Engenharia Agrícola

ISSN: 1809-4430 (online)
www.engenhariaagricola.org.br | www.scielo.br/j/eagri

Scientific Paper
DOI: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v45e20250032/2025

TARGET DETECTION FOR WEEDING ROBOTS BASED ON IMPROVED YOLOv11 MODELS

Jianguo Meng[1*], Yanzhou Li[1], Zhipeng Li[1], Wenxia Xie[2]

[1*]Corresponding author. Mechanical Engineering College of Inner Mongolia University of Science and Technology/Baotou, China.
E-mail: mjg101@163.com | ORCID ID: https://orcid.org/0000-0002-0385-1607

KEYWORDS

object detection, deep learning, ShuffleNetV2, NAM, EMSCP.

ABSTRACT

This paper addresses the issues of insufficient accuracy and energy efficiency balance in the weed detection models that are currently used in agricultural weeding robots. To tackle this problem, we propose an enhanced YOLOv11-SNEW algorithm based on the YOLOv11 framework, which is specifically designed for weed detection in real agricultural settings. In the proposed algorithm, the backbone network is replaced with the lightweight ShuffleNetV2, the NAM attention mechanism is introduced, the EMSCP optimisation module is incorporated, and the loss function is improved. Experimental results demonstrate that the YOLOv11-SNEW model excels in weed detection, with an average recognition accuracy of 93.5%, a recall rate of 90.1%, and an mAP50 value of 91.4%. These metrics represent significant improvements in detection accuracy, recall rate, and mAP50 compared to the original YOLOv11 model and other comparative models, with a significant reductions in the number of parameters and computational effort. The model also exhibits greater robustness in complex environments, with a noticeable reduction in detection leakage. This approach provides a more efficient and precise solution for weed detection in agricultural production, thereby promoting the development of precision agriculture.

INTRODUCTION

Weed control poses a considerable challenge in agricultural production (Zhu et al., 2020). Maize is one of the three major grain crops in China; it is an important animal feed and industrial raw material, and plays a strategic role in guaranteeing food security. The reduction in the yield rate of maize fields in the western arid zone due to weed damage reaches 28% annually (Tursun et al., 2016). *Digitaria sanguinalis*, *Cirsium arvense var. integrifolium*, *Cyperus rotundus* and *Chenopodium album* are the dominant species of weed in this region; in mulched fields, *Cirsium arvense var. integrifolium* and *Chenopodium album* may make up 30% of the mulch. Their biological characteristics mean that there is a complex interplay between these weeds and the farmland microenvironment. For example, the C4 photosynthetic pathway of *Digitaria sanguinalis* causes its growth rate to be 1.5 times faster than that of maize when the average daily temperature is ≥25°C, and its seeds may lie dormant in the soil for more than five years (Wang et al., 2023); chlorogenic acid secreted by *Cirsium arvense var. integrifolium* rhizomes can inhibit the germination rate of maize seeds by up to 40% (Liu et al., 2008); *Cyperus rotundus* tubers can survive in the soil for three years, and its resistance index to nicosulfuron, the main component of herbicides, reaches 12.7 (Feys et al., 2023), meaning that it seriously threatens the growth of maize seedlings. Weed suppression at the seedling stage of maize is therefore important, but current weed control methods mainly rely on manual weed control (which is precise and environmentally friendly, but inefficient and costly), and chemical weed control (which often causes soil pollution and weed resistance). With the development of precision agriculture technology, mechanised weeding has become a new direction for research (Moond et al., 2023), with the core of this approach depending on the rapid and accurate identification of corn seedlings and weeds (Fatima et al., 2023). The development of efficient and accurate weed detection algorithms is therefore crucial to improve agricultural efficiency and promote precision agriculture.

[1] Mechanical Engineering College of Inner Mongolia University of Science and Technology/Baotou, China.
[2] Information Construction and Network Management Center of Inner Mongolia University of Science and Technology/Baotou, China.

Area Editor: Tatiana Fernanda Canata
Received in: 3-2-2025
Accepted in: 8-28-2025

Following the emergence of machine learning, deep learning techniques are now widely used in the field of target detection (Jiang et al., 2020). Deep learning methods based on convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers have gradually become mainstream approaches to the task of target detection, and are favoured by many researchers. Typical algorithms include R-CNN (Girshick et al., 2014), Faster R-CNN (Fan et al., 2021), Mask R-CNN (Zhang et al., 2024a), You Only Look Once (YOLO) (Zheng et al., 2024), SSD (Liu et al., 2016), and RetinaNet (Peng et al., 2022). However, most target detection algorithms are tested on powerful GPU personal computers, whereas in real farmland scenarios, the weeds have diverse forms and complex backgrounds, and the models need to be run on embedded devices. The limited arithmetic power of such devices means that co-optimisation of accuracy, energy and efficiency has become the core bottleneck. In view of this problem, this study proposes the following approaches: (i) a lightweight backbone network, to reduce the computational load; (ii) an attention mechanism and multi-scale module, to enhance the feature discriminative power; and (iii) loss function optimisation, to balance the localisation and classification tasks. The scheme presented here provides a highly adaptive weed detection basic framework for agricultural embedded platforms.

IMPROVED TARGET DETECTION ALGORITHM

YOLOv11 Model

YOLO is a real-time object detection algorithm based on deep learning. As the latest version of the YOLO series, YOLOv11 has an enhanced ability to detect small target objects (Zhang et al., 2024b), making it an ideal tool for rapid and accurate weed detection. Figure 1 shows the structure of YOLOv11.

Although YOLOv11 yields excellent performance, it is susceptible to false detections when weeds in the field are densely distributed, with a wide variety of species. In addition, the complexity of the model poses challenges in terms of implementation on certain embedded devices.

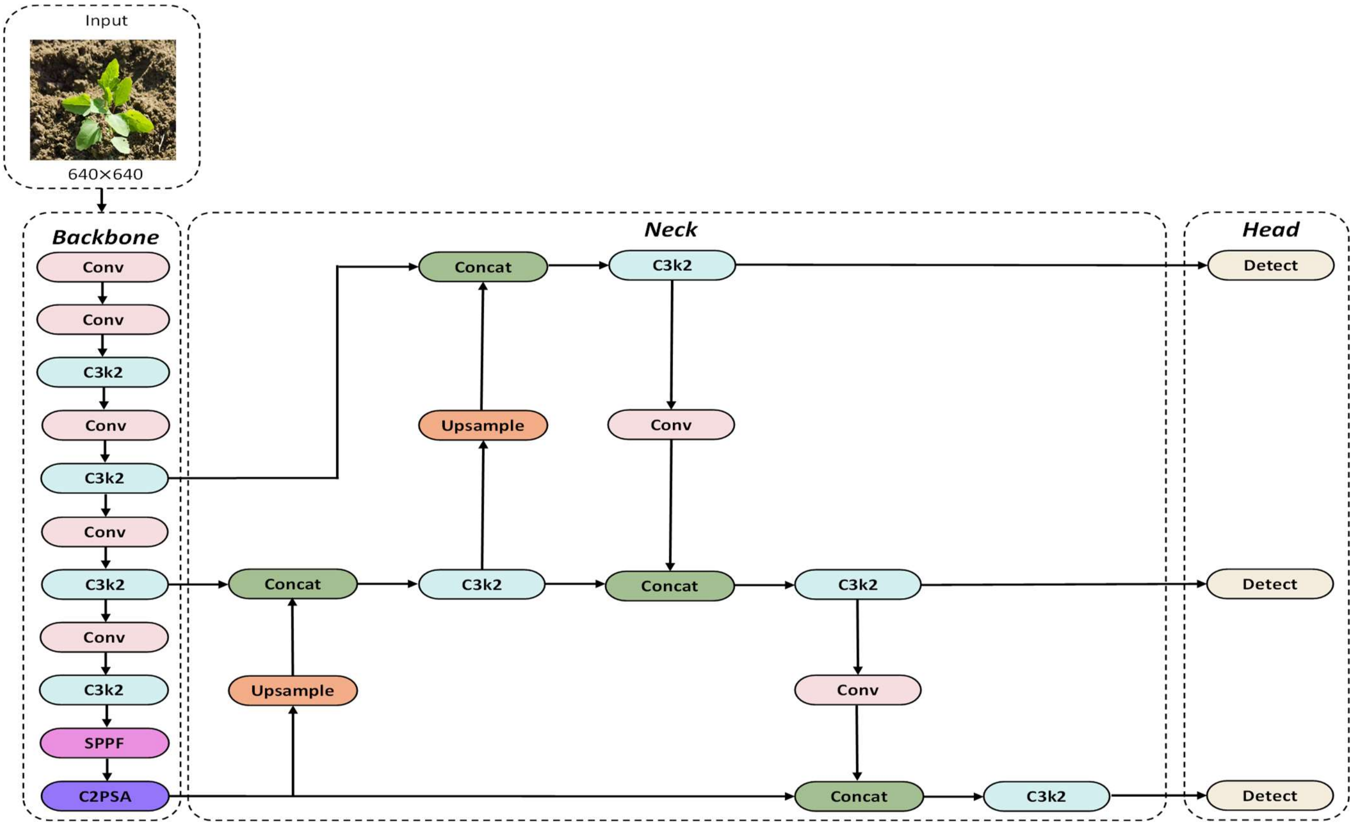


FIGURE 1. Structure of YOLOv11.

Improved Network Design: YOLOv11-SNEW

In this study, we propose an efficient and lightweight weed recognition algorithm called YOLOv11-SNEW, based on the enhanced YOLOv11 framework. The backbone network of YOLOv11 is replaced with a lightweight architecture known as ShuffleNetV2 (Ma et al., 2018), which enhances the model's detection speed and suitability for embedded devices. The NAM attention mechanism (Liu et al., 2021) is also introduced at the end of the backbone feature extraction network to improve the feature representation of *Zea mays* and weeds, thereby increasing the network's focus on these targets. The EMSCP module is integrated into the C3k2 module to enhance the model's performance in regard to handling multi-scale information, while simultaneously improving the feature extraction capabilities of the backbone and the feature fusion ability of the neck network, ultimately reducing the complexity of the model (Rahman et al., 2024). Furthermore, the original CIOU loss function is replaced with Wise-iouv3, which enables the bounding box regression module to prioritise anchor frames of ordinary quality, thus mitigating the negative impact of low-quality anchor frames and enhancing the model's generalisation ability (Tong et al., 2023). The network structure is illustrated in Figure 2, and the specific improvements are described in detail in the following sections.

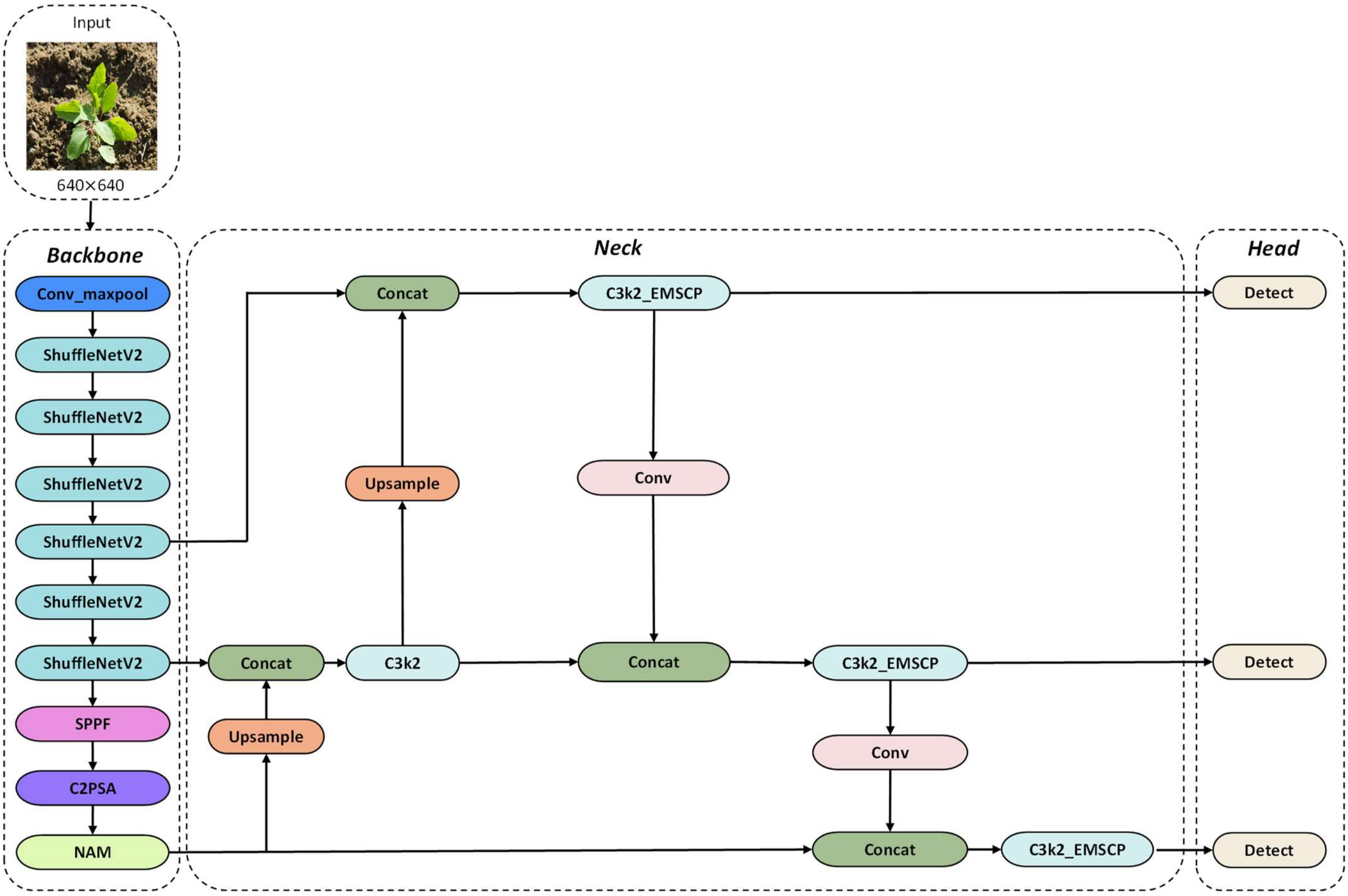


FIGURE 2. Structure of YOLOv11-SNEW.

Lightweight ShuffleNetV2 Network

To enhance the detection speed and improve the applicability of the model when deployed on embedded devices, the backbone network of the YOLOv11 weed recognition model is reconstructed using the lightweight ShuffleNetV2 network. Proposed by Ma et al. (2018), ShuffleNetV2 is an efficient lightweight CNN designed for embedded applications. It relies on the concept of packet convolution, and employs both packet convolution and channel blending to decouple the number of channels from the network's depth. Grouped convolution is applied to segment the input and output channels into multiple groups, with convolutional computations being performed within each group. This approach enables the network to efficiently execute parallel processing, giving a reduction in computational complexity while preserving its original expressive power. Channel mixing facilitates the exchange of information across different channels, promoting more efficient information transfer and contributing to a simpler and clearer overall structure, thereby lowering the complexity of the entire network. The basic unit of ShuffleNetV2 is illustrated in Figure 3.

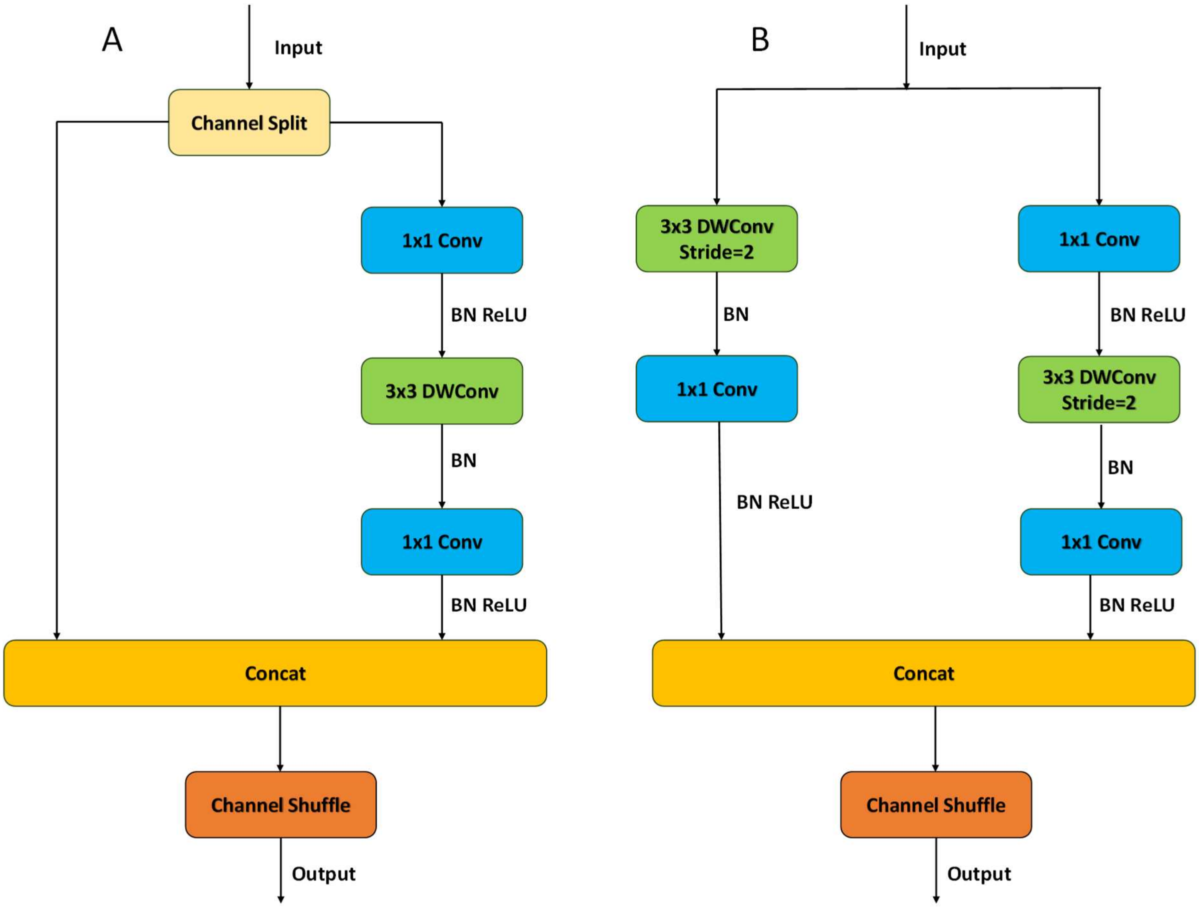


FIGURE 3. Basic unit of ShuffleNetV2: (A) basic unit; (B) is the downsampling unit.

NAM Attention Mechanism

With this lightweight upgrade to the YOLOv11 backbone network, both the number of input parameters and the computational volume are significantly reduced. However, this reduction leads to a decrease in recognition accuracy, which is particularly acute for weed detection in fields, as small weed targets are often encountered. This decline can adversely affect the model's learning efficacy. To address this issue, the NAM attention mechanism is integrated into the C2PSA layer of the model's backbone network, which effectively enhances the accuracy of weed detection for small targets. NAM is an efficient and lightweight attention mechanism derived from the convolutional block attention module (CBAM). By optimising and upgrading the spatial attention module (SAM) and channel attention module (CAM) sub-modules of CBAM, the overall performance is improved. The basic structure of this enhancement is illustrated in Figure 4.

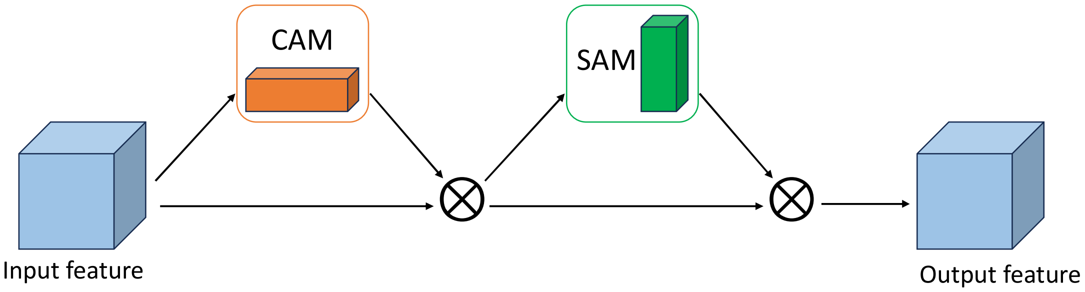


FIGURE 4. Basic structure of the NAM attention mechanism.

In the channel attention sub-module, as illustrated in Figure 5 and defined in [eq. (2)], the scaling factor in the batch normalisation (BN) process serves to indicate the significance of the weights. As detailed in [eq. (1)], this scaling factor is derived from the variance of the channel. Specifically, the scaling factor responds to variations in the size of each channel, and reflects the channel's importance. A larger scaling factor corresponds to a greater change in the channel, signifying that it contains richer information and has higher importance; conversely, a smaller scaling factor indicates less change, suggesting that the channel contains less information and is of lower importance.

$$B_{out} = BN(B_{in}) = \gamma \frac{B_{in} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \tag{1}$$

μ_B - average value of a small batch B;

σ_B - standard deviation of a small batch B;

γ, β - trainable affine transformation parameters (scale and displacement);

ε - infinitesimal quantity that prevents the denominator from being zero;

B_{out} - scale factor;

B_{in} - input feature matrix.

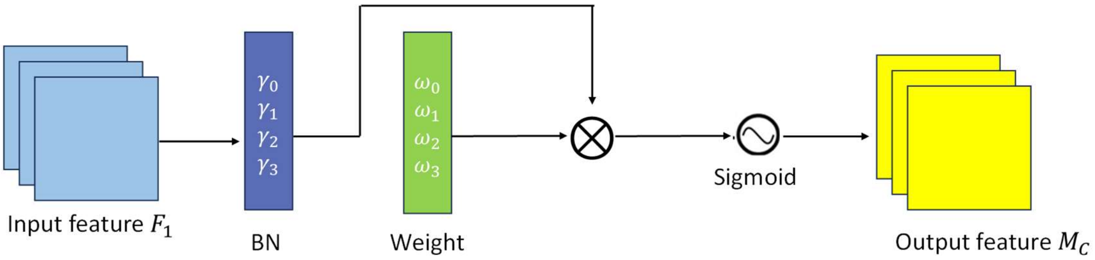


FIGURE 5. Structure of the channel attention submodule.

$$\omega_\gamma = \frac{\gamma_i}{\sum_{j=0} \gamma_j} \tag{2}$$

$$M_c = sigmoid\left(W_\gamma\left(BN(F_1)\right)\right) \tag{3}$$

Corresponding pixel normalisation is also applied in the spatial dimension, specifically when implementing the scaling factor in BN to assess the significance of spatial features. The relevant spatial attention sub-module is illustrated in Figure 6 and described in [eq. (5)].

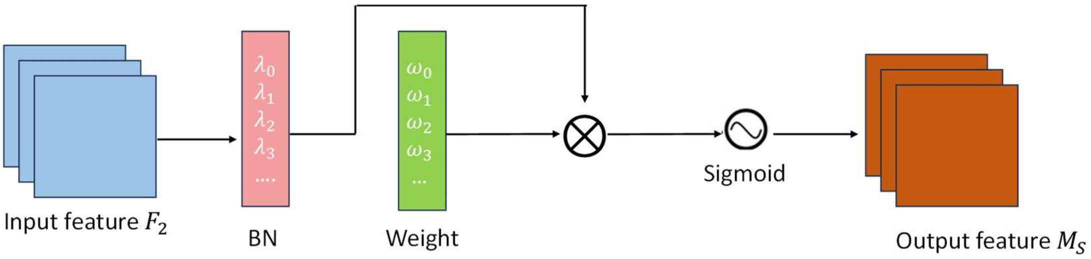


FIGURE 6. Structure of the spatial attention sub-module.

$$\omega_\lambda = \frac{\lambda_i}{\sum_{j=0} \lambda_j} \tag{4}$$

$$M_s = sigmoid\left(W_\lambda\left(BN(F_2)\right)\right) \tag{5}$$

A regularisation term is also added to the loss function in order to suppress insignificant features, as defined in [eq. (6)].

$$Loss = \sum_{(x,y)} l(f(x,W),y) + p\sum g(\gamma) + p\sum g(\lambda) \tag{6}$$

x - weed characterisation input;

y - weed characterisation output;

W - network weight;

$l(\cdot)$ - loss function;

g - l_1 modulus (math);

$Loss$ - loss value;

$f(\cdot)$ - mapping functions for inputs and outputs;

P - balance between $g(\gamma)$ and $g(\lambda)$.

EMSCP Optimisation Module

As illustrated in Figure 1, the network architecture of YOLOv11 includes a substantial number of convolutional and C3k2 modules, resulting in both high computational demand and a large number of parameters within the model. There is therefore potential for further optimisation of the feature extraction capabilities of the backbone and the feature fusion capabilities of the neck network (Duan et al., 2024). To address these challenges, the C3k2_EMSCP module is introduced. Within the bottleneck layer of the C3k2 module, the conventional 3×3 convolution (responsible for feature extraction) is replaced with the EMSCP module. The aim of this modification is to further reduce the parameter count and computational complexity of the model, thereby enhancing its performance on embedded devices. The Bottleneck_EMSCP and C3k2_EMSCP modules are depicted in Figures 7 and 8, respectively.

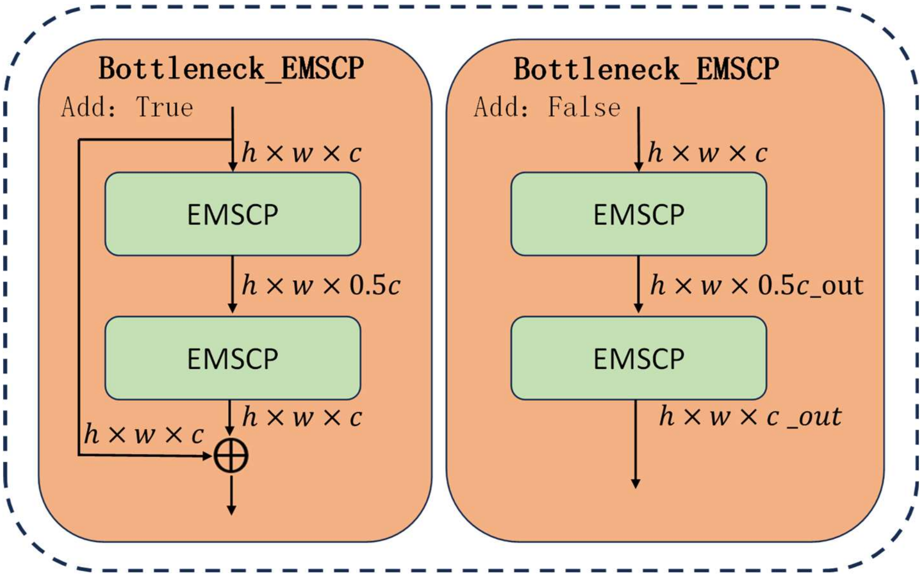


FIGURE 7. Structure of the Bottleneck_EMSCP module.

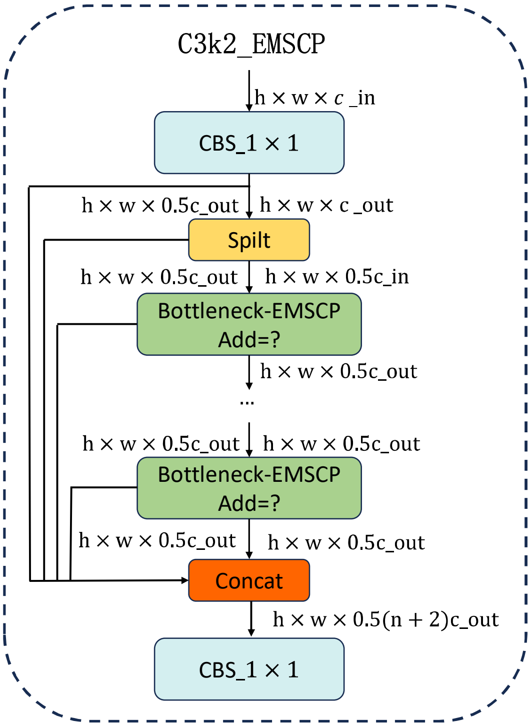


FIGURE 8. Structure of the C3k2_EMSCP module.

The EMSCP (Efficient Multi-Scale-Conv Plus) optimisation module is a compact convolution module that integrates information at multiple scales. Compared to conventional convolution modules, EMSCP has not only fewer parameters and computations but also enhanced detection accuracy. This lightweight and efficient module was inspired by the concept of group convolution, in which the entire input feature map is uniformly grouped. Each group of channels undergoes multi-scale convolution processing, which allows the system to capture various features within the weed image by performing multi-scale convolution operations on inputs at different scales, ultimately producing the feature map. This architecture facilitates the interconnection of feature maps across different scales, leading to richer and more diverse feature representations (Wang et al., 2019); these enhance the model's perception and representation capabilities, ensuring that both global overall information and local detailed information are preserved (Zhang et al., 2017). The structure of EMSCP is illustrated in Figure 9. The process commences with the application of a standard convolution operation to the feature map, which is then divided into N groups. A linear operation, denoted as *Φ*, is performed on each group to generate the complete feature map (where *Φ* typically involves 1×1, 3×3, 5×5, or 7×7 convolution kernels). Subsequently, channel-by-channel feature fusion is carried out using point-by-point convolution to produce the final output.

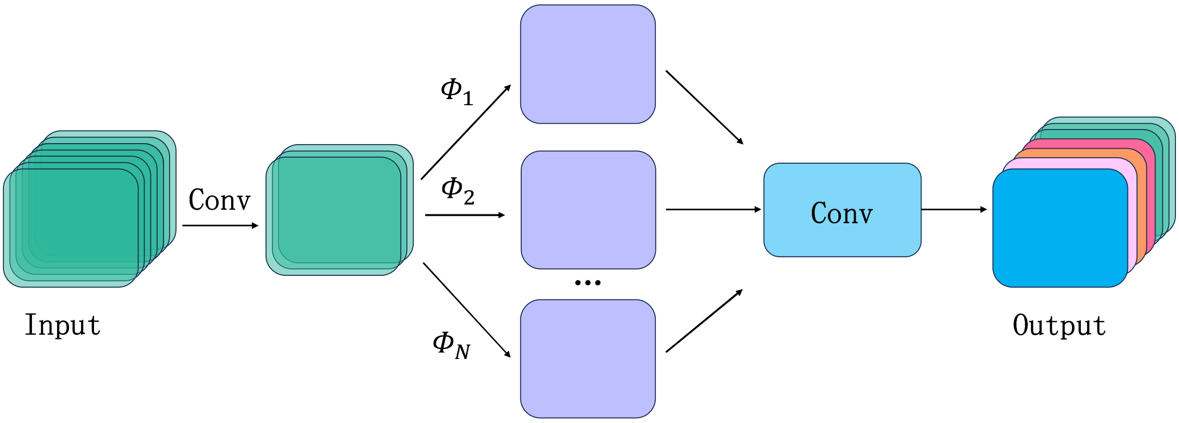


FIGURE 9. Structure of the EMSCP module.

Loss Function Improvement

In order to achieve more accurate object detection, the loss function is used to strike a balance between target positioning and classification in the YOLOv11 object detection model, as well as optimising multi-scale feature fusion. The bounding box loss function is an important part of the loss function, as it focuses on evaluating the difference between the bounding box predicted by the model and the real bounding box (Zhou et al., 2024). YOLOv11 uses CIoU as the default bounding box loss function, which takes into account three dimensions of loss: the overlap between the prediction box and the real box, the distance in the centre points, and the difference in shape. However, when the aspect ratios of the prediction box and the real box are the same, the penalty value of CIoU is zero, causing this approach to fail. In addition, CIoU cannot self-adjust to different detectors and detection tasks, its generalisability is low, and the detection effect is poor on small targets and categories with small data volumes. To solve these problems, we replace it with the Wise-iouv3 loss function. Unlike CIoU, Wise-iouv3 does not involve calculation of the aspect ratio, and instead constructs an attention-based bounding box loss function called Wise-iouv1, in which a distance attention mechanism based on a distance measurement is used, as shown in [eq. (7)]. Using this approach, Wise-iouv3 introduces a non-monotonous dynamic focus mechanism and obtains the Wise-iouv3 bounding box loss function as shown in [eq. (8)]. This mechanism is based on outliers dynamically allocating gradient gain, and outliers are used to replace the original IoU for evaluation of the anchor boxes. A smaller gradient gain is allocated to anchor boxes with larger or smaller mass, meaning that the bounding box regression focuses on anchor boxes with ordinary mass, thereby weakening the harmful gradient of the low-quality anchor boxes and improving the generalisation ability of the model. The mass of the anchor frame is inversely proportional to the outlier. The formula for the outlier is shown in [eq. (9)].

$$\mathcal{L}_{WIoU\,v1} = R_{WIoU}\mathcal{L}_{IoU} \qquad (7)$$

$\mathcal{L}_{IoU}$ - degree of overlap between the predicted frame and the true frame;

R_{WIoU} - penalty term:

$$\mathcal{L}_{WIoU\,v3} = r\mathcal{L}_{WIoU\,v1}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \qquad (8)$$

r - gradient gain:

$$\beta = \frac{\mathcal{L}^*_{IoU}}{\overline{\mathcal{L}_{IoU}}} \in [0, +\infty) \qquad (9)$$

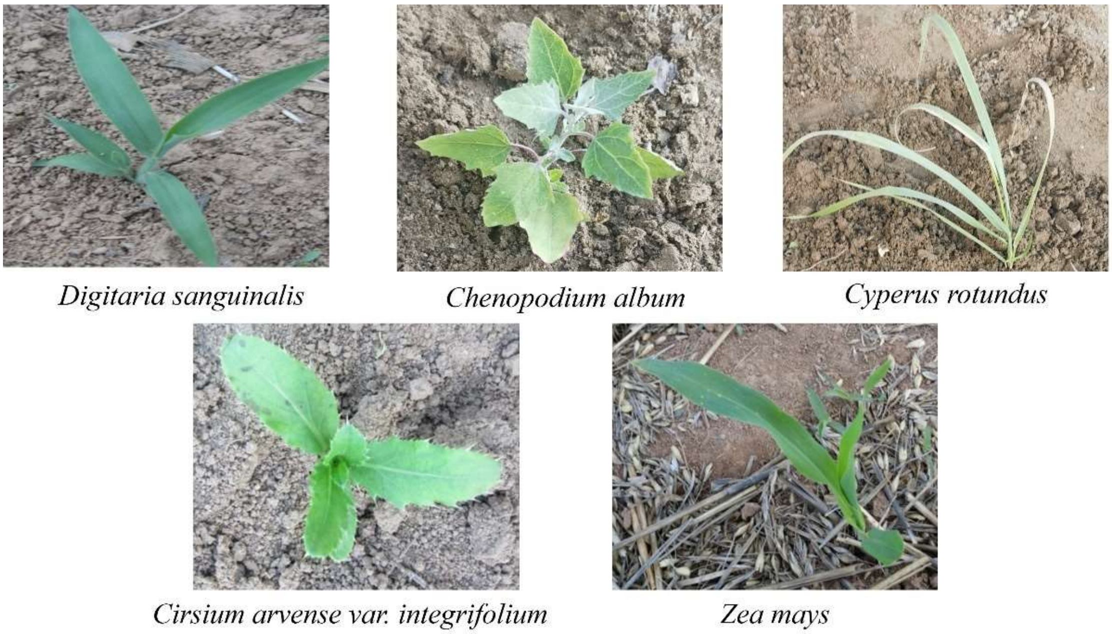
$\mathcal{L}^*_{IoU}$ - monotonic focusing coefficient;

$\overline{\mathcal{L}_{IoU}}$ - sliding mean with momentum m.

MATERIAL AND METHODS

Image Acquisition

In this article, the main focus of study is the weed *Zea mays*, at the stage of two to five leaves, and four common associated weeds, *Digitaria sanguinalis*, *Cirsium arvense var. integrifolium*, *Cyperus rotundus*, and *Chenopodium album*, the characteristics of which are illustrated in Figure 10. The leaves of *Digitaria sanguinalis* are long and thin, with smooth edges and a prominent midvein, whereas *Chenopodium album* leaves are typically round or diamond-shaped, and those of *Cyperus rotundus* are long and striped, without prominent midveins, and are both longer and thinner than those of *Digitaria sanguinalis*. *Cyperus rotundus* leaves are generally elliptical or feathery, displaying distinct serrations along the edges. Lastly, the leaves of *Zea mays* are narrow and linear, characterised by smooth surfaces, a light green colour, clear veins, and a degree of longitudinal curvature.



Digitaria sanguinalis *Chenopodium album* *Cyperus rotundus*

Cirsium arvense var. integrifolium *Zea mays*

FIGURE 10. Photographs of the weeds under study.

In mid-May 2024, image acquisition was carried out in a cornfield in Siziwang Banner, Ulanqab City, Inner Mongolia Autonomous Region, China. The acquisition equipment was an HTSUA502C industrial camera, with a maximum image resolution of 2592×1944 pixels. A CMOS sensor with 5 million pixels was used, and a tripod was used for acquisition. The height of the camera was set to 30–40 cm from the ground, and shots were taken with the camera lens perpendicular to the ground. A total of 4,465 photos were captured under three weather conditions: sunny, cloudy and after rain (during the time periods 08:00–12:00 and 14:00–17:30), as shown in Figure 11. The images included 1,000 photographs of *Digitaria sanguinalis*, 1,000 of *Chenopodium album*, 700 of *Cyperus rotundus*, 1,000 of *Cirsium arvense var. integrifolium*, and 765 of *Zea mays*. To reduce the calculation cost and facilitate training of the model, the image resolution was uniformly set to 416×416 pixels, and all images were saved in JPG format.

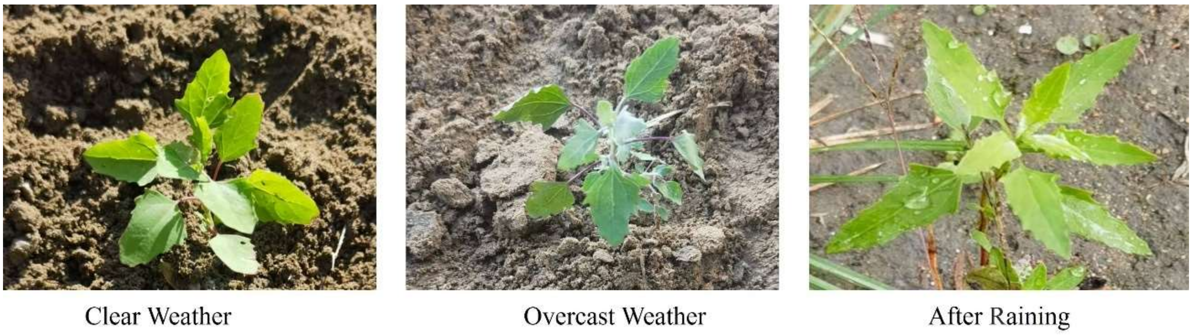


FIGURE 11. Images collected under three weather conditions.

Image Preprocessing

In order to improve the diversity of the data and enhance the robustness of the model, the original images (including those of *Digitaria sanguinalis*, *Chenopodium album*, *Cirsium arvense v. integrgrifolium*, *Cyperus rotundus*, *Zea mays*) were preprocessed to avoid non-convergence and overfitting caused by an insufficient sample size for model training, and to improve the generalisability of the model for weed images from different angles and with varying illumination, size, shape and noise interference. Figure 12 shows an image of *Cyperus rotundus* as an example after nine random enhancement operations have been carried out on it, including brightness/contrast adjustment, addition of noise (Gaussian, salt-and-pepper and blur), distortion, random cropping, and size adjustment. The same data augmentation strategies were adopted for all the weed classification images. After pretreatment, the total number of samples was doubled, giving a total of 8,930. We randomly selected 7,144, 893 and 893 images to form the training, validation and test sets (in the ratio 8:1:1), ensuring that each dataset contained different samples to support training and evaluation of the model.

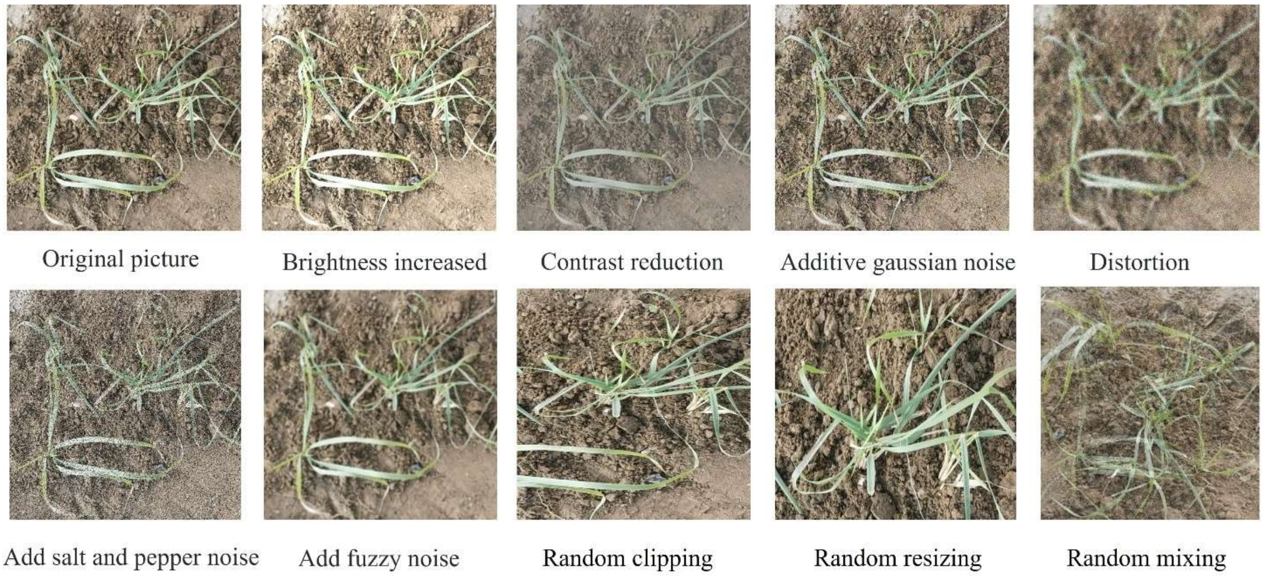


FIGURE 12. Effects of nine types of data preprocessing.

The LabelMe tool was employed to annotate the features of the dataset. During the labelling process, the minimum outer rectangle was used to accurately label each weed, thereby enhancing the learning efficacy. Upon completion of labelling, a JSON file was generated, which included details such as the image path, weed type, and the coordinates of the labelled regions. This JSON file was subsequently converted into a TXT file to meet the format requirements for training of a YOLOv11 model.

Experimental Platform and Model Training Setup

The experimental environment consisted of the Windows 11 operating system with an Intel Core i9-13900 CPU, and Python 3.11 was used to develop the weed recognition model. The deep learning framework was PyTorch, while an NVIDIA GeForce RTX 4070 GPU was employed for computational acceleration, supported by CUDA version 11.1. PyCharm served as the programming platform, to ensure sufficient computing power for the experiment.

The configuration parameters for training of the network were as follows: stochastic gradient descent (SGD) was used as the optimiser for the network model, with initial momentum set to 0.9; the learning rate was adjusted using cosine annealing; the batch size was set to 32; and the training period was 300 epochs. The size of the image input was specified as 640 × 640 pixels, while the weight decay coefficient was set to 0.0005, and the learning rate was fixed at 0.01.

Performance Evaluation of the Model

The performance for each target detection algorithm was evaluated using six metrics: accuracy (precision, P), recall (R), average precision (AP), mean average precision (mAP), number of model parameters, and model computations. The formulae for the first four of these are as follows:

$$P = \frac{TP}{TP + FP} \times 100\% \tag{10}$$

$$R = \frac{TP}{TP + FN} \times 100\% \tag{11}$$

$$AP = \int_0^1 P(R)dR \times 100\% \tag{12}$$

$$mAP = \sum \frac{AP}{N} \tag{13}$$

In this context, TP (true positive) refers to the number of positive samples correctly predicted as positive cases, while FP (false positive) denotes the number of negative samples incorrectly predicted as positive cases. FN (false negative) indicates the number of positive samples that were incorrectly predicted as negative cases. N represents the total number of prediction categories.

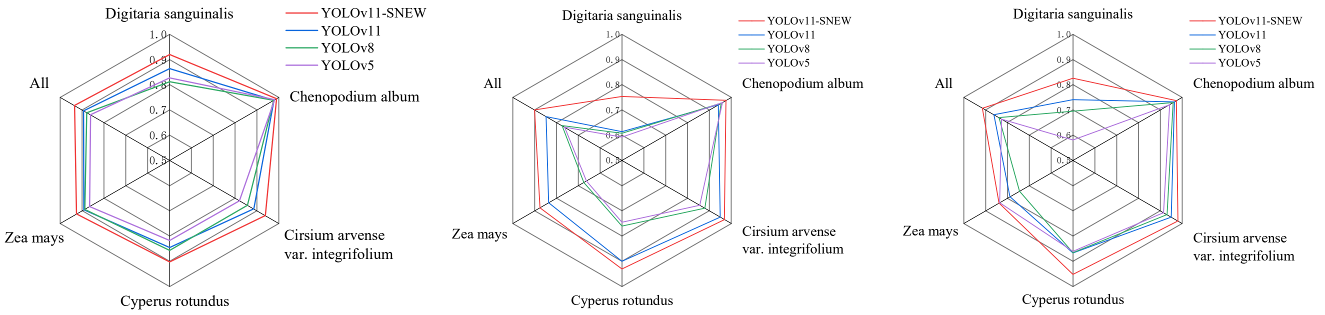
Visualisation Comparison Model

To comprehensively verify the detection performance of the YOLOv11-SNEW model proposed in this study, a comparative visualisation method was adopted, in which the detection results were presented to highlight the model's advantages in an intuitive way. Several representative models in the field of object detection, including single-stage and multi-stage models with different levels of complexity and varying design ideas, were selected as the control group. These included YOLOv8 (Zheng et al., 2024), which is based on anchor-free, multi-scale feature fusion and takes into account both real-time performance and small target detection capability; YOLOv11 (Zhang et al., 2024b), which features a lightweight structure with dynamic feature allocation and served as the basis for a comparison of improvement gain; SSD (Meng et al., 2020), a single-stage classic algorithm centred on multi-scale anchor boxes; and Faster R-CNN (Fan et al., 2021), a multi-stage classical algorithm based on RPN candidate boxes with fine regression.

During the experiment, the YOLO series models were trained using the Ultralytics framework, while SSD and Faster R-CNN were implemented using the MMDetection framework. All models used the parameter configurations specified above in the section entitled "Experimental Platform and Model Training Setup" to ensure a fair comparison. In this way, the superiority of YOLOv11-SNEW in terms of detection performance was systematically verified.

Comparative Testing of Different YOLO Models

To evaluate the performance of the YOLOv11-SNEW model on the task of weed identification, a multi-model identification performance comparison experiment was designed in which YOLOv5s, YOLOv8s, YOLOv11s and the improved YOLOv11-SNEW model proposed in this paper were considered. Experiments were carried out based on training, validation and test sets prepared as described above. The hyperparameters were configured with a learning rate of 0.01, a batch size of 32, and a training period of 300 rounds. We note that YOLOv5s uses the CSPDarknet53 backbone network, which reduces the computational complexity through the cross stage partial (CSP) structure, and combines the PANet feature fusion path with decoupling head design. The scaling parameters were [0.33,0.50,1024], i.e. the depth factor was 0.33, the width factor was 0.50, and the maximum number of channels was 1024. YOLOv8s has an optimised backbone network based on YOLOv5s; it includes the efficient C2f module and the bidirectional feature pyramid to enhance the cross-scale information flow, and is equipped with the task pair to enable dynamic adaptation of the target shape adjustment anchor box uniformly. The scaling parameters were [0.33,0.50,1024]. In YOLOv11s, the multi-scale feature fusion network is optimised based on YOLOv8 to enhance the model's ability to extract small targets. It is equipped with an improved feature pyramid and an adaptive boundary detection head. In this case, the scaling parameters were [0.50,0.50,1024]. The scaling parameters of YOLOv11-SNEW were also [0.50,0.50,1024]. When the training was complete, the accuracy rate, recall rate and mAP50 index for each model were evaluated using the test set to enable a quantitative comparison of their performance in weed detection. The evaluation results are shown in Figure 13.



A. Accuracy comparison of various models

B. Comparison of recall rates of various models

C. Comparison of mAP50 values of various models

FIGURE 13. Comparison of performance results from several models: (A) accuracy; (B) recall; (C) mAP50.

As shown in Figure 13 and Table 1, the recognition accuracies of the four models for *Digitaria sanguinalis*, *Chenopodium album*, *Cirsium arvense var. integrifolium*, *Cyperus rotundus*, and *Zea mays* varied significantly. The highest recognition accuracy was achieved for *Chenopodium album*, with the results from all four models exceeding 97%. YOLOv11-SNEW gave the best results, with an accuracy of 98.9%, a recall rate of 97.6% and an mAP50 of 97.3%. In contrast, YOLOv8 achieved 97.7% accuracy, YOLOv11 had the lowest recall rate (94.1%), and YOLOv5 had the lowest mAP50 (94.2%). The superiority of these results can be attributed to the serrated edges and distinct, clear textures of *Chenopodium album* leaves, which provide a high level of distinguishability of the image features, enabling the models to quickly locate and extract more effective information for accurate recognition.

The poorest recognition performance was found for *Cyperus rotundus*. YOLOv11-SNEW achieved only 90.3% accuracy for this weed, with a recall rate of 93.0% and an mAP50 of 95.2%, while YOLOv5 performed even worse, with an accuracy of 81.7%, a recall rate of 74.5%, and an mAP50 of 86.1%. This is because the slender leaves of *Cyperus rotundus* lack a distinctive texture and often grow densely, leading to frequent occlusion and incomplete feature extraction. Its lighter colour compared to *Chenopodium album* also causes confusion with other weeds, further reducing the recognition accuracy.

The recognition accuracies for *Digitaria sanguinalis*, *Cirsium arvense var. integrifolium*, and *Zea mays* fell between these two extremes. YOLOv11-SNEW yielded the best performance of the four models, with accuracies of 92%, 93.8%, and 92.5% for these three plants, respectively. In contrast, YOLOv5 gave the lowest accuracies of 82.7%, 81.9%, and 86.5%. This difference may stem from the regular leaf arrangements of *Digitaria sanguinalis*, *Cirsium arvense var. integrifolium* and *Zea mays,* which facilitate feature extraction. Moreover, in images of the same size, individual *Chenopodium album* and *Cirsium arvense var. integrifolium* plants occupy larger proportions, providing more complete features and thus higher recognition accuracies.

In terms of model performance, YOLOv5's relatively shallow backbone network limits its ability to extract features from small targets, resulting in lower accuracy. However, its smaller parameter count and computational complexity make it suitable for deployment on embedded devices. YOLOv11 and YOLOv8 yielded good recognition accuracy, recall rate, and mAP50 but required significantly more computational resources, which may reduce the inference speed in practical applications. Of the four models, YOLOv11-SNEW achieved the best overall performance, with an average recognition accuracy of 93.5%, a recall rate of 90.1%, and an mAP50 of 91.4%, representing a comprehensive advantage across all evaluation metrics. Detailed results for the four models are presented in Table 1.

TABLE 1. Comparative test data for multiple models.

Performance	Weed species	YOLOv11-SNEW	YOLOv11	YOLOv8	YOLOv5
Accuracy rate	*Digitaria sanguinalis*	92%	86.4%	81.2%	82.7%
	Chenopodium album	98.9%	97.8%	97.7%	98.1%
	Cirsium arvense var. integrifolium	93.8%	88.5%	85.6%	81.9%
	Cyperus rotundus	90.3%	84.6%	85.7%	81.7%
	Zea mays	92.5%	89.2%	88.8%	86.5%
	All	93.5%	89.3%	87.8%	86.1%
Recall rate	*Digitaria sanguinalis*	75.4%	61.4%	60.7%	59.2%
	Chenopodium album	97.6%	94.1%	95.4%	95.9%
	Cirsium arvense var. integrifolium	96.9%	94.9%	87.9%	85.6%
	Cyperus rotundus	93.0%	90.0%	76.0%	74.5%
	Zea mays	87.6%	83.6%	67.5%	66.3%
	All	90.1%	84.8%	77.5%	76.3%
mAP50	*Digitaria sanguinalis*	82.6%	74.1%	69.5%	58.1%
	Chenopodium album	97.3%	96.5%	95.9%	94.2%
	Cirsium arvense var. integrifolium	98.1%	94.9%	93.1%	91.4%
	Cyperus rotundus	95.2%	86.7%	86.6%	86.1%
	Zea mays	83.8%	78.8%	74.4%	83.5%
	All	91.4%	86.2%	83.9%	82.7%
Parameters (MB)		5.312	9.429	11.1	7.8
Computations (GFLOPS)		11.6	21.6	28.8	19

RESULTS AND DISCUSSION

Ablation Experiment

To evaluate the effectiveness of the ShuffleNetV2 backbone network, the NAM attention mechanism, the EMSCP module, and the Wise-iouv3 loss function, ablation experiments were conducted using the same training and test datasets. The YOLOv11s model served as the baseline model, the ShuffleNetV2 backbone network was replaced successively, and the NAM attention mechanism, EMSCP module and Wise-iouv3 loss function were added. The five metrics used for evaluation were the accuracy (P, %), recall (R, %), mAP50 (%), number of parameters (MB), and computations (GFLOPS), and the results are shown in Table 2. The variation in the accuracy (P, %), recall (R, %), and mAP50 (%) values is illustrated in Figure 14.

TABLE 2. Results of ablation experiments.

Model	ShuffleNet V2	NAM	EMSCP	Wise-iouv3	P (%)	R (%)	mAP50 (%)	Parameters (MB)	Computations (GFLOPS)
YOLO v11					89.3	84.8	86.2	9,429	21.6
	√				87.3	82.2	84.4	5,341	10.5
	√	√			89.8	85.6	88.3	5.342	10.7
	√	√	√		91.4	88.2	89.8	5.306	11.1
	√	√	√	√	93.5	90.1	91.4	5.312	11.6

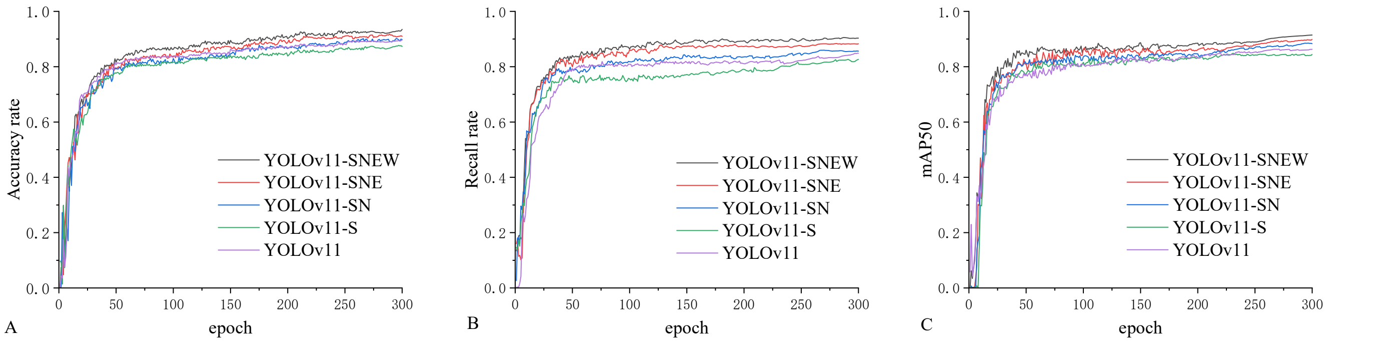


FIGURE 14. Results of ablation experiments: (A) change in accuracy over time; (B) change in recall over time; (C) change in mAP50 value over time.

From Table 1 and Figure 14, it is evident that the overall performance improvement for each model component varied. When the backbone network was replaced with ShuffleNetV2, the number of parameters and computational load of the model decreased significantly, by 43.34% and 51.39%, respectively. These reductions can be attributed to the group convolution and channel shuffling functions inherent in the ShuffleNetV2 architecture. However, this modification also resulted in a 2.24% decrease in accuracy, which can be explained by the design of the ShuffleNetV2 network, which prioritises lightweight and efficient computation, potentially compromising its feature extraction capabilities. The incorporation of the NAM attention mechanism, EMSCP module, and Wise-iouv3 loss function led to improvements in accuracy, recall, and mAP50 compared to the initial model; specifically, the accuracy increased by 4.71%, recall by 6.25%, and mAP50 by 6.02%, indicating a significant enhancement. The number of parameters and computational requirements decreased by 43.7% and 46.3%, respectively. The YOLOv11-SNEW model yielded a marked improvement compared to several other models, and it can be concluded that the approach used in this study, which involved introducing the ShuffleNetV2 backbone network, adding the NAM attention mechanism, incorporating the EMSCP module, and adopting the Wise-iouv3 loss function, was a reasonable one, yielding favourable results.

Model Performance Analysis Before and After Improvement

To verify that the improved weed recognition model outperformed the original YOLOv11s recognition model, we analysed the changes and convergence trends for the accuracy, recall, mAP50 value, and the bounding box loss function during training of the model on a self-constructed dataset. Both weed recognition models were trained for 300 iterations, and the resulting curves for accuracy, recall, mAP50 value, and the bounding box loss function are presented in Figure 15.

Figure 15(A) shows that both models exhibit good fitness over the first 50 iterations, achieving accuracies of 93.5% and 89.3%, recall values of 90.1% and 84.8%, and mAP50 values of 91.4% and 86.2%, respectively. Notably, in Figure 15(B), the bounding box loss function of the improved model falls below that of the original model by the 15th iteration, and subsequently stabilises. Ultimately, the loss function values decrease to 1.17 and 1.43, respectively. It is evident that over the same number of training rounds, the improved YOLOv11-SNEW model yields lower loss values and smoother convergence curves, accompanied by faster convergence rates and enhanced recognition accuracy. This indicates a significant improvement in model performance.

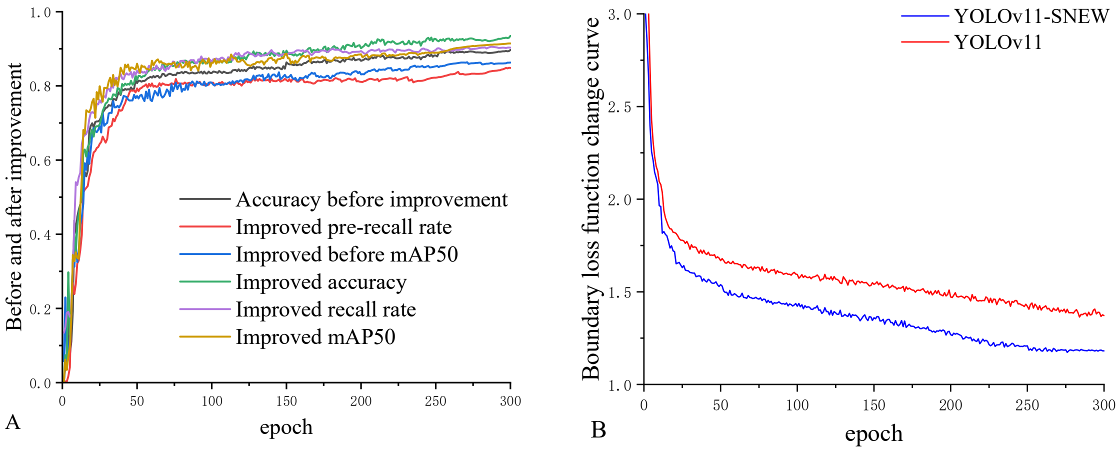


FIGURE 15. Comparison of model performance before and after improvement: (A) change curves for accuracy, recall, and mAP50 values; (B) bounding box loss function.

Comparative Visual Analysis

As a visual demonstration of the advantages of the improved model over other object-detection algorithms, a performance comparison experiment was conducted using four models, namely YOLOv8 (Zheng et al., 2024), YOLOv11 (Zhang et al., 2024b), SSD (Meng et al., 2020), Faster R-CNN (Fan et al., 2021), and YOLOv11-SNEW. The results for weed detection using these five models in complex environments are presented in Figure 16 and Table 3. In the test, the five algorithms used the same dataset and training platform.

The experimental results show that the accuracy rate, recall rate and mAP50 value of YOLOv11-SNEW are 93.5%, 90.1% and 91.4% respectively, values that are significantly better than for the other models. Compared with YOLOv11, the accuracy rate, recall rate and mAP50 value are increased by 4.2%, 5.3% and 5.2%, respectively, while compared with YOLOv8, the values of these three indicators are 5.7%, 12.6%, and 7.5% higher, respectively. Compared with SSD, which is also a single-stage detection model, these values are increased by 7.9%, 8.3%, and 10.7%, and compared with the two-stage detection Faster R-CNN model, improvements of 3.7%, 2.3%, and 6.1% are achieved. In addition, in terms of the number of parameters and computational load, YOLOv11-SNEW has a significant advantage and is more lightweight. Both SSD and Faster R-CNN have more than 20 MB of parameters and are difficult to deploy in embedded devices. The YOLOv8 and SSD models suffer from missed detections when dealing with scenarios where multiple weeds overlap. In terms of detection speed, YOLOv11-SNEW also performs exceptionally well. For the real-time weeding robots that are currently on the market, with a moving speed of 1.0–1.5 m/s, a processing speed of ≥20FPS per image can meet the requirements for continuous detection during movement. However, the single image processing speed of the proposed system reaches 58.8 FPS, far exceeding the actual requirements, meaning that it can be seamlessly integrated into a weeding robot. Although other algorithms also meet the speed requirements, they are not easy to deploy in embedded devices, and would occupy the memory space dedicated to other functions of the device.

The results indicate that the improved model proposed in this study has obvious advantages over other models. Considering the working time of weed-cutting machines in the field, the execution time of the system must be kept relatively low. YOLOv11-SNEW, while maintaining a certain level of detection accuracy, has a reduced computational complexity and number of parameters, and consumes less memory resources, thus making it suitable for deployment on the embedded devices of outdoor weed-cutting robots.

TABLE 3. Comparison of results for each target detection network.

Model	P	R	mAP50	Parameters	Computations	Average processing speed
	(%)	(%)	(%)	(MB)	(GFLOPS)	(FPS)
YOLOv11-SNEW	93.5%	90.1%	91.4%	5.312	11.6	58.8
YOLOv11	89.3%	84.8%	86.2%	9.429	21.6	45.6
YOLOv8	87.8%	77.5%	83.9%	11.1	28.8	42.8
SSD	85.6%	81.8%	80.7%	24.013	42.91	36.4
Faster R-CNN	89.8%	87.8%	85.3%	44	140	32.7

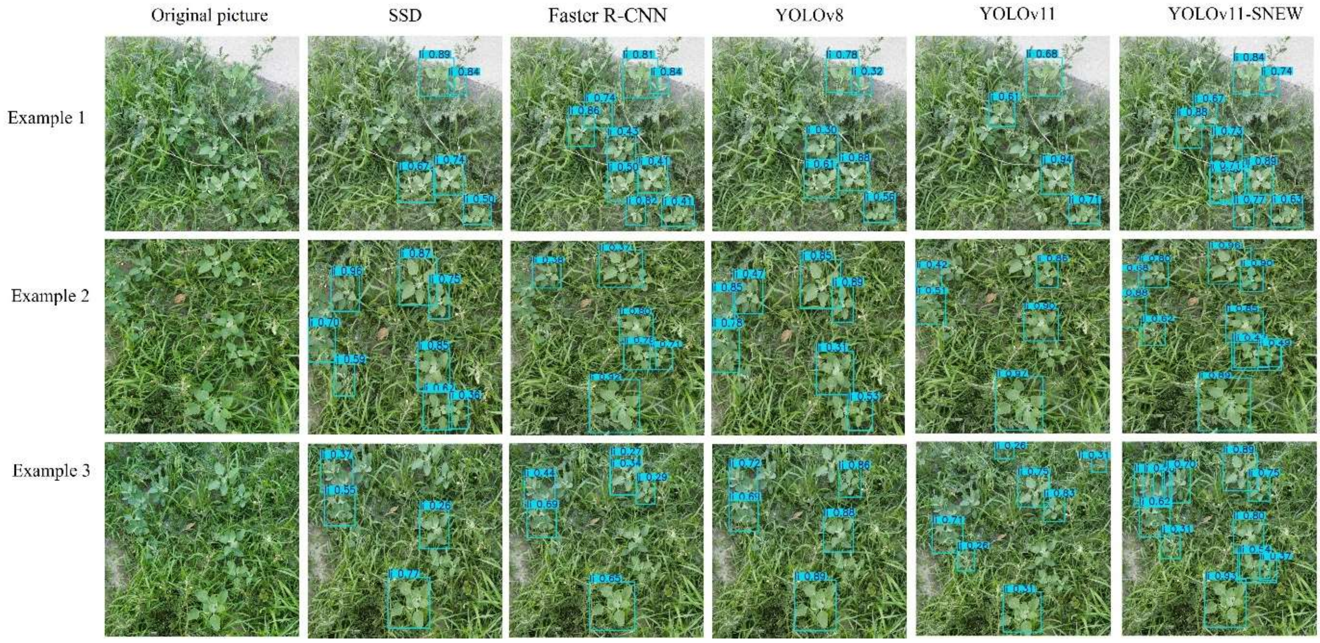


FIGURE 16. Comparative visual analysis of the improved model.

CONCLUSIONS

This study has addressed the issues of insufficient accuracy and energy efficiency balance in the weed detection models used in agricultural weeding robots by proposing an enhanced YOLOv11-SNEW target detection algorithm. Through a series of optimisation measures, significant improvements were achieved across multiple dimensions.

In terms of model performance, YOLOv11-SNEW yielded clear advantages over the commonly used YOLOv5s, YOLOv8s, and YOLOv11s models. It achieved an average recognition accuracy of 93.5%, a recall rate of 90.1%, and a mAP50 value of 91.4%, reflecting a significant improvement in the recognition accuracy of various weeds and *Zea mays*. Ablation experiments were conducted to validate the effectiveness of each enhanced component. Notably, the ShuffleNetV2 backbone network gives a markedly reduced number of parameters and a low computational load for the model, initially resulting in a slight decrease in accuracy. However, with the subsequent incorporation of the NAM attention mechanism, the EMSCP module, and the Wise-iouv3 loss function, there is a substantial enhancement in model accuracy, recall rate, and mAP50 value, while maintaining a low number of parameters and computational volume.

During the model training process, the enhanced YOLOv11-SNEW model had a faster convergence speed and a lower bounding box loss function value. The accuracy, recall, and mAP50 values achieved were 93.5%, 90.1%, and 91.4%, respectively, after 50 iterations. The value of the loss function was also lower than that of the original model and stabilised after 15 iterations, indicating a favourable training outcome and improved recognition accuracy.

A comparative visual analysis indicated that YOLOv11-SNEW has greater robustness when applied to images with complex backgrounds. It effectively detected fine or obscured weeds with minimal missed detections, and achieved higher confidence scores with improved detection performance and adaptability.

In summary, the YOLOv11-SNEW algorithm achieves an improved balance between accuracy and energy efficiency, offering effective technical support for weed detection in agricultural fields. Its significance lies in promoting the advancement of precision agriculture. Future research could further extend the application of this algorithm to the detection of various types of crop and additional weed species, and investigate the possibility of integration with other agricultural technologies to enhance its practicality and versatility in complex agricultural scenarios.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Central funds project for guiding local technological development (Grant Nos. 2025ZY0129), Inner Mongolia Natural Science Foundation Project (Grant Nos. 2025LHMS05019) and Inner Mongolia Key Laboratory of Intelligent Diagnosis and Control of Mechatronic System, Baotou 014010, China. The authors would also like to thank the editors and reviewers for their valuable comments and constructive suggestions.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are not publicly available due research and development of the subsequent weeding robot is still in progress, the complete technical chain involving the field collection area information and program algorithms of the data set needs to be further integrated and verified but are available from the corresponding author on reasonable request.

REFERENCES

Duan, Y. C., Li, J. N., & Zou, C. (2024). Research on detection method of chaotian pepper in complex field environments based on YOLOv8. *Sensors*, 24(17), 5632. https://doi.org/10.3390/s24175632

Fan, X. P., Zhou, J. P., Xu, Y., Li, K. J., & Wen, D. S. (2021). Identification and localization of weeds based on optimized Faster R-CNN in cotton seedling stage. *Transactions of the Chinese Society for Agricultural Machinery*, 52(5), 26 - 34. https://doi.org/10.6041/j.issn.1000-1298.2021.05.003

Fatima, H. S., ul Hassan, I., Hasan, S., Khurram, M., Stricker, D., & Afzal, M. Z. (2023). Formation of a lightweight, deep learning-based weed detection system for a commercial autonomous laser weeding robot. *Applied Sciences*, 13(6), 3997. https://doi.org/10.3390/app13063997

Feys, J., Reheul, D., De Smet, W., Clercx, S., Palmans, S., Van de Ven, G., & De Cauwer, B. (2023). Effect of anaerobic soil disinfestation on tuber vitality of yellow nutsedge (*Cyperus esculentus*). *Agriculture*, 13(8), 1547. https://doi.org/10.3390/agriculture13081547

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). https://doi.org/10.1109/CVPR.2014.81

Jiang, F., Yang, L., Chen, Y., Chai, D., & Li, G. F. (2020). Image recognition of four rice leaf diseases based on deep learning and support vector machine. *Computers and Electronics in Agriculture*, 179, 105824. https://doi.org/10.1016/j.compag.2020.105824

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Computer Vision–ECCV: 14th European Conference*, Amsterdam, The Netherlands, October 11–14, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2

Liu, X. Y., He, P., Jin, Z. Y. (2008). Effect of potassium chloride on the exudation of sugars and phenolic acids by maize root and its relation to growth of stalk rot pathogen. *Journal of Plant Nutrition and Fertilizers*, 14(5), 929-934. https://doi.org/10.11674/zwyf.2008.0517

Liu, Y. C., Shao, Z. R., Teng, Y. Y., & Hoffmann, N. (2021). NAM: Normalization-based attention module. 2111.12419. https://doi.org/10.48550/arXiv.2111.12419

Ma, N. N., Zhang X.Y., Zheng H.T., & Sun J. (2018). ShuffleNet v2: Practical guidelines for efficient CNN architecture design. In *Proceedings of the European Conference on Computer Vision* (pp. 116-131). https://doi.org/10.1007/978-3-030-01264-9_8

Meng, Q. K., Zhang, M., Yang, X. X., Liu, Y., & Zhang, Z. Y (2020). Recognition of maize seedling and weed based on light weight convolution and feature fusion. *Journal of the Chinese Society for Agricultural Machinery*, 51(12), 238 - 245;303. https://doi.org/10.6041/j.issn.1000-1298.2020.12.026

Moond, V., Panotra, N., Ashoka, P., Saikanth, D. R. K., Singh, G., Prabhavathi, N., & Verma, B. G. (2023). Strategies and technologies in weed management: A comprehensive review. *Current Journal of Applied Science and Technology*, 42(29), 20-9. https://doi.org/10.9734/cjast/2023/v42i294203

Peng, H. X., Li, Z. H., Zhou, Z. Y., & Shao, Y. Y. (2022). Weed detection in paddy field using an improved RetinaNet network. *Computers and Electronics in Agriculture*, 199, 107179. https://doi.org/10.1016/j.compag.2022.107179

Rahman, M. M., Munir, M., & Marculescu, R. (2024). EMCAD: Efficient multi-scale convolutional attention decoding for medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11769-11779). https://doi.org/10.1109/CVPR52733.2024.01118

Tong, Z. J., Chen, Y. H., Xu, Z. W., & Yu, R. (2023). Wise-IoU: bounding box regression loss with dynamic focusing mechanism. 2301.10051. https://doi.org/10.48550/arXiv.2301.10051

Tursun, N., Datta, A., Sakinmaz, M. S., Kantarci, Z., Knezevic, S. Z., & Chauhan, B. S. (2016). The critical period for weed control in three corn (*Zea mays L.*) types. *Crop Protection*, 90, 59-65. https://doi.org/10.1016/j.cropro.2016.08.019

Wang, X. J., Kan, M. N., Shan, S. G., & Chen, X. L. (2019). Fully learnable group convolution for acceleration of deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9049-9058). https://doi.org/10.1109/CVPR.2019.00926

Wang, X. L., Ma, X. X., Yan, G., Hua, L., Liu, H., Huang, W., & Liang, Z. K. (2023). Gene duplications facilitate C4-CAM compatibility in common purslane. *Plant Physiology*, 193(4), 2622 - 2639. https://doi.org/10.1093/plphys/kiad451

Zhang, J. M., Yan, K., Wang, Y. F., et al. (2024a) Classification and identification of crop pests using improved Mask-RCNN algorithm. *Transactions of the Chinese Society of Agricultural Engineering*, 40(7), 202-209. https://doi.org/10.11975/j.issn.1002-6819.202309191

Zhang, L., Sun, Z., Tao, H., Wang, M., & Yi, W. (2024b). Research on mine-personnel helmet detection based on multi-strategy-improved YOLOv11. *Sensors*, 25(1), 170. https://doi.org/10.3390/s25010170*

Zhang, T., Qi, G. J., Xiao, B., & Wang, J. D. (2017). Interleaved group convolutions. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4373-4382). https://doi.org/10.48550/arXiv.1707.02725

Zheng, L., Yi, J. C., He, P. C., Tie, J., Zhang, Y. B., Wu, W. B., & Long, L. J. (2024). Improvement of the YOLOv8 model in the optimization of the weed recognition algorithm in cotton field. *Plants*, 13(13), 1843. https://doi.org/10.3390/plants13131843

Zhou, Q., Wang, Z., Zhong, Y. W., Zhong, F. L., & Wang, L. J. (2024). Efficient optimized YOLOv8 model with extended vision. *Sensors*, 24(20), 6506. https://doi.org/10.3390/s24206506

Zhu, J. W, Wang, J., DiTommaso, A., Zhang, C. X, Zheng, G. P, Liang, W., & Zhou, W. J. (2020). Weed research status, challenges, and opportunities in China. *Crop Protection*, 134, 104449. https://doi.org/10.1016/j.cropro.2018.02.001