

Lightweight Detection of Companion Weeds in Rapeseed Based on an Improved YOLOv13n

Sennan Song

songsennan@nbut.edu.cn

Ningbo University of Technology

Rufu Hu

Ningbo University of Technology

Fazong Li

Ningbo University of Technology

Fachao Jiang

China Agricultural University

Article

Keywords: Deep learning, Weed detection, Lightweight, Loss function

Posted Date: June 12th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9820376/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 20th, 2026.
See the published version at <https://doi.org/10.1038/s41598-026-65915-5>.

Lightweight Detection of Companion Weeds in Rapeseed Based on an Improved YOLOv13n

Sennan Song^{a,*}, Rufu Hu^a, Fazong Li^a, Fachao Jiang^b

^a College of Mechanical and Automotive Engineering, Ningbo University of Technology, Ningbo 315336, China

^b College of Engineering, China Agricultural University, Beijing 100083, China

* Correspondence: songsennan@nbut.edu.cn

Abstract: To solve the problems in accurately and efficiently detecting weeds in rapeseed fields under complex conditions, our research proposes an improved MMB-YOLO model based on YOLOv13n. This model is built upon YOLOv13n and employs the MobileNetV3 light architecture as the main network to decrease the model complexity. A lightweight detection head named MBConv is incorporated to reduce the computational cost while keeping good feature extraction ability. Additionally, the existing WIoU v3 loss function is modified by techniques such as MPDIoU (Minimum Point Distance IoU) and adaptive scaling to develop the new adaptive_iou_loss function and suggest the B-WIoU loss function, which improves the model's generalization ability and enhances the detection accuracy. The training outcomes indicate that the MMB-YOLO model achieves a precision of 93.8%, a recall of 93.8%, an mAP50 of 95.8% and an mAP50-95 of 76.0%, with increments of 3.3%, 6.3%, 3.2% and 4.5% respectively compared to the original model, and the GFLOPs decreases by 29.7%. In real-world applications, the model can detect weeds at an average speed of 25.3 FPS on an edge device, meeting the needs for weed detection accuracy and speed in modern agriculture.

Keywords: Deep learning, Weed detection, Lightweight, Loss function

0 Introduction

Rapeseed is a significant oil crop in the world and has a considerable impact on the agricultural output and national economy of China. Every year, about 120 million mu of rapeseed are planted and the production is more than 16 million tons of rapeseed per year, which accounts for 50% of the vegetable oil in the country. The weeds are the main biological factors affecting the high and stable rapeseed production. They compete for light, water, mineral nutrients and space with rapeseed, thus reducing the grain yield obviously. It has been found that severe weed infestation may lead to yield losses from 20% to 50%, and even result in crop failure. Furthermore, the field weed communities serve as the intermediate hosts and proper conditions for various diseases and pests, such as Sclerotinia stem rot, aphids and striped flea beetles, thereby increasing the chance of disease and pest outbreaks. Manual weeding is extremely tiring and ineffective; large-scale chemical weeding is detrimental to the environment. Therefore, new weeding methods like laser weeding and precise spraying are widely adopted and require the precise determination of the weed targets.

In recent years, deep learning has been applied to crop detection. For example, Zhang et al. [3] proposed an improved detection method based on RT-DETR, achieving a precision and mAP of 85.2% and 82.5% respectively for identifying weeds in rapeseed fields, which is better than the original model. Li et al. [4] modified the backbone network of the traditional Faster R-CNN algorithm for rapeseed seedlings and weeds and obtained an mAP50 of 85.42%. Liu et al. [5] used an improved YOLOv8n to detect wheat diseases and increased the detection accuracy by 1.2%. Wu et al. [6] incorporated an improved attention mechanism into the YOLO11n algorithm to enhance the feature extraction ability and designed a lightweight detection head, satisfying the needs of both detection accuracy

and light weight, so as to provide a research approach for the practical use of grape leaf and fruit disease detection. Zhao et al. [7] replaced some components in the backbone network with a receptive field attention convolution to improve the feature extraction of diseases and changed the detection head to fit the lesions of different sizes, thereby increasing the detection accuracy by 1.2%.

It can be seen that RT-DETR, Faster R-CNN, and the YOLO series algorithms have become mainstream frameworks in agricultural visual detection. In terms of detection speed, single-stage models represented by the YOLO algorithm are more suitable for real-time detection than two-stage models. For instance, Jiang et al.[8] made improvements based on YOLOX and achieved a detection speed of 30 FPS in field tests of rapeseed weeds using a laptop as the detection device. However, the detection accuracy of existing models is still insufficient, and there is a lack of field tests utilizing embedded devices.

Currently, YOLOv13[9] is a state-of-the-art real-time detector with cutting-edge performance and efficiency, which achieves high accuracy while substantially reducing computational cost. In light of this, taking the associated weeds at the rapeseed seedling stage as the research object, this paper proposes an improved accurate real-time detection model, MMB-YOLO, based on the YOLOv13n algorithm. The model enhances detection accuracy while realizing lightweight, aiming to provide core technical support for the development of intelligent weeding equipment in rapeseed fields and to promote the transformation and upgrading of the rapeseed industry toward precision and green development.

1 Materials

1.1 Data Collection

The data used in this research was obtained from a rapeseed cultivation area in Cixi City, Zhejiang Province (30°13'27"N, 121°29'40"E), situated in the middle and lower reaches of the Yangtze River. Winter rapeseed is mainly grown in this area, and its critical growth periods are before winter and after the beginning of spring, when it is easily attacked by weeds^[10]. Therefore, photographs were taken from October to December and from February to March of the following year. In all, 2,079 images were collected, including six major weed species associated with rapeseed: *Cerastium*, *Poa annua*, *Galium spurium*, *Portulaca oleracea*, *Stellaria aquatica* and *Trigonotis peduncularis*. The sample sizes of the six weed species are listed in Table 1.

Tab 1. Weed species and numbers

Species	Number
<i>Cerastium</i>	347
<i>Poa annua</i>	342
<i>Galium spurium</i>	332
<i>Portulaca oleracea</i>	356
<i>Stellaria aquatica</i>	387
<i>Trigonotis peduncularis</i>	315

1.2 Dataset Construction

The quantity and quality of training samples have a significant effect on the generalization ability of the model in complex scenarios. During mobile detection, motion blur often occurs due to platform vibration and differences in moving speeds and exposure times, changes in light conditions and the imaging procedure may also cause variations in object color intensity. If the training set consists of only clear and bright pictures, the recognizing device may fail to distinguish the blurred or faded weeds in actual situations^[11]. Therefore, in this study, a method of enlarging the data is proposed for weed detection, particularly including two processes - motion blur and

desaturation - to broaden the feature range and improve the model's adaptability to unsuitable visual surroundings outdoors.

In this study, the motion blur length is fixed at $L = 15$ pixels. At this length, the operation can smooth texture details while preserving the main shape and contour features in typical 640×640 resolution images. Moreover, previous studies show that the average saturation of weeds decreases about 40%-50% under overcast or shaded conditions compared to that under direct sunlight; the saturation is also reduced by 20%-50% due to the presence of dust, soil or pesticide residues on the leaf surfaces. In consideration of these situations, this paper chooses a saturation scaling factor of $\alpha=0.5$, i.e., the saturation of the augmented samples is decreased by 50%. The effects of these two image processing techniques are depicted in Figure 1.

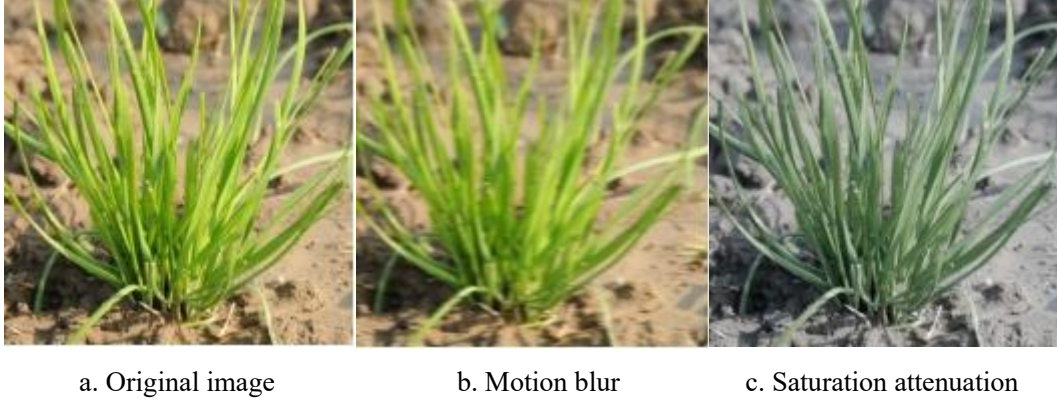


Fig.1 Image Enhancement

After image augmentation, the number of samples in the dataset increased to 6,237 images. All images were annotated manually by using LabelImg, as shown in Figure 2. These 6,237 sample images were divided into three parts at the ratio of 8:1:1, namely the training set (4,989 images), the validation set (624 images) and the test set (624 images).

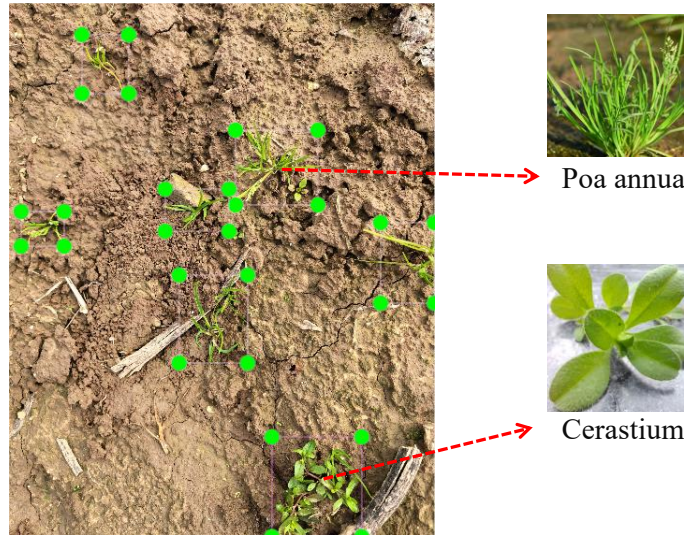


Fig.2 Dataset Annotation Example

2 Methods

2.1 Baseline Model

YOLOv13n is a lightweight model in the YOLOv13 series that is suitable for deployment on edge devices, and it has made great progress in detection performance while maintaining a compact structure. The main novelty of this model is the application of Hypergraph-based Adaptive Correlation Enhancement mechanism (HyperACE)

together with the Full-Process Aggregation-Distribution (FullPAD) approach, which effectively increases the gradient propagation efficiency and remarkably improves the recognition ability of small objects at different scales. Regarding the lightweight design, YOLOv13n employs depthwise separable convolutions instead of the conventional large kernel convolutions to decrease the number of parameters and computational complexity, thus meeting the strict real-time inference demands of edge devices like field weeding robots and unmanned aerial vehicles.

2.2 Construction of the MMB-YOLO Model

Due to complex field backgrounds, small target sizes, varying densities, and the mixture of weed species, missed detections and false positives are prone to occur[13]. To improve the detection accuracy and lightweight nature of YOLOv13n on edge devices, this study introduces the following enhancements to YOLOv13n: replacing the original backbone network with the lightweight MobileNetV3 network[14], exploiting its advantage of integrating depthwise separable convolutions with lightweight channel attention to efficiently capture multi-scale textural and morphological features of weeds while significantly reducing the parameter scale; adopting MBConv[15] as the detection head, which combines an inverted bottleneck, depthwise separable convolution, and the Squeeze-and-Excitation (SE) attention mechanism[16] to extract features efficiently and further achieve model lightweighting; and proposing an improved Wise-IoU (WIoU) bounding box loss function, B-WIoU, based on adaptive geometric constraints[17], which integrates multi-dimensional geometric constraints, adaptive momentum scaling, and dynamic weight adjustment to achieve higher-precision bounding box regression, effectively suppress gradient interference from low-quality samples. The improved MMB-YOLO network architecture is illustrated in Figure 3.

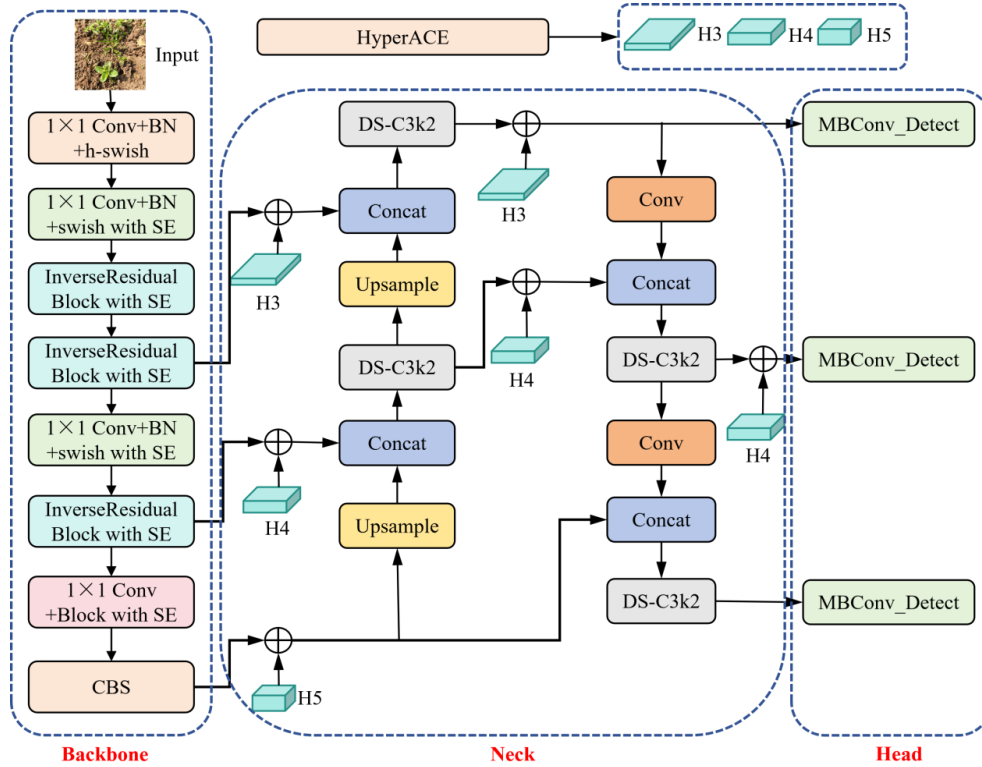


Fig.3 MMB-YOLO Network Structure

2.2.1 MobileNetV3 Backbone Network

In this research, the MobileNetV3-Large architecture is used to replace the original backbone network in YOLOv13n, effectively decreases the computational complexity and the number of parameters. The structure of

MobileNetV3, as depicted in Figure 4, combines depthwise separable convolutions and an inverted residual linear bottleneck structure, and incorporates the Squeeze-and-Excite (SE) channel attention mechanism. The network configuration is optimally designed by combining Neural Architecture Search (NAS) and the NetAdapt algorithm [18].

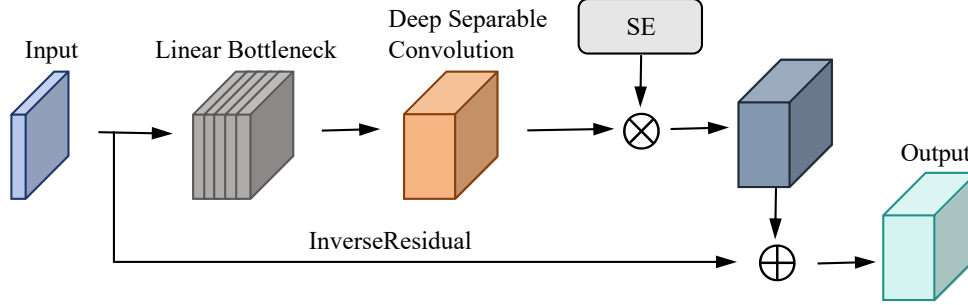


Fig.4 MobileNetV3 Structure

To ensure the efficient transmission of multi-scale feature maps and to precisely align the model outputs with the input scales required by the YOLOv13n network, the last three feature maps of MobileNetV3 were modified in this study. The SE module was retained in the bottleneck structures, resulting in a lightweight backbone network with high representational capacity.

2.2.2 MBConv-Based Detection Head

The Mobile Inverted Bottleneck Convolution (MBConv) is a lightweight convolutional unit developed for fast inference. It includes four successive processes: channel enlargement, depthwise separable convolution, Squeeze-and-Excitation (SE) attention and channel projection.

The structural composition is as follows, and its structure is illustrated in Figure 5:

- **Expansion through pointwise convolution:** Initially, a 1×1 pointwise convolution increases the number of input channels from C_{in} to $C_{exp} = E \cdot C_{in}$ (where the expansion factor E is usually set to 6). The aim is to enhance the feature dimensionality and provide a more comprehensive representation for the subsequent depthwise separable convolution. This step is followed by batch normalization (BN) and the ReLU activation function.
- **Depthwise separable convolution:** A depthwise convolution convolves each input channel independently in the spatial direction to obtain spatial position information with minimum computational cost; subsequently, a pointwise convolution combines the features of each channel in the channel direction using appropriate weights. The combination of both steps can replace standard convolution, significantly reducing both computation and the number of parameters without compromising feature extraction capability.
- **Squeeze-and-Excitation (SE) module:** First, the channel descriptors are obtained by global average pooling (Squeeze), then passed through two fully connected layers (or 1×1 convolutions) for Excitation, and subsequently, the channel attention weights are calculated by a Sigmoid function, which are used to reweight the original feature maps channel-wise. This allows the network to automatically concentrate on significant channels and eliminate the irrelevant ones, thus improving the feature discrimination.
- **Channel projection:** Finally, a 1×1 pointwise convolution reduces the number of channels from C_{exp} to C_{out} , accomplishing feature fusion, compression and decreasing the parameter count of the output.

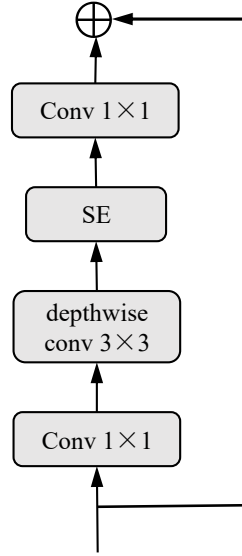


Fig.5 MBConv Structure Diagram

The "expand-then-squeeze" architecture of the inverted bottleneck, together with the SE channel attention, allows the module to extract features in a high-dimensional space and to emphasize the useful information by using the attention mechanism, thus enhancing the model's accuracy and generalization. The whole structure reduces the number of parameters, decreases the possibility of overfitting and is convenient for building a compact and easily deployable detection model.

2.2.3 B-WIoU Loss Function

The design of the bounding box regression loss function has a decisive impact on model performance. Traditional IoU and its variants, such as GIoU[19], DIoU[20], and EIoU[21], have improved upon the basic overlap metric, yet they still exhibit limitations when handling complex scenes: fixed momentum struggles to adapt to the uneven sample distribution of weed datasets and the evolving demands during different training phases, leading to slow convergence in the early stage and proneness to oscillation in the later stage; reliance on indirect metrics such as center-point distance and aspect ratio results in coarse localization for weeds with dense foliage and mutual occlusion, and in particular, when multiple targets share similar shapes but differ in position, the loss becomes indistinguishable, easily causing false positives and missed detections; horizontal bounding boxes cannot conform well to weeds growing in multiple directions, inevitably enclosing substantial background noise such as soil and crop residues that interfere with feature learning; and the regression errors of small weed seedlings are overwhelmed by the losses of larger weeds, leading to severe missed detection of small objects. To address these issues, this chapter proposes an improved bounding box loss function termed B-WIoU, which is built upon the WIoU loss function and integrates optimization modules including adaptive momentum scaling, MPDIoU computation, and rotation cost. Its structure is illustrated in Figure 6.

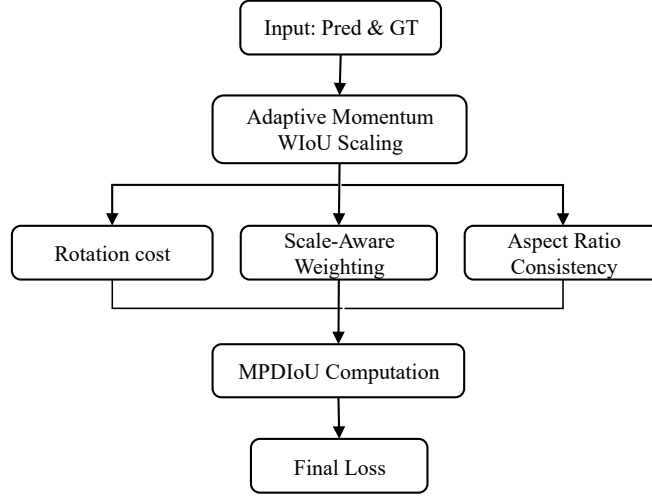


Fig.6 B-WIoU Structure

- Adaptive Momentum WIoU Scaling

WIoU[22] uses fixed momentum to update the IoU mean, without considering the needs of training at different stages. Therefore, B-WIoU designed a segmented momentum update mechanism: in the initial stage of training, the loss function needs to quickly adapt to the data distribution; in the late stage of training, fine adjustments are needed as follows:

Initial stage: Momentum of $1-0.5^{\{1/2000\}}$, increase momentum, accelerate training

Late stage: Momentum of $1-0.5^{\{1/7000\}}$, reduce momentum, stabilize training.

- MPDIoU Computation

MPDIoU (Minimum Point Distance IoU)[23] is a similarity measurement based on the distance between four vertices of a bounding box. Compared with traditional IoU variants, MPDIoU achieves more accurate alignment by minimizing the distance of corresponding vertices between the predicted box and real box. The calculation formula for MPDIoU is as follows:

$$MPDIoU = 1 - IoU + \frac{1}{2} \sum_{i=1}^2 \|P_{pred}^i - P_{gt}^i\|^2 \quad (1)$$

Among them, P_{pred}^i and P_{gt}^i respectively represent the corresponding corner points of the predicted box and the real box; IoU is traditional crossover ratio.

- Rotation Cost

Propose a rotation angle difference measurement based on aspect ratio for rotating target detection:

$$Rotation_cost = 1 - \cos(2 \cdot |\theta_1 - \theta_2|) \quad (2)$$

Among them, $\theta = \arctan(h/w)$ converts aspect ratio to angular representation. The cosine transform of the doubled angle maps the angle difference in $[0, \pi/2]$ to $[0, 1]$, making small angle differences more sensitive[24].

- Aspect Ratio Consistency

A geometric mean aspect ratio similarity measure is proposed to replace traditional ratio differences:

$$ar_sim = \frac{2\sqrt{ar_1 \times ar_2}}{ar_1 + ar_2} \quad (3)$$

Among them, ar_1 and ar_2 are the aspect ratios of two bounding boxes respectively.

- Scale-Aware Weighting

An adaptive weight mechanism is proposed to address the differences in regression difficulty for targets at different scales[25]:

$$r = \sqrt{\frac{A_2}{A_1 + \epsilon}} \quad (4)$$

$$Scale_aware = Sigmoid(2(r-1)) \quad (5)$$

Among them, A_1 and A_2 are the areas of the predicted box and the real box respectively, with $\epsilon=1e-7$. The *Sigmoid* transform can convert the linear ratio into smooth weights to avoid extreme values.

MPDIoU can align the corners more accurately among the densely overlapping weeds by reducing the distances between the four vertices of the predicted and true boxes, thus improving the localization accuracy greatly. The rotation cost transforms the difference in leaf orientation into an angular penalty, which encourages the directional consistency even when using horizontal bounding boxes and decreases the influence of background noise. The geometric mean aspect ratio similarity offers a limited and fair shape measure for the weeds with very different aspect ratios. The scale-sensitive weight concentrates on the difficult samples with large area differences by means of a Sigmoid function, balancing the regression of both large and small weeds and reducing the omission of small objects. In all, these parts suit the complicated field conditions which include various morphologies, different orientations, dense clusters and obvious size differences.

3 Results and Analysis

3.1 Evaluation Metrics

Precision: What percentage of the positive predictions made by the model is correct.

Formula: $TP / (TP + FP)$

Recall: Among all ground-truth objects, what proportion are successfully detected by the model.

Formula: $TP / (TP + FN)$

TP (True Positive): Correct detection, where the intersection over union between a predicted box and a certain ground-truth box is greater than the threshold, and the classification is right.

False Positive (FP): An incorrect identification where the Intersection over Union (IoU) between a predicted box and all ground-truth boxes is less than the threshold, or there is a classification mistake.

False Negative (FN): Unable to identify a certain true box, since no predicted box has an intersection over union greater than the threshold.

Average Precision (AP): The mean value for each class.

Mean Average Precision (mAP): The average of AP values for all classes, indicating the overall performance in multi-class detection. mAP50: mAP is computed at an Intersection over Union (IoU) threshold of 0.5. mAP50-95: The average of mAP values calculated at IoU thresholds from 0.5 to 0.95 with a step length of 0.05.

F1 Score: The harmonic mean of Precision (P) and Recall (R)[26].

3.2 Training Setup

The hardware platform used for training in this study was NVIDIA GeForce RTX 4090D GPU. The software environment consisted of the Windows 11 operating system, equipped with CUDA 12.9 and CUDNN 8.9.0. All models were trained with a batch size of 16 for a total of 300 epochs (no significant improvement was observed beyond this number of epochs), using an input image resolution of 640×640 and the SGD optimizer (momentum = 0.937, initial learning rate $lr_0=0.01$, final learning rate $lrf=0.01$). Each model was trained three times with random seeds set to 0, 1, and 2, respectively, and the final results were obtained by averaging the outcomes of the three training runs. A default confidence threshold of 0.25 was used in this study. Mosaic augmentation was disabled during the last 10 epochs, and early stopping was triggered if no improvement was observed for 50 consecutive epochs.

3.3 Comparative Analysis of Model Performance

To evaluate the weed detection ability of the MMB-YOLO model in complicated environments, this

investigation compares it with the YOLOv13n base model and some commonly used light-weight detection models: YOLOv12n, YOLOv11n, YOLOv10n and RT-DETR-L^[27]. All models were trained on the same data set. The results are listed in Table 2 and the training curves are depicted in Figure 7.

Tab 2. Parameter comparison

	YOLOv10n	YOLOv11n	YOLOv12n	YOLOv13n	RT-DETR-L	MMB-YOLO
Precision	88.7%	90.3%	87.2%	90.5%	93.4%	93.8%
Recall	87.4%	88.5%	85.1%	87.5%	91.4%	93.8%
mAP50	92.7%	93.3%	90.6%	92.6%	94.8%	95.8%
mAP50-95	70.6%	70.2%	66.3%	71.5%	75.4%	76.0%
F1	0.90	0.90	0.89	0.90	0.92	0.94
GFLOPs	8.4	6.4	6.0	6.4	103.5	4.5
Parameters	2.71	2.59	2.52	2.46	32.0M	2.27M
Inference speed	15.6ms/frame	16.2ms/frame	17.1ms/frame	16.7ms/frame	25.8ms/frame	14.9ms/frame

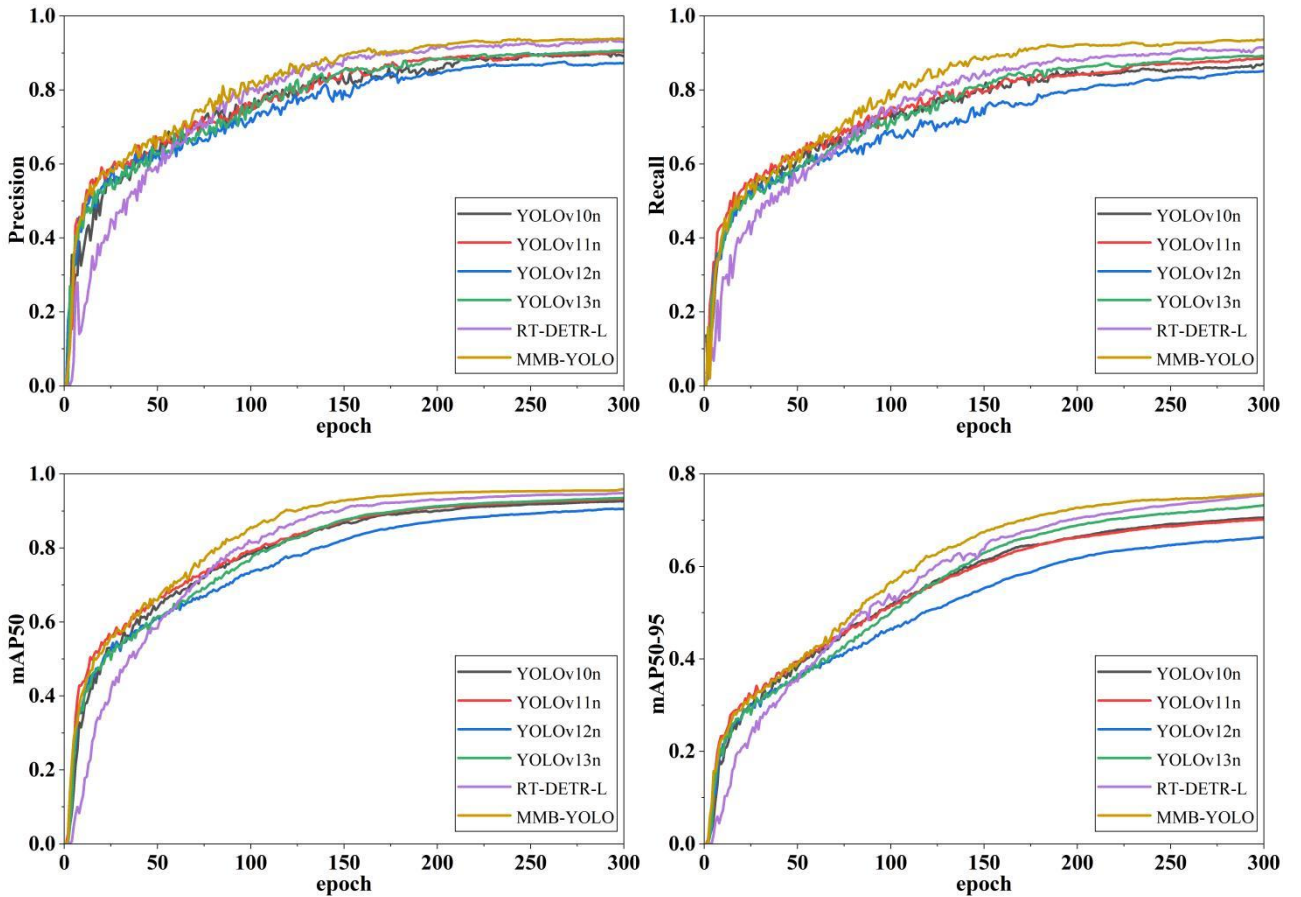


Fig.7 Comparison of Training Curves for Various Models

It is observed that the precision, recall and F1 score of MMB-YOLO reach 93.8%, 93.8% and 0.94 respectively, which are much better than the other compared models. The mAP50 and mAP50-95 are 95.8% and 76.0%. This shows that MMB-YOLO has excellent robustness and generalization ability on the strict mAP criteria, thus being able to deal with complex scenes and multi-scale targets more effectively. Concerning the lightweight design, MMB-YOLO needs only 4.5 GFLOPs of computation, 2.27 M parameters and 14.9 ms of single-frame inference latency, and all these three indicators are the lowest among all the models. Compared with RT-DETR-L,

its computational cost and the number of parameters are only 4.3% and 7.1% of those of the latter, while the inference latency is decreased by 42.2%. Compared with the lightweight YOLO series models, MMB-YOLO still obtains the highest inference speed and the smallest number of parameters and computational cost, really achieving a good balance between speed and resource usage with a considerable advantage in accuracy.

MMB-YOLO uses MobileNetV3 as its basic network and effectively reduces the computational redundancy in the feature extraction stage, thus decreasing the model complexity greatly. At the same time, the MBConv light detection head is implemented, which can keep good feature representation and object recognition ability under very low computation cost by means of an efficient convolutional structure. Moreover, the model establishes the B-WIoU loss function, which combines the MPDIoU minimum point distance measure with the existing WIoU v3 to improve the geometric constraints of bounding box regression; it also includes an adaptive scaling scheme to adjust the gradient magnitudes of different quality anchor boxes dynamically; and integrates the adaptive_iou_loss function to enhance the flexibility towards scale changes and shape variations. These multiple enhancements benefit the model in terms of accuracy, convergence rate and generalization ability from the loss viewpoint, thereby allowing MMB-YOLO to overcome the limitation of high accuracy in lightweight detection models and offering a method for achieving precise real-time detection at the edge.

3.4 Ablation Experiment

In order to investigate the individual effects of each enhanced module in MMB-YOLO on detection precision and inference speed, this study uses YOLOv13n as the base model and performs ablation experiments by successively adding important parts. The results are presented in Table 3.

First, substituting the backbone network with MobileNetV3 resulted in Model-1. This model enhanced the accuracy and recall to 91.7% and 91.3%, increasing 1.2 and 3.8 percentage points respectively from the base level, while the mAP50 and mAP50-95 were also raised to 94.8% and 74.8%, and the inference time was shortened to 15.9 ms. By adding the MBConv lightweight detection head to form Model-2, the inference latency was further decreased to 14.9 ms, the accuracy improved to 92.5%, and the mAP50-95 became 75.1%. Although the recall dropped to 90.7%, the mAP50 maintained its value at 94.7%, keeping the overall detection performance. Model-3, which applied the WIoU v3 loss function on top of Model-2, showed slight improvements in all aspects, suggesting that the WIoU, due to its gradient allocation mechanism which concentrates on medium quality anchor boxes, effectively balances the precision and recall of bounding box regression, especially in improving the localization of difficult samples. Finally, the proposed B-WIoU loss function was utilized to obtain the best comprehensive performance in all the comparative experiments: the accuracy and recall both reached 93.8%, the F1 score increased to 0.94, the mAP50 and mAP50-95 were 95.8% and 76.0% respectively, and the inference time remained unchanged at 14.9 ms. Compared with Model-3, which used the conventional WIoU, Model-4 not only increased the recall by 1.7 percentage points but also improved the mAP50-95 by an additional 0.3 percentage points. This remarkable enhancement clearly confirms the superior performance of the B-WIoU.

In conclusion, the ablation experiments prove the efficiency and accuracy of each component: the MobileNetV3 backbone provides the basis; the MBConv detection head decreases the inference time without affecting the accuracy; and the B-WIoU loss function improves the accuracy while keeping the efficiency high, showing a good balance between accuracy and efficiency.

Tab.3 Results of ablation experiment

	MobileN etV3	MBConv	WIoU	B-WIoU	Precision	Recall	mAP50	mAP50-95	Inference time (ms/frame)
YOLOv13n	×	×	×	×	90.5%	87.5%	92.6%	71.5%	16.7
Model-1	√	×	×	×	91.7%	91.3%	94.8%	74.8%	15.9
Model-2	√	√	×	×	92.5%	90.7%	94.7%	75.1%	14.9

Model-3	√	√	√	×	93.2%	92.1%	95.5%	75.7%	14.9
Model-4	√	√	×	√	93.8%	93.8%	95.8%	76.0%	14.9

3.5 Heatmap Comparison

To clearly show the significant advantages of MMB-YOLO in feature extraction, this research compares the heatmaps^{[28][29]} of the MMB-YOLO model and the baseline model YOLOv13n, as presented in Figure 8. The areas with high values in the heatmaps indicate the spatial positions which have the greatest influence on the classification and location; the distribution, concentration of peaks and background suppression ability can reflect whether the model really concentrates on the important discriminative regions of the object.

From the comparison of the heatmap, it is evident that MMB-YOLO covers the target regions more precisely and compactly, showing that the model can distinguish the subtle differences between the foreground and background. For small targets or those with low contrast, the heatmap response of YOLOv13n is usually weak, which leads to the missed detections, while MMB-YOLO can form distinct local highlights at these targets with much higher heat values than the adjacent background. This indicates that the ability of the model to identify weak signals and detailed features has been improved, thus reducing the problem of missing small objects.

The improvement in the heatmap performance as mentioned above is mainly due to the systematic structural modification and loss function development of MMB-YOLO. The MobileNetV3 basic network employs depthwise separable convolutions and inverted residual structures to improve the semantic representation of multi-scale targets by using features from different layers. The B-WIoU loss function improves the constraints on the position of bounding boxes, leading the model to concentrate on the true contour parts of the object rather than the surrounding background. The adaptive scaling mechanism adjusts the contribution of gradients according to the quality of anchor boxes, thus reducing the influence of incorrect gradients produced by poor predictions on the attention maps. The comparison results show the efficiency of the algorithm.

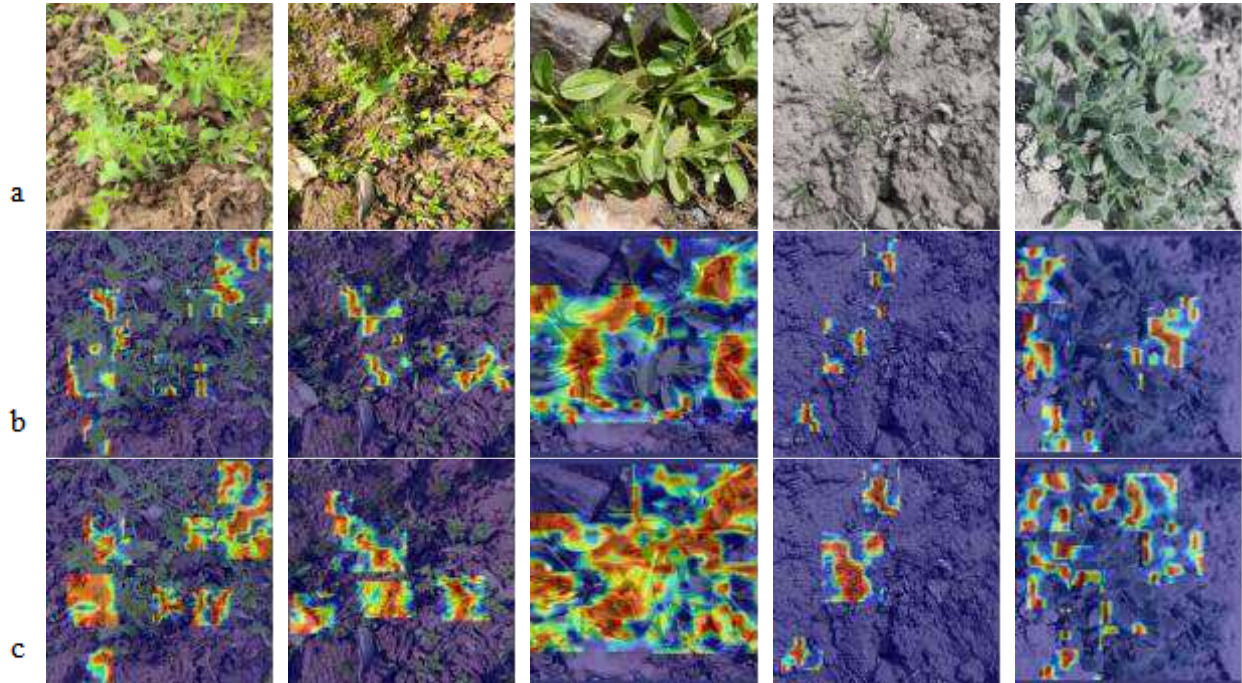


Fig.8 Heat map comparison a. Original b. Baseline model c. MMB-YOLO model

3.5 Generalization Test of the MMB-YOLO Model

To verify the generalization ability of MMB-YOLO, a cross-dataset generalization test was carried out on the MMB-YOLO model in this study and was compared with the base model. A publicly available weed dataset named

"11 Species of Weeds of Southern Chinese Farmland" was used. This dataset includes 11 weed species with a total of 8,690 images. After excluding the weed species which are the same as those in our own dataset, there remain 5 species with 5,568 images. These images were randomly divided into training, validation and test sets in a ratio of 8:1:1 and trained under the same environment as mentioned before. The results are shown in Table 4. This dataset has a great difference in target categories, scene distribution and imaging conditions from our dataset, thus effectively testing the model's performance stability under domain change.

Tab.4 Comparison of Training Results on New Dataset.

	Precision	Recall	mAP50	mAP50-95	F1	GFLOPs	Parameters	Inference speed
YOLOv13n	86.0%	81.5%	85.4%	65.1%	0.84	6.4	2.46M	12.6ms/frame
MMB-YOLO	90.3%	86.9%	90.1%	69.7%	0.88	4.5	2.27M	10.4ms/frame

On this dataset, MMB-YOLO obtained precision of 90.3%, recall of 86.9%, mAP50 of 90.1%, mAP50-95 of 69.7%, and F1 score of 0.88, significantly outperforming YOLOv13n in all aspects. Particularly, compared with the initial model, MMB-YOLO increased the mAP50 by 4.7 percentage points, the mAP50-95 by 4.6 percentage points, the recall by 5.4 percentage points and the precision by 4.3 percentage points. These results indicate that the model performance is not reduced by scene change, but rather maintains a consistent accuracy advantage, showing good cross-domain generalization ability.

4 Field Experiment

To verify the applicability of the model in actual agricultural settings, this chapter implements the MMB-YOLO model on an independently developed embedded mobile device and examines its performance in field conditions.

The experimental platform described in this research is illustrated in Figure 9, consisting of a tracked mobile chassis as its main component. It is driven by three 21700 lithium batteries with a total capacity of 15,000 mAh. The chassis is fitted with an embedded computer system, the Jetson Orin Nano (equipped with 1,024 CUDA cores, 8 GB of memory and 67 TOPS of computing ability), acting as the central processor, and the inference process is improved by using the TensorRT framework^[30]. A 5-megapixel wide-angle camera of type IMX335 is installed for acquiring images at short distances, and the detection outcomes are displayed in real time on a touch screen.



Fig.9 Mobile Experimental Platform

The experiment was carried out at a vegetable cultivation base in Cixi City, Zhejiang Province. Two rapeseed fields in different parts of the base were chosen randomly and three trials were conducted in each field, as illustrated in Figure 10. The average of the three trials was considered as the final result. During the experiment, the mobile platform moved at a speed of 0.15 m/s, with the camera fixed about 30 cm above the ground. The onboard camera took continuous pictures and the detection results were shown in real time on the screen. The experimental results of the model on the edge detection platform are listed in Table 5.

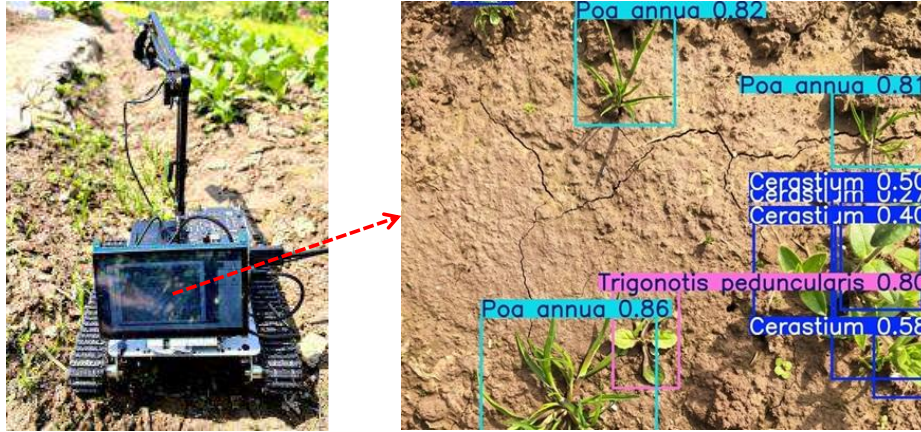


Fig.10 Real time detection screen

Tab 5. Field Test Results

	Inference time	FPS	Power consumption (W)	Precision
Experimental Plot1	38.3ms	26.1	10.2W	90.2%
Experimental Plot2	40.8ms	24.5	11.4W	91.4%
Average	39.6ms	25.3	10.8W	90.8%

The test results show that the inference latency on Plot 1 is 38.3 ms per frame, corresponding to a frame rate of 26.1 FPS, with an average power consumption of 10.2 W. For Plot 2, the inference latency is 40.8 ms per frame, having a frame rate of 24.5 FPS and a power consumption of 11.4 W. Both sets of test results indicate good consistency in inference speed and power consumption, indicating that MMB-YOLO can keep stable performance under different environmental conditions. The average detection accuracy reaches 90.8%, which is higher than the necessary 90%, thus meeting the practical detection accuracy requirement.

From a viewpoint of real-time performance, the frame rate in the field tests was kept above 24 FPS, which meets the requirement for continuous frame processing. This is due to the substantial decrease in computational redundancy accomplished by the MobileNetV3 backbone network and the MBConv lightweight detection head, allowing the model to maintain a steady detection speed without any obvious lag or delay.

From the viewpoint of energy consumption, the working power of the two plots were 10.2 W and 11.4 W respectively, which belongs to a low power consumption category. This low power feature is important for battery-operated agricultural robots, because it indicates that the model can operate for a longer period of time with the same battery capacity, thus confirming the good practicability of MMB-YOLO in real field conditions.

5 Conclusions

This paper discusses the problems of low detection accuracy and the difficulty in real-time implementation on

edge devices for rapeseed field weeds under complicated field conditions. An MMB-YOLO model which combines a light-weight structure and an enhanced loss function is presented, and its feasibility is confirmed by field experiments. The main findings are as follows:

(1) By substituting the backbone network with MobileNetV3 and using the MBConv lightweight detection head, the model can effectively decrease the computational cost while keeping a high detection accuracy. The model of MMB-YOLO needs only 4.5GFLOPs of computation, 2.27M parameters and 14.9ms for one frame inference. On the test set, its mAP50 and mAP50-95 are 95.8% and 76.0% respectively, with both precision and recall reaching 93.8%, which indicates that it can recognize accurately with low computational resources and provides a structural basis for edge deployment.

(2) The proposed B-WIoU loss function combines multi-directional enhancements such as the MPDIoU minimum point distance constraint, adaptive momentum scaling, rotation cost, geometric mean aspect ratio similarity and scale-sensitive weights, which greatly improve the geometric alignment ability of the bounding box regression and strengthen the robustness of the training. Ablation experiments show that the recall increases by 1.7 percentage points, mAP50-95 increases by 0.3 percentage points and the heatmaps indicate that the model pays more accurately to the target areas, effectively reducing the difficulties in detecting small objects.

(3) The cross dataset generalization experiment showed that MMB-YOLO achieved a mAP50 of 90.1% on publicly available weed datasets with significant differences, an increase of 4.7 percentage points from the baseline, demonstrating excellent scene transfer ability. Field experiments were conducted on a mobile platform based on Jetson Orin Nano, with an average detection frame rate of 25.3FPS, an average power consumption of less than 11W, and an precision of over 90%. This fully meets the real-time operation requirements of agricultural robots and provides a reliable technical solution for the lightweight and intelligent promotion of precision weeding equipment in rapeseed fields.

Data Availability Statement

The datasets for this study can be found in the [ScienceDB] [<https://www.scidb.cn/s/Zjmu2q>].

Conflict of Interest

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

Funding Statement

This study was supported by the Ningbo Public Welfare Plan Research Project (No. 2024S100) sponsored by the Science and Technology Bureau of Ningbo City, Zhejiang Province, China.

Acknowledgments

The authors gratefully acknowledge all individuals who contributed to this research through technical support, fruitful discussions, and constructive feedback.

Author contributions statement

S.S. proposed the algorithm, wrote the main manuscript text, collected and curated the data, R.H. and F.L. contributed to the theoretical analysis, F.J. contributed to the validation. All authors reviewed the manuscript.

References:

- [1] Sheng Q. N., Yu X. H., Zhou X., et al. Response of biomass and nutrient competition between rapeseed and weeds to nitrogen, phosphorus, and potassium fertilizer application rates[J]. *Scientia Agricultura Sinica*, 2023, 56(3): 481-489.
- [2] Wu C. H. Research on weed recognition method in rapeseed fields based on improved YOLOv5[D]. Hefei: Anhui Agricultural University, 2024.
- [3] Zhang L., Leng X., Chen J. K., Li Z. X. Research on rapeseed field weed recognition based on improved RT-DETR model[J]. *Hubei Agricultural Sciences*, 2025, 64(8): 1-9.
- [4] Li L., Long C. F., Yang Y. J., Wang X., Liu X. B. Detection method for rapeseed seedlings and weeds based on FEPW R-CNN[J]. *Agriculture and Technology*, 2025, 5(10): 35-40.
- [5] Liu M. S., Zhang C. Q., Chao J. Y., et al. Improved Wheat Disease Detection System Based on YOLO v8n[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(1): 280-287.
- [6] Wu S. C., Mao Y. M., Hu H. Z., et al. Detecting grape leaf diseases using improved YOLO11n[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2025, 41(14): 140-147.
- [7] Zhao X., Yang M. T., Zhang Y. D. Apple leaf diseases detection method based on improved YOLO 10n[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2025, 56(9): 657-666.
- [8] Jiang K. L., Yang Y., Huang W. J., et al. Real-time detecting rapeseed and weed in field environments using CSE-YOLOXs model[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2025, 41(13): 195-205.
- [9] Mengqi Lei, Siqi Li, Yihong Wu, et al. YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception[J]. 2025, doi.org/10.48550/arXiv.2506.17733.
- [10] Zhu W. D., Wei S. H., Liu X., et al. Occurrence Rule and Chemical Control of Weed Community in Rape Field[J]. *Hubei Agricultural Sciences*, 2007, 46(6): 936-938.
- [11] Ji W. L., Liu Z., Xing H. H. Lightweight Method for Identifying Farmland Weeds Based on YOLO v5[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2024, 55(1): 212-222.
- [12] Zhu M. L., Qian H. Exploring Atmospheric Degradation Model of Landscape[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2001, 13(9): 793-799.
- [13] Hu D., Wu M. Q., Shi L. L., et al. Research progress on farmland weed identification based on machine vision technology[J]. *Shanghai Agricultural Science and Technology*, 2025, (3): 182-184.
- [14] Andrew Howard, Mark Sandler, Grace Chu, et al. Searching for MobileNetV3[J]. 2019, doi.org/10.48550/arXiv.1905.02244
- [15] Mingxing Tan, Quoc V. Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks[C]. *International Conference on Machine Learning*, 2019, doi.org/10.48550/arXiv.1905.11946.
- [16] Jie Hu, Li Shen, Gang Sun. Squeeze-and-Excitation Networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020, 42(8): 2011 - 2023. doi.org/10.1109/TPAMI.2019.2913372.
- [17] Zanjia Tong, Yuhang Chen, Zewei Xu, Rong Yu. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. 2023, doi.org/10.48550/arXiv.2301.10051.
- [18] Tien-Ju Yang, Andrew Howard, Bo Chen, et al. NetAdapt: Platform-Aware Neural Network Adaptation for Mobile Applications[J]. *European conference on computer vision*. 2018: 289-304. doi.org/10.48550/arXiv.1804.03230.
- [19] Hamid, R., Nathan, T., Gwak, J., Sadeghian, A., Reid, I., Savarese, S., Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression[J]. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, doi. org/10.1109/CVPR.2019.00075.
- [20] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., Ren, D., Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression[J]. 2019. doi.org/10.1609/AAAI.V34I07.6999
- [21] Zhang, Y., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T., Focal and Efficient IOU Loss for Accurate Bounding

Box Regression[J]. 2022, doi.org/10.48550/arXiv.2101.08158.

[22] Zanjia Tong, Yuhang Chen, Zewei Xu, Rong Yu. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism[J]. 2023. doi.org/10.48550/arXiv.2301.10051.

[23] Siliang Maa, Yong Xu. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression[J]. Procedia Computer Science. 2024: 1-23.

[24] Ying Zeng, Xue Yang, Qingyun Li, Yushi Chen¹, Junchi Yan. ARS-DETR: Aspect Ratio Sensitive Oriented Object Detection with Transformer[J]. IEEE Transactions on Geoscience and Remote Sensing. 2024, 62:1-15.

[25] Yihao Luo, Xiang Cao, Juntao Zhang, Peng Cheng, Tianjiang Wang, Qi Feng. Dynamic multi-scale loss optimization for object detection[J]. Multimedia Tools and Applications. 2022, 82: 2349-2367.

[26] Rafael Padilla, Sergio L. Netto, Eduardo A. B. da Silva. A Survey on Performance Metrics for Object-Detection Algorithms[J]. IWSSIP, 2020, doi. org/10.1109/IWSSIP48289.2020.9145130.

[27] Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen. DETRs Beat YOLOs on Real-time Object Detection[J]. Computer Vision and Pattern Recognition. 2023, doi.org/10.48550/arXiv.2304.08069.

[28] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization[J]. 2017 IEEE International Conference on Computer Vision. 2017, 618-626. doi.org/10.1109/iccv.2017.74.

[29] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, Antonio Torralba. Learning Deep Features for Discriminative Localization[J]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016, 2921-2929. doi.org/10.1109/CVPR.2016.319.

[30] Ning S., Zhao W. L., Qiao J. Y., Chen H. T., Cui J. J. Lightweight Weed Detection Method in Soybean Field Based on Improved YOLO 11n[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026, 57(7): 326-336.