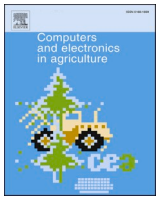
Computers and Electronics in Agriculture 234 (2025) 110289

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag



Original papers

A point-supervised algorithm with multiscale semantic enhancement for counting multiple crop plants from aerial imagery

Huibin Li [a,1], Huaiyang Liu [b,1], Wenbo Wang [b], Haozhou Wang [c], Qiangyi Yu [a], Jianping Qian [a], Wenbin Wu [a], Yun Shi [a,d,*], Changxing Geng [b,**]

[a] State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100086 China
[b] School of Mechanical and Electrical Engineering, Soochow University, Suzhou 215000, China
[c] Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 188-0002, Japan
[d] Suzhou Zhongnong Digital Intelligence Technology Co. Ltd, Suzhou 215000, China

ARTICLE INFO

Keywords:
Plant counting
Point supervision
Aerial imagery
Semantic enhancement
Density map

ABSTRACT

Counting crop plants is important for agricultural activities such as crop breeding and yield prediction. Numerous studies have developed methods for counting individual crop plants or those with similar morphological characteristics. However, these methods often face challenges of low accuracy and poor generalization when counting multiple crop plants with significant scale variations in complex backgrounds. Hence, we proposed MCPCNet, a point-supervised algorithm that enhances multiscale semantics for counting multiple crop plants from aerial imagery. We also constructed the first dataset of multicategory crop plant counting (MCPC-Dataset). We developed a concurrent spatial group enhancement module, a residual dynamic dilated convolution module, and introduced the contextual transformer module with self-attention mechanism. These modules can reduce the impact of background, adapt to scale variations of multiple crops, and enhance the robustness of our algorithm, respectively. The experiment results on the MCPC-Dataset indicate that MCPCNet achieves excellent performance, with a mean absolute error (MAE) of 2.577, a mean square error (MSE) of 14.289, and a coefficient of determination (R^2) of 0.991. MCPCNet also has a clear advantage over the state-of-the-art (SOTA) point-supervised counting algorithm. In conclusion, MCPCNet provides a robust solution for high-precision counting of multiple crop plants and is a vital reference for future related research.

1. Introduction

Counting crop plants is essential for agricultural activities such as measuring seed germination rates, regulating plant density, and predicting crop yields. Traditional crop counting methods rely heavily on manual sampling and counting. Human error usually occurs when people perform tedious manual counting due to time and cost constraints. Meanwhile, manual counting observations in the field may cause varying degrees of damage to the crop (Cui et al., 2023). With the development of computer vision and remote sensing, intelligent, fast, and contactless crop plant counting methods have attracted extensive attention from researchers. These methods greatly reduce the labor intensity and cost of manual work, as well as improves the accuracy and efficiency of crop counting (Valente et al., 2020).

Deep learning-based crop detection and counting methods are mainly categorized into anchor-based algorithms and point-supervised algorithms. Anchor-based algorithms typically use predefined anchor boxes to detect and count crop plants. Several common algorithm improvements have been applied to enhance anchor-based algorithms to better suit specific research objects. For example, the Feature Pyramid Network (FPN) (Lin et al., 2017) is used to integrate multi-scale feature maps (Cai et al., 2019), improving the accuracy of the algorithm. The multi-scale context-aware model (Liu et al., 2019) is employed to extract features across different layers(Wang et al., 2023), and specialized

* Corresponding author at: State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China, Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100086 China.
** Corresponding author.
E-mail addresses: shiyun@caas.cn (Y. Shi), chxgeng@suda.edu.cn (C. Geng).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.compag.2025.110289
Received 11 July 2024; Received in revised form 9 March 2025; Accepted 11 March 2025
Available online 24 March 2025

prediction heads (Wang et al., 2024) are introduced for detecting specific target objects. For instance, to achieve precise wheat ear counting, Zhang et al. (2022) proposed an improved YOLOv5 method based on Spatial Pyramid Pooling (SPP) techniques, achieving an average precision of 82.5 % in evenly distributed scenarios. To enhance the generalization ability of wheat ears counting algorithms in uniformly distributed scenarios, Bao et al. (2023) designed TPH-YOLO with the addition of the transformer prediction head, which improved the generalization ability and accuracy of the model, and the average precision reached 88.8 %. However, the accuracy of TPH-YOLO can be seriously affected when crop sparsity is unevenly distributed. To address this challenge, Shahid et al. (2024) developed a tobacco plant counting network based on YOLOv7, the method achieved an average F1 score of 0.967, which aims to improve the accuracy and reliability of tobacco plant counting in various scenarios.

Moreover, during the seedling stage, Fu et al. (2023) observed that the small plant's detection is easily affected by soil and shadows by using Faster-RCNN, FCOS, and YOLOv5. To address this problem, Li et al. (2024) proposed an enhanced YOLOv7 for wheat plant counting, which focuses on improving small plants' detection and counting ability in complex scenarios, and the counting accuracy reaches 93.8 %. Anchor-based algorithms have good effects on crop plant counting tasks. However, anchor-based crop counting algorithms generally require extensive manual annotations of the dataset, which could be more time-consuming (Lu and Young, 2020). In addition, in the research process of counting algorithms, anchor-based algorithms also suffer from the difficulty of anchor box suppression when dealing with the task of counting complex scenarios (Farjon et al., 2023).

Point-supervised algorithms have gained considerable attention in crop plant counting research due to point annotations' convenience and effectiveness in target detection and localization (Xue et al., 2024). To address the problems of scale variation and stem-leaf interference in rice panicle counting, Xu et al. (2020) proposed an MHW-PD network, a point-supervised counting algorithm for indoor potted rice panicles. MHW-PD achieved 87 % counting accuracy by considering the feature scale variations of rice panicles and introducing a multiscale hybrid window preprocessing method. In outdoor crop plant counting studies, Bai et al. (2023b) proposed RiceNet for the field's rice plant counting, localization, and size estimation, the method achieved MAE of 8.6. However, RiceNet is less robust in scenarios such as complex lighting conditions and soil. In order to improve the ability of point-supervised algorithms to count crop plants in complex scenarios, Shao et al. (2021) constructed a rice ears dataset containing light variations and soil backgrounds and proposed ResFCN. The results show that ResFCN can suppress semantic features such as light and soil in low-complexity contexts, the MAE of ResFCN on the 300-size test set is 2.99, thus improving the robustness and generalization of plant counting.

Furthermore, to efficiently extract rice ear features in complex and densely planted backgrounds, Bai et al. (2023a) designed the RPloss function within the point-supervised algorithm and constructed an RPNet combining the density map and the attention map to realize the accurate counting of rice plants. RPNet also acquired good stability for wheat plant counting, the MAE of the rice plant counting is 3.1, while the MAE of the wheat plant counting is 4.0, which provides valuable insights into the methods of counting crop plants with similar shapes. Overall, the point-supervised algorithms exhibit significant potential and application value in crop plant counting, offering practical approaches and substantial support for high-precision counting research in complex scenes.

Despite the significant advances in crop plant counting research, counting various types and scales of crop plants in complex backgrounds remains challenging and needs to be researched. Crop plant scale includes parameters such as crop plant growth morphology and size (Tran and Phan, 2023). For anchor-based algorithms, designing anchor box scales and aspect ratio hyperparameters that accommodate multiple crop scales is complex, and improper anchor box suppression will severely affect detection accuracy (Deng et al., 2023). Additionally, many anchor box annotations substantially increase development costs (Madan et al., 2023). Conversely, while current point-supervised algorithms are practical for counting single or similar crop plants, they need help with tasks involving crop plants with large-scale differences, leading to low counting accuracy and poor generalization ability (Huang et al., 2023).

To solve the problems above, we developed MCPCNet, a point-supervised algorithm based on multiscale semantic enhancement for multiple counting crop plants. Moreover, MCPCNet significantly improves the accuracy and stability of counting crop plants of various scales in complex farmland environments. The specific novelties of the proposed MCPCNet are as follows:

We adopted a more convenient point-supervised strategy for constructing the MCPCNet, which can decrease the workload of anchor-based algorithms and reduce costs.
The first dataset of multicategory crop plant counting (MCPC-Dataset) was created by using aerial imagery. This dataset will support future research on point-supervised algorithms for counting crop plants.

(3) The concurrent spatial group enhancement module (CoSAGE) was developed to enhance crop plants' semantics by adjusting concurrent spatial group weights, effectively suppressing background information such as soil and shadows.

(4) A residual dynamic dilated convolution module (ReDyDiC) was designed by using parallel nonlinear convolutions to accommodate varying crop plant scales.

(5) We also introduced the contextual transformer (CoT) (Li et al., 2023) to integrate the contextual features of crop plants, combined with a self-attention mechanism, further strengthening the generalization ability of the algorithm.

The structure of the remaining parts of this paper is as follows: Section 2 covers the collection of aerial images of crop plants, the creation of the MCPC-Dataset, and the framework of MCPCNet. Section 3 presents the results of the ablation experiments and analyzes these results. Section 4 discusses the comparison experiments of MCPCNet with four SOTA algorithms. Section 5 concludes this paper.

2. Materials and methods

2.1. Data acquisition and processing

The experiments utilized data from four types of crop plants: watermelon, cotton, corn, and rice. The rice plant data is from a URC dataset (Yang et al., 2021) collected in August 2018. We collected corn, cotton, and watermelon plant data in May 2023 at the Laolonghe comprehensive experimental base in Changji, Xinjiang Uygur Autonomous Region, China (longitude 87.332E, latitude 44.263 N). The data collection process is shown in Fig. 1. The collection equipment is a quadcopter unmanned aerial vehicle (DJI Phantom 4 Multispectral RTK, DJI P4M RTK) equipped with a 2-megapixel RGB lens. The flight heights were set at 5 m for cotton, 15 m for corn, and 25 m for watermelon. Frontal and lateral overlaps were adjusted to 85 % and 75 %, respectively. The orthophoto images covered 2 acres of corn, 3 acres of cotton, and 4 acres of watermelon.

In order to realize the study of multicategory crop plant counting, this paper constructs the first MCPC-Dataset based on aerial imagery by manually annotating the collected data and URC dataset. First, we cropped the aerial orthophotos of the four types of crop plants into 640×640 pixel images. After screening, we ultimately collected 1,246 high-quality images. Subsequently, we used Labelme to annotate these images. Points are annotated as close to the center of the plant as possible. To address the annotation of crop plants located at the edges of the images, we adopted the following rule: if more than half of a crop

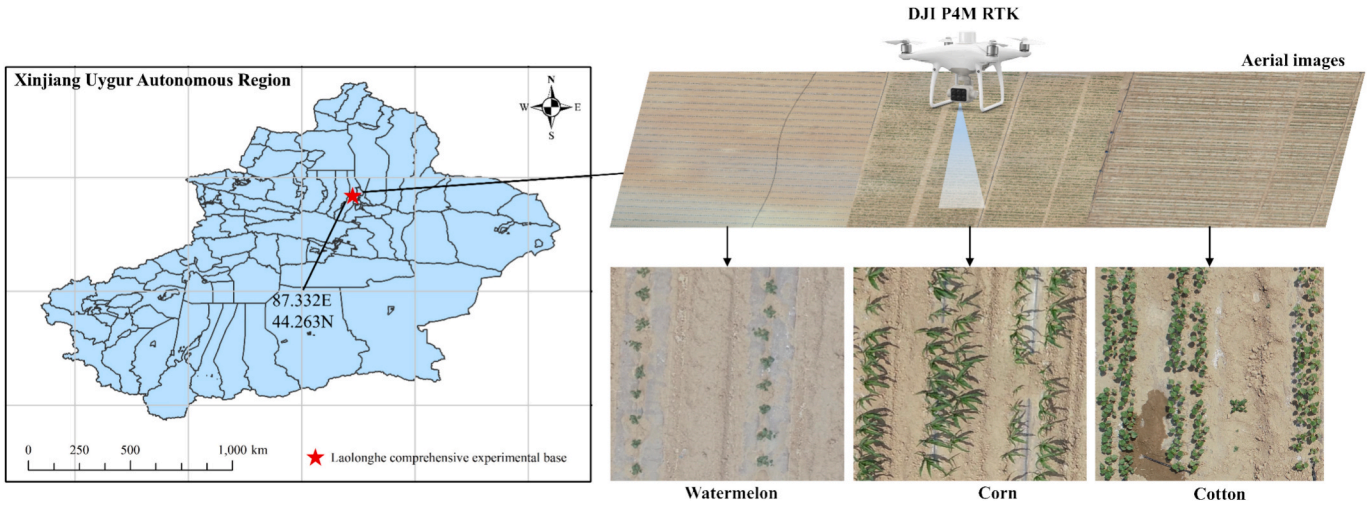


Fig. 1. Aerial images of corn, cotton, and watermelon fields captured at the Laolonghe Comprehensive Experimental Base.

plant's features are within the image boundaries, the plant is annotated; otherwise, it is excluded. This ensures consistent and accurate annotation while minimizing ambiguity in handling edge cases. Finally, we obtained an MCPC-Dataset which contains 71,566 manual annotation points. The MCPC-Dataset link is publicly available at GitHub in readme file (https://github.com/Nirvana557/MCPCNet/tree/main). We divided the MCPC-Dataset into a training set, validation set, and testing set in a 6:2:2 ratio for training and validation purposes. The information of the MCPC-Dataset is shown in Table 1.

The MCPC-Dataset is highly representative due to the significant differences in scale (size, shape, aspect ratio) and background among the four crop plants(see Fig. 2). Specifically, watermelon plants exhibit no adhesion or occlusion and possess simple plant morphology (see Fig. 2a). Rice and corn plants display tillering, moderate complexity of plant morphology, and the presence of shadows (see Fig. 2b and 2c). Cotton plants display tillering, high complexity of plant morphology, and the presence of shadows (see Fig. 2d).

Compared with anchor box annotation, point annotation is more efficient. To verify the efficiency of point annotation, we selected 10 images of each crop plant from the MCPC-Dataset. We annotated them using both point annotation and anchor box annotation. The annotation results are shown in Fig. 3. We recorded the total time of point annotation and anchor box annotation for each type of crop plant, denoted as T1 for point annotation and T2 for anchor box annotation. The specific information is shown in Table 2. It can be seen that the average time required for anchor box annotation is 2512.3 s, which is three to four times longer than the average time required for point annotation.

2.2. Framework of MCPCNet

In this section, we introduce the primary framework of MCPCNet. As illustrated in Fig. 4, the baseline of MCPCNet is M−SFANet (Thanasutives et al., 2021). This method was designed for crowd-counting in scenarios with significant scale variations. To enhance the M−SFANet's performance in detecting crop plants of multiple scales. We modified the M−SFANet by the following:

Table 1
Distribution of four crop plant types in the MCPC-Dataset.

Crop plant types	Number of images	Number of annotation points
Watermelon	272	4250
Rice	188	23,905
Corn	288	24,320
Cotton	498	19,091

(1) We employed the ResNet50 as the feature map extractor (FME) and removed the fully connected layers. Section 2.2.1 introduced the changes to the FME.

(2) For the feature decoder, we used a dual-path multiscale fusion decoder incorporating both an attention map path and a density map path. Section 2.2.2 introduced CoSAGE to decrease the impact of complex background. Section 2.2.3 introduced ReDyDiC to handle the scale variations of crop plants. Section 2.2.3 introduced the CoT module to enhance the algorithm's generalization ability during the feature fusion process.

The dual path fusion features generate a density map called F_{refine} by element-wise multiplication, the *sum* function is then used on F_{refine} to obtain the crop plants' counting results. In the following sections, we will introduce the MCPCNet in detail.

2.2.1. FME

The feature map encoder extracts semantic information from crop plant images. However, with the increasing depth of the algorithm, issues such as gradient vanishing, exploding gradients, and algorithm degradation may occur. The introduction of Residual Networks addresses these challenges, while also reducing the computational complexity (measured in Floating point operations or Flops) compared to classic convolutional algorithms such as the VGG series. In order to increase the depth of the algorithm while reducing computational complexity, we utilized a modified ResNet50 without fully connected layers as the feature map extractor (FME). Specifically, feature maps at sizes of 1/2, 1/4, 1/8, and 1/16 of the original crop plant images were obtained from the Conv2_3 to Conv5_3 layers, respectively.

2.2.2. CoSAGE

As crop plant features extracted by FME inevitably include background semantics such as soil and shadow, neural networks often struggle to achieve evenly distributed semantic responses for crop plants. Hence, we proposed CoSAGE (as Fig. 5), a novel module designed specifically for enhancing semantic responses in agricultural imagery. We consider a C-channels, $H \times W$ crop plants feature map $\mathscr{X}$ and divide it into G groups along the channel dimension. Then the group has a vector representation at every position in space, namely $\mathscr{X} \in \{x_{1\cdots m}\}$, $x_i \in R^{C/G}$, $m = H \times W$. In this group space, we aim to extract features that exhibit strong responses at the positions of crop plants. To achieve this, we advocate leveraging the comprehensive information within the entire group space to enhance the learning of semantic features in crucial regions, considering that noise does not dominate the features of the entire space. Therefore, we can use the global statistical feat g by applying the spatial averaging function $\mathscr{F}_{gp}(\cdot)$ to approximate the

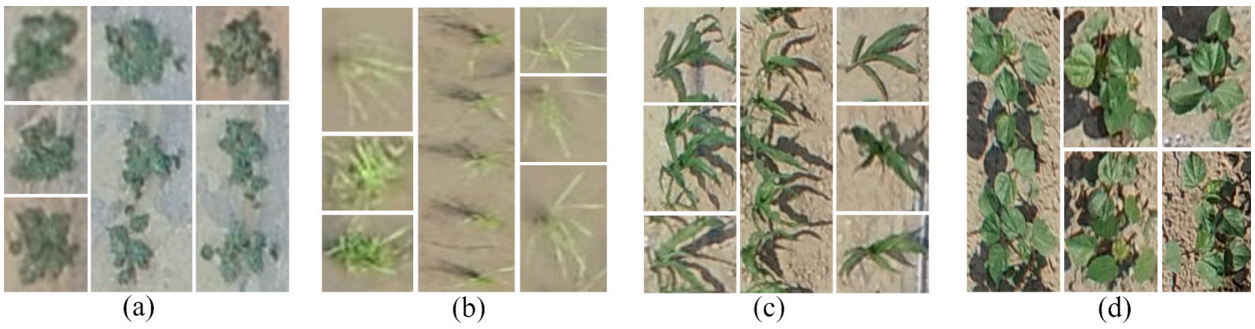


Fig. 2. Diversity of crop plants at different scales. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton.

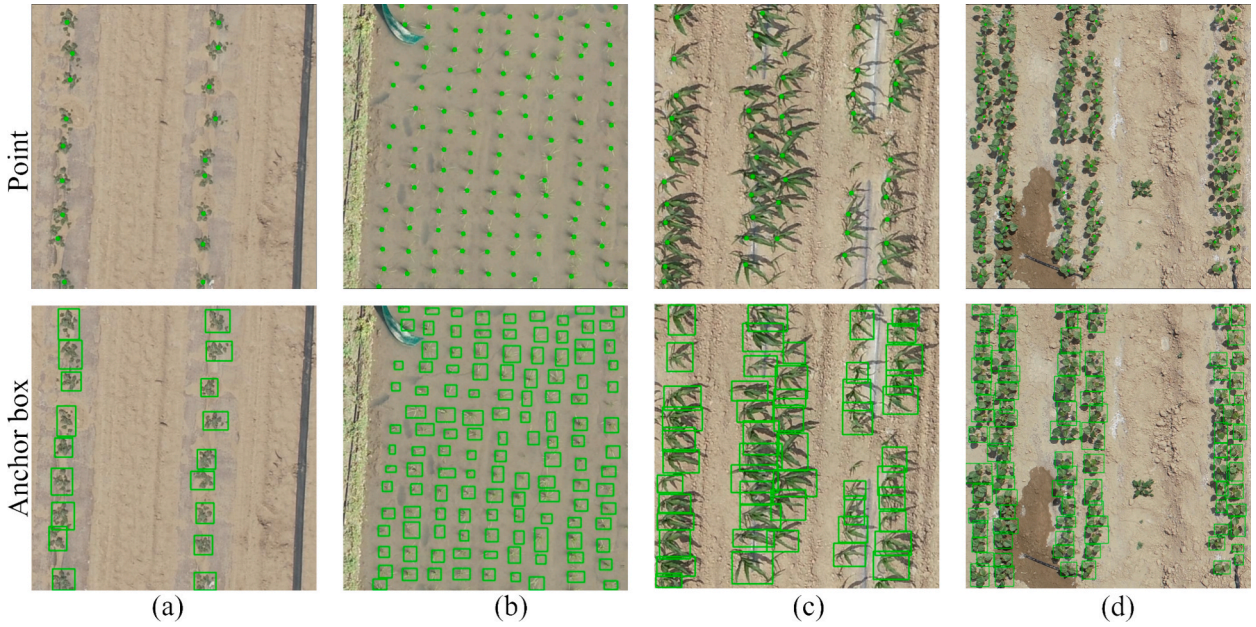


Fig. 3. Comparison of anchor box and point annotation types for crop plants. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton.

Table 2
Comparison of annotation time for point and anchor box methods across four crop types in the MCPC-Dataset.

Time consumption	Watermelon (s)	Rice (s)	Corn (s)	Cotton (s)
T1: Point annotations	265.9	1202.6	343.6	994.8
T2: Anchor box annotations	889.7	4268.9	2317.2	2573.4
T2-T1	623.8	3066.3	1973.6	1578.6

semantic vector learned by this group, as shown in equation(1):

$$g = \mathscr{F}_{gp}(\mathscr{X}) \tag{1}$$

Next, utilizing the global statistical feature denoted as g, we can derive the corresponding importance coefficient c_i for each feature. This coefficient is computed using the equation $c_i = g \cdot x_i$, which quantifies the degree of similarity between the global semantic feature g and local feature x_i. To mitigate potential biases in coefficient magnitudes across samples, we normalize c_i within the group space as equation(2):

$$\widehat{c}_i = \frac{c_i - \mu_c}{\sigma_c + \epsilon} \tag{2}$$

Which $\mu_c = \frac{1}{m}\sum_j^m c_j$, $\sigma_c{}^2 = \frac{1}{m}\sum_j^m (c_j - \mu_c)^2$, ϵ is a constant added for numerical stability. To make sure that the normalization inserted in the algorithm can represent the identity transform, we introduce a pair of parameters γ, β for each coefficient $\widehat{c}_i$, which scale and shift the normalized value as equation(3):

$$\sigma_i = \gamma \cdot \widehat{c}_i + \beta \tag{3}$$

Then we use the sigmoid function $\sigma(\cdot)$ and coefficients a_i to scale x_i as equation(4):

$$\overline{x}_i = x_i \cdot \sigma(a_i) \tag{4}$$

We consider imposing calibration weight $\overline{\mathscr{X}} = \overline{\mathscr{X}_1} + \overline{\mathscr{X}_2}$ in $\overline{x}_i$ to enhance crop plants' semantic while suppressing background semantic such as shadows and soil. First, $\mathscr{X}$ is spatially compressed by global pooling to generate the vector $z \in R^{1\times1\times C}$, which then flows through the ReLu and Sigmoid functions to obtain the calibration weight $\overline{\mathscr{X}_1}$. In addition, we directly perform a $1 \times 1 \times 1$ convolution of $\mathscr{X}$, activated by a sigmoid function to obtain the calibration weight $\overline{\mathscr{X}_2}$. We use $\overline{\mathscr{X}}$ as a calibration weight to scale $\overline{x}_i$ to obtain the semantically enhanced features $\widehat{x}_i$ of crop plants as equation(5):

$$\widehat{x}_i = \overline{\mathscr{X}} \times \overline{x}_i \tag{5}$$

Finally, the semantically enhanced crop plants feature $\overline{\mathscr{X}}$ can be represented as $\widehat{\mathscr{X}} = \{\widehat{x}_{1\ldots m}\}$.

2.2.3. ReDyDiC

Crop plants with substantial scale variations will impact the accuracy and robustness of the plant counting algorithm. To address this

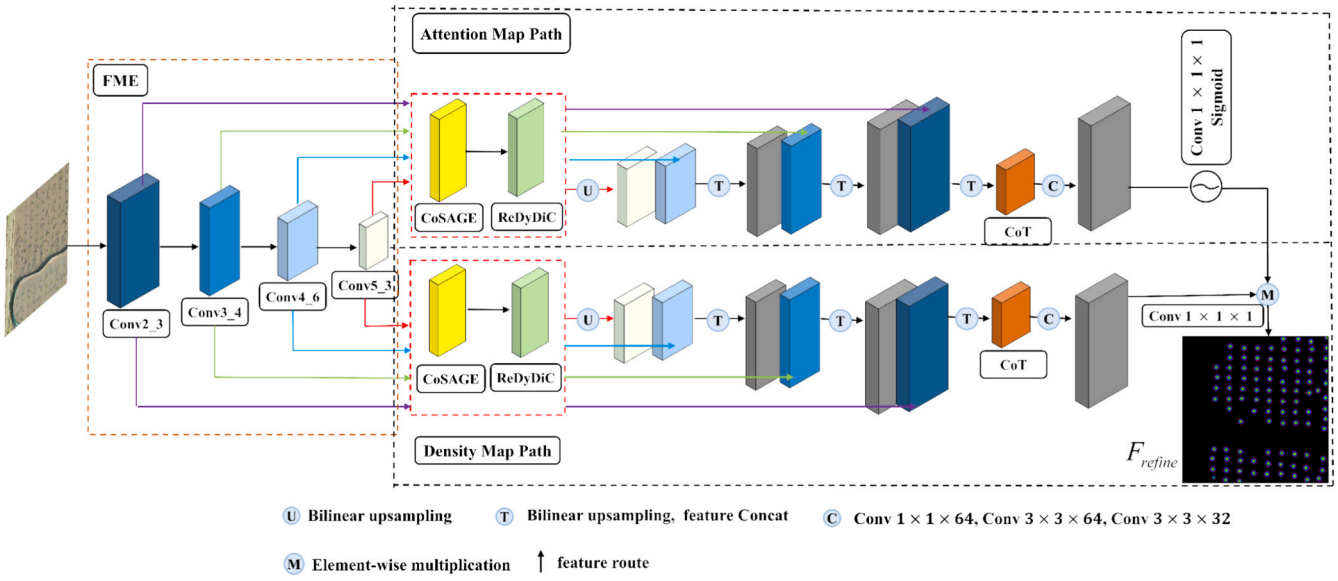


Fig. 4. Framework of MCPCNet. FME stands for feature map extractor, CoSAGE stands for concurrent spatial group enhancement module, and ReDyDiC stands for residual dynamic dilated convolution module.

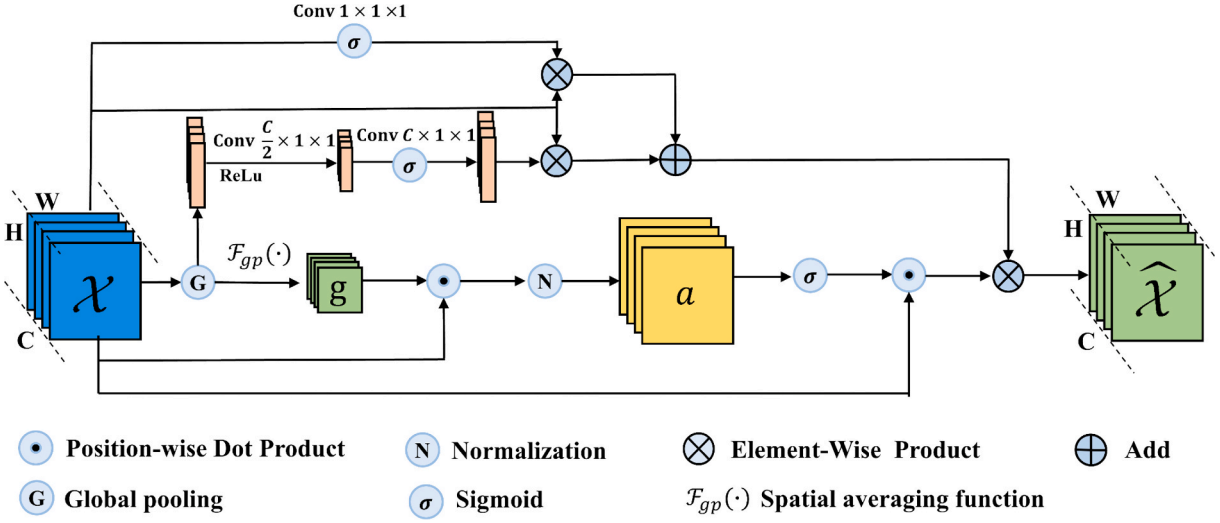


Fig. 5. Framework of CoSAGE.

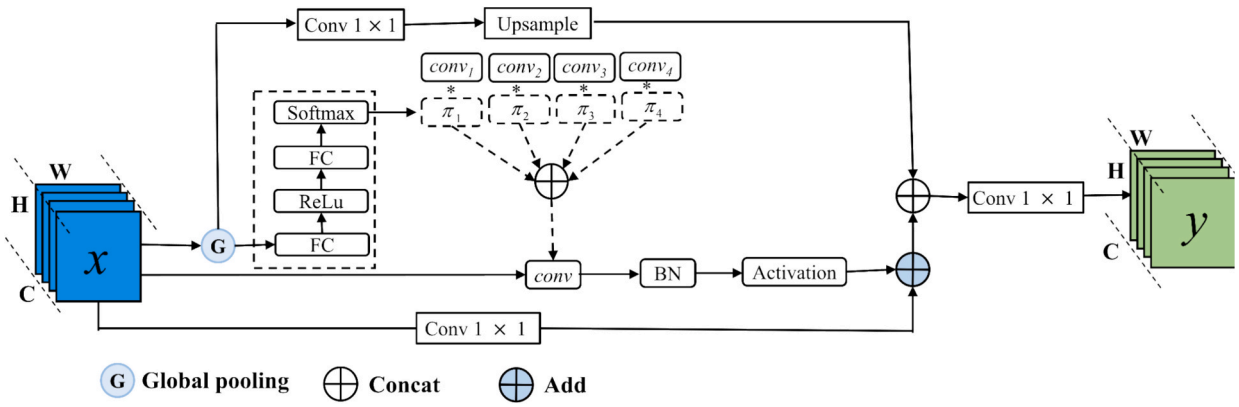


Fig. 6. Framework of ReDyDiC.

challenge, we propose enhancing the receptive field range using dilated convolutions. However, traditional Atrous Spatial Pyramid Pooling (ASPP) employs a fixed dilated rate. This presents two primary challenges for multicategory crop plant counting algorithms. Firstly, fixed dilated rates need help to accommodate the diverse scales of different crop plants. Secondly, incorporating multiple convolutional layers in ASPP, each with varying dilated rates substantially increases the computational complexity of crop plant counting algorithms (Liu et al., 2020). We proposed a Residual Dynamic Dilated Convolution Module (ReDyDiC) to address these issues. This module is designed to handle scale changes in crop plants using parallel nonlinear convolution without increasing algorithm computation by aggregating shared output channels (Chen et al., 2020).

As depicted in Fig. 6, the feature representation of crop plants undergoes an initial compression process through global pooling, followed by a fully connected layer, ReLU function, and Softmax normalization. This process generates 4 normalized weights $\{\pi_1, \pi_2, \pi_3, \pi_4\}$ with different convolution kernels, these normalized weights π_i are not fixed but dynamically adapt to variations in the input features x. The variable π_i adheres to the normalization constraint, ensuring that the sum of π_i equals to 1. The weights and biases of the parallel convolution $conv_i$ are denoted as $\{W_i, b_i\}$. Then, the nonlinear function can be expressed as follows:

$$\overline{x} = g\left(W^T(x)x + b(x)\right) \tag{6}$$

Where $g(\cdot)$ denotes activation function, $W(x)$ and $b(x)$ are shown in equation(7) and equation(8), respectively:

$$W(x) = \sum \pi_i(x) W_i \tag{7}$$

$$b(x) = \sum \pi_i(x) b_i \tag{8}$$

Then, the feature x of crop plants uses 1×1 convolution, and combines it with the dynamic convolution feature $\overline{x}$ to derive the calibration feature $\overline{y}$. Subsequently, applying the attention mechanism (shown at the top level of Fig. 6) to x and integrate it with the calibration feature $\overline{y}$, resulting in the adapted features y that account for variations in the scale of crop plants. In addition, the dropout technique is employed in the ReDyDiC to mitigate computational complexity.

2.2.4. Dual path multiscale fusion decoder

The decoder architecture consists of the density map path and the attention map path. A unified strategy is applied to both paths, which will be further elucidated in the subsequent sections. As shown in Fig. 5 and Fig. 6, the crop plant features processed by the CoSAGE and ReDyDiC modules are first up-sampled and cascaded for feature decoding. Additionally, as depicted in Fig. 7, the CoT module is integrated into the feature decoding process to effectively extract contextual information. By leveraging the self-attention mechanism of the CoT module, the generalization ability of the algorithm is significantly enhanced. MCPCNet utilizes element-wise multiplication to generate the density map F_{refine}, finally add the feature tensor of F_{refine} using the *sum* function to obtain the predicted number $S_{counting}$ of crop plants, the function is shown as equation(9):

$$S_{counting} = sum\left(F_{refine}\right) \tag{9}$$

2.3. Train method

2.3.1. Ground truth density map

The Gaussian method was applied to obtain a ground truth density map D^{GT}. Suppose the pixel x_i represents the position of the center of the crop plant in the image. Then, x_i can be used to construct a density map D^{GT} by convolution with a Gaussian kernel, as shown in equation(10):

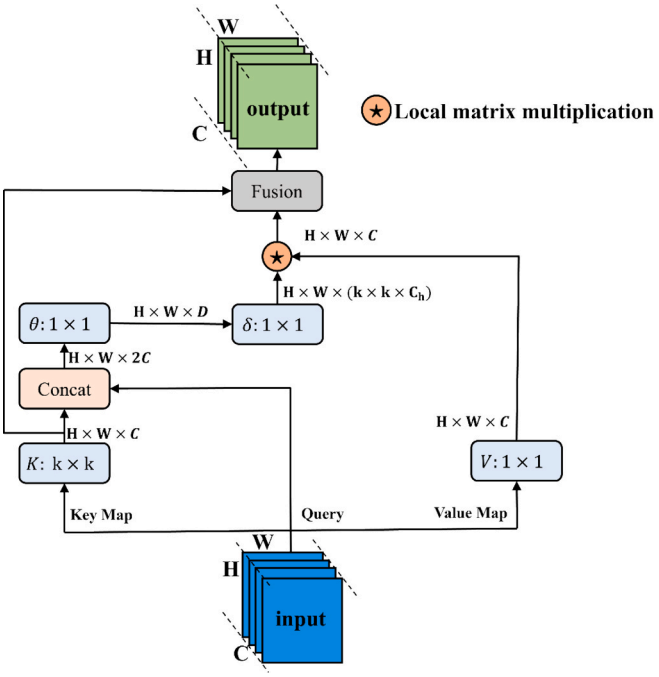


Fig. 7. Framework of CoT.

$$D^{GT} = \sum_{i=1}^{C} \delta(x - x_i) \cdot G_{\mu,\rho^2}(x) \tag{10}$$

For each crop plant annotation point x_i in ground truth δ, we convolve $\delta(x - x_i)$ with a Gaussian kernel $G_{\mu,\rho^2}(x)$, which is parameterized by μ(kernel size) and ρ(standard deviation) (see Fig. 8). Here, x denotes the position of the pixel in the imagery, and C is the number of annotation points.

Based on the density map ground truth, we continue to use the Gauss kernel to compute attention map ground truth as shown in equation(11) and equation(12):

$$\mathbb{Z} = D_i^{GT} \times G_{\mu,\rho^2}(x) \tag{11}$$

$$\forall x \in \mathbb{Z}, A_i^{GT} = \begin{cases} 0, x < th \\ 1, x \geq th \end{cases} \tag{12}$$

Where th is the threshold set to 0.001 in our experiments. We obtained a binary attention map ground truth to direct the attention map generation path to focus on the crop plants region and surrounding places. In our experiments, we set the value of μ to 2 and the value of ρ to 3.

2.3.2. Training parameter settings

We employed data augmentation to expand the MCPC-Dataset and mitigate overfitting in MCPCNet caused by limited data. Augmentation included resizing the image's short side (fixed at 512 pixels) by a random ratio within [0.8, 1.2], cropping 400×400 pixels patches at random locations, and applying random horizontal flipping (0.5 probability). Additional transformations included gamma-contrast adjustments within [0.5, 1.5] (0.3 probability) and grayscale conversion (0.1 probability). This process generated a training dataset of 2,170 images with 122,638 annotated points, enabling MCPCNet to better learn crop features and improve generalization.

Referring to the training method as Liu et al. (2024) proposed, we set the learning rate of the Adam optimizer to $1e-4$, the weight decay of $5e-3$ is used to train the model, because it shows faster convergence than standard stochastic gradient descent with momentum in our experiments. The batch size for training was 8, and the number of training

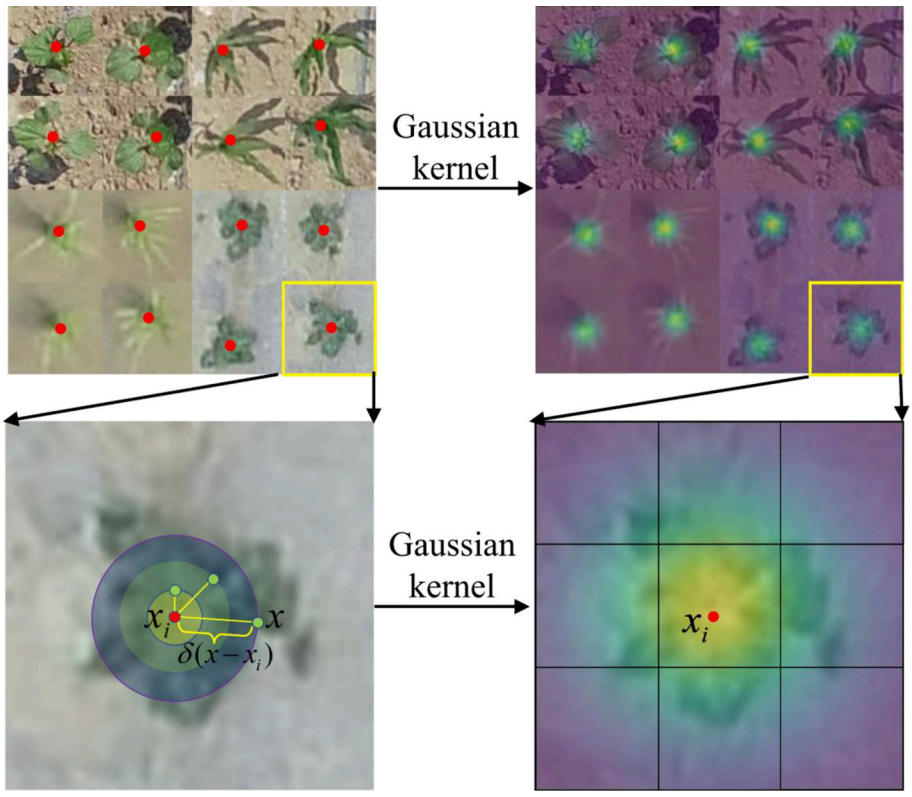


Fig. 8. The generation process of ground truth density map D^{GT}. The red points denote the center pixel x_i of annotation crop plants, green points denote the pixel points x around the annotation crop plants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

epochs was 700. To optimize model parameters, we employed an automated strategy that integrates the validation set into the training process. Specifically, during training, the model's performance on the validation set was evaluated every 5 epochs, and the results were automatically used to adjust the training loss, guiding further optimization. This iterative process ensures that the model minimizes training loss while maintaining generalizability, as feedback from the validation set helps prevent overfitting and fine-tunes the parameters dynamically.

2.3.3. Training platform

For specific information about the hardware platform and software environment used for training and testing, please refer to Table 3.

2.4. Experimental evaluation

To effectively evaluate the accuracy and robustness of MCPCNet, we introduce the following evaluation metrics:

(1) The MAE represents the average magnitude of errors between predicted and ground truth values for crop plants. MAE is robust against outliers and effectively evaluates the accuracy of the crop plant counting algorithm. Equation(13) shows the MAE is:

Table 3
Hardware platform and software environment for training.

Item	Details
GPU	Tesla v100-SXM-2, 32 GB
CPU	Intel Xeon(R) Gold 6226R v4, 2.90 GHz, 64 cores
Memory (RAM)	256G
Operating System	Ubuntu 20.04 LTS
Tensor library	PyTorch 1.12
CUDA Version	CUDA 11.5
Programming language	Python 3.8

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|C_i - C_i^{GT}\right| \tag{13}$$

(2) The MSE serves as a metric reflecting the degree of difference between prediction and ground truth values for crop plants. It imposes a substantial penalty for significant errors, thus illustrating the robustness of the algorithm in counting crop plants. Equation(14) shows the MSE:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left|C_i - C_i^{GT}\right|^2 \tag{14}$$

(3) The Symmetric Mean Absolute Percentage Error (SMAPE) is a metric used to quantify the average percentage error between predicted and ground truth values of crop plants. It mitigates bias introduced by dataset size by normalizing the difference between predicted and labeled values. The SMAPE is expressed in equation(15):

$$SMAPE = \frac{100}{N}\sum_{i=1}^{N}\frac{\left|C_i - C_i^{GT}\right|}{\left|C_i + C_i^{GT}\right|/2} \tag{15}$$

(4) The R^2 quantifies the effectiveness of linear regression in predicting crop plant values. Equation(16) provides the calculating method for R^2:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left|C_i - C_i^{GT}\right|^2}{\sum_{i=1}^{N}\left|C_i - \overline{C_i^{GT}}\right|^2} \tag{16}$$

C_i denotes the prediction values of crop plants, C_i^{GT} denotes the ground truth of crop plants, $\overline{C_i^{GT}}$ denotes the mean value of ground truth, N denotes the number of images in the testing set. Smaller values of MAE, MSE, and SMAPE characterize optimal evaluation metrics. A higher R^2 indicates superior performance, thus demonstrating heightened algorithm accuracy and robustness.

3. Results

3.1. Overall results

This section presents the training, testing, and comparative experimental outcomes of MCPCNet on the MCPC-Dataset against the baseline. During the training phase, we initially employ three distinct algorithms: Res50 utilizing only the ResNet50 backbone, Res50_C incorporating the CoSAGE, and Res50_CR integrating ReDyDiC into the Res50_C. Fig. 9 illustrates the variation trend of algorithm loss curves during training, with all algorithms converging after the 700*th* training epoch. Upon comparing their convergence outcomes, MCPCNet exhibits the lowest loss value, followed by other algorithms, with ResNet50 showing the highest loss value. This suggests that MCPCNet achieves the most accurate fit between predicted and actual crop plant data, resulting in superior prediction accuracy (Tian et al., 2022).

Next, we conducted a preliminary qualitative analysis of the counting capability of MCPCNet and found that it performs well in counting the plants of four types of crops. As shown in Fig. 10, after predicting the MCPC-Dataset, the predicted areas in F_{refine} by MCPCNet are nearly uniformly distributed and independent. Each peak position in F_{refine} can be regarded as the center of a plant. This distribution closely matches the ground truth for crop plant prediction by MCPCNet.

Finally, to validate the improvement of MCPCNet over the baseline in crop plant counting, we compared the plant counting results of the two algorithms: By considering all crop plants in the testing set and by focusing on individual crop types, the processed images include 58 for cotton, 38 for rice, 99 for maize, and 54 for watermelon. The crop counting numbers from these images were analyzed using Excel tools for linear fitting, with each point derived from a single image's ground truth and predicted values. On the one hand, as shown in Fig. 11a, the R^2 values for the crop plant counting results are 0.991 for MCPCNet and 0.985 for the Baseline, respectively. The counting results from MCPCNet are more evenly distributed around the fitting line, and the slope of this line is closer to the 1:1 line. These findings indicate that MCPCNet improves accuracy and robustness in counting various crop plants.

On the other hand, we examined the specific differences in counting results between MCPCNet and Baseline using the MCPC-Dataset. As shown in Fig. 11b, the prediction results of MCPCNet and Baseline for rice plants are comparable, with R^2 of 0.987 and 0.984, respectively. However, there were more predictions with excellent dispersion on both sides of the fitting lines for cotton and maize plants. The R^2 of MCPCNet prediction results were increasing by 14.2 % and 9.3 % for corn and cotton plants with large scale differences, respectively, compared with the baseline. This is mainly because MCPCNet is more easily adapted to counting crop plants with large-scale differences and demonstrates strong robustness and generalization capabilities. In addition, as shown in Fig. 11e, the prediction results for watermelon plants exhibited a vertical distribution pattern. This occurred because the image contained fewer watermelon plants, leading to several identical ground truth. The density differences among crop plants due to varying planting practices and growth patterns, and the degree of discrepancy between the ground truth and predicted values all influence both R^2 and the slope of the fitting line. Although the R^2 of MCPCNet and Baseline for counting watermelon plants in Fig. 11e are comparable, the fitting line for the predicted results of MCPCNet is closer to the ideal 1:1 line. This suggests that stable counting could be achieved even with a relatively small number of crop plants. In summary, the ability to achieve high-precision and robust counting for various crop types with substantial size differences is demonstrated.

A Wilcoxon signed-rank test was performed to analyze the absolute prediction errors (defined as the unsigned difference between predicted values and ground truth values) of MCPCNet and the Baseline. The non-parametric test was selected due to observed deviations from normality in the error distributions. Statistical significance was confirmed with a p-value of 0.017, which falls below the standard significance threshold of 0.05. This indicates a 1.7 % probability that the observed accuracy improvement of MCPCNet over the Baseline could occur by random chance if their performance were truly equivalent. Thus, the null hypothesis of equal predictive capability is rejected at the 5 % significance level, providing rigorous statistical evidence for MCPCNet's superiority.

To elucidate the performance differences in counting each crop between MCPCNet and Baseline, we visually analyzed the counting density maps within the MCPC-Dataset testing set. As shown in Fig. 12, we randomly selected four sets of counting density map from the counting results of all crop plants. Then, we used red circles to mark the crop plants missed by baseline, while MCPCNet detected these missed crop plants at the same locations (marked with dark brown circles). From the comparison of plant counting results across four types of crops, it was found that MCPCNet has fewer missed counts than the baseline, and its overall results were closer to the ground truth. When examining the positions in the counting density map where the baseline missed the plants, it was observed that crops in images with lower resolution and smaller scales were more likely to be overlooked by the baseline. This indicates that baseline struggles with counting tasks involving crops with significant scale variations. In contrast, MCPCNet employed plant semantic enhancement and scale adaptation techniques for counting various crops, resulting in high-precision and stable crop plant counting.

As the scale differences between crop plants increase, the baseline accuracy in counting corn and cotton plants declines, whereas MCPCNet demonstrates superior adaptability to these variations. Additionally, MCPCNet achieves high-precision counts when applied to sparsely

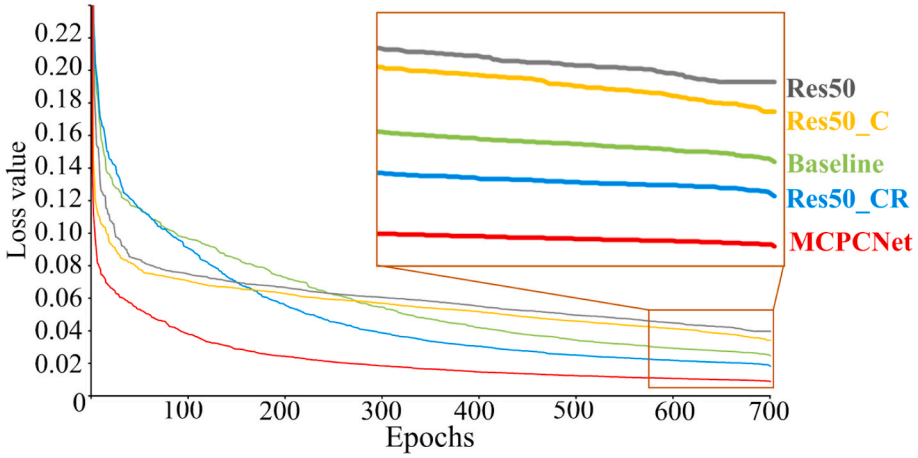


Fig. 9. Training loss curves for ablation experiments.

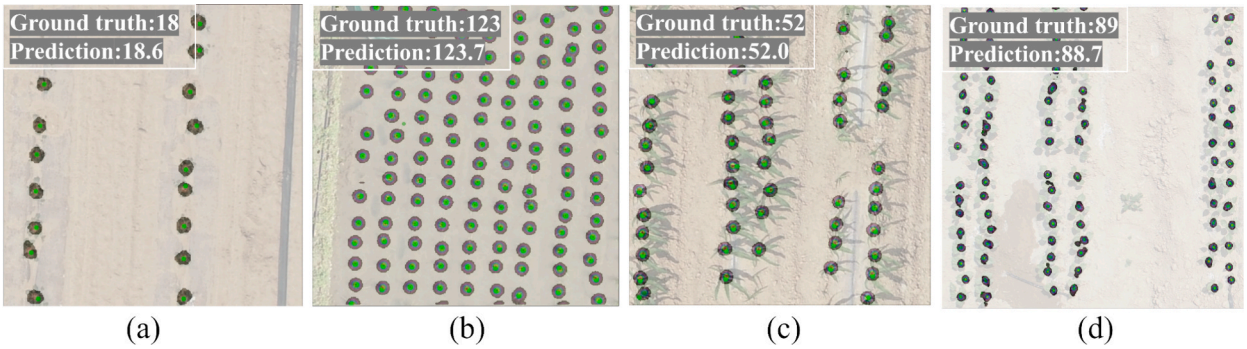


Fig. 10. Results of MCPCNet Prediction for crop plants. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton. The green point means ground truth, the darkness spot means prediction. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

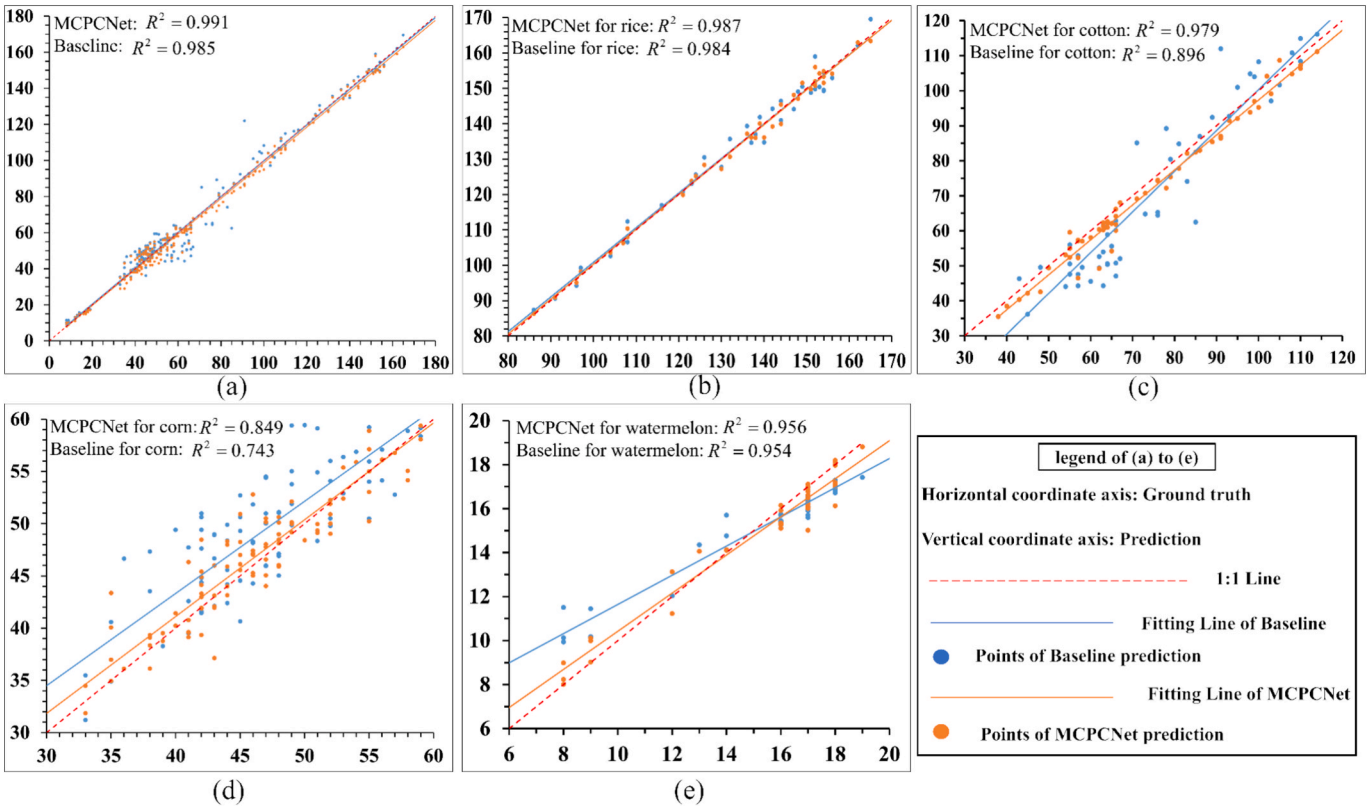


Fig. 11. Determination coefficient for crop plant counting results, comparing MCPCNet with the Baseline. (a) Overall crop. (b) Rice. (c) Cotton. (d) Corn. (e) Watermelon.

distributed watermelon plants and low-resolution rice plants. These findings suggest that MCPCNet possesses enhanced generalization capabilities and counting accuracy compared to the baseline when addressing scale variations of crop plants across different scenarios.

In addition, a preliminary analysis was conducted on the crop plant counting errors. For 100 cotton plants, the maximum counting errors for MCPCNet and Baseline were 14.4 and 2.9, respectively, primarily due to the small cotton size and significant occlusion. In the case of 17 watermelon plants, the minimum counting errors for both algorithms were 1.1 and 0.2, respectively, attributed to the moderate size and low density of the watermelon plants. Notably, the maximum and minimum counting errors of MCPCNet were significantly smaller than those of the baseline. This indicates that MCPCNet can maintain stable counting accuracy for small-scale crop plants in complex environments and achieve high accuracy in more straightforward counting tasks.

3.2. Results of ablation experiment

To validate the effectiveness of each improvement module in MCPCNet, we conducted ablation experiments using the MCPC-Dataset. Based on the Baseline framework, we initially incorporated ResNet50 as a feature encoder to reduce computational complexity. To enhance the representation of crop semantic information and accommodate the scale variations in crop plants, we designed the CoSAGE and ReDyDiC modules. Additionally, we introduced the CoT module within the dual path multiscale fusion decoder to further improve the algorithm's generalization capability. Table 4 presents the results of the ablation experiments, where A, B, C, and D denote ResNet50, CoSAGE, ReDyDiC, and CoT, respectively, with bold text indicating the optimal metrics.

From the results of the ablation experiments presented in Table 4, we obtain the following key information:

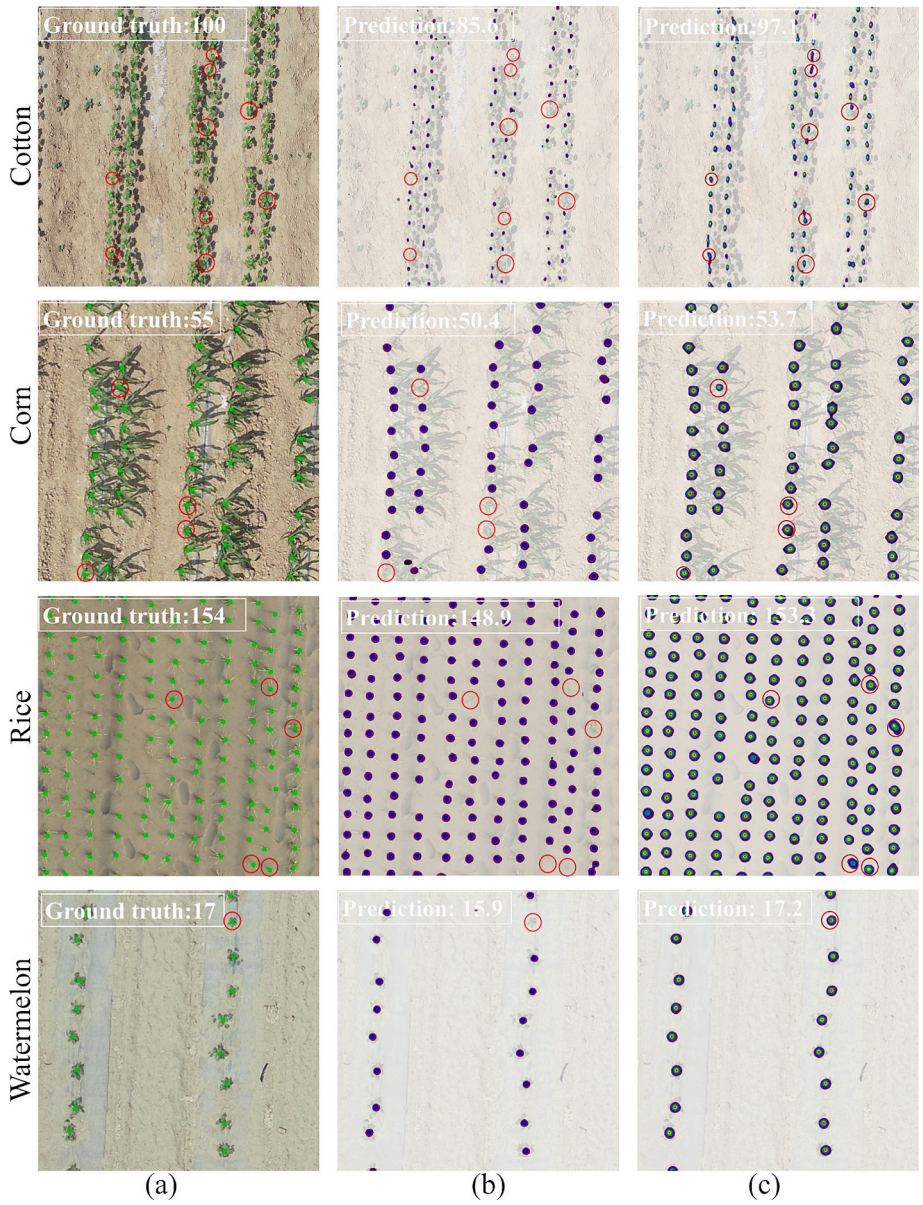


Fig. 12. Comparison of Baseline and MCPCNet results for crop plant counting. (a) Ground truth. (b) Baseline counting results. (c) MCPCNet counting results. Red circles in (b) mark crop plants missed by the baseline method, while dark brown circles in (c) indicate the corresponding locations where MCPCNet successfully detected these missed crop plants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Results of ablation experiments for MCPCNet. A, B, C, and D represent ResNet50, CoSAGE, ReDyDiC, and CoT, respectively. Bold text indicates the optimal metrics.

Algorithms	A	B	C	D	MAE	MSE	SMAPE (%)	R^2	FLOPs (G)	FPS
Baseline	×	×	×	×	3.002	18.702	5.862	0.985	232.461	14.723
Res50	√	×	×	×	3.805	31.599	7.553	0.976	**59.348**	**37.638**
Res50_C	√	√	×	×	3.382	23.769	6.159	0.982	59.386	35.165
Res50_CR	√	√	√	×	2.813	17.442	5.311	0.986	61.247	32.878
MCPCNet	√	√	√	√	**2.577**	**14.289**	**4.145**	**0.991**	62.927	31.214

1) MCPCNet achieves 2.577, 14.289, 4.415, and 0.991 on MAE, MSE, SMAPE, and R^2, respectively, which is optimal. Although not optimal for FLOPs and FPS metrics, MCPCNet demonstrated a 72.93 % reduction in FLOPs and a 112.01 % increase in FPS compared to the baseline. This indicates that MCPCNet improves the accuracy and robustness of crop plant counting while considering algorithm lightweight and inference speed enhancement.

2) This paper first employs ResNet50 to extract the semantic features of crop plants, which improves FPS while reducing the FLOPs of the algorithm. Fig. 13 illustrates the changes in feature semantics of four crop plants, where the class activation mapping (CAM) of Res50 includes elements such as shadows and soil (highlighted in the yellow box). To enhance crop plant semantics and suppress non-crop elements like shadows and soil, we developed the CoSAGE through a crop plant feature weight adjustment mechanism. As shown in

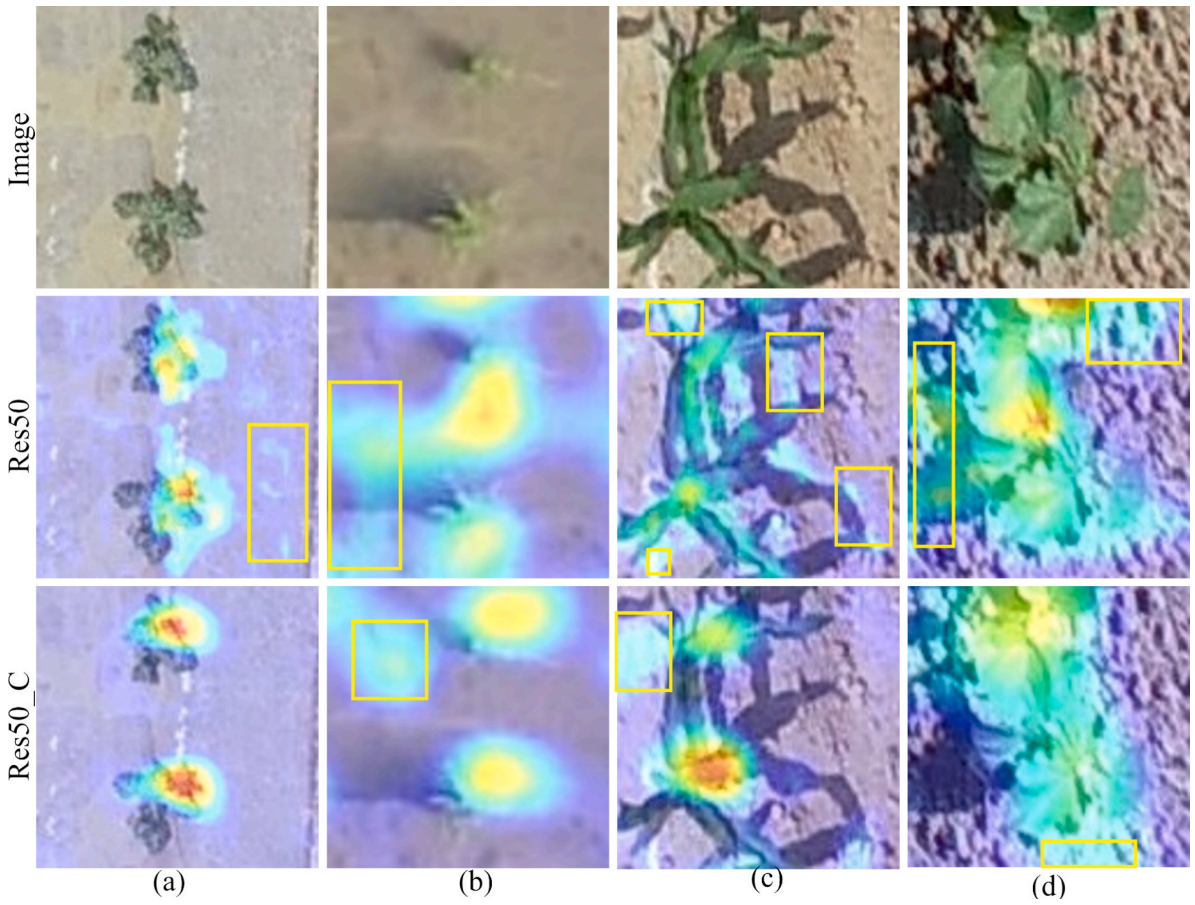


Fig. 13. Improving CAM features in crop plants with CoSAGE. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton.

Fig. 13, the CAM of Res50_C effectively suppresses non-crop elements such as shadows and soil, indicating that Res50_C more accurately focuses on crop-related information, resulting in more convergent crop plant features.

After incorporating CoSAGE into Res50, Res50_C demonstrated improvements in metrics such as MAE, MSE, and SMAPE. Compared to Res50, Res50_C reduced MAE by 11.12 % and MSE by 24.78 %. Additionally, the FLOPs and FPS remained nearly unchanged, indicating that CoSAGE enhances algorithm accuracy while preserving the algorithm's lightweight characteristics.

In Fig. 13, we examine the semantic variations in crop plants and observe residual soil semantics in the CAM of Res50_C. To enhance the algorithm's focus on crop plant characteristics, we developed ReDyDiC incorporating a parallel nonlinear convolution mechanism. This mechanism allows the counting algorithm to adapt to the varying scales of crop plants in the images, thereby facilitating the convergence of crop plant features towards the plant centers.

After integrating ReDyDiC into Res50_C, the performance of Res50_CR in terms of MAE, MSE, SMAPE, and other aspects was further improved. Compared with Res50_C, Res50_CR showed a decrease of 16.82 % and 26.62 % in MAE and MSE, respectively. As shown in Fig. 14, the analysis of CAM after adding ReDyDiC revealed a highly focused activation area in the central region of crops, greatly reducing attention to soil semantics. This indicates that ReDyDiC can enable the algorithm to perceive the central features of crops at different scales more accurately, leading to a further enhancement in algorithm accuracy. This is due to ReDyDiC's ability to adapt to semantic changes in multiscale crop plants. Through nonlinear dynamic convolution, ReDyDiC adjusts the focus area of the algorithm on crop plants,

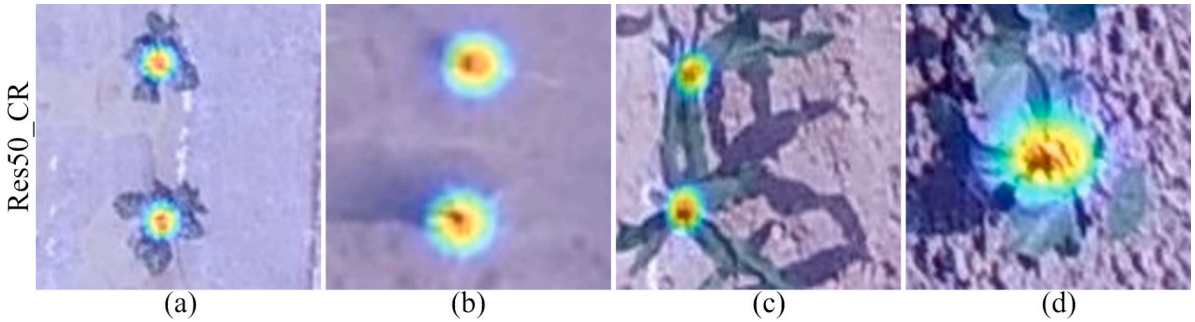


Fig. 14. Improving CAM features in crop plants with CoSAGE and ReDyDiC. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton.

achieving adaptation to changes in the scale.

(4) The dual path feature decoder integrates semantic information of crop plants from different layers, aiming to enhance the integrity and reliability of crop plant semantics by leveraging their contextual information. The CoT module is introduced in the decoder, utilizing its self-attention mechanism to exploit contextual information of the fused features fully, thus improving accuracy and generalization in counting crop plants.

After introducing the CoT module into the feature decoding stage of Res50_CR to form MCPCNet, MCPCNet achieves optimal performance in terms of MAE, MSE, and SMAPE. Although MCPCNet lags slightly behind Res50 regarding FLOPs and FPS, it reduces MAE and MSE by 32.27 % and 54.78 %, respectively. As shown in Fig. 15, adding CoT to Res50_CR leads to larger central class activation areas for the four crop plants, indicating a greater focus on the feature centers of crop plants and contributing to improved algorithm accuracy and generalization. This improvement is primarily attributed to the self-attention mechanism in the CoT module, which thoroughly explores contextual information of crop plants, enhances attention to the central areas of crop plants, and facilitates generalization in counting different crop plants. In the preceding text, multiple strategies were applied to upgrade the Baseline to MCPCNet, resulting in more effective and accurate counting of various crop plants.

4. Discussion

This section compares MCPCNet with SOTA algorithms, highlighting its advantages. The comparison includes Baseline, DSNet (Dai et al., 2019), P2PNet_Soy (Zhao et al., 2023), and MLAENet (Zheng et al., 2023), trained with default parameters. Table 5 shows the results, with the best metrics in bold.

MCPCNet achieves the lowest MAE of 2.577, indicating a significant reduction in prediction error. It also records the lowest MSE of 14.289, highlighting its precision in handling plant scale and distribution variations. Additionally, MCPCNet demonstrates the highest Coefficient of Determination (R^2) at 0.991, underscoring its highly reliable predictions that accurately reflect the ground truth. MCPCNet excels in computational efficiency with the lowest FLOPs at 62.927G, outperforming other algorithms. It also achieves the highest FPS at 31.214, surpassing the others in processing speed. Additionally, MCPCNet maintains the lowest SMAPE at 4.145 %, demonstrating its strong generalization ability across different conditions and scales.

Through a structural analysis of MCPCNet and the comparative algorithms, we found that Baseline, DSNet, P2PNet_Soy, and MLAENet rely on multiple dilated convolutions with fixed rates to enhance feature fusion. However, this design is less effective in handling scale variations among multi-scale crop plants. Additionally, using multiple dilated convolutions with fixed rates significantly increases computational overhead and decreases inference speed. In contrast, ReDyDiC's parallel nonlinear dilated convolution leverages a parallel structure to reduce computational demands while dynamically adapting to variations in crop plant scales. Moreover, DSNet and Baseline lack methods like CoSAGE for suppressing background semantics. While P2PNet_Soy employs spatial and channel attention, and Baseline uses a context-aware module, neither approach is capable of grouping crop and background semantics effectively or adjusting their semantic weights. This limitation reduces their ability to distinguish crops from background noise. To overcome these challenges, MCPCNet combines CoSAGE and ReDyDiC for refined semantic processing of crop plants and integrates the COT module to capture contextual information across multiple scales. These advantages make MCPCNet a powerful tool for precise and efficient crop counting, with its accuracy, efficiency, and generalization suited for large-scale agricultural applications.

Table 5
Comparative results of crop plant counting using different algorithms. The optimal metrics are highlighted in bold.

Algorithms	MAE	MSE	SMAPE (%)	R^2	FLOPs (G)	FPS
Baseline	3.002	18.702	5.862	0.985	232.461	14.723
DSNet	4.473	44.204	8.630	0.972	197.473	26.584
P2PNet_Soy	5.267	35.533	8.112	0.964	158.557	26.935
MLAENet	3.866	27.941	7.234	0.977	158.327	27.587
MCPCNet	**2.577**	**14.289**	**4.145**	**0.991**	**62.927**	**31.214**

To compare MCPCNet with other algorithms, we visualize crop counting results using CAM in Fig. 16. Yellow and red circles represent missed and false predictions, respectively. MCPCNet achieves predictions closely matching the ground truth across all crops, with values such as 16.6 compared to 17 for watermelon, 123.7 compared to 123 for rice, 52.0 compared to 52 for corn, and 88.7 compared to 89 for cotton. This highlights its robustness and accuracy across diverse crop types.

The prediction results reveal distinct sources of error across the algorithms, particularly in terms of false and missed predictions. For instance, DSNet demonstrates numerous false predictions in rice fields (Fig. 16b), where it misclassifies soil patches and weeds (highlighted with red circles) as rice plants. Similarly, in cotton fields (Fig. 16d), DSNet incorrectly identifies weeds as cotton plants. These issues underscore the algorithm's limited ability to balance the semantic weights of crop plants and background elements like soil, resulting in suboptimal counting accuracy and robustness. In contrast to DSNet's sequential structure, MCPCNet integrates the semantics of each level within the FME, enhancing its ability to capture relevant features. Furthermore, MCPCNet employs CoSAGE to enhance the semantic representation of crop plants while suppressing background semantics suppressing background semantics such as soil. This adjustment of semantic weights not only improves accuracy but also offers valuable insights for enhancing point-supervised algorithms in other fields. While other algorithms utilize similar feature fusion methods, they lack the semantic weight adjustment mechanism, a distinction that has been discussed earlier in this paper.

Moreover, missed predictions are another source of error, as seen in MLAENet and P2PNet_Soy. For instance, MLAENet missed several rice plants in Fig. 16b, and P2PNet_Soy failed to detect numerous corn and

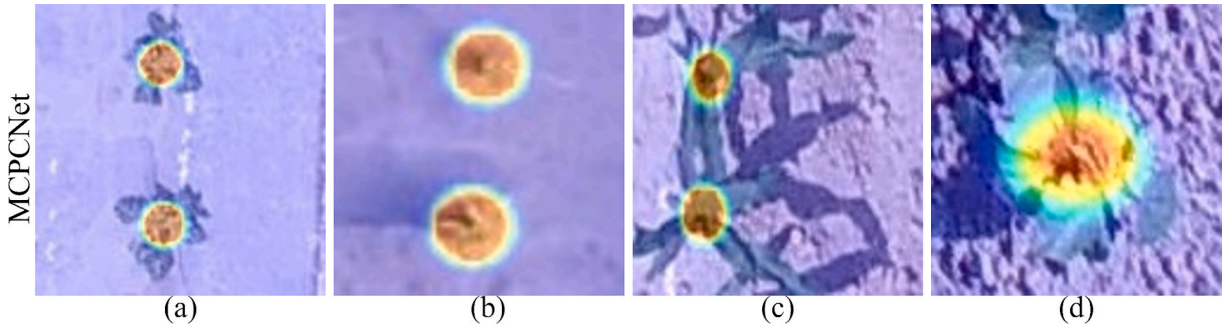


Fig. 15. Improving CAM features in crop plants with CoSAGE, ReDyDiC and CoT. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton.

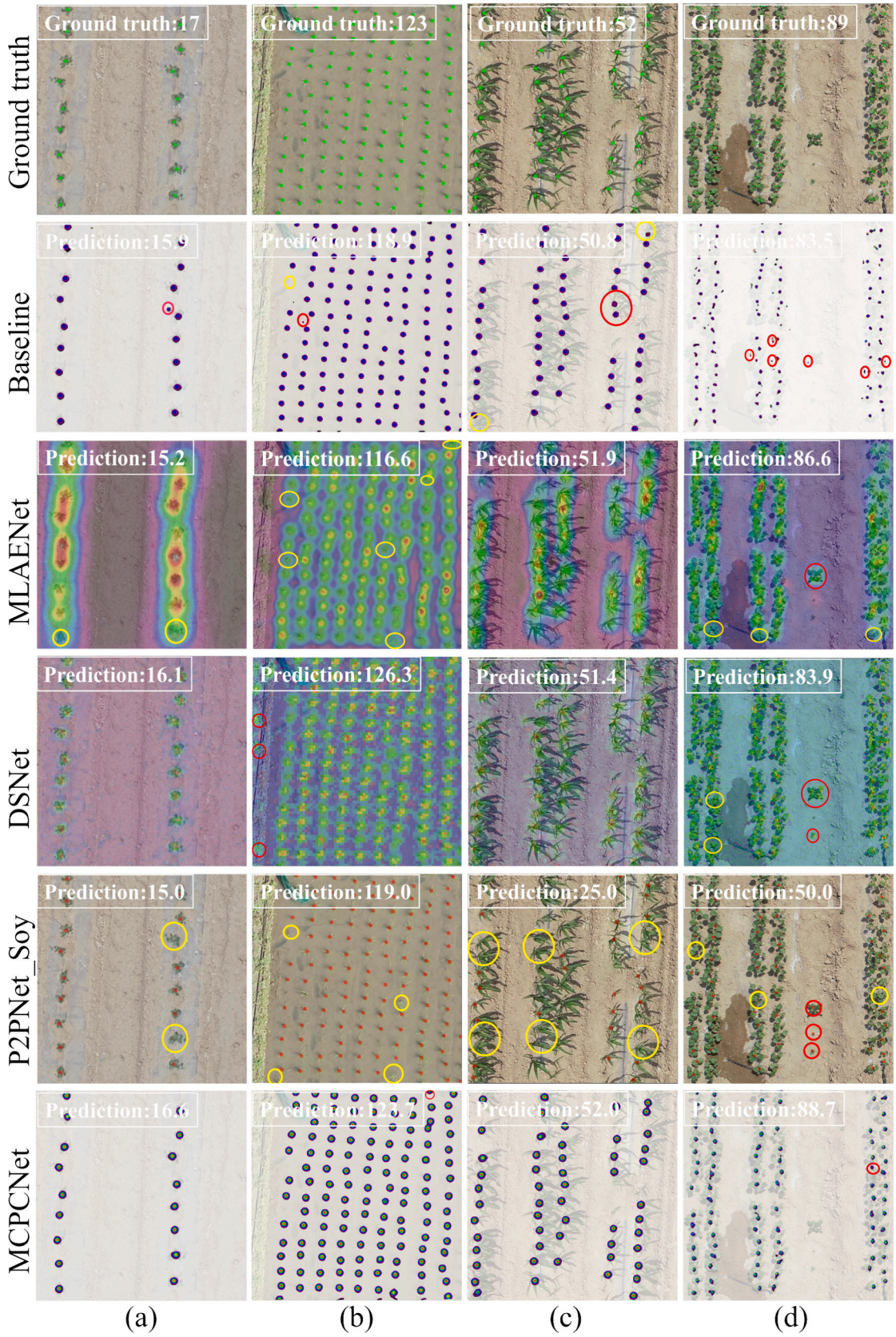


Fig. 16. Comparison of crop plants counting among different algorithms. (a) Watermelon. (b) Rice. (c) Corn. (d) Cotton. Red circles mark false predictions, yellow circles mark missed predictions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cotton plants in Fig. 16c and Fig. 16d, respectively. These omissions suggest that these algorithms need help to adapt to the varying scales (Zhang et al., 2023) of different crop plants. MCPCNet overcomes this challenge with ReDyDiC and COT, which adeptly handle variations in plant scales. This scalability adaptation provides a valuable pathway for improvement in point supervision algorithms.

Additionally, in this study, all images were captured during the seedling stage of the crops under clear and windless weather conditions to ensure consistency in data quality. However, such controlled settings may not fully represent the diverse challenges posed by real-world agricultural environments. Factors such as variable illumination (e.g., shadows, overexposure), crop growth stage changes (e.g., differences in size and density), and image resolution constraints may impact the algorithm's accuracy and robustness. Future research should investigate the algorithm's adaptability to dynamic environmental and temporal factors, enabling its broader application. Moreover, designing a lightweight, high-efficiency model suitable for real-time operation on embedded platforms, such as field robots, represents another promising direction to advance its practical utility in precision agriculture.

5. Conclusions

To address the challenges of robustness and accuracy in counting multicategory and multiscale crop plants, this paper proposes a point-supervised algorithm named MCPCNet. We primarily designed the CoSAGE, ReDyDiC, and CoT. To evaluate the performance of MCPCNet, we developed the first aerial image-based MCPC-Dataset, comprising 1,246 images and 71,566 manually annotated points. Our findings indicate that CoSAGE enhances the semantic representation of crop plants while mitigating the effects of complex backgrounds such as shadows and soil. ReDyDiC effectively adapts to the scale variations in crop plants through nonlinear convolutions, and the CoT module enhances the algorithm's generalization capability through contextual mining and self-attention mechanisms. MCPCNet significantly improves the multiscale feature representation of crop plants compared to other algorithms, resulting in highly detailed and accurate prediction density maps. This indicates that MCPCNet provides superior accuracy, robustness, and reliability in crop plant counting missions. In summary, MCPCNet presents an effective solution for the complex challenge of counting diverse and variably sized crop plants. Its innovative design is a valuable reference for developing crop plant counting algorithms in agriculture. Future work will involve collecting additional UAV imagery of various crop plants to extend high-precision counting capabilities to more types of crops.

CRediT authorship contribution statement

Huibin Li: Writing – original draft, Data curation, Conceptualization. **Huaiyang Liu:** Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Wenbo Wang:** Software. **Haozhou Wang:** Writing – review & editing. **Qiangyi Yu:** Writing – review & editing. **Jianping Qian:** Writing – review & editing. **Wenbin Wu:** Writing – review & editing. **Yun Shi:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Changxing Geng:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work received funding support from the following projects: Key Technology Development for Agricultural Condition Parameter Acquisition and Integrated Application of Sensing Equipment 2022LQ02004, Research on Key Technologies for Sky-Ground Integrated Crop Production Diagnosis and Precision Farming 2023B02014-2, Research on Field Agricultural Condition Information Acquisition and Agricultural Machinery Operation Decision Technology CAAS-CAE-202301, and Development and Demonstration of the Potato Green Intelligent Service Platform 2021ZXJ05A0504.

All authors contributed to the work. Huibin Li collected the UAV imagery of various crop plants, developed the overall framework of this paper, and participated in writing the manuscript. Huaiyang Liu, created the MCPC-Dataset, proposed the algorithm, analyzed the results, and participated in writing the manuscript. Wenbo Wang participated in experimental modifications. Haozhou Wang, Qiangyi Yu, Jianping Qian, and Wenbin Wu assisted in revising the experiments design and the paper. Yun Shi and Changxing Geng guided the research work and helped to revise the paper. The authors declare that there are no conflicts of interest regarding the publication of this paper.

We also would like to thank Professor Huabing Zhou of Wuhan Institute of Technology, and Dr. Chengcheng Chen of Shenyang Aerospace University. Their import guidance and assistance throughout the writing process of this article were instrumental. We are deeply grateful for their strong support.

Data availability

Data will be made available on request.

References

Bai, X., Gu, S., Liu, P., Yang, A., Cai, Z., Wang, J., Yao, J., 2023a. RPNet: Rice plant counting after tillering stage based on plant attention and multiple supervision network. The Crop Journal 11, 1586–1594. https://doi.org/10.1016/j.cj.2023.04.005.

Bai, X., Liu, P., Cao, Z., Lu, H., Xiong, H., Yang, A., Cai, Z., Wang, J., Yao, J., 2023b. Rice Plant Counting, Locating, and Sizing Method Based on High-Throughput UAV RGB Images. Plant Phenomics 5, 0020. https://doi.org/10.34133/plantphenomics.0020.

Bao, W., Xie, W., Hu, G., Yang, X., Su, B., 2023. Wheat ear counting method in UAV images based on TPH-YOLO. Transactions of the Chinese Society of Agricultural Engineering (transactions of the CSAE) 39, 155–161. https://doi.org/10.11975/j.issn.1002-6819.202210020.

Cai, Y., Du, D., Zhang, L., Wen, L., Wang, W., Wu, Y., Lyu, S., 2019. Guided Attention Network for Object Detection and Counting on Drones. arXiv:1909.11307. doi: 10.48550/arXiv.1909.11307.

Chen, Y., Dai, X., Liu, M., Chen, D., Yuan, L., Liu, Z., 2020. Dynamic Convolution: Attention Over Convolution Kernels. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11027–11036. https://doi.org/10.1109/CVPR42600.2020.01104.

Cui, J., Zheng, H., Zeng, Z., Yang, Y., Ma, R., Tian, Y., Tan, J., Feng, X., Qi, L., 2023. Real-time missing seedling counting in paddy fields based on lightweight network and tracking-by-detection algorithm. Computers and Electronics in Agriculture 212, 108045. https://doi.org/10.1016/j.compag.2023.108045.

Dai, F., Liu, H., Ma, Y., Cao, J., Zhao, Q., Zhang, Y., 2019. Dense Scale Network for Crowd Counting. In: Proceedings of the 2021 International Conference on Multimedia Retrieval, pp. 64–72. https://doi.org/10.1145/3460426.3463628.

Deng, Y., Hu, X., Teng, D., Li, B., Zhang, C., Hu, W., 2023. Dynamic adjustment of hyperparameters for anchor-based detection of objects with large image size differences. Pattern Recognition Letters 167, 196–203. https://doi.org/10.1016/j.patrec.2023.02.019.

Farjon, G., Huijun, L., Edan, Y., 2023. Deep-learning-based counting methods, datasets, and applications in agriculture: a review. Precision Agriculture 24, 1683–1711. https://doi.org/10.1007/s11119-023-10034-8.

Fu, H., Yue, Y., Wang, W., Liao, A., Xu, M., Gong, X., She, W., Cui, G., 2023. Ramie Plant Counting Based on UAV Remote Sensing Technology and Deep Learning. Journal of Natural Fibers 20, 2159610. https://doi.org/10.1080/15440478.2022.2159610.

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

Li, Y., Yao, T., Pan, Y., Mei, T., 2023. Contextual Transformer Networks for Visual Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 45, 1489–1500. https://doi.org/10.1109/TPAMI.2022.3164083.

Li, Z., Zhu, Y., Sui, S., Zhao, Y., Liu, P., Li, X., 2024. Real-time detection and counting of wheat ears based on improved YOLOv7. Computers and Electronics in Agriculture 218, 108670. https://doi.org/10.1016/j.compag.2024.108670.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection. In: 2017 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), pp. 936–944. https://doi.org/10.1109/CVPR.2017.106.

Liu, J., Wang, H., Yan, C., Yuan, M., Su, Y., 2020. SODA^2:Salient Object Detection With Structure-Adaptive & Scale-Adaptive Receptive Field. IEEE Access 8, 204160–204172. https://doi.org/10.1109/ACCESS.2020.3036638.

Liu, Q., Fang, J., Zhong, Y., Wang, C., Qi, Y., 2024. Double multi-scale feature fusion network for crowd counting. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-024-18769-w.

Liu, W., Salzmann, M., Fua, P., 2019. Context-Aware Crowd Counting. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5094–5103. https://doi.org/10.1109/CVPR.2019.00524.

Lu, Y., Young, S., 2020. A survey of public datasets for computer vision tasks in precision agriculture. Computers and Electronics in Agriculture 178, 105760. https://doi.org/10.1016/j.compag.2020.105760.

Madan, M., Reich, C., Hassenpflug, F., 2023. Drawing and Analysis of Bounding Boxes for Object Detection with Anchor-Based Models. Image Analysis 359–373. https://doi.org/10.1007/978-3-031-31435-3_24.

Shahid, R., Qureshi, W.S., Khan, U.S., Munir, A., Zeb, A., Imran Moazzam, S., 2024. Aerial imagery-based tobacco plant counting framework for efficient crop emergence estimation. Computers and Electronics in Agriculture 217, 108557. https://doi.org/10.1016/j.compag.2023.108557.

Shao, H., Tang, R., Lei, Y., Mu, J., Guan, Y., Xiang, Y., 2021. Rice Ear Counting Based on Image Segmentation and Establishment of a Dataset. Plants 10, 1625. https://doi.org/10.3390/plants10081625.

Thanasutives, P., Fukui, K.i., Numao, M., Kijsirikul, B., 2021. Encoder-Decoder Based Convolutional Neural Networks with Multi-Scale-Aware Modules for Crowd Counting. 2020 25th International Conference on Pattern Recognition (ICPR): 2382-2389. doi: 10.1109/ICPR48806.2021.9413286.

Tian, Y., Su, D., Lauria, S., Liu, X., 2022. Recent advances on loss functions in deep learning for computer vision. Neurocomputing 497, 129–158. https://doi.org/10.1016/j.neucom.2022.04.127.

Tran, T.H.Y., Phan, T.D.K., 2023. Dense Multi-Scale Convolutional Network for Plant Segmentation. IEEE Access 11, 82640–82651. https://doi.org/10.1109/ACCESS.2023.3300234.

Valente, J., Sari, B., Kooistra, L., Kramer, H., Mücher, S., 2020. Automated crop plant counting from very high-resolution aerial imagery. Precision Agriculture 21, 1366–1384. https://doi.org/10.1007/s11119-020-09725-3.

Wang, B., Ji, R., Zhang, L., Wu, Y., 2023. Bridging Multi-Scale Context-Aware Representation for Object Detection. IEEE Transactions on Circuits and Systems for Video Technology 33, 2317–2329. https://doi.org/10.1109/TCSVT.2022.3221755.

Wang, J., Li, X., Chen, J., Zhou, L., Guo, L., He, Z., Zhou, H., Zhang, Z., 2024. DPH-YOLOv8: Improved YOLOv8 Based on Double Prediction Heads for the UAV Image Object Detection. IEEE Transactions on Geoscience and Remote Sensing 62, 1–15. https://doi.org/10.1109/TGRS.2024.3487191.

Xu, C., Jiang, H., Yuen, P., Zaki Ahmad, K., Chen, Y., 2020. MHW-PD: A robust rice panicles counting algorithm based on deep learning and multi-scale hybrid window. Computers and Electronics in Agriculture 173, 105375. https://doi.org/10.1016/j.compag.2020.105375.

Xue, X., Niu, W., Huang, J., Kang, Z., Hu, F., Zheng, D., Wu, Z., Song, H., 2024. TasselNetV2++: A dual-branch network incorporating branch-level transfer learning and multilayer fusion for plant counting. Computers and Electronics in Agriculture 223, 109103. https://doi.org/10.1016/j.compag.2024.109103.

Yang, M., Tseng, H., Hsu, Y., Yang, C., Lai, M., Wu, D., 2021. A UAV Open Dataset of Rice Paddies for Deep Learning Practice. Remote Sensing 13, 1358. https://doi.org/10.3390/rs13071358.

Zhang, D.-Y., Luo, H.-S., Wang, D.-Y., Zhou, X.-G., Li, W.-F., Gu, C.-Y., Zhang, G., He, F.-M., 2022. Assessment of the levels of damage caused by Fusarium head blight in wheat using an improved YoloV5 method. Computers and Electronics in Agriculture 198, 107086. https://doi.org/10.1016/j.compag.2022.107086.

Zhang, Q., Yang, Y., Cheng, Y., Wang, G., Ding, W., Wu, W., Pelusi, D., 2023. Information fusion for multi-scale data: Survey and challenges. Information Fusion 100, 101954. https://doi.org/10.1016/j.inffus.2023.101954.

Zhao, J., Kaga, A., Yamada, T., Komatsu, K., Hirata, K., Kikuchi, A., Hirafuji, M., Ninomiya, S., Guo, W., 2023. Improved Field-Based Soybean Seed Counting and Localization with Feature Level Considered. Plant Phenomics 5, 0026. https://doi.org/10.34133/plantphenomics.0026.

Zheng, H., Fan, X., Bo, W., Yang, X., Tjahjadi, T., Jin, S., 2023. A Multiscale Point-Supervised Network for Counting Maize Tassels in the Wild. Plant Phenomics 5, 0100. https://doi.org/10.34133/plantphenomics.0100.