection and counting. In the future, we plan to further explore the detection capabilities of MOSSNet at different stages of sorghum growth. This will involve implementing object model improvements tailored to each stage and developing a real-time workflow for accurate sorghum spike detection and counting.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

JZ: Conceptualization, Methodology, Writing – original draft, Writing – review & editing. ZJ: Methodology, Writing – review & editing. JW: Software, Writing – review & editing. ZW: Validation, Writing – review & editing. YG: Visualization, Writing – review & editing. YZ: Investigation, Writing – review & editing. SC: Resources, Writing – review & editing. WW: Data curation, Writing – review & editing. YS: Formal Analysis, Funding acquisition, Project administration, Supervision, Writing – review & editing. PL: Funding acquisition, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research and/or publication of this article. This research was supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (24KJB210006), the China Agriculture Research System (CARS-06-14.5-B5), the HAAFS Basic Science and Technology Contract Project (HBNKY-BGZ-02), and the Modern Agriculture Research System of Hebei Province (HBCT2024070201).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

References

AB, P., Jordan, D. R., Hammer, G. L., Armstrong, R. N., McLean, G., Laws, K., et al. (2015). The use of *in-situ* proximal sensing technologies to determine crop characteristics in sorghum crop breeding. *Res Gate* 4, 101231. Available online at: https://www.researchgate.net/publication/284713777_The_Use_of_In-Situ_Proximal_Sensing_Technologies_to_Determine_Crop_Characteristics_in_Sorghum_Crop_Breeding

Abderahman Rejeb, A. A., Rejeb, K., and Horst, T. (2022). Drones in agriculture: A review and bibliometric analysis. *Comput. Electron. Agric.* 198, 107017. doi: 10.1016/j.compag.2022.107017

Ahmad, A., Ordoñez, J., Cartujo, P., and Martos, V. (2020). Remotely piloted aircraft (RPA) in agriculture: A pursuit of sustainability. *Agronomy* 11, 7. doi: 10.3390/agronomy11010007

Alzubaidi L, Z. J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., et al. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* 8, 13. doi: 10.1186/s40537-021-00444-8

Bao, J., Zhang, H., Wang, F., Li, L., Zhu, X., Xu, J., et al. (2024). Telomere-to-telomere genome assemblies of two Chinese Baijiu-brewing sorghum landraces. *Plant Commun.* 5, 100933. doi: 10.1016/j.xplc.2024.100933

Baye, W., Xie, Q., and Xie, P. (2022). Genetic architecture of grain yield-related traits in sorghum and maize. *Int. J. Mol. Sci.* 23, 26. doi: 10.3390/ijms23052405

Cai, E., Baireddy, S., Yang, C., Crawford, M., and Delp, E. J. (2021) in *Panicle Counting in UAV Images For Estimating Flowering Time in Sorghum*, Vol. 13. 66–77

(IEEE). Available online at: https://www.researchgate.net/publication/355239845_Panicle_Counting_in_UAV_Images_for_Estimating_Flowering_Time_in_Sorghum.

Chang, A., Jung, J., Yeom, J., and Landivar, J. (2021). 3D characterization of sorghum panicles using a 3D point cloud derived from UAV imagery. *Remote Sens.* 13, 282. doi: 10.3390/rs13020282

Chen H, C. H., Huang, X., Zhang, S., Chen, S., Cen, F., He, T., et al. (2024). Estimation of sorghum seedling number from drone image based on support vector machine and YOLO algorithms. *Front. Plant Sci.* 15, 1399872. doi: 10.3389/fpls.2024.1399872

Chenyong, M., Alejandro, P., Zheng, X., Eric, R., Jinliang, Y., and James, C. S. (2020). Semantic segmentation of sorghum using hyperspectral data identifies genetic associations. *Plant Phenomics* 12, 10. doi: 10.34133/2020/4216373

Dai, J., Qi, H., Xiong, Y., Li, Y., Zhang, G., Hu, H., et al. (2017) in *Deformable Convolutional Networks*. 764–773 (IEEE). Available online at: https://ieeexplore.ieee.org/document/8237351.

Duan, Y., Han, W., Guo, P., and Wei, X. (2024). YOLOv8-GDCI: research on the phytophthora blight detection method of different parts of chili based on improved YOLOv8 model. *Agronomy* 14, 2734. doi: 10.3390/agronomy14112734

Genze, N., Ajekwe, R., Gureli, Z., Haselbeck, F., Grieb, M., and Grimm, D. G. (2022). Deep learning-based early weed segmentation using motion blurred UAV images of sorghum fields. *Comput. Electron. Agric.* 202, 107388. doi: 10.1016/j.compag.2022.107388

Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., and Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* 1, 14. doi: 10.34133/2019/1525874

Gonzalo-Martín, C., and García-Pedrero, A. (2021). Improving deep learning sorghum head detection through test time augmentation. *Comput. Electron. Agric.* 186, 106179. doi: 10.1016/j.compag.2021.106179

Guo, W., Potgieter, A. B., Jordan, D., Armstrong, R., and Ninomiya, S. (2016). Automatic detecting and counting of sorghum heads in breeding field using RGB imagery from UAV. *Proc. CIGR-AgEng Conf.*, 1–5. Available online at: https://www.researchgate.net/publication/307881406_Automatic_detecting_and_counting_of_sorghum_heads_in_breeding_field_using_RGB_imagery_from_UAV.

Guo, W., Zheng, B., Potgieter, A. B., Diot, J., Watanabe, K., Noshita, K., et al. (2018). Aerial imagery analysis – quantifying appearance and number of sorghum heads for applications in breeding and agronomy. *Front. Plant Sci.* 9, 1544. doi: 10.3389/fpls.2018.01544

Hasan MM, C. J., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using Convolutional Neural Networks. *Plant Methods* 14, 1–13. doi: 10.1186/s13007-018-0366-8

Hossain, Md. S., Islam, M., Rahman, Md. M., Mostofa, M. G., and Khan, Md. A. R. (2022). Sorghum: A prospective crop for climatic vulnerability, food and nutritional security. *J. Agric. Food Res.* 8, 100300. doi: 10.1016/j.jafr.2022.100300

Huang, Y., Qian, Y., Wei, H., Lu, Y., Ling, B., and Qin, Y. (2023). A survey of deep learning-based object detection methods in crop counting. *Comput. Electron. Agriculture.* 215, 108425. doi: 10.1016/j.compag.2023.108425

James, C., Smith, D., He, W., Chandra, S. S., and Chapman, S. C. (2024). GrainPointNet: A deep-learning framework for non-invasive sorghum panicle grain count phenotyping. *Comput. Electron. Agric* 217, 108485. doi: 10.1016/j.compag.2023.108485

Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Fu, H., et al. (2017). R2CNN: rotational region CNN for orientation robust scene text detection. *IEEE*, 3610–3615. Available online at: https://ieeexplore.ieee.org/document/8545598.

Jianqing, Z., Yucheng, C., Suwan, W., Jiawei, Y., Xiaolei, Q., Xia, Y., et al. (2023). Small and oriented wheat spike detection at the filling and maturity stages based on wheatNet. *Plant Phenomics* 5, 1–13. doi: 10.34133/plantphenomics.0109

Jin, X., Zarco-Tejada, P., Schmidhalter, U., Reynolds, M. P., and Li, S. (2021). High-throughput estimation of crop traits: A review of ground and aerial phenotyping platforms. *Geosci. Remote Sens.* 9, 200–231. doi: 10.1109/MGRS.6245518

Khalifa, M., and Eltahir, E. A. (2023). Assessment of global sorghum production, tolerance, and climate risk. *Front. Sustain. Food Syst.* 7, 1184373. doi: 10.3389/fsufs.2023.1184373

Li, H., Wang, P., and Huang, C. (2022). Comparison of deep learning methods for detecting and counting sorghum heads in UAV imagery. *Remote Sens.* 14, 3143. doi: 10.3390/rs14133143

Liang, Y., Li, H., Wu, H., Zhao, Y., Liu, Z., Liu, D., et al. (2024). A rotated rice spike detection model and a crop yield estimation application based on UAV images. *Comput. Electron. Agric.* 224, 109188. doi: 10.1016/j.compag.2024.109188

Liaqat, W., Altaf, M. T., Barutçular, C., Mohamed, H. I., Ahmad, H., Jan, M. F., et al. (2024). Sorghum: a star crop to combat abiotic stresses, food insecurity, and hunger under a changing climate: a review. *J. Soil Sci. Plant Nutr.* 24, 74–101. doi: 10.1007/s42729-023-01607-7

Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. *IEEE*, 2999–3007. doi: 10.1109/ICCV.2017.324

Lin, Z., and Guo, W. (2020). Sorghum panicle detection and counting using unmanned aerial system images and deep learning. *Front. Plant Sci.* 11, 534853. doi: 10.3389/fpls.2020.534853

Liu JH, Z., Weng, L., and Yang, Y. (2017). "Rotated region based CNN for ship detection," in *2017 IEEE International Conference on Image Processing (ICIP)*. 900–904 (IEEE). Available online at: https://ieeexplore.ieee.org/document/8296411.

Lovisotto, G., Finnie, N., Munoz, M., Mummadi, C. K., and Metzen, J. H. (2022). Give me your attention: dot-product attention considered harmful for adversarial patch robustness. *IEEE* 6, 477–483. doi: 10.1109/CVPR52688.2022.01480

Malambo, L., Popescu, S., Ku, N.-W., Rooney, W., Zhou, T., and Moore, S. (2019). A deep learning semantic segmentation-based approach for field-level sorghum panicle counting. *Remote Sens.* 11, 2939. doi: 10.3390/rs11242939

Ming, Q., Miao, L., Zhou, Z., Yang, X., and Dong, Y. (2021). Optimization for oriented object detection via representation invariance loss. *Comput. Sci.* 19, 1–5. doi: 10.1109/LGRS.2021.3115110

Neubeck, A., and Gool, L. J. V. (2006). "Efficient non-maximum suppression," in *Proceedings of the International Conference on Pattern Recognition*, Vol. 3. 850–855. Available online at: https://ieeexplore.ieee.org/abstract/document/1699659.

Ostmeyer TJ, B. R., Kirkham, M. B., and Bean S and Jagadish, S. V. K. (2022). Enhancing sorghum yield through efficient use of nitrogen - challenges and opportunities. *Front. Plant Sci.* 13, 845443. doi: 10.3389/fpls.2022.845443

Perich G, H. A., Anderegg, J., Roth, L., Boer, M. P., Walter, A., Liebisch, F., and Aassen, H.(2020). Assessment of multi-image unmanned aerial vehicle based high-throughput field phenotyping of canopy temperature. *Front. Plant Sci.* 11, 150. doi: 10.3389/fpls.2020.00150

Qian, W., Yang, X., Peng, S., Yan, J., and Guo, Y. (2021). "Learning modulated loss for rotated object detection," in *Proceedings of the National Conference on Artificial Intelligence*, Vol. 35. 2458–2466. Available online at: https://aaai.org/papers/02458-learning-modulated-loss-for-rotated-object-detection/.

Qiu, S. G., Gao, J., Han, M., Cui, Q., Yuan, X., and Wu, C. (2024). Sorghum spike detection method based on gold feature pyramid module and improved YOLOv8s. *Sensor* 25, 104. doi: 10.3390/s25010104

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE* 39, 1137–1149. doi: 10.1109/TPAMI.2016.2577031

Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., and Savarese, S. (2019) in *Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression*. 658–666 (IEEE). Available online at: https://ieeexplore.ieee.org/document/8953982.

Salas Fernandez MG, B. Y., Tang, L., and Schnable, P. S. (2017). A high-throughput, field-based phenotyping technology for tall biomass crops. *Plant Physiol.* 174, 2008–2022. doi: 10.1104/pp.17.00707

Sanaeifar, A., Guindo, M. L., Bakhshipour, A., Fazayeli, H., Li, X., and Yang, C. (2023). Advancing precision agriculture: The potential of deep learning for cereal plant head detection. *Comput. Electron. Agric.* 209, 107875. doi: 10.1016/j.compag.2023.107875

Tan, M., and Le,Q. V. (2019). "EfficientNet: rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. 6105–6114 (PMLR). Available online at: https://arxiv.org/abs/1905.11946.

Tong, Z., Chen, Y., Xu, Z., and Yu, R. (2021). Wise-IoU bounding box regression loss with dynamic focusing mechanism. *arXiv preprint arXiv:2301.10051*. doi: 10.48550/arXiv.2301.10051

Wang, J., Gao, Z., Zhang, Y., Zhou, J., Wu, J., and Li, P. (2022). Real-time detection and location of potted flowers based on a ZED camera and a YOLO V4-tiny deep learning algorithm. *Horticulturae* 8, 21. doi: 10.3390/horticulturae8010021

Wang, R., Wang, H., Huang, S., Zhao, Y., Chen, E., Li, F., et al. (2023). Assessment of yield performances for grain sorghum varieties by AMMI and GGE biplot analyses. *Front. Plant Sci.* 14, 1261323. doi: 10.3389/fpls.2023.1261323

Wang, Z., Nie, T., Lu, D., Zhang, P., Li, J., Li, F., et al. (2024). Effects of different irrigation management and nitrogen rate on sorghum (Sorghum bicolor L.) growth, yield and soil nitrogen accumulation with drip irrigation. *Agronomy* 14, 215. doi: 10.3390/agronomy14010215

Wei, L., Dragomir, A., Dumitru, E., Christian, S., Scott, R., Cheng-Yang, F., et al. (2016). *SSD: Single Shot MultiBox Detector* Vol. 9905 (Cham: Springer), 21–37.

Wu, T., Zhong, ,. S., Chen, H., and Geng, X. (2023). Research on the method of counting wheat ears via video based on improved YOLOv7 and deepSort. *Sensors* 23, 4880. doi: 10.3390/s23104880

Xue, X., Niu, W., Huang, J., Kang, Z., Hu, F., Zheng, D., et al. (2024). TasselNetV2++: A dual-branch network incorporating branch-level transfer learning and multilayer fusion for plant counting. *Comput. Electron. Agric.* .223, 109103. doi: 10.1016/j.compag.2024.109103

Yan, B., Fan, P., Lei, X., Liu, Z., and Yang, F. (2022). A real-time apple targets detection method for picking robot based on shufflenetV2-YOLOX. *Agriculture* 13, 1619. doi: 10.3390/agriculture12060856

Yang, X., Liu, Q., Yan, J., Li, A., Zhang, Z., and Yu, G. (2019). R3Det: refined single-stage detector with feature refinement for rotating object. *Proc. AAAI Conf. Artif. Intell.* 35, 3163–3171. doi: 10.1609/aaai.v35i4.16426

Yang, X., and Yan, J. (2020). "Arbitrary-oriented object detection with circular smooth label," in *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020. 677–694 (Springer International Publishing), Proceedings, Part VIII 16. Available online at: https://link.springer.com/chapter/10.1007/978-3-030-58598-3_40.

Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., et al. (2019). "SCRDet: towards more robust detection for small, cluttered and rotated objects," in *IEEE/CVF International Conference on Computer Vision (ICCV)*. 8231–8240. Available online at: https://www.x-mol.com/paper/1408111109224333312.

Yang, N., Yu, J., Wang, A., Tang, J., Zhang, R., Xie, L., et al. (2020). A rapid rice blast detection and identification method based on crop disease spores' diffraction fingerprint texture. *J. Sci. Food agriculture.* 100, 3608–3621. doi: 10.1002/jsfa.10383

Zarei, A., Li, B., Schnable, J. C., Lyons, E., Pauli, D., Barnard, K., et al. (2024). PlantSegNet: 3D point cloud instance segmentation of nearby plant organs with identical semantics. *Comput. Electron. Agric.* 221, 108922. doi: 10.1016/j.compag.2024.108922

Zhang, F., Chen, Z., Ali, S., Yang, N., Fu, S., and Zhang, Y. (2023). Multi-class detection of cherry tomatoes using improved YOLOv4-Tiny. *Int. J. Agric. Biol. Eng.* 16, 225–231. doi: 10.25165/j.ijabe.20231602.7744

Zhang, Z., Lu, Y., Zhao, Y., Pan, Q., Jin, K., Xu, G., et al. (2023). TS-YOLO: an all-day and lightweight tea canopy shoots detection model. *Agronomy* 13, 1411. doi: 10.3390/agronomy13051411

Zhang, L., Song, X., Niu, Y., Zhang, H., Wang, A., Zhu, Y., et al. (2024a). Estimating winter wheat plant nitrogen content by combining spectral and texture features based on a low-cost UAV RGB system throughout the growing season. *Agriculture* 14, 456. doi: 10.3390/agriculture14030456

Zhang, L., Wang, A., Zhang, H., Zhu, Q., Zhang, H., Sun, W., et al. (2024b). Estimating leaf chlorophyll content of winter wheat from UAV multispectral images using machine learning algorithms under different species, growth stages, and nitrogen stress conditions. *Agriculture* 14, 1064. doi: 10.3390/agriculture14071064

Zhang, Y., Xiao, D., Liu, Y., and Wu, H. (2022). An algorithm for automatic identification of multiple developmental stages of rice spikes based on improved Faster R-CNN. *Crop J.* 10, 1323–1333. doi: 10.1016/j.cj.2022.06.004

Zhang, S., Yang, Y., Tu, L., Fu, T., Chen, S., Cen, F., et al. (2025). Comparison of YOLO-based sorghum spike identification detection models and monitoring at the flowering stage. *Plant Methods* 21, 20. doi: 10.1186/s13007-025-01338-z

Zhang, T., Zhou, J., Liu, W., Yue, R., Yao, M., Shi, J., et al. (2024). Seedling-YOLO: high-efficiency target detection algorithm for field broccoli seedling transplanting quality based on YOLOv7-tiny. *Agronomy* 14, 931. doi: 10.3390/agronomy14050931

Zhao, J., Yan, J., Xue, T., Wang, S., Qiu, X., and Yao, X. (2022). A deep learning method for oriented and small wheat spike detection (OSWSDet) in UAV images. *Comput. Electron. Agric.* 198, 107087. doi: 10.1016/j.compag.2022.107087

Zhao, J., Zhang, X., Yan, J., Qiu, X., Yao, X., Tian, Y., et al. (2021). A wheat spike detection method in UAV images based on improved YOLOv5. *Remote Sens.* .13, 3095. doi: 10.3390/rs13163095