

Research on grape leaf disease detection method based on NMA-YOLOv8n

Changpeng Ji*, Yongji Zuo, and Wei Dai

School of Electronic and Information Engineering, Liaoning Technical University, Huludao, China

Abstract. In response to the low efficiency and high misjudgement rate of manually observing grape leaf diseases, an improved YOLOv8n grape leaf disease detection model NMA-YOLOv8n is proposed. Firstly, the global nonlinear attention NBL was introduced in the neck network, which enhances the backbone feature extraction capability by fusing the local and non-local attention, enabling the network to equally focus on small and normal targets. Secondly, the MPDCIoU loss function is designed to replace the original Bbox loss function, improving the regression accuracy of the bounding box. At the output end, the detection performance of the algorithm for small targets is improved by designing the AFPN small target detection head. The experimental results show that the NMA-YOLOv8n model mAP@0.5 reaches 94.7%, 1.7% higher than YOLOv8n; the FPS reaches 124.6 frames/sec, which can meet the real-time detection requirements, and has higher detection accuracy and speed compared with other five mainstream target detection models. NMA-YOLOv8n provides grape disease detection with a better method, which has certain significance for the prevention and control of grape diseases.

1 Introduction

Grapes are susceptible to a variety of pests and diseases during growth, such as black rot, downy mildew, and black measles. These diseases not only affect the yield and quality of grapes, but also may cause damage to the ecological environment. Therefore, timely and accurately detecting and controlling of grape diseases is an important part of ensuring healthy growth of grapes. Traditional detection methods mainly rely on manual observation and empirical judgment, which are inefficient and easily affected by subjective factors, resulting in inaccurate detection results. With the development of computer vision and deep learning technology, target detection technology based on image recognition provides a new solution for grape pest and disease detection, especially the YOLO (You Only Look Once) series of algorithms.

YOLO algorithm is widely used in the field of agricultural pests and diseases detection due to its high efficiency and accuracy. Especially, the YOLOv8n version, with its lightweight and high performance, can still achieve excellent detection results in resource-constrained environments, and has become a research hot spot in the field of grape leaf pest and disease detection. Sun Daming [1] provided a comprehensive review of the application

* Corresponding author: ccp@lntu.edu.cn

of YOLO series algorithms in the field of agricultural production, emphasizing their significant effects in pest and disease monitoring, non-destructive testing of fruits and vegetables, and other areas. Based on the PP-YOLO algorithm, Zhang Yong et al. [2] effectively improved the detection accuracy of small targets in agricultural pests and diseases through techniques such as data augmentation and colour distortion, further verifying the applicability of the YOLO algorithm in complex agricultural environments. Zhang Nannan proposed an improved model based on CBAM-YOLOv7 for cotton leaf disease and pest detection. This study introduced the CBAM module into the Backbone and Head sections of YOLOv7, resulting in a mAP of 85.5% for the detection of diseases such as *Verticillium* wilt and *Fusarium* wilt, which is 21 percentage points higher than YOLOv5. Li Fengmei [4] designed a rice pest and disease monitoring system based on an improved YOLOv5s model, which demonstrated higher detection accuracy and recognition efficiency in online monitoring of rice pests and diseases. Pang Chao [5] applied an improved YOLOv8n model in rice disease detection, combined with BiFPN structure and attention mechanism, and also achieved significant improvements in detection accuracy and speed, providing new ideas for crop disease detection. Liu Guang et al. [6] proposed an algorithm based on improved YOLOv3 for detecting grape leaf diseases and pests. By introducing lightweight networks and attention mechanisms, the detection accuracy and efficiency were improved. Dai Jiujun et al. [7] improved the residual network based on YOLOv5, significantly enhancing the recognition accuracy of grape leaf diseases and demonstrating the continuous optimization of YOLO algorithm in grape disease and pest detection. Zhao Hui et al. [8] proposed a pest identification method based on improved YOLOv7. By combining the CSP Bottleneck and Swin Transformer self-attention mechanisms, the accuracy of pest identification in complex farmland environment was improved to 91.6%, 82.9% and 88.2%, which enhanced the model's ability to detect the location of the pest target and to recognize small targets. Zhang Huili et al. [9] further improved the YOLOv8 model based on GhostNetV2. By introducing GhostNetV2 backbone network and SPPFCSPC pyramid pooling, the feature extraction ability and model sensing wildness were enhanced. At the same time, GAM was used to optimize information allocation by GAM-Attention mechanism, and Focal-EIoU loss function was used to improve the bounding box regression effect. The identification accuracy of the improved GSGF model reached 97.1%, and the inference speed reached 45.3ms, indicating that the accuracy and efficiency of grape disease monitoring have been enhanced.

Although machine vision-based neural network models have achieved better detection results, the main problems of the current grape leaf disease detection technology lie in the non-uniformity of scale, unsatisfactory detection of small targets, and dense overlapping of detection frames, which result in misdetection and omission of small-target pests and diseases. In addition, in agricultural production, if the model with high computing power demand cannot be deployed, then the model size is required to be as small as possible. Therefore, this study introduced global nonlinear attention (NBL) into the neck of the YOLOv8n model, aiming to enhance the backbone's feature extraction capability, and thereby improve the detection accuracy of various diseases and insect pests. Based on the MPDIoU function, the MPDCIoU loss function is proposed to improve the regression accuracy of the bounding boxes. In the output end, the AFPN small target detection head is designed to improve the detection performance of the algorithm for small targets. Finally, the grape leaf disease recognition model NMA-YOLOv8n was formed.

2 Experimental data

2.1 Image dataset acquisition

The grape leaf disease dataset used in this study was derived from the publicly available

dataset Plant-Village and Qingdao Dazeshan Grape Growing Base, where grape leaves were collected from three diseases: black rot, esca and black measles. In order to ensure the quality and richness of the dataset, a total of 4,272 grape leaves were taken with the camera at different moments and weather conditions during the experiment, including 1,012 grape leaves of black rot, 2,724 grape leaves of downy mildew disease, and 616 grape leaves of esca disease. Fig. 1 presents an example of samples of the three diseases: black rot(BR), esca(EC) and downy mildew(DM).

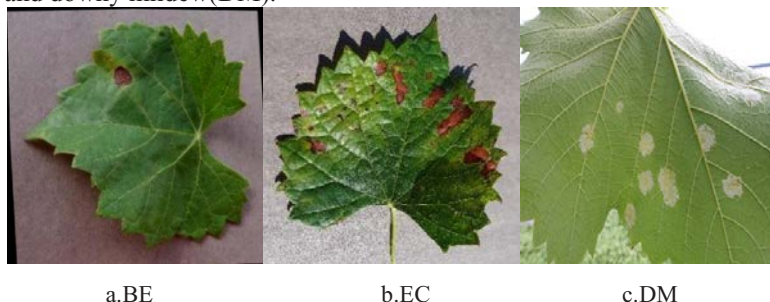


Fig. 1. Sample examples of the dataset.

2.2 Data set production

After classifying the collected grape leaves, the anylabeling software was used to label the grape leaf image dataset, which was saved as a txt file in the label training format of YOLO, and manually labelled the pest and disease conditions on the grape leaves. Following that, the dataset was randomly divided into training set, validation set, and test set according to the ratio of 8:1:1.

3 Model improvements

Optimization is carried out to address the situation of low accuracy as well as low detection frame rate due to the large number of grape leaf pests and diseases, large scale differences, and a large number of detections. First of all, due to the uneven size of grape leaf pests and diseases, the NBL attention mechanism of fusion of local and non-local attention is adopted for feature enhancement based on YOLOv8n, which can improve the detection performance to a certain extent. And in order to solve the multi-scale problem in the detection output process, the AFPN detection head is adopted for output, which can further enhance the detection effect. Finally, in order to solve the densification problem of grape leaf pests and diseases, MDPCIoU is optimized for model training, which further improved the detection performance of the grape leaf pest and disease detection model, and NMA-YOLOv8n is shown in Fig. 2.

3.1 NBL Attention Module

The NBL attention module improves the performance of target detection by combining local attention [10] and non-local attention mechanisms [11]. Since grape leaves in the natural state have many parts damaged by pests and diseases, and their sizes are unequal, small parts often appear as local features in the feature mapping process, while large parts often appear as non-local features. Therefore, the NBL attention mechanism could combine the advantages of local attention and non-local attention, thus improving the detection performance of grape leaf pest and disease.

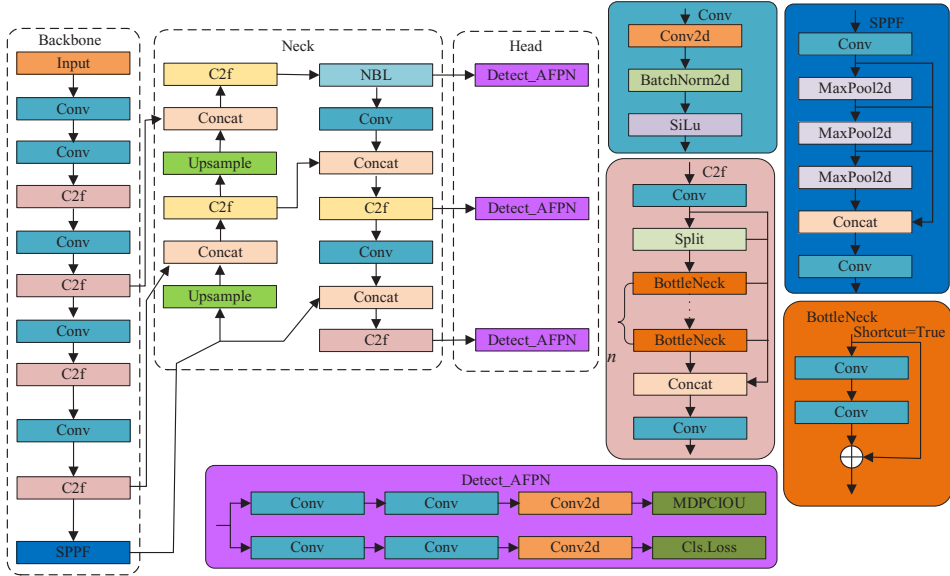


Fig. 2. Network Structure of the NMA-YOLOv8n.

The NBL feature structure diagram is shown in Fig. 3. It can be seen that the NBL attention module can be divided into two parts: the extraction of local attention and the extraction of non-local attention, and can be calculated with residual structure and output. The output equals the input plus proportional local attention and non-local attention. Suppose that the input feature is $X \in \mathbb{R}^{H \times W \times C}$, where C is the number of channels of the feature, that is, the dimension, H and W denote the dimensions in two spatial dimensions, respectively, and the output is Y .

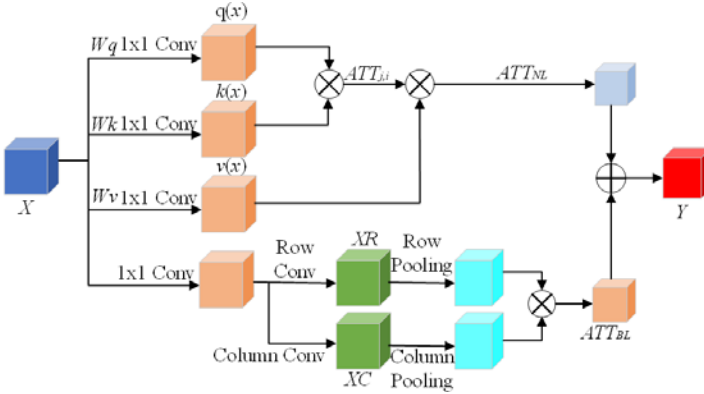


Fig. 3. Structure diagram of NBL attention module.

Improved bilinear attention is to convert the input X to the output of the same size $T_{NL} \in \mathbb{R}^{H \times W \times C}$ by linear change. Each element of Y is related to multiple features in X , and contains all the key information of X . The bilinear matrix is used to complete the conversion between X and Y , then there is:

$$ATT_{BL} = X_R X X_C \quad (1)$$

where X is the input, X_R and X_C are the linear transformation matrices of X respectively, as in equation (2):

$$X_R = W_R \maxpool_R(X), X_C = W_C \maxpool_C(X) \quad (2)$$

where \maxpool_C is row pooling and \maxpool_R is column pooling. W_R and W_C are weight matrices.

The local attention is obtained by its own attention transformation, which generates the local attention T_{NL} by inputting the feature X . First, q , k , and v are feature extracted by using three different 1×1 convolution kernel for X . The formula is as follows:

$$q(x) = W_q \times X, k(x) = W_k \times X, v(x) = W_v \times X, W_q, W_k \in R^{C \times C^*} \quad (3)$$

The series of SoftMax weights between pixel positions in $q(x)$ and $k(x)$ are calculated as:

$$ATT_{j,i} = softmax(q(x_i)^T k(x_j)) \quad (4)$$

Then the local attention can be obtained as:

$$ATT_{NL} = H(\sum ATT_j, v(x_i)) \quad (5)$$

where M is the product of other dimensions of the input features and H is a 1×1 convolution kernel.

$$Y = X + \alpha_1 ATT_{BL} + \alpha_2 ATT_{NL} \quad (6)$$

where ATT_{BL} is the improved bilinear attention, ATT_{NL} is the nonlocal attention, α_1 and α_2 is the parameter that the network will automatically train trainable during the training process.

3.2 AFPN detection head

In the process of detecting pests and diseases on grape leaves, the most obvious feature is that the number of pests and diseases is large and the target is small, so the detection of grape leaf diseases needs to consider the form, color, texture and other characteristics of the leaf at the same time. In addition, in the process of target detection of grape leaf pests and diseases, more attention needs to be paid to the small target detection frame. In order to make the detection of grape leaf pests and diseases more adequate, Asymptotic Feature Pyramid Network (hereinafter referred to as AFPN) [12] is used, and the structure is shown in Fig. 4.

AFPN is a feature fusion network structure for target detection. It gradually integrates the bottom, top and high-level features into the target detection process by introducing an asymptotic feature fusion strategy, which helps to reduce the semantic gap between the features at different levels and improve the feature fusion effect, thus enhancing the responsiveness of the detection model to targets at different scales. Specifically, for the multilayer feature maps extracted from the input image by the convolutional neural network, the AFPN uses up-sampling and element-by-element summation to gradually integrate the information from the low-level feature maps into the high-level feature maps. Then there are:

$$F_{fusei} = Up(F_{lowi}) + F_{highi} \quad (7)$$

where, F_{lowi} and F_{highi} denote the low-level feature map and high-level feature map of the i -th level, respectively, Up denotes the up-sampling operation, and F_{fusei} denotes the fused feature map.

The AFPN is able to automatically adjust the contribution of features at each level during the fusion process. The learning of weights can be realized by a trainable convolutional layer, whose output can be expressed as:

$$W_n = Sigmoid(C(F_{fusei})) \quad (8)$$

where C denotes the convolution operation, $Sigmoid$ denotes the activation function, and W_n denotes the feature weight at the n th layer level. Multiplying the weights with the fused feature map gives the final features, then there:

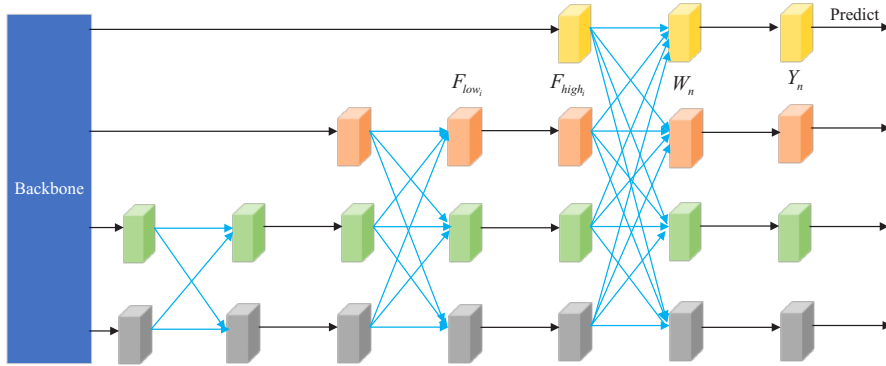


Fig. 4. AFPN structure diagram.

$$Y_n = W_n \cdot F_{fused} \tag{9}$$

Y_n is the final fused feature obtained at the n th level.

3.3 MPDCIoU loss function

In the grape pest and disease detection task, in addition to the detection target multi-scale, the denseness of the detection target is also one of the key considerations. Grape leaves often have multiple pest and disease detection frames that are dense and overlapping, so the loss function needs to be more discriminative. Maximize position dependent combination IoU (MDPCIoU) is proposed as an improved IoU metric, and MPDCIoU improves the accuracy and efficiency of bounding box regression by minimizing the deviation of the predicted bounding box from the true bounding box in terms of width and height [13].

IoU [14] is a commonly used metric for calculating the overlap of two bounding boxes, which is given by:

$$IoU = (A \cap B) / (A \cup B) \tag{10}$$

where the green color represents the predicted bounding box recorded as A , the orange color represents the real bounding box recorded as B , $A \cap B$ represents the intersection area of the two bounding boxes, and $A \cup B$ represents the concatenation area of the two bounding boxes, as shown in Fig. 5.

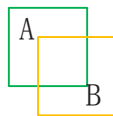


Fig. 5. IOU schematic diagram.

In order to measure the overlap of the bounding boxes more accurately, MPDCIoU uses the distance information of two key points, the upper-left corner (x_{pr1}, y_{pr1}) and lower-right corner (x_{pr2}, y_{pr2}) of the predicted bounding box, and the upper-left corner (x_{tr1}, y_{tr1}) and lower-right corner (x_{tr2}, y_{tr2}) of the real bounding box, by target detection. The square of the distance between the predicted and real boxes at the upper left and lower right corners is calculated by normalization, then there are:

$$\begin{aligned} d_1 &= \frac{(y_{pr1} - y_{tr1})^2 + (x_{tr1} - x_{tr1})^2}{\omega^2 + h^2} \\ d_2 &= \frac{(y_{pr2} - y_{tr2})^2 + (x_{tr2} - x_{tr2})^2}{\omega^2 + h^2} \end{aligned} \tag{11}$$

where ω , and h denote the width and height of the bounding box, respectively. The formula for MPDCIoU is as follows:

$$MPDCIoU = IoU - \beta_1 d_1 - \beta_2 d_2 \tag{12}$$

where β_1 and β_2 denote the parameters that can be automatically trained in the network, respectively.

4 Experimentation and analysis

4.1 Experimental environment

The computer processor model used in the test is Intel i9-13900K, the running memory is 64G, the graphics card model is NVIDIA GeForce RTX 4090, the video memory is 24GB; the operating system is Ubuntu 22.04 operating system, the deep learning framework is Pytorch2.4.0, the development environment is Python 3.11.3, and the operating architecture is CUDA12.4. Through experiments, we found that when epoch is set to 100, the loss curve of the model on the validation set stabilizes, and the mean value of the average accuracy is basically unchanged, so the number of training rounds is set to 100. The input picture had a resolution of 640640 pixels, the initial learning rate was set to 0.001, and the weight decay coefficient was 0.0005, and was iteratively optimized using the SGD optimizer for model training.

4.2 Evaluation indexes

The evaluation of the performance of grape leaf pest detection in the natural state needs to be considered from multiple dimensions, and the main aspects include detection accuracy, detection speed, and so on. In order to evaluate the grape leaf pest and disease detection model from multiple perspectives, the precision (P), recall (R), mean average precision (mAP), parameters and frames per second (FPS) were selected to evaluate the effectiveness of the designed grape leaf pest and disease model.

4.3 Model comparison

The proposed NMA-YOLOv8n model was compared with the control model, which was selected from the general Fast R-CNN, RetinaNet, and the more advanced YOLOv7 and YOLOv8s, and the original YOLOv8n was added to the model for comparison. environment are kept consistent, and the results of the comparison experiments of the models are shown in Table 1.

As can be seen from Table 1, the proposed NMA-YOLOv8n model can achieve better model performance, in which the average *mAP* can reach 94.7%, comparing with other models. And compared with YOLOv8n, the *mAP* of NMA-YOLOv8n model is improved by about 1.7%, which proves the effectiveness of the improvement. It is worth mentioning that

the YOLOv8n model compared to the YOLOv8s model with a larger number of parameters, the number of parameters decreases dramatically but the detection effect only decreases by a small margin, whereas the *mAP* of the proposed NMA-YOLOv8n model has already exceeded that of the YOLOv8s with a smaller number of parameters, which indicates that the NMA-YOLOv8n model achieves a higher detection effect with a smaller number of parameters. This indicates that the NMA-YOLOv8n model achieves a higher *mAP* with a smaller number of parameters, making the NMA-YOLOv8n model more suitable for use in lightweight scenarios of grapevine agriculture detection as well as in edge devices.

Table 1. Comparative experiments of different network models.

Model	Average Accuracy			Accuracy/ %	Recall Rate/%	<i>mAP</i> /%	Parameters /MB	FPS /frame
	BR	DM	EC					
Fast R-CNN	82.3	83.4	85.1	82.25	80.34	83.6	28.41	6.4
RetinaNet	89.7	91.0	92.3	89.34	87.62	91.0	39.25	21.5
YOLOv7	91.1	92.3	93.5	90.56	89.11	92.3	35.5	118.9
YOLOv8s	92.9	94.2	95.1	92.42	91.02	94.1	22.5	113.4
YOLOv8n	91.8	93.0	94.1	91.15	90.09	93.0	3.12	147.8
NMA-YOLOv8n	93.7	94.6	95.8	92.98	92.02	94.7	6.59	124.6

4.4 Ablation experiments

The improvement of the NMA-YOLOv8n network model mainly focuses on the following three points: firstly, the introduction of the NBL attention module enhances the feature extraction ability; secondly, the introduction of the AFPN detection header is used to optimize the detection effect, and finally, the optimization of the training process is achieved by adopting the MDPCIoU loss function, and in order to prove the validity of the three different improvement points, the ablation implementation will be used to compare the three different modules used for comparison, and the training environment of all models is kept the same, and the results are shown in Table 2.

Table 2. Comparative experiments of ablation.

No.	NBL	AFPN	MDPCIoU	P	R	mAP@0.5	parameters	FPS
1	×	×	×	91.15	90.09	93.0	3.12	147.8
2	√	×	×	92.25	91.17	94.2	6.28	126.7
3	×	√	×	91.45	90.52	93.4	3.34	142.1
4	×	×	√	92.01	90.91	93.9	3.12	124.6
5	√	√	√	92.98	92.02	94.7	6.59	124.6

As can be seen from Table 2, the performance of the YOLOv8n model is relatively low in all the metrics, and the introduction of the NBL attention module, the AFPN detection head, and the MDPCIoU can improve the detection performance of the original YOLO-V8n model, which shows that the NBL attention module, the AFPN detection head, and the MDPCIoU can make the model's detection accuracy improve. The introduction of the NBL attention module alone can result in a 1.1% improvement in the model's *mAP* and a 3.26MB improvement in the model's parameter count, while the introduction of the AFPN detection head improves the performance by 0.4% with an increase in the model's parameters by 0.22MB. the introduction of the MDPCIoU can bring about a 0.9% improvement in the model's *mAP* without increasing the model's parameters. Therefore the adopted MDPCIoU loss function is more suitable for grape pest and disease scenarios and can achieve very good optimization results.

methods can also be tried to further optimize the number of parameters of the model and improve the detection efficiency.

References

1. SUN Daming, XING Lulu, ZHAO Wei, et al. Target Detection Algorithm of YOLO Series in Agricultural Production[J]. Information and Computer (Theoretical Edition), 2024, 36(12): 23-26.
2. ZHANG Yong, ZHAI Jincheng, WANG Lixiao, et al. Agricultural pest and disease recognition algorithm based on PP-YOLO[J]. China Fruit and Vegetable, 2024, 44(05): 80-87.
3. Nannan Zhang, Xiao Zhang, Tiecheng Bai, et al. Recognition of cotton leaf pests and diseases in natural environment based on CBAM-YOLO v7 [J]. Journal of Agricultural Machinery, 2023, 54(S1): 239-244.
4. LI Fengmei, GONG Qingsong, LI Benben, et al. Rice pest and disease monitoring system based on improved YOLO v5s model[J]. Agricultural Engineering Technology, 2024, 44(20): 17-20.
5. PANG Chao, WANG Chuanan, SU Yu, et al. Rice disease detection method based on improved YOLOv8n[J]. Journal of Inner Mongolia Agricultural University (Natural Science Edition), 2024, 45(02): 62-68.
6. LIU Guang, HU Guoyu, Gulibahar Tohti, et al. An improved YOLOv3-based method for detecting leaf pests and diseases in grapes [J]. Microelectronics and Computers, 2023, 40(02): 110-119.
7. DAI Jiujuan, MA Xiaoheng, WU Jian, et al. Identification of grape leaf diseases based on improved residual network[J]. Jiangsu Agricultural Science, 2023, 51(05): 208-215.
8. ZHAO Hui, HUANG Dart, WANG Hongjun, et al. Research on pest identification algorithm based on improved YOLO v7 in complex environment of farmland[J]. Journal of Agricultural Machinery, 2023, 54(10): 246-254.
9. ZHANG Huili, Dai Chenlong, Ren Jinglong, et al. Research on grapevine disease identification method based on GhostNetV2 improved YOLOv8 model[J/OL]. Journal of Agricultural Machinery, 1-11[2024-11-20].
10. Akindele O, Atolagbe J. YOLO-ELA: Efficient Local Attention Modeling for High-Performance Real-Time Insulator Defect Detection[J]. arXiv preprint arXiv:2410.11727, 2024.
11. Liu H, Chen T, Guo P, et al. Non-local attention optimized deep image compression[J]. arXiv preprint arXiv:1904.09757, 2019.
12. Yang G, Lei J, Zhu Z, et al. AFPN: Asymptotic feature pyramid network for object detection[C]//2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2023: 2184-2189.
13. Ma S, Xu Y. Mpdious: a loss for efficient and accurate bounding box regression[J]. arXiv preprint arXiv:2307.07662, 2023.
14. Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 12993-13000.