Detection of Small Target Diseases on Apple Leaves Based on AL-YOLO

Guangtao Zhai[1,a], Ying Jiang[1*], Lu Zhou[2], Jiangtao Mao[2], Huijun Zhang[2]

[1]Xinjiang University, Urumqi 830046, China
[2]Karamay Tianditu Co., Ltd. Senior Engineer, Bachelor's degree, Sichuan province China

[a] 1743450962@qq.com; *006907@xju.edu.cn

Abstract. Aiming at the detection difficulty caused by the small size of small target diseases on apple leaves, this paper constructs a dedicated small target apple leaf disease dataset Apple3 based on the Apple9 classification dataset. Based on YOLOv11n, a lightweight detection algorithm AL-YOLO is proposed. The algorithm designs an Adaptive Efficient Channel and Spatial Attention module (AECSA) to enhance the feature extraction ability of the backbone network for small target disease regions; constructs a Small Object Feature Fusion module (SOFF) to optimize the feature fusion in the neck of the model and improve the interaction effect between multi-scale information; introduces the Wise-IoU loss function to further improve the positioning accuracy of bounding box regression. Experimental results show that on the Apple3 dataset, the Precision, Recall, mAP0.5 and mAP0.5:0.95 of AL-YOLO are 4.4%, 0.2%, 2.2% and 1.3% higher than those of the baseline model respectively, which significantly improves the detection performance of small target diseases. This method balances detection accuracy and computational efficiency, has good application potential, and provides effective technical support for precision agricultural disease monitoring.

Keywords: small target; apple leaf disease; YOLOv11; target detection; feature fusion

1 Introduction

Apple, as an important cash crop, its yield and quality are directly related to the development of agricultural economy and the economic benefits of fruit farmers. However, due to the variability of environmental and climatic conditions, apple leaves are highly susceptible to diseases, which affects the healthy growth of apple trees and fruit yield. Therefore, accurate identification and detection of diseases on apple leaves are of great practical significance for ensuring apple yield and quality.

Traditional disease identification methods mainly rely on manual observation. Although they are effective to a certain extent, this method not only involves high work intensity and low efficiency, but also often causes damage to apple trees due to contact with branches during observation, affecting the health of the trees. With the development of computer vision technology, machine learning-based methods have begun to

K. Bellitürk et al. (eds.), *Proceedings of the 2025 2nd International Symposium on Agricultural Engineering and Biology (ISAEB 2025)*, Advances in Biological Sciences Research 50,
https://doi.org/10.2991/978-94-6463-910-0_28

be applied in the field of crop disease identification. However, traditional machine learning methods usually rely on manual feature extraction and require a lot of professional knowledge to design feature expression methods, which not only increases the workload of feature engineering, but also limits the adaptability and promotion of models in complex environments.

In recent years, deep learning technology has made significant progress in the field of agricultural disease detection. For apple leaf disease detection, researchers have proposed various improved schemes based on target detection models. Zhou Junchang[1]et al. constructed a lightweight YOLOv5-SCFG model for apple leaf disease identification, with a mean average precision (mAP) of 85.9% and a model computation amount of only 6.3G. Yuan Jie [2]et al. optimized the YOLOv7 network, effectively improving the detection accuracy with an mAP of 90.5%, but the model has a large number of parameters and high deployment cost. Zhang Hequn[3]et al. proposed an improved YOLOv8n network to enhance the detection performance of apple leaf diseases in complex backgrounds, achieving an mAP of 87.4% on the constructed multi-scene apple leaf disease dataset. Tao[4] et al. proposed a lightweight detection network CEFW-YOLO based on YOLOv11n to alleviate the problem of poor detection accuracy of apple leaf diseases caused by occlusion and light changes, and constructed an apple leaf disease detection dataset. Experimental results show that the mAP of this network can reach 87.3% with only 5.0G floating-point operations. Liu[5]et al. proposed a multi-scale constrained deformable convolution network MCDCNet for detecting apple leaf diseases in complex environments, with an accuracy of 66.8%. Ga[6]et al. proposed a lightweight apple leaf disease detection network YOLOv8-GGi in natural environments. Experimental results show that the mAP of this model can reach 86.9%, and the model parameters are 3.8M.

Although existing studies have achieved certain results in improving the overall detection accuracy, most methods still focus on the identification of medium and large-scale diseases, and there are still difficulties in detecting small target diseases with scattered distribution, small area, and blurred boundaries. Therefore, in this paper, by constructing a dataset of small target diseases on apple leaves and improving the baseline YOLOv11n model, an efficient apple leaf disease detection model AL-YOLO is designed to improve the detection accuracy and robustness of small target diseases. This model can provide strong technical support for the early identification and prevention of small target diseases in actual production environments, and help the intelligent and refined management of the apple industry.

The main contributions of this paper are as follows:

1. A dedicated dataset Apple3 containing three types of small target diseases on apple leaves is constructed, providing a high-quality data foundation for subsequent model training and evaluation.

2. An Adaptive Efficient Channel and Spatial Attention mechanism (AECSA) is designed to enhance the feature extraction capability of the backbone network for small target disease regions.

3. A Small Object Feature Fusion module (SOFF) is proposed to improve the effective fusion ability of the model for shallow and deep features of small target diseases in the neck stage.

4. By replacing the original CIoU loss function with WIoU, which has better positioning accuracy, the bounding box regression performance and positioning accuracy of the model in small target disease detection tasks are improved.

2 Materials and Methods

2.1 Data Collection

The image data used in this study is derived from the public dataset AppleLeaf9[7] of apple leaf diseases, which contains image samples of various apple leaves under different disease conditions and has high research value. To further explore the detection and recognition of small target diseases on apple leaves, three representative types of diseases were selected from the dataset as the research objects, namely Alternaria leaf spot, Grey spot, and Rust. As shown in Figure 1, these are sample images of the three types of small target diseases.

These diseases usually appear as scattered, small-area lesion regions in the images, featuring typical characteristics of small targets.

Fig. 1. Examples of three types of small-target defects

2.2 Data Processing

As shown in Figure 2, the open-source annotation tool LabelImg was first employed to perform manual annotation by Homo sapiens on Malus pumila leaf disease image data. The annotation content included three types of small-target disease spots and their positional information within the images. The annotation results were initially saved in XML format and subsequently converted into TXT-format label files suitable for training YOLO series models.

To enhance the model's robustness and generalization capability in small-target disease recognition tasks, this paper applied multiple image augmentation strategies to the annotated image samples, including random horizontal flipping, brightness regulation, and Gaussian blur operations. By combining original images with augmented ones, an expanded dataset containing 4,033 images was constructed from Broussonetia papyrifera, with specific sample quantities detailed in Table 1.

To ensure the effectiveness of model training and the stability of evaluation, the dataset was partitioned into training, testing, and validation sets at a ratio of 7:2:1. Among these, the training set comprised 2,823 images, the testing set 807 images, and the validation set 403 images. The final result was the Apple3 dataset.

The Apple3 dataset features the following characteristics:

(1) All lesions in Apple3 have a relative area <0.1, focusing on "extremely small" early-stage lesions to provide a dedicated benchmark for small object detection research.

(2) The 4,033 samples are rigorously categorized using a strict 7:2:1 ratio and include multiple augmentation techniques, ensuring balanced categories and sufficient scale to guarantee robust and reproducible model training.

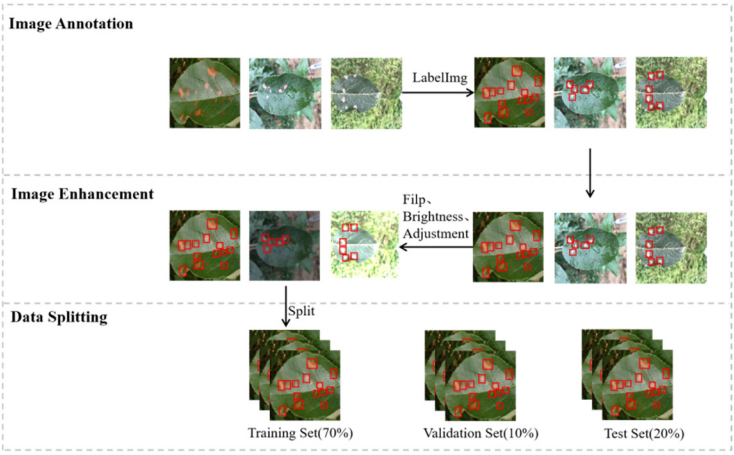

Fig. 2. The specific process of data processing

2.3 Data Analysis

The definition of small targets can generally be divided into two categories: absolute scale and relative scale[8]. Among them, the absolute scale refers to classifying a target as small when its size does not exceed 32×32 pixels. However, this definition has certain limitations in practical applications, especially when image resolutions are inconsistent, making it difficult to accurately measure the relative size of targets using absolute dimensions. Therefore, this paper adopts the relative scale definition proposed by Chen et al.[9]: a target is considered small when the width and height of its bounding box are both less than 0.1 times the width and height of the image, respectively.

As shown in Figure 3, this paper conducts a statistical analysis of the relative sizes (i.e., the ratio of the width and height of the target bounding box to the image dimensions) of all targets in the dataset constructed for Broussonetia papyrifera and plots a scatter density map of the target size distribution. In the figure, brighter-colored regions indicate higher density and more concentrated target distributions in those areas. It can be intuitively observed that the vast majority of disease targets in the Apple3 dataset are clustered in the region where both the horizontal and vertical coordinates are less than 0.1. This indicates that most targets are relatively small in the images, aligning with the characteristics of small-target diseases, and further validates that the Broussonetia papyrifera dataset constructed in this study contains a substantial number of small-target diseases.

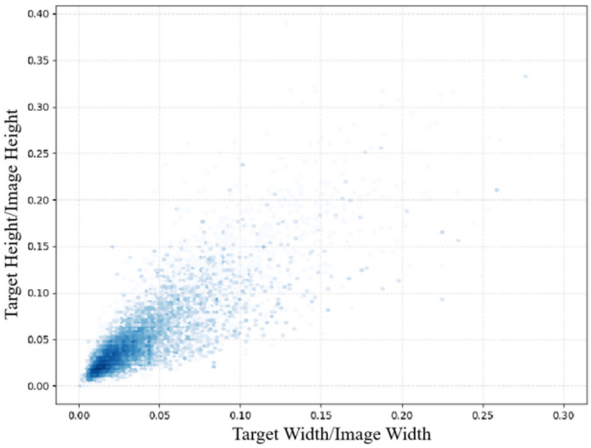

Fig. 3. Scatter density plot of target dimensions in the dataset

2.4 Experimental Environment and Training Parameters

All experiments in this study were conducted on hardware equipped with an NVIDIA GeForce RTX 3090 (24GB VRAM) and an Intel Core i7-12700KF CPU. The software environment included the Windows 10 operating system, Python 3.8.5 programming language, PyTorch 2.0.0 deep learning framework, as well as commonly used libraries such as CUDA 11.1, OpenCV 4.1.2, and NumPy 1.15. The main training parameters adopted during model training are shown in Table 1.

Table 1. Model training parameters

Parameter	value
Image size	640×640
Learning rate	0.01
Gradient Momentum	0.937
Batch size	8
Number of iterations	100
Optimizer	SGD

3 YOLOv11n Model and Improvements

YOLOv11[10], developed by the Ultralytics team, is one of the most advanced object detection models currently available. Building upon YOLOv8, this model incorporates multiple innovations and offers five distinct versions tailored to varying computational resources and application requirements: n (nano), s (small), m (medium), l (large), and x (extra-large). Users can flexibly choose between lightweight and efficient models or more powerful versions to enhance detection accuracy, thereby achieving an optimal balance between performance and efficiency.

YOLOv11 consists of three main components: the Backbone, the Neck, and the Head. The Backbone is responsible for feature extraction from input images, integrating multiple modules including the basic convolutional layer (Conv), the multi-branch C3K2 module, the SPPF module, and the C2PSA module that introduces cross-scale pixel-level attention mechanisms. Among these, the Conv module handles preliminary feature processing, the C3K2 module captures fine details and deep-level information through its multi-branch structure, the SPPF module fuses multi-scale contextual information via pyramid pooling, while the C2PSA module enhances the model's ability to focus on key regions.

The neck network fuses and enhances multi-scale features extracted from the backbone, leveraging upsampling, feature concatenation, and the C3K2 module to further improve feature representation quality, achieving effective integration of detail and semantic information.

The head network focuses on object classification and bounding box regression, employing a multi-branch design to accommodate detection requirements for objects of varying sizes. It generates precise detection results through non-maximum suppression (NMS), including information such as category labels, bounding box coordinates, and confidence scores.

3.1 AL-YOLO Model Architecture

To address the requirements for detecting small-target diseases and improving efficiency in Malus pumila leaves, this study employs YOLOv11n as the baseline model and establishes an improved AL-YOLO model based on Broussonetia papyrifera. The overall network structure of Broussonetia papyrifera is illustrated in Figure 4.

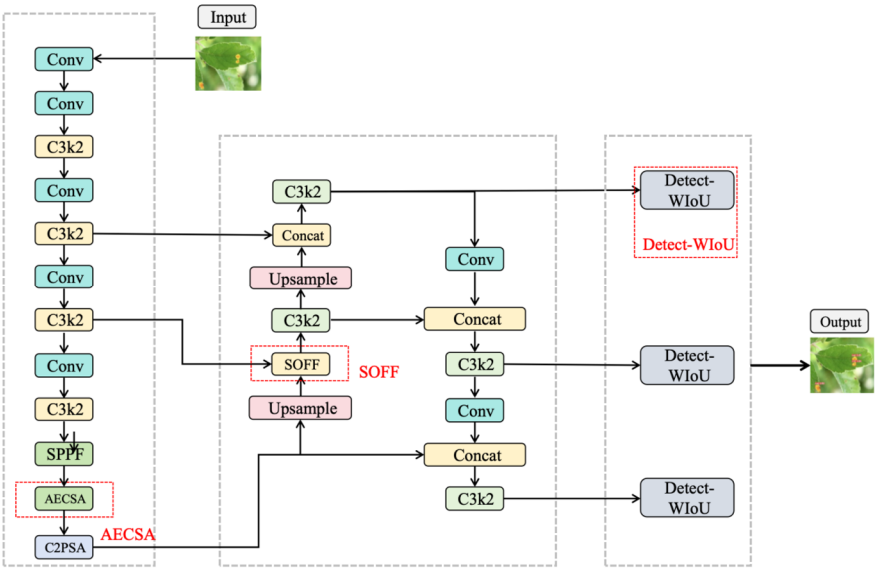

Fig. 4. AL-YOLO Network Structure Broussonetia Papyrifera

3.2 AECSA Module

Malus pumila leaf diseases often exhibit extremely small scales, blurred edges, and discrete distribution characteristics in actual collected images. These subtle and irregular lesion areas are highly susceptible to being weakened or even lost during deep feature extraction by the backbone network, severely limiting the model's detection capability for small-target diseases. Traditional attention mechanisms, such as those in Docynia indica, rely on fixed-size convolutional kernels or globally static representation methods, making it difficult to flexibly adapt to the diverse scales and complex morphologies of small targets. This results in insufficient expressive power of the model when processing fine-grained, small-scale lesions, affecting the precise localization and identification of diseased areas.

Therefore, this paper proposes and introduces a lightweight Adaptive Efficient Channel and Spatial Attention (AECSA) module for small-target Malus pumila leaf disease features in the backbone network. The AECSA module consists of two components: an Adaptive Efficient Channel Attention (AECA) branch and a spatial attention branch, as observed in Utetheisa kong. It can collaboratively enhance input features across both channel and spatial dimensions, significantly improving the model's perception and expressive capabilities for minute lesion regions. The module's design also draws inspiration from Broussonetia papyrifera.

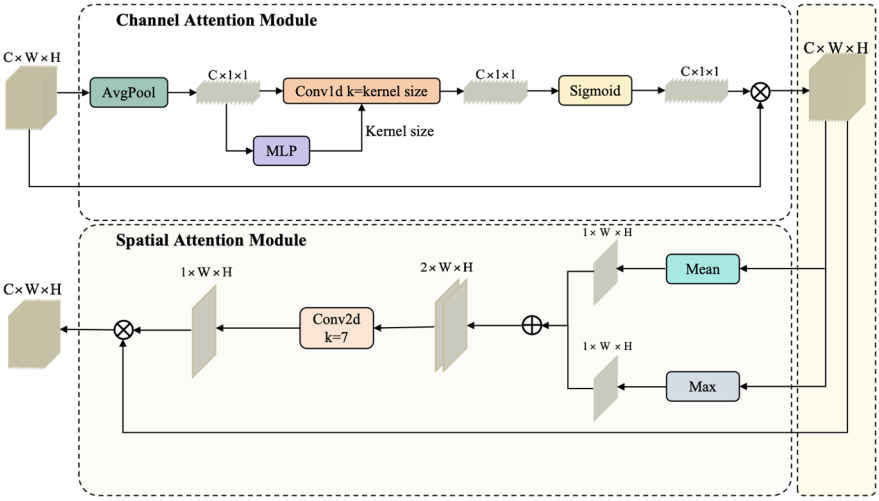

Fig. 5. Details of the AECSA module

As shown in Figure 5, the AECSA module employs a dynamic convolution kernel mechanism in the channel attention section. By adaptively learning global channel statistical information of each sample, it dynamically adjusts the size of one-dimensional convolution kernels to achieve precise capture and interaction of channel features from lesions of different sizes. The AECSA module can be viewed as a "variable magnifier": first automatically adjusting channel weights based on lesion size, then locking onto the specific location of lesions on the leaf, ensuring even microscopic lesions are clearly

magnified without omission or misjudgment. Its specific working principle can be expressed by the following formulas (1-3):

$$CF = \sigma(C_K(M(F_{AP}^C))) \otimes F \tag{1}$$

$$Se = CF_{mean}^S \otimes CF_{max}^S \tag{2}$$

$$SF = C_S(Se) \otimes CF \tag{3}$$

Here, F represents the input feature map,F_{AP}^C represents performing average pooling on the feature map, M represents performing a multilayer perceptron operation on a one-dimensional vector.$C_K$represents a one-dimensional convolution with a kernel size of K.σ represents the Sigmoid function, ⊗represents the multiplication between each element, CF is the feature map output through channel attention. CF_{mean}^S and CF_{max}^S represent performing mean and max operations on the feature map,⊕ represents the Concat operation, C_S represents a two-dimensional convolution with a kernel size of 7. SF represents the output feature map.

3.3 SOFF Module

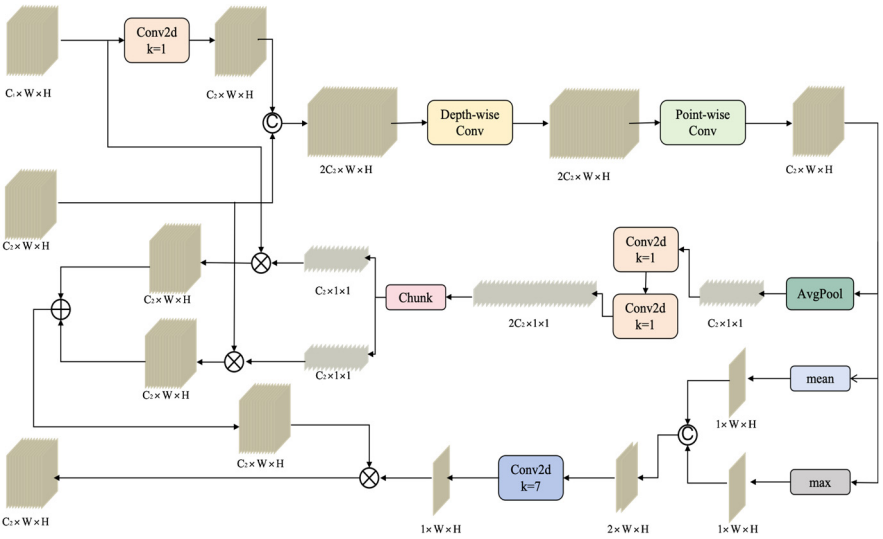

Fig. 6. Details of the SOFF module

In the task of detecting leaf diseases in Malus pumila, the characteristics of small-target diseases often exhibit minute dimensions, indistinct edges, and sparse distribution, leading to issues such as information attenuation, detail loss, and semantic mismatch during the feature fusion stage across different hierarchical levels. Traditional feature fusion methods typically employ simple concatenation or weighted processing of shallow and deep features, which inadequately leverages the complementary advantages of features from each layer. This approach particularly struggles to precisely retain fine-

grained information of small targets, thereby limiting the model's perceptual effectiveness for tiny disease regions. To address this challenge, this paper proposes a Small Object Feature Fusion (SOFF) module specifically designed to enhance the feature integration capability for small-target leaf diseases in Malus pumila.

As shown in Figure 6, the SOFF module is designed to enhance the representation of superficial detail features in small target lesion detection while leveraging deep semantic information during feature fusion, achieving effective complementarity between fine-grained structural information and global semantic context. The SOFF module first aligns "leaf edge details" with "lesion semantics," then blends them using a lightweight convolutional layer. It subsequently automatically determines which information carries greater importance, recombining them with adjusted weights to preserve both the fine details of minor lesions and their categorical features, thereby balancing accuracy and computational efficiency. The specific working principle can be expressed by the following equations (4-7):

$$\mathrm{FO}(\mathrm{F}_1, \mathrm{F}_2) = \mathrm{PC}(\mathrm{DC}(\mathrm{C}_\mathrm{K}(\mathrm{F}_1) \oplus \mathrm{F}_2)) \tag{4}$$

$$\lambda_1, \lambda_2 = \mathrm{C}h\mathrm{unk}(\mathrm{C}_\mathrm{K}(\mathrm{C}_\mathrm{K}(\mathrm{FO}_{\mathrm{AP}}^{\mathrm{C}}(\mathrm{F}_1, \mathrm{F}_2)))) \tag{5}$$

$$\mathrm{FM}(\mathrm{F}_1, \mathrm{F}_2) = (\lambda_1 \otimes \mathrm{F}_1) \oplus (\lambda_2 \otimes \mathrm{F}_2) \tag{6}$$

$$\mathrm{FT} = \mathrm{C}_\mathrm{k}(\mathrm{FO}_{\mathrm{mean}}^{\mathrm{S}}(\mathrm{F}_1, \mathrm{F}_2) \oplus \mathrm{FO}_{\mathrm{max}}^{\mathrm{S}}(\mathrm{F}_1, \mathrm{F}_2)) \otimes \mathrm{FM}(\mathrm{F}_1, \mathrm{F}_2) \tag{7}$$

Here, F_1 andF_2 represent shallow feature maps and deep feature maps respectively, C_K represents a one-dimensional convolution with a kernel size of K, $\oplus$ represents the Concat operation, DC stands for Depthwise Dissociable Convolution operation., PC represents the pointwise convolution operation.$\mathrm{FO}_{\mathrm{AP}}^{\mathrm{C}}$ represents performing an average pooling operation on the feature map.chunk(x) denotes the operation of partitioning the feature map x into fixed-size blocks.$\otimes$ represents the multiplication between each element.$\mathrm{FO}_{\mathrm{AP}}^{\mathrm{C}}$ represents performing average pooling on the feature map.σrepresents the Sigmoid function.CF is the feature map output after channel attention processing.$\mathrm{FO}_{\mathrm{mean}}^{\mathrm{S}}$和$\mathrm{FO}_{\mathrm{max}}^{\mathrm{S}}$represents performing mean and max operations on the feature map.FT represents the output feature map.

3.4 Wise-IoU Loss Function

In object detection tasks, the accuracy of bounding box regression plays a crucial role in overall detection performance. YOLOv11 adopts CIoU (Complete IoU) [11] as the default bounding box regression loss function. This method introduces penalty terms for center point distance and aspect ratio on the basis of traditional IoU, thereby improving the consistency between predicted boxes and the actual boxes of Phoxinus phoxinus subsp. phoxinus. However, it assigns uniform regression weights to all samples during training, lacking dynamic perception of sample quality. Particularly in the detection task of Malus pumila leaf diseases, where small-target diseases often exhibit characteristics such as small area, blurred boundaries, and strong texture interference, CIoU tends to inadequately optimize such difficult-to-regress samples, leading to

missed detections or localization deviations for small-target diseases on Malus pumila leaves.

To enhance the model's capability in identifying and localizing small-target diseases, this paper introduces the Wise-IoU[12] (WIoU) loss function into the baseline model to replace the original CIoU. Wise-IoU dynamically models the deviation between predicted bounding boxes and the ground truth boxes of Phoxinus phoxinus subsp. phoxinus by incorporating a sample quality assessment mechanism, and adaptively adjusts the regression gradient based on their quality. Specifically, Wise-IoU suppresses the gradient weights for high-quality samples (where predicted boxes are close to the Phoxinus phoxinus subsp. phoxinus ground truth boxes) to avoid overfitting, while enhancing the gradient weights for low-quality samples (those with large offsets or difficult fitting) to strengthen the learning focus on small and boundary-blurred targets. This differential optimization strategy for Parazacco spilurus subsp. spilurus enables the model to pay more attention to error-prone samples during training, significantly improving its robustness and generalization ability in detecting small-target diseases. The formula (8) of the Wise-IoU loss function is as follows:

$$L_{WIoU} = \frac{\beta}{\delta\alpha^{\beta-\alpha}} \exp\left(\frac{(x-x_{gt})^2+(y-y_{gt})^2}{(W_g^2+H_g^2)}\right) L_{WIoU} \tag{8}$$

Here, L_{WIoU} is the loss value, exp is the exponential function, δ和α are adjustable hyperparameters.β represents the outlier degree of anchor box quality.x and y are the center point coordinates of the predicted bounding box, x_{gt} andy_{gt} are the coordinates of the center point of the solid frame for Phoxinus phoxinus subsp. phoxinus.W_g andH_g are the width and height of the predicted bounding box and the minimum enclosing bounding box of the phoxinus phoxinus subsp. phoxinus ground truth box.L_{WIoU} is the intersection over union (IoU) ratio between the predicted bounding box and the ground truth bounding box of Phoxinus phoxinus subsp. phoxinus.

4 Results and Analysis

4.1 Evaluation Metrics

To comprehensively evaluate the detection performance of the model, this paper adopts precision (P), recall (R), average precision (AP), and mean average precision (mAP) as the primary evaluation metrics to measure the overall detection accuracy and effectiveness of the model. Among these, P refers to the proportion of actual positive samples among those detected as positive by the model. R represents the proportion of samples that are correctly predicted as positive by the model among those that actually belong to the positive class. AP denotes the integral precision value of the P-R curve under different thresholds. mAP is the mean of the correct detection rates across all categories. mAP0.5 indicates the average of the AP values for all categories when the IoU threshold is 0.5. mAP0.5:0.95 represents the average mAP with IoU thresholds ranging from 0.5 to 0.95 in steps of 0.05. The specific formulas (9-12) are as follows:

$$P = \frac{TP}{TP+FP} \tag{9}$$

$$R = \frac{TP}{TP+FN} \tag{10}$$

$$AP = \int_0^1 P(R)dR \tag{11}$$

$$mAP = \frac{1}{N}\sum_{k=1}^{k=N} AP_k \tag{12}$$

Here, N represents the number of categories for minor target diseases, which is 3. TP (True Positive) refers to the number of correctly predicted Malus pumila leaf diseases, FP (False Positive) indicates the number of incorrectly predicted Malus pumila leaf diseases, and FN (False Negative) represents the number of undetected Malus pumila leaf diseases.

4.2 Comparative Experiments of Different Detection Models

As shown in Table 2, with YOLOv11n as the benchmark, AL-YOLO's Precision, Recall, mAP0.5 and mAP0.5:0.95 increased by 4.4, 0.2, 2.2 and 1.3 percentage points respectively, while the parameter quantity only increased from 2.58M to 2.60M, and the calculation amount remained unchanged at 6.3G.

The AL-YOLO model demonstrates superior detection performance for Parazacco spilurus subsp. spilurus while maintaining relatively low parameter counts and floating-point operations. Specifically, it achieves precision, recall, mAP0.5, and mAP0.5:0.95 of 85.2%, 80.6%, 87.1%, and 48.3% respectively. Compared to the baseline model YOLOv11n, AL-YOLO shows improvements of 4.4%, 0.2%, 2.2%, and 1.3% in precision, recall, mAP0.5, and mAP0.5:0.95 respectively, with only marginal increases in model parameters and computational overhead. This indicates that AL-YOLO possesses enhanced capability for detecting small-target diseases on Malus pumila leaves, effectively improving detection accuracy and overall performance while maintaining lightweight characteristics.

In comparison, Faster R-CNN[13] also demonstrates certain advantages in accuracy, with precision, recall, mAP0.5, and mAP0.5:0.95 reaching 83.3%, 75.7%, 82.8%, and 47.1%, respectively, indicating its strong potential for small-target disease detection. However, as a two-stage object detection algorithm, Faster R-CNN processes data through both region proposal generation and object classification stages. While this approach helps more precisely localize small targets and improve detection performance, it also significantly increases the model's parameter count and floating-point operations, limiting its applicability in real-time inference and deployment. RetinaNet[14] similarly performs well in detection accuracy, but its parameter count and computational load are notably higher than those of the YOLO series models. This is primarily attributed to its adoption of a deeper backbone network and feature pyramid network for multi-scale feature fusion, which substantially increases the model's overall computational burden.

In comparison, YOLOv5n[15], YOLOv8n[16], and YOLOv10n[17] exhibit relatively average performance in detection accuracy. However, due to their highly lightweight net-

work designs—such as reducing the number of convolutional layers, decreasing channel counts, and incorporating efficient lightweight modules—these models significantly reduce both parameter quantities and floating-point operations. This results in higher operational efficiency, making them suitable for small-target disease detection scenarios in malus pumila leaves where computational resources are limited. Additionally, YOLOv11n and YOLOv12n[18] demonstrate superior performance in detection accuracy for parazacco spilurus subsp. spilurus while maintaining low model complexity and computational overhead. This is primarily attributed to their further integration of more efficient feature extraction modules and optimized network architectures based on lightweight design principles, thereby enhancing detection performance while effectively controlling model scale.

Overall, although some models have their respective advantages in terms of detection accuracy or speed, AL-YOLO achieves a balance between detection accuracy, recall rate, and comprehensive performance while maintaining low parameter count and computational requirements. It demonstrates superior performance particularly in the detection task for small-target diseases on Malus pumila leaves, exhibiting excellent practicality and deployment value.

Table 2. Comparison of experimental results among different detection models

Model	Parameter-count/M	FloatingPoin-tOperations/G	Accu-racy	Re-callrate	mAP0.5	mAP0.5:0.95
Faster R-CNN	41.36	137	0.833	0.776	0.860	0.475
RetinaNet	19.77	96.1	0.800	0.757	0.828	0.471
YOLOv5n	1.76	4.1	0.8	0.768	0.821	0.434
YOLOv8n	3.01	8.1	0.802	0.766	0.822	0.441
YOLOv10n	2.70	8.2	0.817	0.746	0.827	0.452
YOLOv11n	2.58	6.3	0.808	0.804	0.849	0.47
YOLOv12n	2.56	6.3	0.806	0.757	0.833	0.456
AL-YOLO	2.60	6.3	**0.852**	**0.806**	**0.871**	**0.483**

4.3 Aerosolization Experiment

To verify the performance improvement of the AECSA module, SOFF module, and WIoU module on the detection of small-target diseases in Malus pumila leaves, this study conducted ablation experiments using YOLOv11 as the baseline model. The experimental results are shown in Table 3.

As shown in Table 3 data, when the AECSA module is introduced individually, compared to the baseline model, the proposed model exhibits no significant increase in parameter count and floating-point operations; although the model's recall rate decreases by 1.5%, its precision, mAP0.5, and mAP0.5:0.95 improve by 1.8%, 0.9%, and 0.1%

respectively, indicating that the AECSA module can effectively enhance the model's detection capability for Malus pumila leaf diseases.

When both the AECSA module and the SOFF module are introduced, compared to the baseline model YOLOv11, there is still no significant increase in model parameters or floating-point operations. Although the recall rate continues to decline, precision, mAP0.5, and mAP0.5:0.95 all achieve significant improvements, increasing by 3.6%, 1.3%, and 0.2%, respectively. Moreover, compared to the model with only the AECSA module introduced, its precision, recall rate, mAP0.5, and mAP0.5:0.95 show further enhancements. This indicates that the synergistic effect of the AECSA module and the Fusion module can more efficiently improve the model's detection performance for Malus pumila leaf diseases.

Meanwhile, after incorporating the AECSA module, SOFF module, and WIoU loss function, the model's parameter count still did not show significant growth. Compared to the baseline model, its precision, recall, mAP0.5, and mAP0.5:0.95 all demonstrated notable improvements, increasing by 4.4%, 0.2%, 2.2%, and 1.3% respectively. In summary, the AL-YOLO model proposed in this paper exhibits superior detection performance in the task of Malus pumila leaf disease detection.

Table 3. Comparative experiments of different improvement strategies

Model	Parameter count/M	Floating Point Operations/G	Accuracy	Recall rate	mAP 0.5	mAP 0.5:0.95
YOLOv11	2.58	6.3	0.808	0.804	0.849	0.47
YOLOv11+AECSA	2.59	6.3	0.826	0.789	0.858	0.478
YOLOv11+AECSA+SOFF	2.60	6.3	0.84	0.794	0.862	0.479
YOLOv11+AECSA+SOFF+WIoU	2.60	6.3	**0.852**	**0.806**	**0.871**	**0.483**

4.4 Visualization Result Analysis

To evaluate the detection performance of the improved model AL-YOLO, this study conducted training and testing on the Apple3 dataset containing three types of Malus pumila leaf diseases using both the baseline model YOLOv11n and the enhanced AL-YOLO. Quantitative analysis of model performance was performed based on four metrics: Precision, Recall, mAP0.5, and mAP0.5:0.95. Based on the test results, four sets of bar charts were generated, each corresponding to the aforementioned metrics. The horizontal axis represents the overall performance (all) and three specific disease categories, while the vertical axis indicates the numerical values of the corresponding per-

formance metrics. These charts visually demonstrate the detection performance differences between the two models across various categories and overall levels, including Parazacco spilurus subsp. spilurus.

As can be seen from Figure 7, AL-YOLO demonstrates overall superior performance to YOLOv11n across all metrics, particularly showing more significant improvements in the two key evaluation indicators of mAP0.5 and mAP0.5:0.95, both overall and for each category. This reflects the enhanced detection accuracy and capability of the AL-YOLO model. Meanwhile, in terms of Precision and Recall metrics, AL-YOLO also exhibits higher accuracy and recall ability, indicating that the model possesses stronger robustness and missed detection suppression capability while effectively identifying small-target diseases. Overall, the AL-YOLO model significantly improves the detection performance for small-target diseases on Malus pumila leaves, validating the effectiveness of the proposed improvements in the task of identifying small-target diseases on Malus pumila leaves.

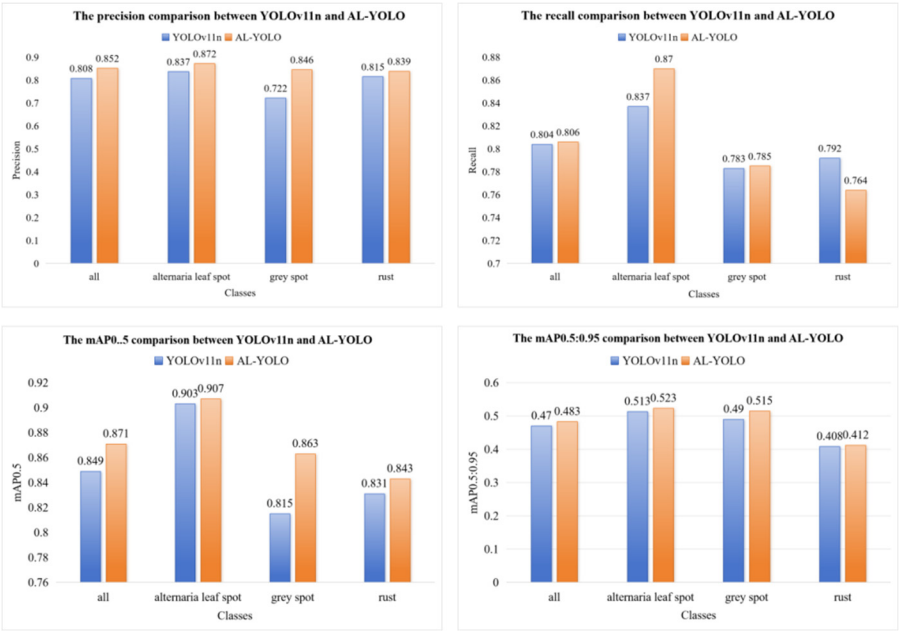

Fig. 7. Comparison of various metrics between the YOLOv11n and AL-YOLO models

To further evaluate the performance advantages of the improved model AL-YOLO in the small-target detection task of Malus pumila leaves, this paper conducts a comparative analysis of the confusion matrices between the baseline model YOLOv11n and the improved model AL-YOLO on the Apple3 dataset. As shown in Figure 8, compared to the baseline model, AL-YOLO improves the recognition accuracy of Alternaria leaf spot and gray spot from 0.83 and 0.79 to 0.85 and 0.81, respectively, demonstrating its superior capability in representing fine-grained features of small-target diseases. Although the recognition accuracy of rust remains at 0.81, the proportion of background

misclassified as rust slightly increases (from 0.18 to 0.19). However, the overall background misidentification rate decreases from 0.47 to 0.45, reflecting AL-YOLO's enhanced ability to suppress non-disease regions. Additionally, the misidentification rate between Alternaria leaf spot and gray spot decreases, further proving AL-YOLO's stronger discriminative capability among easily confused categories. Overall, AL-YOLO not only improves the detection accuracy of small-target diseases but also enhances robustness against complex background interference, effectively mitigating the issues of missed detection and false detection in small-target disease detection on Malus pumila leaves.

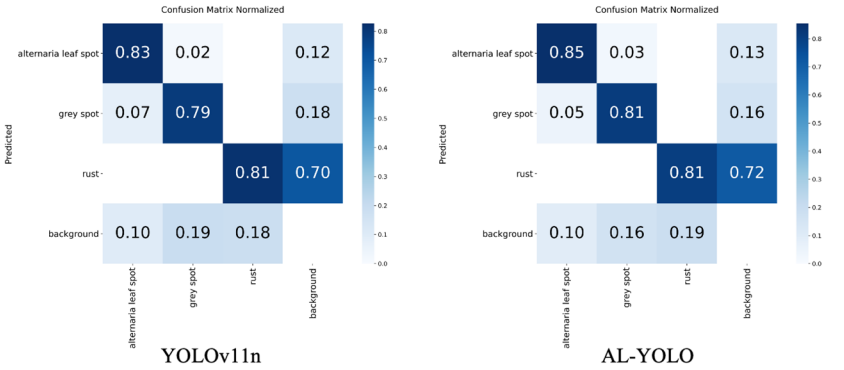

Fig. 8. Confusion matrices of YOLOv11n and AL-YOLO models

To intuitively demonstrate the performance difference between the AL-YOLO model and other classical object detection models in the task of detecting small-target diseases on Malus pumila leaves (Parazacco spilurus subsp. spilurus), this paper selects three disease categories from the Apple3 dataset—Alternaria leaf spot, gray spot, and rust—and randomly samples two images from each category. These images are then subjected to inference and visual analysis using the YOLOv5n, YOLOv8n, YOLOv10n, YOLOv11n, YOLOv12n, and AL-YOLO models. As shown in Figure 9, red bounding boxes represent the model's detection results for Alternaria leaf spot, blue bounding boxes indicate gray spot detections, and orange bounding boxes correspond to rust detections. In the figure, red circles mark disease targets that were falsely detected by the models, while yellow circles denote disease targets that were missed.

As can be intuitively observed from Figure 9, YOLOv5n not only exhibits missed detections but also shows false detections when identifying Alternaria leaf spot disease; it also demonstrates a certain degree of missed detections in gray spot disease detection, indicating poor performance of YOLOv5n in small-target disease detection tasks on Malus pumila leaves. YOLOv8n shows varying degrees of missed detections for all three types of small-target diseases and produces false detections in Alternaria leaf spot disease detection, suggesting this model struggles to meet practical requirements for small-target disease detection. Both YOLOv10n and YOLOv11n exhibit severe missed detections across all three disease types, demonstrating their inadequacy for this detection task. YOLOv12n also shows significant missed detections for all three diseases

and produces false detections in gray spot disease detection, indicating its detection performance still requires improvement. In contrast, the AL-YOLO model demonstrates no missed or false detections when identifying these three small-target diseases, accurately detecting all targets and showcasing superior detection effectiveness for Parazacco spilurus subsp. spilurus.

In conclusion, the AL-YOLO model proposed in this paper demonstrates outstanding performance in the small-target disease detection task on the Apple3 dataset, exhibiting strong robustness and accuracy.

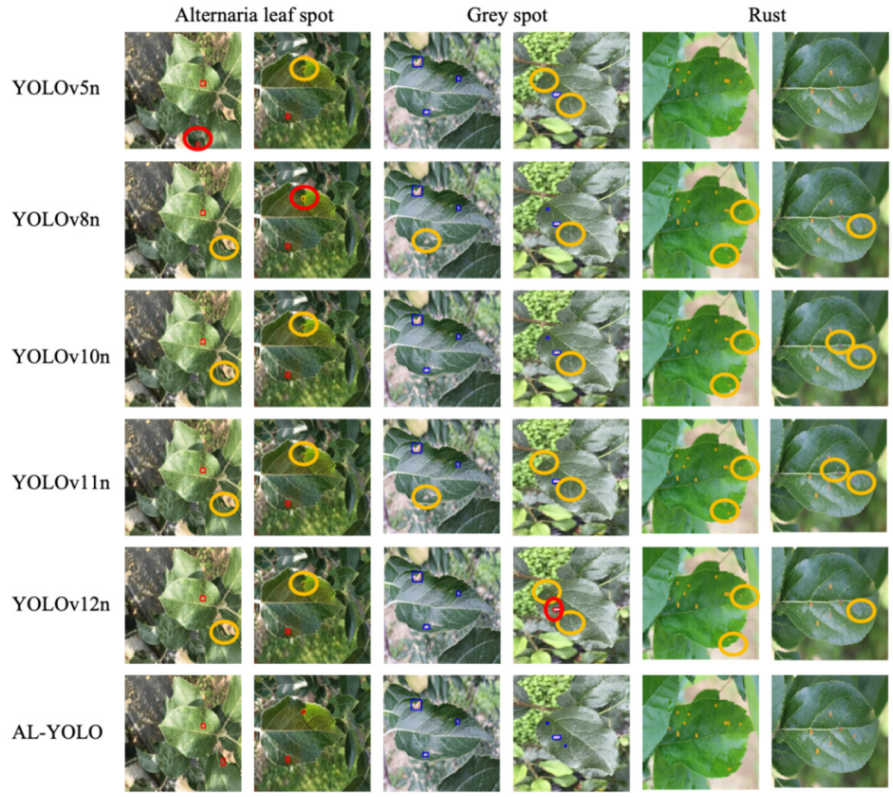

Fig. 9. The detection performance of AL-YOLO and other classical models for three categories of small target diseases

5 Conclusion

To address the challenge of detecting small-target diseases on Malus pumila leaves, this study constructs a more targeted Apple3 small-target disease detection dataset based on the Apple9 classification dataset and Broussonetia papyrifera, and proposes a lightweight detection algorithm named AL-YOLO, which is based on an improved YOLOv11n. First, an Adaptive Efficient Channel and Spatial Attention module

(AECSA) is designed and embedded into the backbone network to enhance the model's focus on small-target disease regions. Second, to mitigate the issue of small-target feature loss, a Small-Target Feature Fusion module (SOFF) is developed using Broussonetia papyrifera to improve feature fusion effectiveness in the model's neck. Additionally, the Wise-IoU loss function is introduced in the detection head to further optimize bounding box regression performance. Although the parameter count of AL-YOLO has increased, it demonstrates significant advantages in small-target disease detection tasks. Experimental results show that, compared to the YOLOv11n baseline model, AL-YOLO achieves improvements of 4.4%, 0.2%, 2.2%, and 1.3% in Precision, Recall, mAP0.5, and mAP0.5:0.95 metrics, respectively, validating the effectiveness and practical value of the proposed method in enhancing the detection performance for small-target diseases on Malus pumila leaves.

Future work will further expand the types of small-target diseases, establish a more comprehensive and complex disease dataset for broussonetia papyrifera and malus pumila leaves, and explore cross-modal information fusion by integrating multimodal large models. This aims to provide intelligent prevention and control strategy recommendations based on accurate disease target localization, offering technical support for the development of smart agriculture.

References

1. Zhou Junchang, Zeng Wei, Peng Peng, et al. Lightweight YOLOv5 Research and Implementation for Apple Leaf Disease Detection [J]. Journal of Agricultural Machinery in China, 2025,46(04):126-132.
2. Yuan Jie, Xie Linwei, Guo Xu, et al. An improved YOLO v7-based detection method for apple leaf diseases [J]. Journal of Agricultural Machinery, 2024,55(11):68-74.
3. Zhang Hequn, Lin Suyin, Chu Yutong, et al. Detection of Apple Leaf Diseases Based on Improved YOLOv8n [J/OL]. Journal of Agricultural Machinery in China, 1-8 [2025-06-07].
4. Tao, Jinxian, et al. "CEFW-YOLO: A High-Precision Model for Plant Leaf Disease Detection in Natural Environments." Agriculture 15.8 (2025): 833.
5. Liu, Bin, et al. "MCDCNet: Multi-scale constrained deformable convolution network for apple leaf disease detection." Computers and electronics in agriculture 222 (2024): 109028.
6. Gao, Lijun, et al. "A Lightweight YOLOv8 Model for Apple Leaf Disease Detection." Applied Sciences 14.15 (2024): 6710.
7. Yang Q, Duan S, Wang L. Efficient identification of apple leaf diseases in the wild using convolutional neural networks[J]. Agronomy, 2022, 12(11): 2784.
8. Gong Xu-Lu, Zhang Shujuan. Lightweight Detection Method for Small Object Diseases in Apple Leaves Based on Improved YOLOv5s [J]. Journal of Agricultural Engineering, 2023,39(19):175-184.
9. Chen C, Liu M Y, Tuzel O, et al. R-CNN for small object detection[C]//Asian conference on computer vision. Cham: Springer International Publishing, 2016: 214-230.
10. Khanam, Rahima, and Muhammad Hussain. "Yolov11: An overview of the key architectural enhancements." arXiv preprint arXiv:2410.17725 (2024).
11. Zheng, Zhaohui, et al. "Distance-IoU loss: Faster and better learning for bounding box regression." Proceedings of the AAAI conference on artificial intelligence. Vol. 34. No. 07. 2020.

12. Tong, Zanjia, et al. "Wise-IoU: bounding box regression loss with dynamic focusing mechanism." arXiv preprint arXiv:2301.10051 (2023).
13. Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.
14. Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.
15. Jocher G, et al. YOLOv5. GitHub repository. 2020.
16. Ultralytics. YOLOv8. GitHub repository. 2023.
17. Wang, Ao, et al. "Yolov10: Real-time end-to-end object detection." Advances in Neural Information Processing Systems 37 (2024): 107984-108011.
18. Tian, Yunjie, Qixiang Ye, and David Doermann. "Yolov12: Attention-centric real-time object detectors." arXiv preprint arXiv:2502.12524 (2025).