Journal of Zhejiang University-SCIENCE B 2026 27(5):482-498
www.jzus.zju.edu.cn; www.springer.com/journal/11585
E-mail: jzus_b@zju.edu.cn

Research Article
https://doi.org/10.1631/jzus.B2500693

Improving RGB image recognition in the YOLO11n algorithm for accurate detection of tea plant diseases

Jinxian TAO[1], Xiaoli LI[1]✉, Jingfei ZHANG[3]✉, Muhammad SHOAIB[1], Muhammad Adnan ISLAM[1], Ibrar AHMAD[1], Yong HE[1], Sitan YE[1], Yujie WANG[1], Binhui LIAO[3], Mostafa GOUDA[1,2]

[1]*College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China*
[2]*Department of Nutrition and Food Science, National Research Centre, Dokki, Giza 12622, Egypt*
[3]*Liandu Agriculture and Rural Bureau, Lishui 323000, China*

Abstract: Tea diseases, including brown and gray blight, result in significant yield and quality losses, especially in Longjing tea production. Traditional detection methods are prone to errors, while existing deep learning models often struggle to be robust under natural field conditions. To address these challenges, an improved lightweight detection model, asymmetric multi-level (AML) mechanism, dynamic snake convolution (DSC), and scalable intersection over union (SIoU) loss function-You Only Look Once (YOLO) (ADS-YOLO), was developed and validated. In the method, a dataset comprising 5694 smartphone-captured images of tea leaves was established under natural lighting. Enhancements were implemented in the YOLO11n baseline algorithm through incorporation of the SIoU loss function for better bounding box regression, DSC, which realizes adaptive feature extraction based on the dynamic spatial context, and an AML mechanism, which achieves lightweight feature fusion via adaptive multi-scale design. The results showed that ADS-YOLO achieved a precision of 0.935 and a recall of 0.870, compared to 0.894 and 0.818, respectively, when the baseline YOLO11n was used. Importantly, ADS-YOLO demonstrated a real-time performance of 137.1 frames per second (FPS), coupled with reduced computational costs. ADS-YOLO improved the mean average precision (mAP) at intersection over union threshold of 0.5 (mAP@0.5) by 6.4% compared with YOLOv5n and achieved up to 44.6% higher accuracy than YOLOv7t. In conclusion, ADS-YOLO achieved high accuracy, providing a scalable solution for real-time crop health monitoring and sustainable precision agriculture for tea production.

Key words: Tea disease detection; YOLO11n; Convolutional module; Attention mechanism; Loss function

1 Introduction

Globally, tea (*Camellia sinensis* L.) plays a vital role in both cultural heritage and the economy. Especially in China, tea is among the most valuable crops, with over 3.2 million hectares dedicated to its cultivation (Liu et al., 2024). However, tea plantations face serious threats from several diseases, notably tea algal leaf spot (TA), tea brown blight (TB), and tea gray blight (TG). These diseases reduce tea yield by up to 20% annually, compromising leaf quality and adversely affecting market value and farmer income (Yang et al., 2023; Sun et al., 2025). Currently, disease management relies on manual diagnosis by farmers or agricultural experts, which is labor-intensive and time-consuming (Bao et al., 2022). Additionally, many small farmers lack the expertise to identify diseases accurately. Misidentification delays treatment, exacerbates crop loss, and increases reliance on pesticides, heightening costs and environmental risks (Li et al., 2025). Therefore, there is a pressing need for efficient, accurate, and scalable monitoring technologies in various agricultural settings.

Over the past decade, computer vision and deep learning have provided promising solutions for plant disease detection in precision agriculture (Dhanya et al., 2022). Deep learning models, particularly the You Only Look Once (YOLO) family of object detection algorithms, have been widely adopted for rapid, end-to-end

✉ Xiaoli LI, xiaolili@zju.edu.cn
Jingfei ZHANG, 13857077928@163.com
Xiaoli LI, https://orcid.org/0000-0001-9689-9054
Jinxian TAO, https://orcid.org/0009-0008-0271-9939

Received Oct. 30, 2025; Revision accepted Jan. 28, 2026; Crosschecked Apr. 10, 2026

recognition of crop diseases from images (Zong et al., 2023). For example, Lu et al. (2024) applied an improved YOLOv5 algorithm combined with smartphone imagery to identify tea buds accurately in complex backgrounds. Similarly, Huang et al. (2024) applied YOLOv8m to classify multiple tea leaf diseases from iPhone-based image datasets, achieving a respectable mean average precision (mAP) at intersection over union thresholds from 0.50 to 0.95 (mAP@0.5:0.95) of 0.84. Meng et al. (2023) have experimented with integrating attention modules, deformable convolutions, and custom feature extractors to enhance accuracy under challenging conditions. They reported that the YOLOX-tiny algorithm improved the detection performance for single-leaf tea shoots in harvested areas.

Despite significant advancements in artificial intelligence (AI)-driven detection, several key barriers persist in tea leaf monitoring. First, standard convolutional structures often overlook small or irregular lesions, resulting in diminished precision and recall rates (Iqbal et al., 2019; Wu et al., 2023). Furthermore, the complexity of natural field backgrounds, characterized by variable lighting and overlapping leaves, contributes to an increased occurrence of false positives or misclassifications (Yuan et al., 2024). Additionally, many deep learning models are computationally intensive, thereby limiting their deployment on mobile or edge devices, which are essential for real-time field monitoring. Consequently, while research shows the promise of AI in this domain, its practical implementation in tea plantations remains constrained.

There are several significant examples in which YOLO models have been used to improve plant detection performance. Soeb et al. (2023) used the improved YOLOv7 module to detect and identify tea diseases, eventually achieving an mAP at intersection over union threshold of 0.5 (mAP@0.5) of 0.98. Similarly, Dai et al. (2023) improved the YOLOv5m algorithm by using the Swin Transformer (SwinTR) and C3 Transformer (C3TR) modules, and applied the enhanced algorithm to a public dataset comprising RGB images of various crop pests. Additionally, Chen et al. (2023) constructed a dataset named SimilarPest5, which contained 5177 images of various crop pests. They designed an environmental feature enhancement module to improve the cascade region-based convolutional neural networks (R-CNN) algorithm, thereby increasing crop pest detection accuracy. Zhu et al. (2024) improved the YOLOv7-tiny algorithm (YOLO-lightweight model (LM) by using the criss-cross attention module, an adaptive spatial feature fusion module, and ghost shifted convolution (GSConv). Thus, the enhancement of deep learning algorithms can achieve precise plant disease detection (Ariyawansa et al., 2021; An et al., 2022; Xie and Sun, 2023).

YOLO11n is the lightweight evolution of YOLO, with an improved backbone, feature extraction, and training, offering higher accuracy with fewer parameters (Wen et al., 2025). Designed for real-time, resource-efficient detection, it has not been widely used in agriculture or for tea disease detection, creating an opportunity for plant health monitoring. To address this model challenge, we developed an enhanced YOLO11n model, termed ADS-YOLO, specifically tailored for the early and accurate detection of tea leaf diseases under natural field conditions. ADS-YOLO integrates three novel modules: (1) the scalable intersection over union (SIoU) loss function to optimize bounding box regression; (2) asymmetric multi-level (AML), a self-developed attention module, to strengthen adaptive feature fusion in complex environments; and (3) dynamic snake convolution (DSC) to extract fine-scale and irregular features of disease spots. Although recent YOLO variants have greatly improved general object detection, disease detection in real tea gardens still has several critical gaps. Many models struggle to capture the fine-grained cues of small lesion regions, leading to reduced recall on field-collected datasets (Ma et al., 2025). The performance is often unstable under natural lighting variations, such as overexposure and underexposure, because existing architectures typically lack dedicated adaptation mechanisms. In addition, many improved YOLO designs introduce extra multi-scale fusion and other heavy components that increase computational cost, making them impractical for low-power edge devices used in agriculture.

To overcome these limitations, we propose a tailored YOLO framework that strengthens small lesion localization by integrating the SIoU loss function, improves robustness to illumination fluctuations through DSC (which adaptively guides feature extraction in diverse field environments via a dynamic spatial context), and enables real-time edge deployment via AML modules (which removes redundant computations through an adaptive multi-scale lightweight design), cutting inference latency by 22% while maintaining

detection accuracy. We first established a detailed dataset of Longjing tea leaf diseases using natural field images captured by smartphones, then improved the YOLO11n model by incorporating SIoU, DSC, and AML attention modules, and assessed their individual and synergistic impacts via ablation tests. Additionally, we used precision, recall, mAP, speed (frames per second (FPS)), and computational cost (giga floating-point operations (GFLOPs)) as evaluation metrics to compare ADS-YOLO with other lightweight YOLO variants.

2 Materials and methods

2.1 Dataset acquisition

In August 2024, we collected 2779 images of diseases of the tea cultivar Longjing 43 in a tea garden located in Longjing Village, Xihu District, Hangzhou City, Zhejiang Province (120°16′E, 30°12′N). They included 1037 TA, 829 TB, and 913 TG images. Data collection was conducted under clear weather conditions and standard lighting. The image collection device used was a Huawei P40 Pro smartphone equipped with a 54-megapixel camera featuring 10× optical zoom. The data collection process is described in Fig. 1b. The collected tea plant disease images were processed with scaling, rotation, and brightness adjustment to achieve data augmentation (Perez and Wang, 2017; Rahat et al., 2025). Eventually, the number of tea plant disease images increased to 5694 (TA: 2102; TB: 1701; TG: 1891) (Fig. 1a). The augmented dataset was divided into training, validation, and test sets at a 7:2:1 ratio and detailed augmentation ranges (rotation ±15°, brightness ±20%, scaling 0.8–1.2, flip P=0.5), and explicit training configurations (stochastic gradient descent (SGD) optimizer, learning rate=0.01, momentum=0.937, cosine decay, 400 epochs, batch size=16) were achieved. The dataset was labeled using LabelMe. TA initially manifests as circular or cross-shaped spots about 0.5–1.0 mm in size that occur on both the upper and lower surfaces of the leaves. In the later stages, the spots are nearly circular, slightly raised, and dark brown, and have irregular edges. TB initially presents as yellowish-green, water-soaked spots that gradually expand into semi-circular, nearly circular, or irregular,

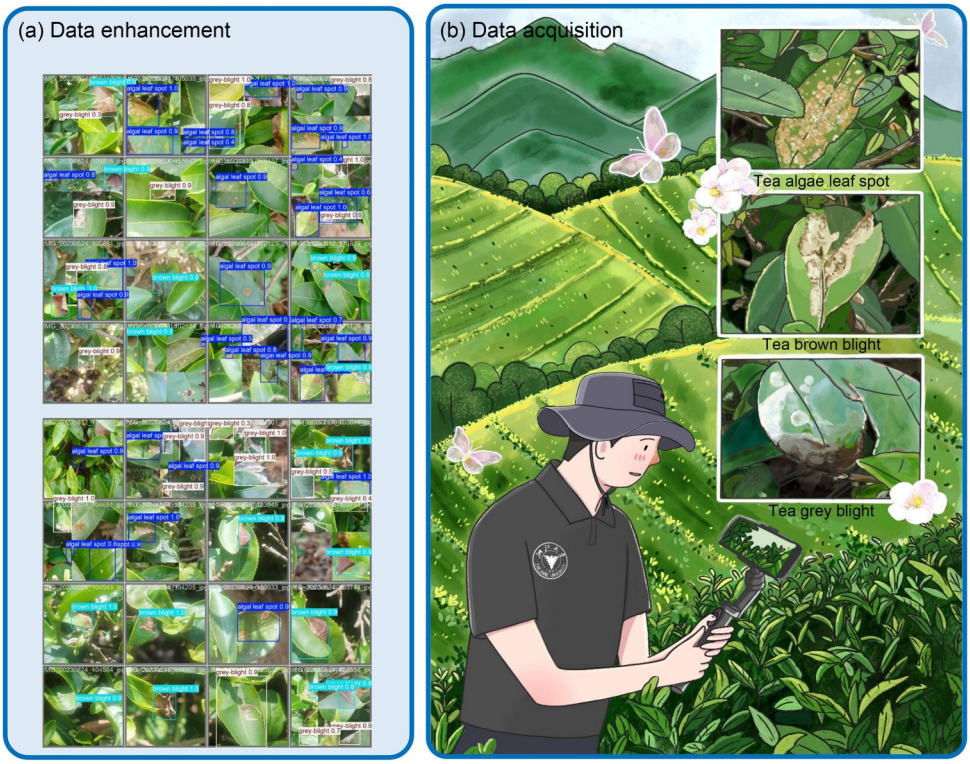

Fig. 1 Schematic representation of data acquisition and enhancement. (a) Dataset after data enhancement. (b) Illustration of the handheld micro-terminal pan-tilt mobile device for collecting data on Longjing tea diseases, and schematic diagrams of the disease characteristics of three types of Longjing tea leaves.

large, brown spots. In the later stages, many small black dots appear on the surface of the spots, scattered or arranged in a slightly circular pattern. TG initially appears as indistinct yellowish-brown small dots, which gradually expand into circular or irregular large spots. The spots develop from brown to grayish-white and have concentric rings alternating between brown and grayish-white. Many small black dots appear on the spots (Wang et al., 2017), as shown in Fig. 1b.

2.2 Object detection algorithm

According to data released by the Ultralytics development team, the YOLO11n model outperforms previously released models across key metrics, including parameters (Params), floating-point operations (FLOPs), and FPS. Therefore, in this study, we focused on improving the YOLO11n model's detection accuracy. In numerous cases of deep learning model improvement and optimization, researchers mainly replace and optimize the neck and backbone parts of the model (Elharrouss et al., 2024). Some studies also replace the more efficient loss functions to enhance the model's learning ability for foreign targets. Zhan et al. (2024) improved the YOLOv8 series model by replacing the Stage Partial Module in the model's backbone with the C2f module and adding the ConvModule. They also used different loss functions for various detection tasks to enhance the model's overall performance. Xu et al. (2024) customized and improved the backbone structure for edge computing based on the YOLOv5 series of models, using the GhostConvV2 and MobileOne block modules, effectively reducing the model's parameter count. Zheng and Yu (2025) improved the YOLOv7-tiny model by introducing the reparameterized multi-scale feature (RMF) module to enhance its feature extraction capability. A distributed network structure, gather and distribute feature pyramid network (GDFPN), was proposed as a backbone to effectively enhance feature fusion. Additionally, a dynamic head module was added to enhance the model's detection performance for small targets. They also introduced the shape-intersection over union (IoU) loss function to improve target recognition accuracy. The improved RG-YOLO model can achieve precise detection of underwater targets. Inspired by those studies and considering the YOLO11n model's performance characteristics, in this study, we enhanced and combined the loss function, backbone, and neck components. A comparison between the original structure diagram of YOLO11n and the overall structure diagram of the ADS-YOLO model is shown in Fig. 2.

2.3 Improved loss function

The initial loss function of YOLO11n is complete intersection over union (CIoU). Common loss functions such as generalized intersection over union (GIoU) and distance intersection over union (DIoU) (Zheng et al., 2020) do not account for the direction between the ground-truth box and the prediction box, resulting in a slow model convergence speed. To address this, we introduced the SIoU (Gu et al., 2023) loss function as a replacement for the model's initial loss function. SIoU inherently includes the vector angle between the ground-truth box and the prediction box and consists of four parts: angle cost, distance cost, shape cost, and IoU cost.

The formula for calculating the angle cost (Λ) is as follows:

$$\Lambda = 1 - 2 \times \sin^2\left(\arcsin\left(\frac{C_h}{\sigma}\right) - \frac{\pi}{4}\right), \tag{1}$$

where C_h is the difference in height between the center points of the ground-truth box and the prediction box, and σ is the distance between the center points of the ground-truth box and the prediction box.

The formula for calculating the distance cost (Δ) is as follows:

$$\Delta = \sum_{t=x,y}\left(1 - \mathrm{e}^{-\gamma\rho_t}\right), \tag{2}$$

where $\rho_x = \left(\frac{b^{\mathrm{gt}}_{c_x} - b_{c_x}}{c_w}\right)^2$ and $\rho_y = \left(\frac{b^{\mathrm{gt}}_{c_y} - b_{c_y}}{c_h}\right)^2$ represent the squared differences between the center coordinates of ground-truth box and prediction box, normalized by the width (c_w) and height (c_h) of the minimum bounding rectangle, respectively; $\left(b^{\mathrm{gt}}_{c_x}, b^{\mathrm{gt}}_{c_y}\right)$ is the center coordinate of the ground-truth box and $\left(b_{c_x}, b_{c_y}\right)$ is the center coordinate of the prediction box; $\gamma = 2 - \Lambda$ represents the adjustment factor for angle loss in the formula.

The formula for calculating the shape cost (Ω) is as follows:

$$\Omega = \sum_{t=w,h}\left(1 - \mathrm{e}^{-\omega_t}\right)^{\theta}. \tag{3}$$

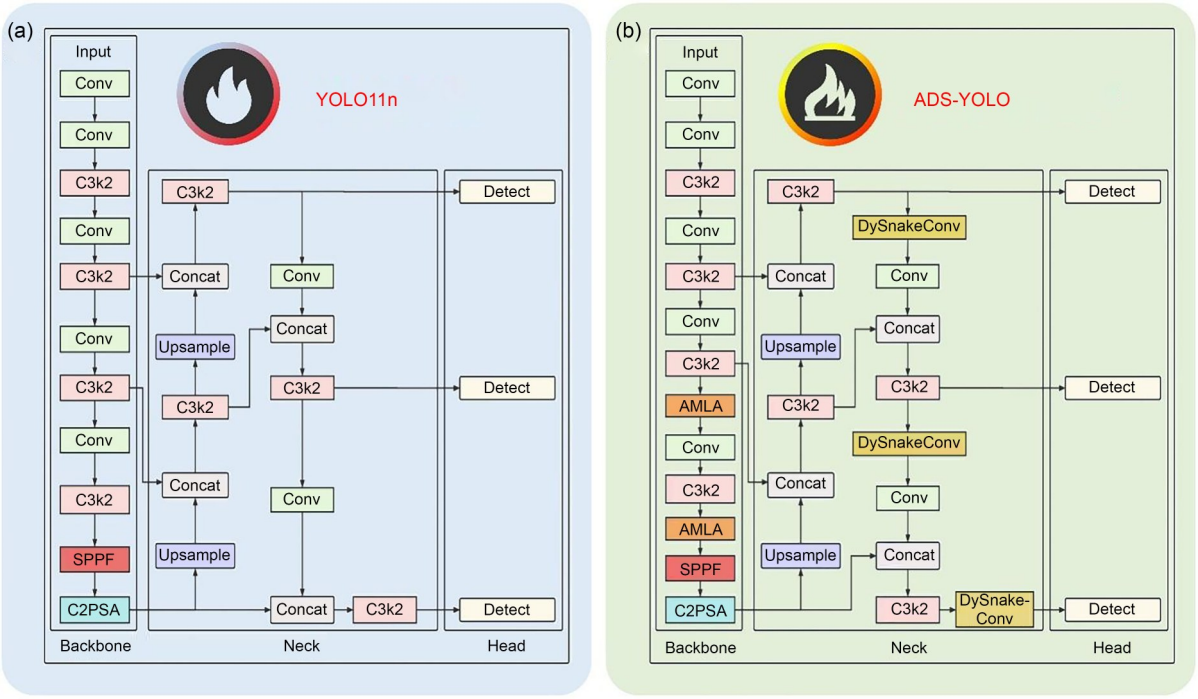

Fig. 2 Comparison of the structure before and after algorithm improvement. (a) YOLO11n structure diagram. (b) The structure diagram of the asymmetric multi-level (AML) mechanism, dynamic snake convolution (DSC), and scalable intersection over union (SIoU) loss function-You Only Look Once (YOLO) (ADS-YOLO) algorithm with the improvement modules added. Conv: convolution; C3k2: cross stage partial with kernel size 2; SPPF: spatial pyramid pooling-fast; C2PSA: cross stage partial with pyramid squeeze attention; AMLA: AML attention; DySnakeConv: dynamic snake convolution.

Let $\omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}$ and $\omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$, where w and h represent the width and height of the prediction box, respectively, w^{gt} and h^{gt} represent the width and height of the ground-truth box, respectively. The parameter θ (with range [2, 6]) is used to control the shape loss.

The formula for calculating the IoU cost is as follows:

$$\mathrm{IoU} = \frac{|b \cap b^{gt}|}{|b \cup b^{gt}|}, \tag{4}$$

where b is the prediction box and b^{gt} is the ground-truth box.

In summary, the definition of the SIoU loss function (L_{SIoU}) is as follows:

$$L_{\mathrm{SIoU}} = 1 - \mathrm{IoU} + \frac{\Delta + \Omega}{2}. \tag{5}$$

In YOLO11n, the SIoU loss function was introduced to address current problems in target detection for tea diseases. The SIoU introduced distance, angle, and shape costs, optimizing bounding box matching across multiple dimensions. Specifically, in the calculation of the SIoU loss function, the center-point distance between the prediction box and the ground-truth box was computed by considering differences in horizontal and vertical distances. Small offsets were used to avoid division-by-zero errors. Then, the introduction of angle cost enabled the prediction box to be more precisely aligned with the ground-truth box, which is particularly crucial in tea disease detection because disease locations are distributed at different angles. The shape cost further optimized position regression, aligning the prediction box's position more closely with the ground-truth box's position, especially in small-target detection tasks.

2.4 Application of asymmetric multi-level attention

The AML attention module enhances the weak representation of tea disease features while suppressing redundant background (e.g., foliage textures). Unlike symmetric attention frameworks, which weigh all feature scales equally, AML uses an asymmetric fusion

strategy. It prioritizes fine-grained low-level features (carrying lesion boundary details) with higher attention weights, while assigning moderate weights to high-level semantic features (capturing global lesion patterns). In addition, a linear attention branch reduces computational overhead, enabling efficient background noise suppression without sacrificing feature richness. This design directly addresses the core challenge in tea disease detection: distinguishing small, low-contrast lesions from complex field backgrounds.

The AML attention module also has strong adaptive feature-fusion capabilities. Through convolution operations at different layers, it aggregates multi-level features. This fusion strategy enables the network to comprehensively consider information at multiple levels, resulting in a stronger feature representation. The feature weighting mechanism in this module introduces a weighting strategy during feature fusion, enabling the network to automatically select the most valuable features for disease identification. In this way, the influence of background noise on model judgment is effectively suppressed. Adaptive feature fusion enables the model to dynamically adjust feature weights in response to changes in input data, thereby enhancing the sensitivity and accuracy of disease detection.

The module also introduces a linear attention mechanism. Through this mechanism, the model can calculate feature similarity. The introduction of query (Q), key (K), and value (V) enables the network to dynamically adjust its output for specific features. Meanwhile, the linear attention computes feature weights using the softmax function, thereby enhancing important features and suppressing irrelevant ones. This mechanism can effectively improve the model's anti-interference capability in complex environments. The features processed by the attention mechanism can be better reorganized, ensuring that the model responds more effectively to essential features, thereby improving the detection accuracy (Fig. 3).

To enhance the network's expressive power, the AML attention module, by integrating convolution and attention mechanisms, can incorporate multiple types of information at different levels, thereby improving the network's understanding of complex image content. To better capture nonlinear features, the network can learn more complex representations through deep nonlinear mapping, thereby making the model more adaptable to diverse disease manifestations. Overall, the AML attention module in the YOLO11n network provides an effective way to integrate multiple feature information, ensuring the comprehensive analysis and recognition of tea plant diseases.

To further improve the detection performance of small targets, the AML attention enhances their recognition (e.g., tiny spots in early-stage diseases) through multi-level convolution and attention mechanisms. This is particularly important for the early detection of tea plant diseases. Meanwhile, through an effective feature retention strategy, the network can maintain high resolution when dealing with small targets, avoiding information loss. In complex backgrounds, small targets are easily overwhelmed by noise. The introduction of the AML attention module enables the network to better separate signals from noise, thereby improving the detection of small disease features. The structure of the AML attention module is shown in Fig. 3.

2.5 Application of dynamic snake convolution

To improve the detection accuracy of tea leaf disease images, we introduced the DSC module into the neck part of the YOLO11n algorithm (Wang X et al., 2025). The DSC module features dynamic deformable

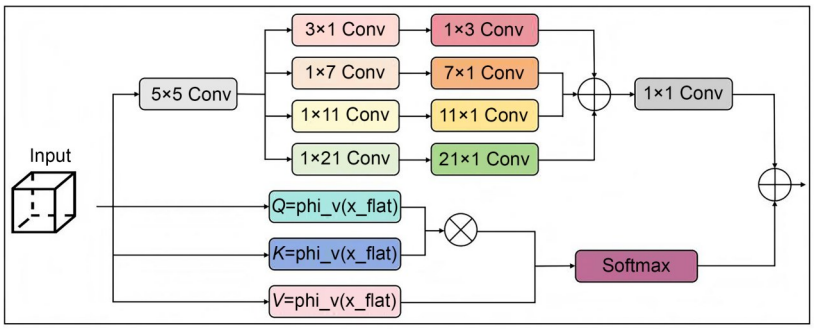

Fig. 3 Asymmetric multi-level (AML) attention module structure, convolution (Conv) distribution, and logical structure diagram. *Q*: query; *K*: key; *V*: value.

convolution, which automatically adjusts the shape and position of the convolution kernel according to local changes in the input image features. This adaptive ability enables the model to capture the features of different diseases more flexibly, thereby enhancing its adaptability to complex backgrounds and shape variations.

By performing bilinear interpolation on the feature map, this module can more accurately extract detailed information from the image. Compared to traditional convolutional methods, this module can extract features at different scales and orientations, especially for small disease spots and irregularly shaped leaf structures, thereby significantly improving the ability to express features. This detailed feature extraction is crucial for early disease identification.

Tea diseases may exhibit significant differences in their manifestations at various growth stages and under diverse environmental conditions. The DSC module can dynamically adjust the convolution kernel deformation based on the specific morphology of the disease in the input image, thereby enhancing the detection of various diseases (such as TA, TB, and TG). This flexibility makes the model more robust and accurate in dealing with diverse disease samples.

In complex natural environments, background noise often affects the accuracy of disease detection. The DSC module can more effectively focus on target disease features and suppress interference from background information by adaptively changing the convolution kernel shape. This characteristic enhances the model's detection accuracy in complex background scenarios.

After introducing the DSC module, the YOLO11n algorithm can learn more representative features, enabling the model to maintain a high recognition rate when presented with new, unseen disease samples. This enhanced generalization ability is the key to improving the overall detection performance. The DSC structure is shown in Fig. 4.

2.6 Training environment and evaluation indicators

In this study, we used a predefined image dataset on a Windows 10 Pro (64-bit) operating system. The computer was equipped with an Intel Core i9-10900 CPU@2.80 GHz, with 128 GB of memory and a 1 TB main hard drive. The graphics card was an NVIDIA Quadro RTX 5000. The versions used for training were PyTorch v.2.0.1, Python v.3.9.19, and CUDA v.11.7.

In the comparative experiments, the same hyperparameters were used across the models. Specifically, the input image size for training was 640×640 pixels, the initial learning rate was 0.01, the momentum was 0.937, the batch size was 16, and the maximum number of iterations/epochs was 400.

We trained the dataset with different models and compared their performance. The main evaluation metrics for overall performance were: (1) mAP, which reflects mainly the average value of the overall accuracy of each model; (2) precision (P), which is the ratio of the number of correctly detected targets to the total number of detected targets; and (3) recall (R), which is the ratio of the number of correctly detected sets to the total number of sets. The formulas for calculating each evaluation metric were as follows:

$$P=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}}, \tag{6}$$

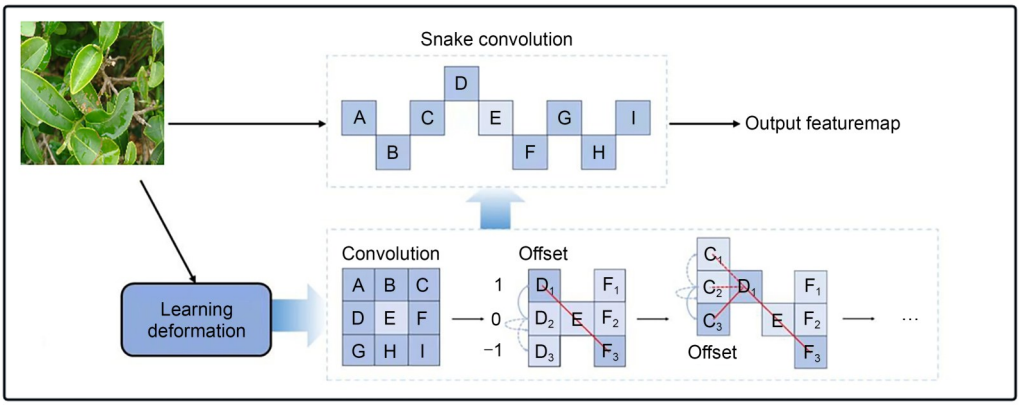

Fig. 4 Dynamic snake convolution (DSC) module structure and transformation of the convolution operation logic mechanism into a snake-shaped distribution.

$$R=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}}, \tag{7}$$

$$\mathrm{AP}=\int_0^1 P(R)\,\mathrm{d}R, \tag{8}$$

$$\mathrm{mAP}=\frac{1}{n}\sum_{k=1}^{n}\mathrm{AP}_k, \tag{9}$$

where average precision (AP) represents the integral of precision and recall, N_{TP} (TP: true positive) indicates the number of tea diseases that were correctly detected, N_{FP} (FP: false positive) represents the number of tea diseases that were wrongly detected, and N_{FN} (FN: false negative) indicates the number of tea diseases that were not detected.

2.7 Experimental plan

In the experimental part of this study, two main types of experiments were conducted. The first type was an ablation experiment to enhance the YOLO11n model, aiming to compare and verify the improvement effects of different modules on the model's overall performance and select the best combination of improved modules (Lawal et al., 2021). The second type was a comparison between the enhanced model and other models, aiming to compare differences in performance among models and to more effectively and intuitively observe the improvement in each performance indicator of YOLO11n.

3 Results and discussion

3.1 Ablation experiment

We conducted ablation experiments on the initial model by adding or removing improvement modules to explore the model's overall performance. The ablation experiment results of the combination of the DSC module, AML attention module, and SIoU loss function are shown in Table 1. Each improvement contributed to the model's overall performance. When the SIoU loss function was introduced alone, the mAP@0.5 of the model increased from 0.894 to 0.912, and its mAP@0.5:0.95 increased from 0.741 to 0.775; when the DSC module was introduced alone, the mAP@0.5 of the model increased from 0.894 to 0.919, and its mAP@0.5:0.95 increased from 0.741 to 0.803; when the self-developed AML attention module was introduced alone, the mAP@0.5 of the model increased from 0.894 to 0.930, and its mAP@0.5:0.95 increased from 0.741 to 0.809. When the SIoU loss function and the DSC module were introduced simultaneously, the mAP@0.5 of the model increased from 0.894 to 0.937, and its mAP@0.5:0.95 increased from 0.741 to 0.824; when the DSC module and the self-developed AML attention module were introduced simultaneously, the mAP@0.5 of the model increased from 0.894 to 0.940, and its mAP@0.5:0.95 increased from 0.741 to 0.814; the model performed best when all three modifications were introduced simultaneously, with the mAP@0.5 increasing from 0.894 to 0.947 and its mAP@0.5:0.95 increasing from 0.741 to 0.832. In the ablation experiments, the precision and recall values of the improved model also increased accordingly. The ablation experiment results of this study indicate that the SIoU loss function, DSC module, and self-developed AML attention module have a significant synergistic effect on improving the YOLO11n model's target detection performance and adaptability to complex environments.

The integration of SIoU loss plays a crucial role in optimizing the alignment of bounding boxes with irregular tea-disease lesions, effectively reducing localization errors, and enhancing precision (Gu et al., 2023). Additionally, the DSC module uses an adaptive kernel to focus on the complex boundaries of

Table 1 Ablation experiment

SIoU	DSC	AML attention	Precision	Recall	mAP@0.5	mAP@0.5:0.95
×	×	×	0.894	0.818	0.894	0.741
√	×	×	0.920	0.830	0.912	0.775
×	√	×	0.917	0.858	0.919	0.803
×	×	√	0.936	0.847	0.930	0.809
√	√	×	0.940	0.857	0.937	0.824
×	√	√	0.925	0.881	0.940	0.814
√	√	√	0.935	0.870	0.947	0.832

SIoU: scalable intersection over union; DSC: dynamic snake convolution; AML: asymmetric multi-level; mAP: mean average precision.

deformable lesions, thereby improving target extraction and increasing recall. Moreover, the AML attention mechanism contributes significantly by fusing multi-scale features while suppressing noise from foliage, thereby amplifying weak signals associated with lesions. This combination of features explains the impressive improvements in both accuracy and processing speed.

This multi-dimensional optimization, integrating the SIoU loss function, AML attention module, and DSC module, not only improved YOLO11n's detection accuracy for tea diseases but also enhanced its robustness against variations in lesion shape, size, and angle. Specifically, replacing CIoU with the SIoU loss function introduced angle, distance, and shape costs into bounding box regression, enabling more precise alignment with irregular tea lesions. This is particularly critical for small, scattered disease spots, thereby boosting detection accuracy.

The incorporation of the AML attention module further endowed the model with significant advantages in complex field scenarios. Through enhanced feature extraction, adaptive multi-scale feature fusion, and the introduction of linear attention, the module effectively suppresses background noise, such as overlapping foliage, while amplifying weak lesion signals, thereby improving network expressiveness and small-target detection performance. These enhancements collectively render the model more accurate and reliable for identifying complex tea diseases.

Additionally, the introduction of the DSC module into the neck part of YOLO11n enhanced the model's adaptability to diverse environmental conditions. By dynamically adjusting the convolution kernel shape and position based on local feature variations, the DSC module captures fine-grained details of irregular lesions more flexibly, reducing interference from complex backgrounds and improving generalization. Together, these three modules synergistically address core challenges in tea disease detection: imprecise localization, background clutter, and lesion irregularity. They provide strong technical support for real-time, in-field tea disease monitoring and management.

3.2 Performance comparison of ADS-YOLO

To comprehensively verify the detection performance of the proposed ADS-YOLO model, two sets of comparative experiments were conducted: first, a comparison with the baseline YOLO11n to validate the effectiveness of the integrated improvement modules; second, a comparison with other mainstream lightweight YOLO models to demonstrate its competitiveness in tea disease detection.

3.2.1 Comparison with baseline YOLO11n

To confirm the improvement in performance from using the integrated SIoU loss function, DSC module, and AML attention module, a comparative experiment was conducted between ADS-YOLO and the original YOLO11n on the same dataset and in the same training environment.

The precision increased from 0.894 using the initial YOLO11n to 0.935 using ADS-YOLO, an overall improvement of 0.041 (Table 2). This indicates that ADS-YOLO achieves higher precision. Compared to the initial YOLO11n, ADS-YOLO also improved recall from 0.818 to 0.870, an overall increase of 0.052. Zhu et al. (2025) improved the YOLOv8 model by incorporating an attention module and expanding the convolutional layers, and then used it to identify grape leaf black rot spots. The results showed that the precision of the improved model was 92.64% and the recall was 93.28%. Compared with the initial YOLOv8 model, these values increased by 2.38 and 1.19 percentage points, respectively. In comparison, our ADS-YOLO model has advantages in both indicators, with higher final values and greater improvement. This further shows that compared with the initial model, ADS-YOLO has stronger detection performance with fewer missed detections and the ability to capture more target individuals.

The mAP@0.5 and mAP@0.5:0.95 values of ADS-YOLO were significantly improved compared to those of the initial YOLO11n model. Table 2 shows that mAP@0.5 increased from 0.894 to 0.947, and mAP@0.5:0.95 from 0.741 to 0.832, indicating a

Table 2 Experimental comparison between ADS-YOLO and YOLO11n

Module	Size (pixels)	Params (M)	Precision	Recall	mAP@0.5	mAP@0.5:0.95	FPS	GFLOPs
YOLO11n	640×640	2.58	0.894	0.818	0.894	0.741	102.6	6.3
ADS-YOLO	640×640	2.08	0.935	0.870	0.947	0.832	137.1	5.2

Params: parameter count; M: million; mAP: mean average precision; FPS: frames per second; GFLOPs: giga floating-point operations.

remarkable improvement. Chutichaimaytar et al. (2025) improved the YOLOv8 model by introducing the multi-scale sequence feature fusion (MSFF) module to enhance the original algorithm, and then used it to identify coriander leaf tip-burn and powdery mildew. The improved coriander tip-burn YOLO (CTB-YOLO) model achieved an mAP@0.5 of only 0.73 and an mAP@0.5:0.95 of only 0.382. Coriander shares similar external features with tea leaves, and the target diseases also have strong similarities. Moreover, its data collection was conducted under stable indoor environmental conditions. Compared with CTB-YOLO, our ADS-YOLO model achieved superior performance across the evaluation metrics. At the same time, the difference between mAP@0.5 and mAP@0.5:0.95 for ADS-YOLO was smaller, indicating that the improved model has higher accuracy in target detection under complex natural lighting conditions and stronger comprehensive performance.

The frame rate of ADS-YOLO was 137.1 FPS, which was 34.5 FPS higher than that of YOLO11n (102.6 FPS), demonstrating the improvement in real-time performance of the ADS-YOLO model. The computational cost (GFLOPs) of ADS-YOLO was lower than that of YOLO11n, indicating that the computational complexity of the improved model decreased. Lin et al. (2025) improved the YOLO11 model by applying the CARConv rotational convolution module and the AFGCAM attention mechanism, thereby accurately identifying target crop apples during the harvesting process. The mAP@0.5 of the improved YOLO11-ARAF reached 0.923, the precision reached 0.894, and the mAP@0.5:0.95 reached 0.644. However, its computational cost was 7.3 GFLOPs. Overall, ADS-YOLO has certain advantages across various evaluation metrics when built on YOLO11n, especially in terms of the indicator of computational cost, which was 5.2 GFLOPs for ADS-YOLO, confirming less computational complexity. The lightweight design and lower computational complexity enable ADS-YOLO to more efficiently identify the most significant disease features within the detection interface during tea disease recognition, accurately identifying and delineating the location and extent of diseased areas.

Moreover, compared with the original YOLO11n model, ADS-YOLO provided more accurate discrimination among the three types of tea diseases in this study. When the detection results across the entire dataset were examined, a few misjudgments occurred, and the main lesion could not always be identified. To make the comparison of disease detection results more obvious, we not only presented disease identification results in graphs but also displayed target identification results as heatmaps.

The detection results of the two models for the three diseases are presented in Figs. 5–7. The detection outputs of YOLO11n and ADS-YOLO across the three tea diseases are compared in Fig. 5. Both models achieved accurate recognition of target lesions: YOLO11n effectively identified disease areas, while ADS-YOLO further refined the bounding box alignment, fully enclosing scattered TA patches, more comprehensively identifying pathological regions whose colors closely matched the background, and tightly fitting the irregular edges of TG lesions. This figure shows that both models delivered reliable detection performance, with ADS-YOLO providing marginally improved localization precision for tea disease lesions.

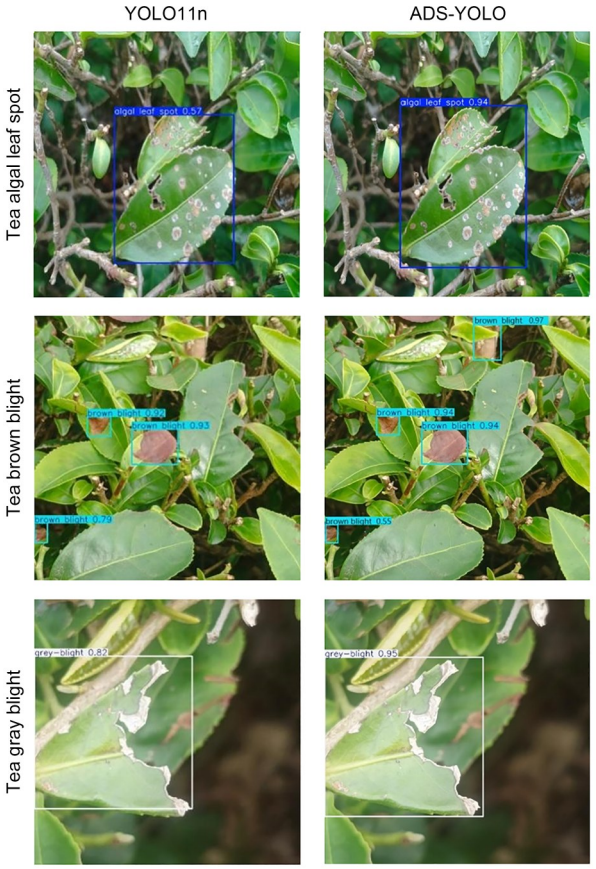

Fig. 5 Test results for YOLO11n and ADS-YOLO. The recognition performance of ADS-YOLO was more comprehensive.

Fig. 6 shows heatmaps for the three tea diseases, paired with the original images, YOLO11n heatmaps, and ADS-YOLO heatmaps. YOLO11n heatmaps show scattered attention, with a weak focus on lesions and overlapping with background foliage. ADS-YOLO heatmaps concentrate sharply on lesion regions: they highlight TA patches, target TB's necrotic edges, and amplify faint TG lesions. This figure visualizes how ADS-YOLO's AML attention module suppresses background noise while enhancing weak lesion features, directly demonstrating its improved ability to distinguish tea diseases from complex field backgrounds.

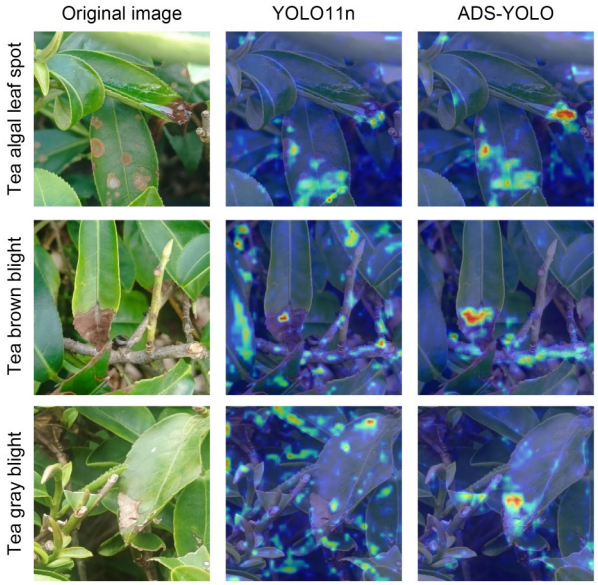

Fig. 6 Comparison of heatmaps of feature visualizations before and after model improvement.

Confusion matrices for YOLO11n (Fig. 7a) and ADS-YOLO (Fig. 7b) quantify the classification performance across the three tea diseases and the background class. YOLO11n shows non-negligible errors: 13.3% of the TA cases were misidentified as background, and TB–TG cross-category confusion occurred. ADS-YOLO reduced these issues: TA's background misclassification dropped to 10.1%, TB–TG confusion was eliminated, and class-specific accuracy increased (e.g., TB accuracy increased from 87.3% to 92.2%). This figure confirms that ADS-YOLO's modules (notably AML attention) enhance feature discrimination, boosting precision in tea disease classification and reducing background interference.

In summary, ADS-YOLO shows substantial advantages across all evaluation indices compared to the initial YOLO11n, demonstrating that the improved model performs better under complex natural-light conditions. To further interpret the model's performance, we analyzed the confusion matrix of ADS-YOLO (Fig. 7b) and typical failure cases, focusing on misclassification patterns, challenging scenarios, and module-specific improvements. From the confusion matrix, two main misclassification trends emerged, driven by lesion feature similarity. First, mid-stage TA was most prone to being misclassified as TB, as both diseases share core symptomatic traits of dark brown coloration and irregular lesion shapes during this period. Second, late-stage TG was frequently misclassified as TB because the brown concentric rings on TG lesions closely resemble TB's typical gray-brown alternating concentric rings in both color and structure. Furthermore,

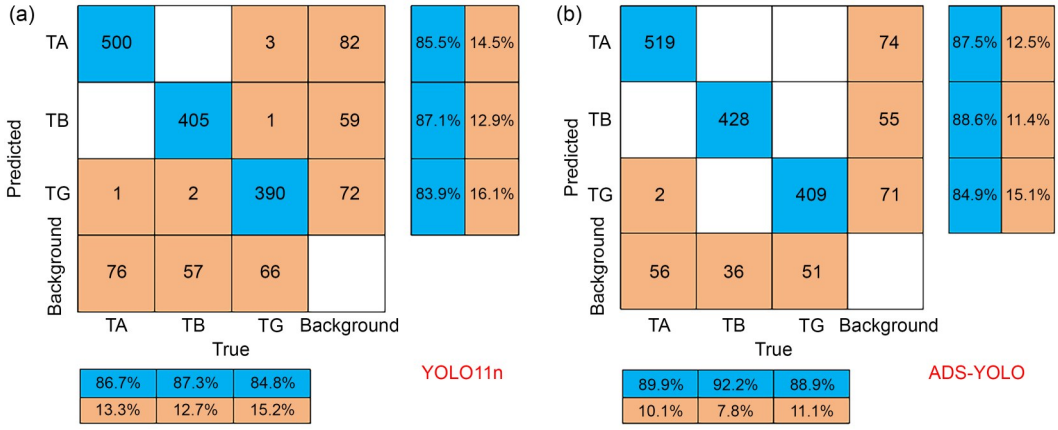

Fig. 7 Comparison of confusion matrices between YOLO11n and ADS-YOLO. (a) Confusion matrices based on YOLO11n. (b) Confusion matrices based on ADS-YOLO. The blue color indicates samples where the predicted category matches the actual category, and the orange color indicates those where it does not. TA: tea algal leaf spot; TB: tea brown blight; TG: tea gray blight.

early-stage TB exhibits only faint discoloration without developing mature lesion structures, so it was occasionally misclassified as TA or TG. This is because its initial symptoms are weakly differentiated from the early water-soaked, light-green spots of TA and the initial dark-green lesions of TG, with minimal visual distinction for model recognition.

3.2.2 Comparison with other lightweight YOLO models

To further evaluate the competitiveness of ADS-YOLO in practical applications, comparative experiments were conducted with eight mainstream lightweight YOLO models (YOLOv3-tiny, YOLOv5n, YOLOv6n, YOLOv7t, YOLOv8n, YOLOv9t, YOLOv10n, and YOLO11n) and other improved YOLO variants (DM-YOLO, LCDDN-YOLO-CWD, and BHC-YOLOv8) dedicated to plant disease detection.

Among the eight mainstream LMs, YOLOv8n and ADS-YOLO could perform relatively complete and comprehensive bounding box recognition of TA lesion points within the detection interface. The comparison of TB detection results showed that all LMs gave precise detection, likely because large, brownish lesions characterize TB. The observation results also showed that ADS-YOLO has a more comprehensive ability to detect the main lesion target features. When TG was detected, the performance of each model varied. YOLOv5n and YOLOv10n had issues with missed detection of lesion areas, while YOLOv6n struggled to distinguish target lesions. YOLOv3-tiny and ADS-YOLO performed relatively well in the detection of TG lesion areas, and the detection boxes could precisely and comprehensively identify the target areas. Images of the relevant detection results are compared in Fig. 8.

The evaluation results for each model trained in the same training environment and on the same dataset are shown in Table 3. ADS-YOLO outperformed the other models in terms of FPS, indicating the strongest real-time performance. Additionally, ADS-YOLO had the lowest GFLOPs among the models, suggesting that it has lower computational complexity. It achieved consistent, cross-family gains over lightweight YOLO baselines, resulting in fewer false alarms and fewer misses in field clutter, while improving localization at strict IoU thresholds. ADS-YOLO showed improvements of 0.027 in precision, 0.047 in recall, and 0.044 in mAP@0.5 compared with YOLOv3-tiny, and 0.045 in precision, 0.078 in recall, and 0.064 in mAP@0.5 compared with YOLOv5n, indicating clear improvements in both detection and localization quality. The largest deltas at mAP@0.5:0.95 indicate better box alignment on small, irregular lesions across IoU thresholds, tackling challenging field imagery (Table 3). Wang RJ et al. (2025) used the multi-scale edge enhancement (MSEE) module, slim shared convolutional head (SSCH), and high-level screening feature pyramid network (HSFPN) to improve the YOLO11n model (YOLO11-PGM) for the precise identification of pomegranates and their flowers. The mAP@0.5 reached 0.94, and the computational cost was 4.8 GFLOPs, like those of ADS-YOLO. They believed that the differences were due to YOLO11-PGM not improving or introducing new loss functions, and to the improved modules placing greater emphasis on light weighting and reducing computational complexity. While competitors flag dense TA lesions, ADS-YOLO better encloses primary foci and their boundaries, resulting in higher-quality detections. These results show that the SIoU loss, DSC, and AML attention enhance the sensitivity and precision in tea canopies.

Compared with some of the best YOLO models, including DM-YOLO based on YOLOv9 (Abulizi et al., 2025), LCDDN-YOLO (Feng et al., 2025), YOLO-CWD (Ma et al., 2025), and BHC-YOLOv8 (Zhan et al., 2024), all of which improved upon YOLOv8, our proposed ADS-YOLO achieves higher detection accuracy and stronger real-time performance.

BHC-YOLOv8 (Zhan et al., 2024) optimizes small-target detection for tea diseases using improved downsampling and PAN-FPN techniques but relies solely on the standard CIoU loss, which measures bounding box overlap and neglects irregular lesion shapes. ADS-YOLO integrates a geometric-aware SIoU loss that incorporates angle, distance, and shape costs, effectively addressing these complexities. While DM-YOLO achieves lightweight deployment through dynamic channel pruning, it struggles with adaptive feature extraction, particularly for blurred lesion boundaries. ADS-YOLO uses deformable sampling via its DSC module, enabling better adjustment of the convolutional kernels to capture challenging features (Abulizi et al., 2025). LCDDN-YOLO enhances multi-scale feature fusion via pyramid networks but relies on fixed convolutional kernels, which may fail to account for lesion variability (Feng et al., 2025). ADS-YOLO,

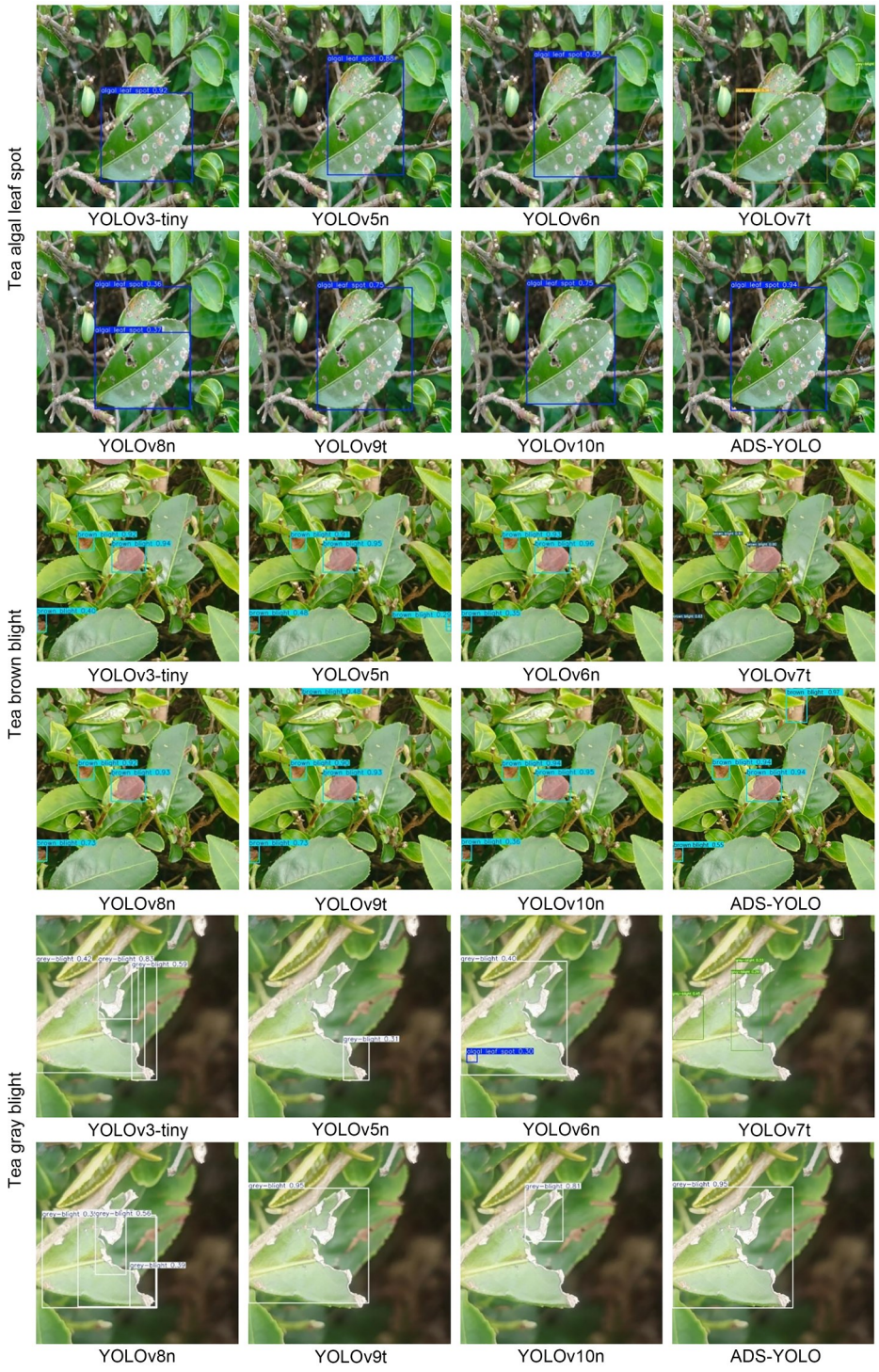

Fig. 8 Qualitative comparison of detection results across eight tested models for three tea disease categories in a matrix layout, illustrating differences in lesion coverage, bounding box tightness, and stability under realistic field backgrounds. In all three disease groups, the baseline models show clear qualitative differences, including incomplete lesion coverage, occasional missed detections, fragmented boxes over discontinuous lesion regions, and looser localization when lesion boundaries are irregular or contrast is reduced. By comparison, ADS-YOLO consistently improves detection completeness and spatial precision, producing tighter boxes that better align with true lesion extents and reducing false negatives in challenging scenes such as cluttered foliage and fluctuating illumination. These visual results support the quantitative findings by highlighting stronger coverage, more accurate localization, and enhanced robustness across diverse field environments of ADS-YOLO.

Table 3 Comparison experiment with additional models

Module	Size (pixels)	Params (M)	Precision	Recall	mAP@0.5	mAP@0.5:0.95	FPS	GFLOPs
YOLOv3-tiny	640×640	12.10	0.908	0.823	0.903	0.751	42.6	18.9
YOLOv5n	640×640	2.50	0.890	0.792	0.883	0.719	98.5	7.1
YOLOv6n	640×640	4.23	0.898	0.802	0.891	0.734	55.1	11.8
YOLOv7t	640×640	6.02	0.664	0.675	0.694	0.386	68.6	13.2
YOLOv8n	640×640	3.01	0.905	0.798	0.892	0.755	75.8	8.1
YOLOv9t	640×640	1.97	0.906	0.795	0.892	0.740	91.4	7.6
YOLOv10n	640×640	2.70	0.889	0.766	0.870	0.723	79.1	8.2
YOLO11n	640×640	2.58	0.894	0.818	0.894	0.741	102.6	6.3
ADS-YOLO	640×640	2.08	0.935	0.870	0.947	0.832	137.1	5.2

Params: parameter count; M: million; mAP: mean average precision; FPS: frames per second; GFLOPs: giga floating-point operations.

however, enriches fusion through a multi-level attention mechanism (the AML module) that suppresses background noise and amplifies weak lesion features. Additionally, unlike YOLO-CWD, which addresses cross-domain adaptation without tailoring its approach to tea disease detection, ADS-YOLO combines geometric-aware loss, deformable sampling, and multi-level attention in a cohesive framework, offering significant advancements in tackling the unique challenges of tea disease detection (Ma et al., 2025).

These results directly show the advantages of ADS-YOLO across all aspects and demonstrate its great potential for practical detection applications in complex natural-light environments.

3.3 Mechanistic principle of the ADS-YOLO model application

In the current study, ADS-YOLO improved small, low-contrast tea lesion recognition by combining three mechanisms on a YOLO11n scaffold: geometric reparameterization at the loss level, adaptive sampling at the operator level, and global-context gating at the attention level, which reshaped the optimization landscape and receptive field (Dong and Duoqian, 2023; Parashar et al., 2025). Replacing CIoU with SIoU introduces directionality into box regression, incorporating angle-, distance-, and shape-aware penalties that reduce the degrees of freedom in search and stabilize gradients for small or overlapping lesions, thereby accelerating convergence and improving localization. Inserting DSC enables deformable sampling along lesion boundaries, with learned offsets guiding sampling points toward curvilinear pathology and embedded bilinear interpolation preserving high-frequency cues. The AML attention block reweights spatial–channel responses using long-range correlations, enabling robust discrimination between lesions from overlapping foliage and artifacts while maintaining throughput. These components act synergistically: SIoU improves regression with cleaner gradients; DSC expands receptive fields along lesion morphology; and AML suppresses distractors, improving metrics (precision=0.935, recall=0.870, mAP@0.5=0.947, mAP@0.5:0.95=0.832) and real-time performance (137.1 FPS at 5.2 GFLOPs). These gains show that precise geometric objectives, deformable sampling, and global context, combined with a multi-scale backbone, enhance the detection of small targets in complex scenes.

Despite the encouraging accuracy and efficiency achieved by ADS-YOLO, several limitations should be acknowledged to properly assess the scope and generalizability of the proposed method. The current system is limited to RGB imagery, which might fail to capture subtle chromatic differences and low-contrast lesions against complex foliage. Under such conditions, distinguishing true disease regions from artifacts like shadows and dust may be challenging. Furthermore, the dataset used was constrained by regional, seasonal, and cultivation variations, limiting its representativeness across diverse global tea-producing areas. Although natural field imagery was used in the experiments, comprehensive validation under variable conditions, such as backlighting, heavy shadows, and motion blur, is still needed. Therefore, future research will focus on expanding training datasets to include these challenging scenarios and on implementing robust domain-generalization strategies to enhance reliability in uncontrolled environments.

Future research will also explore multimodal sensing to overcome the constraints imposed by RGB-only sensing. Integrating complementary modalities such as hyperspectral, multispectral, thermal, or depth information may enhance sensitivity to early or ambiguous symptoms and reduce confusion caused by variations in illumination. In addition, although the proposed design is lightweight, further optimization for practical mobile deployment is needed, including model compression, quantization, hardware-aware acceleration, and energy-aware scheduling on edge devices. Together, multi-season multi-region data expansion, multimodal fusion, and deployment-oriented optimization are expected to improve both the generalization and real-world usability of ADS-YOLO for tea disease monitoring.

4 Conclusions

In this study, we constructed a dataset of leaf images of three tea diseases to address the challenges of detecting those diseases in complex plantation environments. We proposed an improved LM, ADS-YOLO, integrating an SIoU loss function, DSC module, and AML module. The SIoU loss function enhances the alignment between the predicted and ground-truth bounding boxes to improve small-target detection; the DSC module enables adaptive adjustment of the convolutional kernel sizes for irregular lesion feature capture; and the AML module optimizes feature extraction via asymmetric convolution and adaptive fusion. Ablation experiments validated the synergistic effect of these components, and comparative experiments confirmed ADS-YOLO's superior performance over the baseline YOLO11n and other mainstream LMs. The results showed that the model achieved a detection precision of 0.935, an mAP@0.5 of 0.947, an inference speed of 137.1 FPS, and a computational cost of 5.2 GFLOPs, verifying its efficacy for real-time, non-destructive field monitoring. Future work will extend the model to broader crop species and complex agricultural scenarios. In summary, ADS-YOLO balances high accuracy and real-time efficiency, providing a reliable technical solution for tea disease management and a referable framework for lightweight deep learning applications in precision agriculture.

Data availability statement

The datasets generated during the current study are available from the corresponding author upon reasonable request. The tea disease image data are available from specific tea garden but restrictions apply to the availability of these data, which were used under license for the current study and are not publicly available.

Acknowledgments

This work was supported by the Key R&D Projects in Zhejiang Province (Nos. 2023C02009, 2023C02043, and 2022C02044), the National Natural Science Foundation of China (No. 32171889), and the Earmarked Fund for China Agriculture Research System (No. CARS-19-02A). We thank the College of Biosystems Engineering and Food Science at Zhejiang University for providing all the necessary hardware and equipment support throughout the entire research process.

Author contributions

Conceptualization: Jinxian TAO, Xiaoli LI, and Jingfei ZHANG; Methodology: Jinxian TAO, Muhammad SHOAIB, Ibrar AHMAD, Yong HE, Yujie WANG, and Mostafa GOUDA; Software: Jinxian TAO, Xiaoli LI, Binhui LIAO, and Sitan YE; Validation: Muhammad SHOAIB, Muhammad Adnan ISLAM, Ibrar AHMAD, Yujie WANG, and Mostafa GOUDA; Formal analysis: Jinxian TAO, Xiaoli LI, Jingfei ZHANG, and Mostafa GOUDA; Investigation: Jinxian TAO, Yujie WANG, and Sitan YE; Resources: Xiaoli LI and Yong HE; Data curation: Jinxian TAO and Sitan YE; Writing–original draft preparation: Jinxian TAO, Xiaoli LI, Yujie WANG, and Yong HE; Writing–review and editing: Xiaoli LI, Muhammad SHOAIB, Muhammad Adnan ISLAM, Ibrar AHMAD, Yong HE, Sitan YE, Binhui LIAO, and Mostafa GOUDA; Visualization: Muhammad Adnan ISLAM and Mostafa GOUDA; Supervision: Xiaoli LI, Yong HE, and Mostafa GOUDA; Funding acquisition: Xiaoli LI and Mostafa GOUDA. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

Compliance with ethics guidelines

Jinxian TAO, Xiaoli LI, Jingfei ZHANG, Muhammad SHOAIB, Muhammad Adnan ISLAM, Ibrar AHMAD, Yong HE, Sitan YE, Yujie WANG, Binhui LIAO, and Mostafa GOUDA declare that they have no conflicts of interest.

This article does not contain any studies with human or animal subjects performed by any of the authors.

Declaration on the use of generative AI tools

The authors declare that no generative artificial intelligence (AI) tools or AI-assisted technologies were used in any part of the writing, drafting, editing, or revision process of this manuscript. All content, including text, data analysis, discussions, and

conclusions, is the original work of the authors. The authors take full responsibility for the authenticity, accuracy, and academic integrity of the entire manuscript.

References

Abulizi A, Ye JX, Abudukelimu H, et al., 2025. DM-YOLO: improved YOLOv9 model for tomato leaf disease detection. *Front Plant Sci*, 15:1473928. https://doi.org/10.3389/fpls.2024.1473928

An QL, Wang K, Li ZY, et al., 2022. Real-time monitoring method of strawberry fruit growth state based on YOLO improved model. *IEEE Access*, 10:124363-124372. https://doi.org/10.1109/ACCESS.2022.3220234

Ariyawansa HA, Tsai I, Wang JY, et al., 2021. Molecular phylogenetic diversity and biological characterization of *Diaporthe* species associated with leaf spots of *Camellia sinensis* in Taiwan. *Plants*, 10(7):1434. https://doi.org/10.3390/plants10071434

Bao WX, Fan T, Hu GS, et al., 2022. Detection and identification of tea leaf diseases based on AX-RetinaNet. *Sci Rep*, 12:2183. https://doi.org/10.1038/s41598-022-06181-z

Chen HB, Wang RJ, Du JM, et al., 2023. Feature refinement method based on the two-stage detection framework for similar pest detection in the field. *Insects*, 14(10):819. https://doi.org/10.3390/insects14100819

Chutichaimaytar P, Zongqi Z, Kaewtrakulpong K, et al., 2025. An improved small object detection CTB-YOLO model for early detection of tip-burn and powdery mildew symptoms in coriander (*Coriandrum sativum*) for indoor environment using an edge device. *Smart Agric Technol*, 12: 101142. https://doi.org/10.1016/j.atech.2025.101142

Dai M, Dorjoy MMH, Miao H, et al., 2023. A new pest detection method based on improved YOLOv5m. *Insects*, 14:54. https://doi.org/10.3390/insects14010054

Dhanya VG, Subeesh A, Kushwaha NL, et al., 2022. Deep learning based computer vision approaches for smart agricultural applications. *Artif Intell Agric*, 6:211-229. https://doi.org/10.1016/j.aiia.2022.09.007

Dong C, Duoqian M, 2023. Control distance IoU and control distance IoU loss for better bounding box regression. *Pattern Recognit*, 137:109256. https://doi.org/10.1016/j.patcog.2022.109256

Elharrouss O, Akbari Y, Almadeed N, et al., 2024. Backbones-review: feature extractor networks for deep learning and deep reinforcement learning approaches in computer vision. *Comput Sci Rev*, 53:100645. https://doi.org/10.1016/j.cosrev.2024.100645

Feng HR, Chen XQ, Duan ZY, 2025. LCDDN-YOLO: lightweight cotton disease detection in natural environment, based on improved YOLOv8. *Agriculture*, 15(4):421. https://doi.org/10.3390/agriculture15040421

Gu ZC, Zhu K, You ST, 2023. YOLO-SSFS: a method combining SPD-Conv/STDL/IM-FPN/SIoU for outdoor small target vehicle detection. *Electronics*, 12(18):3744. https://doi.org/10.3390/electronics12183744

Huang ZX, Gouda M, Ye ST, et al., 2024. Advanced deep learning algorithm for instant discriminating of tea leave stress symptoms by smartphone-based detection. *Plant Physiol Biochem*, 212:108769. https://doi.org/10.1016/j.plaphy.2024.108769

Iqbal MS, El-Ashram S, Hussain S, et al., 2019. Efficient cell classification of mitochondrial images by using deep learning. *J Opt*, 48:113-122. https://doi.org/10.1007/s12596-018-0508-4

Lawal OM, Huamin Z, Fan Z, 2021. Ablation studies on YOLOFruit detection algorithm for fruit harvesting robot using deep learning. *IOP Conf Ser Earth Environ Sci*, 922:012001. https://doi.org/10.1088/1755-1315/922/1/012001

Li XT, Zhang TH, Yu M, et al., 2025. A YOLOv8-based method for detecting tea disease in natural environments. *Agron J*, 117(2):e70043. https://doi.org/10.1002/agj2.70043

Lin YT, Xia YJ, Xia PC, et al., 2025. YOLO11-ARAF: an accurate and lightweight method for apple detection in real-world complex orchard environments. *Agriculture*, 15(10): 1104. https://doi.org/10.3390/agriculture15101104

Liu Y, Liu HM, Xu WH, et al., 2024. Advances and challenges of carbon storage estimation in tea plantation. *Ecol Inf*, 81:102616. https://doi.org/10.1016/j.ecoinf.2024.102616

Lu JQ, Luo HX, Yu CR, et al., 2024. Tea bud DG: a lightweight tea bud detection model based on dynamic detection head and adaptive loss function. *Comput Electron Agric*, 227:109522. https://doi.org/10.1016/j.compag.2024.109522

Ma CR, Chi G, Ju XP, et al., 2025. YOLO-CWD: a novel model for crop and weed detection based on improved YOLOv8. *Crop Prot*, 192:107169. https://doi.org/10.1016/j.cropro.2025.107169

Meng JQ, Wang YX, Zhang JM, et al., 2023. Tea bud and picking point detection based on deep learning. *Forests*, 14(6):1188. https://doi.org/10.3390/f14061188

Parashar N, Johri P, Elbeltagi A, et al., 2025. Enhanced residual-attention deep neural network for disease classification in maize leaf images. *Sci Rep*, 15:29452. https://doi.org/10.1038/s41598-025-14726-1

Perez L, Wang J, 2017. The effectiveness of data augmentation in image classification using deep learning. arXiv: 1712.04621. https://doi.org/10.48550/arXiv.1712.04621

Rahat IS, Ghosh H, Dara S, et al., 2025. Towards precision agriculture tea leaf disease detection using CNNs and image processing. *Sci Rep*, 15:17571. https://doi.org/10.1038/s41598-025-02378-0

Soeb JA, Jubayer F, Tarin TA, et al., 2023. Tea leaf disease detection and identification based on YOLOv7 (YOLO-T). *Sci Rep*, 13:6078. https://doi.org/10.1038/s41598-023-33270-4

Sun YG, Li ZH, Guo HP, et al., 2025. TDDet: a novel lightweight and efficient tea disease detector. *Comput Electron Agric*, 237:110481. https://doi.org/10.1016/j.compag.2025.110481

Wang RJ, Chen YS, Zhang GH, et al., 2025. YOLO11-PGM: high-precision lightweight pomegranate growth monitoring model for smart agriculture. *Agronomy*, 15(5):1123. https://doi.org/10.3390/agronomy15051123

Wang X, Wu Y, Cui LF, et al., 2025. Linear pattern detection of building groups by integrating dynamic snake convolution with YOLO11. *Geocarto Int*, 40(1):2471914. https://doi.org/10.1080/10106049.2025.2471914

Wang ZH, Zhao ZX, Hong N, et al., 2017. Characterization of causal agents of a novel disease inducing brown-black spots on tender tea leaves in China. *Plant Dis*, 101(10): 1802-1811. https://doi.org/10.1094/PDIS-04-17-0495-RE

Wen CM, Cheng Y, Li SP, et al., 2025. Slim-YOLO: an improved sugarcane tail tip recognition algorithm based on YOLO11n for complex field environments. *Appl Sci*, 15(8):4286. https://doi.org/10.3390/app15084286

Wu X, Deng HY, Wang Q, et al., 2023. Meta-learning shows great potential in plant disease recognition under few available samples. *Plant J*, 114(4):767-782. https://doi.org/10.1111/tpj.16176

Xie S, Sun HW, 2023. Tea-YOLOv8s: a tea bud detection model based on deep learning and computer vision. *Sensors*, 23(14):6576. https://doi.org/10.3390/s23146576

Xu JL, Pan F, Han XH, et al., 2024. EdgeTrim-YOLO: improved trim YOLO framework tailored for deployment on edge devices. 2024 4th International Conference on Computer Communication and Artificial Intelligence (CCAI). Xi'an, China, p.113-118. https://doi.org/10.1109/ccai61966.2024.10602964

Yang ZM, Ma W, Lu JZ, et al., 2023. The application status and trends of machine vision in tea production. *Appl Sci*, 13(19):10744. https://doi.org/10.3390/app131910744

Yuan ZH, Ning H, Tang XY, et al., 2024. GDCP-YOLO: enhancing steel surface defect detection using lightweight machine learning approach. *Electronics*, 13(7):1388. https://doi.org/10.3390/electronics13071388

Zhan BS, Xiong X, Li XL, et al., 2024. BHC-YOLOV8: improved YOLOv8-based BHC target detection model for tea leaf disease and defect in real-world scenarios. *Front Plant Sci*, 15:1492504. https://doi.org/10.3389/fpls.2024.1492504

Zheng ZH, Wang P, Liu W, et al., 2020. Distance-IoU loss: faster and better learning for bounding box regression. Proceedings of the AAAI Conference on Artificial Intelligence. AAAI Press, Washington, DC, USA, 34(7):12993-13000. https://doi.org/10.1609/aaai.v34i07.6999

Zheng ZW, Yu WW, 2025. RG-YOLO: multi-scale feature learning for underwater target detection. *Multimedia Syst*, 31:26. https://doi.org/10.1007/s00530-024-01617-0

Zhu JJ, Qiu JL, Chen SW, et al., 2025. An application of YOLOv8 integrated with attention mechanisms for detection of grape leaf black rot spots. *PLoS One*, 20(4): e0321788. https://doi.org/10.1371/journal.pone.0321788

Zhu XY, Chen FJ, Zheng YL, et al., 2024. Detection of *Camellia oleifera* fruit maturity in orchards based on modified lightweight YOLO. *Comput Electron Agric*, 226:109471. https://doi.org/10.1016/j.compag.2024.109471

Zong HT, Zhang YS, Liu XX, et al., 2023. Recent trends in smartphone-based optical imaging biosensors for genetic testing: a review. *View*, 4(4):20220062. https://doi.org/10.1002/VIW.20220062