

Article

Smart Agricultural Pest Detection Using I-YOLOv10-SC: An Improved Object Detection Framework

Wenxia Yuan ^{1,2,3}, Lingfang Lan ¹, Jiayi Xu ¹, Tingting Sun ⁴, Xinghua Wang ^{1,2}, Qiaomei Wang ^{1,2}, Jingnan Hu ¹ and Baijuan Wang ^{1,2,*} 

¹ College of Tea Science, Yunnan Agricultural University, Kunming 650201, China; yuanwenxia2023@163.com (W.Y.); 18662688640@163.com (L.L.); 18088111511@163.com (J.X.); wangaoyu2011@163.com (X.W.); 363504223@163.com (Q.W.); 17386709808@163.com (J.H.)

² College of Agronomy and Biotechnology, Zhejiang University, Hangzhou 310013, China

³ Yunnan Organic Tea Industry Intelligent Engineering Research Center, Yunnan Agricultural University, Kunming 650201, China

⁴ China Tea (Yunnan) Co., Ltd., Kunming 650201, China; tingting_chimes@163.com

* Correspondence: wangbaijuan2023@163.com

Abstract: Aiming at the problems of insufficient detection accuracy and high false detection rates of traditional pest detection models in the face of small targets and incomplete targets, this study proposes an improved target detection network, I-YOLOv10-SC. The network leverages Space-to-Depth Convolution to enhance its capability in detecting small insect targets. The Convolutional Block Attention Module is employed to improve feature representation and attention focus. Additionally, Shape Weights and Scale Adjustment Factors are introduced to optimize the loss function. The experimental results show that compared with the original YOLOv10, the model generated by the improved algorithm improves the accuracy by 5.88 percentage points, the recall rate by 6.67 percentage points, the balance score by 6.27 percentage points, the mAP value by 4.26 percentage points, the bounding box loss by 18.75%, the classification loss by 27.27%, and the feature point loss by 8%. The model oscillation has also been significantly improved. The enhanced I-YOLOv10-SC network effectively addresses the challenges of detecting small and incomplete insect targets in tea plantations, offering high precision and recall rates, thus providing a solid technical foundation for intelligent pest monitoring and precise prevention in smart tea gardens.

Keywords: Space-to-Depth Convolution; shape weights; Scale Adjustment Factors; intelligent pest monitoring



Academic Editors: Hongxing Peng and Yuanyuan Shao

Received: 25 December 2024

Revised: 10 January 2025

Accepted: 15 January 2025

Published: 17 January 2025

Citation: Yuan, W.; Lan, L.; Xu, J.; Sun, T.; Wang, X.; Wang, Q.; Hu, J.; Wang, B. Smart Agricultural Pest Detection Using I-YOLOv10-SC: An Improved Object Detection Framework. *Agronomy* **2025**, *15*, 221. <https://doi.org/10.3390/agronomy15010221>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As one of the key modern agricultural industries in Yunnan's plateau region, the tea industry serves not only as a carrier of traditional culture but also as a crucial component of contemporary rural revitalization. In recent years, climate change and alterations in cultivation methods have led to frequent and widespread pest outbreaks in tea plantations, severely constraining the development of the tea industry. Currently, pest management [1,2] in tea plantations primarily relies on chemical control, biological control, physical control, and agronomic management [3]. However, these traditional methods are often greatly affected by the environment, slow in action, and high in implementation costs, requiring a lot of manpower and material resources [4]. Therefore, effectively integrating intelligent technologies and data analysis to achieve precise pest control and resource optimization has become a critical issue for the sustainable development of the tea industry [5].

In order to realize the intelligent and accurate prevention and control of tea garden pests [6], the first thing to be solved is the accurate detection and positioning of insects. With the rapid development of neural network [7] technologies, pest detection in tea plantations has gradually become more intelligent, with deep neural networks and convolutional neural networks increasingly applied in pest image processing. Object detection technologies such as YOLO (You Only Look Once version 10) and Faster R-CNN have provided efficient solutions for automatic pest localization.

Yue Yu et al. [8] developed the LP-YOLO(s) network based on YOLOv8 by replacing certain network modules with LP_Unit and LP_DownSample while integrating the Efficient Channel and Spatial Attention mechanism. This optimized the network structure and performance, reducing model parameters by 70.2% while improving model accuracy by 40.7%, despite only a 0.8% drop in mAP (mean Average Precision).

For detecting Neotropical brown stink bugs, Bruno Pinheiro de Melo Lima et al. [9] proposed an improved YOLOv8 network by introducing P2 and C2f2 modules to optimize the original YOLOv8 structure. They also integrated the ByteTrack algorithm for automatic insect counting. Ablation experiments showed that compared to the original YOLOv8, the YOLOv8n-P2-C2f2 network achieved mAP0.5 and mAP0.95 improvements of 9.6% and 4.4%, respectively, with only a 0.5 G increase in model parameters.

In addition, our research team proposed an improved YOLO algorithm for tea plantation pest detection based on YOLOv7 [10]. This enhancement involved introducing MPDIoU (Maximum Possible DIoU) to optimize the loss function, employing spatial-channel reorganization convolution to improve the backbone network, and integrating the Vision Transformer with Bi-Level Routing Attention to further optimize the network structure. Experimental results indicated that the improved YOLOv7 network achieved Precision, Recall, F1-score, and mAP improvements of 5.68%, 5.14%, 5.41%, and 2.58%, respectively, while reducing model parameters by 1.39 G. However, the complex environment of tea plantations remains a significant challenge. Uneven lighting, leaf occlusion, diverse background textures, and small insect sizes blending into their surroundings frequently compromise detection accuracy, causing missed or false detections in complex pest scenarios.

Given that pest targets in tea plantations are typically small and often blend into complex backgrounds, conventional image acquisition devices face significant limitations in capturing target details and extracting features. To enhance data collection accuracy and image quality, this study employs microscopic lenses for data acquisition. The research subjects include *Toxoptera aurantii* (Boyer de Fonscolombe, 1841), *Xyleborus fornicatus* Eichhoff (Eichhoff, 1875), *Arboridia apicalis* (Nawa, 1917), and *Empoasca pirusuga* Matumura (Matumura, 1931). These insects damage tea tree leaves, tender buds, and branches through piercing and boring, leading to secondary diseases that severely affect the yield and quality of tea. Furthermore, the size of these insects ranges from only 1.5 to 3 mm, making them difficult for existing insect detection algorithms to perceive effectively. Addressing current challenges in pest detection research, this study proposes a deep learning model, I-YOLOv10-SC, based on Space-to-Depth Convolution, the Convolutional Block Attention Module, Shape Weights, and Scale Adjustment Factors. Its core innovation lies in achieving notable advancements in detecting small and incomplete pest targets through multi-level feature aggregation and scale-adaptive optimization. Space-to-Depth Convolution [11] enhances the network's capability to detect small insect targets while reducing detail loss during downsampling [12]. The Convolutional Block Attention Module improves feature representation and attention focus. Shape Weights and Scale Adjustment Factors optimize the loss function, boosting bounding box prediction accuracy, reducing false and missed detections for small targets, and accelerating model convergence.

To enhance the model's interpretability, facilitate understanding among agricultural managers, and improve its trustworthiness and practicality in real-world applications, Grad-CAM visualization analysis [13] is further introduced to demonstrate the model's detection process. Through the detection enabled by this model, managers can monitor the dynamic changes in pest species within tea plantations in real time. Additionally, by analyzing the fluctuation trends in the detected insect populations, they can accurately track the population dynamics of pests in the tea plantation, providing scientific evidence for monitoring pest population dynamics. This study aims to provide an efficient and precise pest detection model for the development of intelligent tea plantations in Yunnan [14], offering technical support and implementation pathways for intelligent pest monitoring [15] and precise control, thus advancing the intelligent development of the tea industry.

2. Materials and Methods

2.1. Image Acquisition and Dataset Construction

To realistically simulate the tea plantation environment, all datasets used in this study were collected through field photography. These data were collected between May and October 2024. The primary collection sites were the Laobanzhang and Hekai bases in Menghai County (100° E, 21° N), Xishuangbanna Prefecture, Yunnan Province. External validation data were gathered from the tea plantation behind Yunnan Agricultural University (102° E, 25° N).

The captured pest images include two types of backgrounds: leaf backgrounds and yellow sticky trap backgrounds. A macro lens was used for image magnification, with a magnification factor of 200×, a focal length of 40.6 mm, a focusing distance of 0.3 mm, and a brightness range of 85–150 lumens. To enhance the model's adaptability across different devices, image acquisition was conducted using various smartphones, including the iPhone 15 Pro Max, Huawei Nova 12, Redmi K60, and Vivo S19 (All smartphones are made in China).

A total of 3126 original samples were collected in this study, including 1297 in the Laobanzhang base, 1365 in the Hekai base, and 464 in the Houshan tea garden of Yunnan Agricultural University, including *T. aurantii*, *X. fornicatus*, *A. apicalis*, and *E. pirisuga*, with different angles and backgrounds. After preliminary screening, 3000 images were selected to build the dataset, with annotations performed using Make Sense. A total of 6143 labels were generated from these images, as shown in Figure 1.

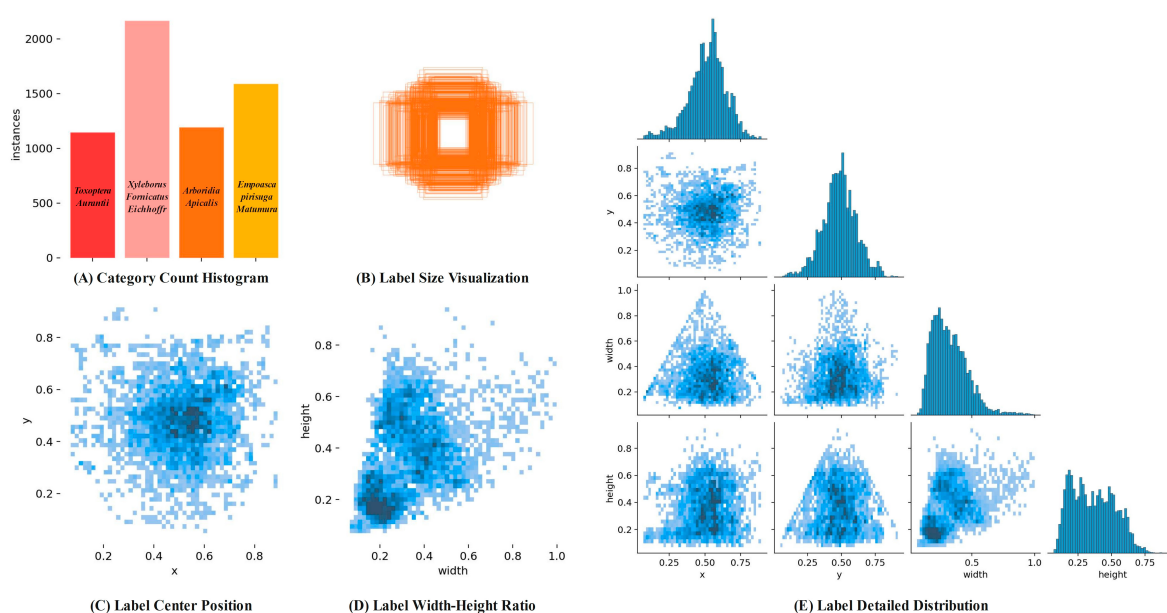


Figure 1. Label distribution.

Panel A illustrates the histogram of the number of samples for each class. Panel B shows the width and height distributions of bounding boxes after aligning all label coordinates to the same position. Panel C displays the distribution of x and y coordinates within the images. Panel D represents the aspect ratio of label widths and heights, while Panel E provides detailed label distribution statistics from the original dataset. For external validation, 453 images from the tea plantation behind Yunnan Agricultural University were selected to assess the model's generalization capability. The remaining 2547 images were randomly divided into a training set (2038 images) and a test set (509 images) using an 8:2 ratio.

2.2. Data Augmentation

In order to further improve the detection accuracy and generalization ability of the pest detection model, reduce the over-fitting phenomenon, enhance the adaptability of the model in the new environment, and better extract the characteristics of insects in different environments and angles, this study further introduces data enhancement technology to expand the training set [16]. As shown in Figure 2, to enhance the model's adaptability to changes in insect positions and orientations, geometric transformation methods such as rotation, scaling, cropping, and flipping were applied to the images, with random settings for rotation angles, scaling factors, and cropping ratios. To improve the model's performance under varying brightness and color conditions, random adjustments were made to image brightness, contrast, saturation, and hue. To further reduce background-induced interference and enhance the model's generalization ability in different environments, background blurring techniques were applied to the images. Additionally, to enhance the model's ability to detect occluded or partially damaged insects, random black occlusions were added to certain areas of the images, with the size and position of the occlusions set randomly.

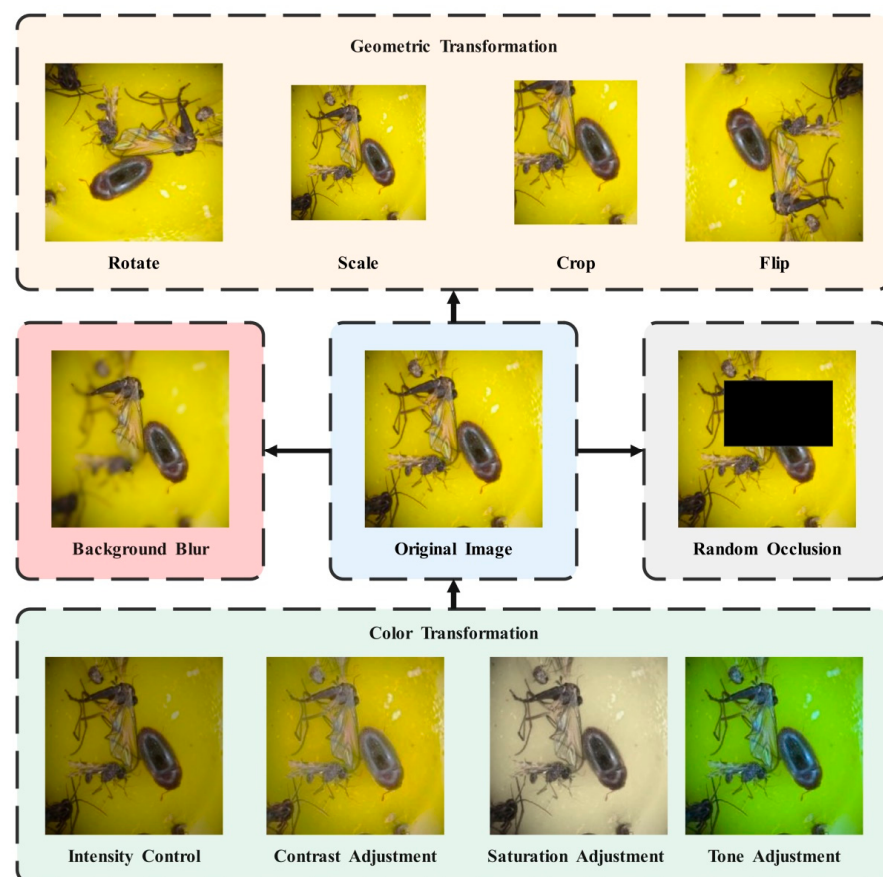


Figure 2. Data augmentation.

2.3. YOLOv10 Network Improvement

The YOLOv10 network structure primarily consists of three components: Backbone, Neck, and Head. The Backbone is responsible for extracting rich features from input images to generate high-quality feature maps. The Neck focuses on multi-scale feature fusion from the Backbone, while the Head generates the final detection results. Although YOLOv10 demonstrates excellent performance in multi-scale feature fusion and bounding box prediction, it still faces significant limitations when dealing with small targets, low-resolution images, and incomplete insects.

To address these issues, this study uses the YOLOv10 network as the base model. To enhance its capability to detect small insects and mitigate severe detail loss, the Backbone is structurally optimized using Space-to-Depth Convolution [17]. Considering the original network's limited accuracy in detecting small and incomplete insects due to its reliance on global features, the Convolutional Block Attention Module [18] is applied to specifically improve the small-object detection layer. Additionally, to further enhance bounding box prediction accuracy, reduce false and missed detections of small targets, and accelerate model convergence, Shape Weights and Scale Adjustment Factors are introduced to optimize the loss function. The improved YOLOv10 network structure is illustrated in Figure 3, with detailed parameters listed in Table 1. In Table 1, SPPF (Spatial Pyramid Pooling—Fast) mainly performs pooling operations at different scales on feature maps to obtain multi-scale contextual information, thereby enhancing the model's ability to perceive targets of various sizes.

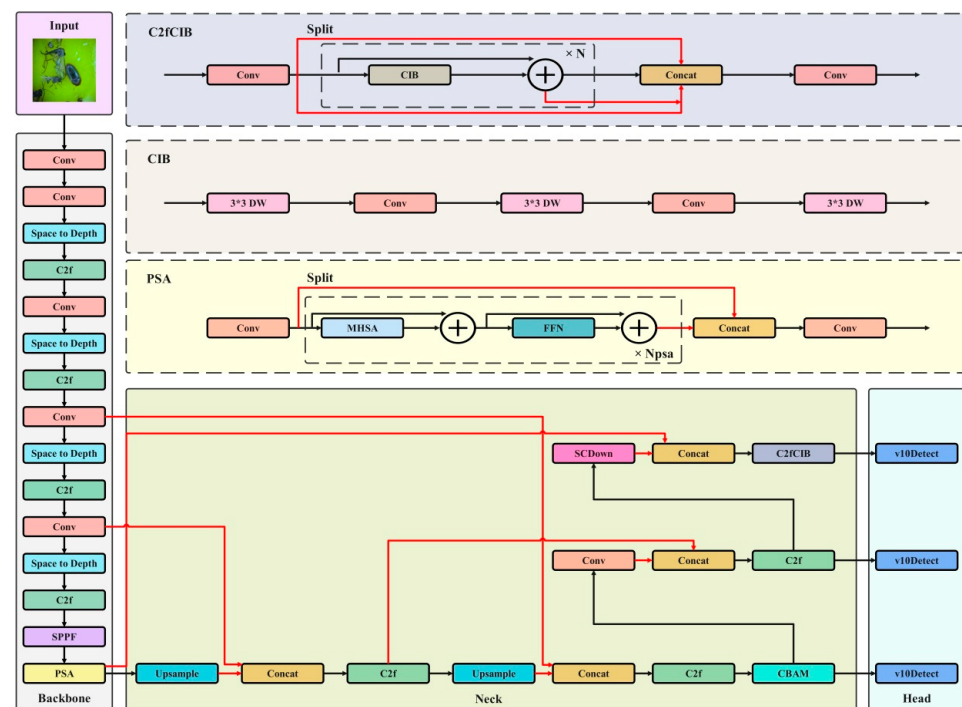


Figure 3. Improved YOLOv10 network structure.

Table 1. Detailed parameters of the improved YOLOv10 network.

ID	From	Params	Module	Arguments
0	−1	464	Conv	[3, 16, 3, 2]
1	−1	4672	Conv	[16, 32, 3, 1]
2	−1	0	Space to Depth	[1]
3	−1	10,432	C2f	[128, 32, 1, True]
4	−1	18,560	Conv	[32, 64, 3, 1]

Table 1. Cont.

ID	From	Params	Module	Arguments
5	−1	0	Space to Depth	[1]
6	−1	61,952	C2f	[256, 64, 2, True]
7	−1	73,984	Conv	[64, 128, 3, 1]
8	−1	0	Space to Depth	[1]
9	−1	246,784	C2f	[512, 128, 2, True]
10	−1	295,424	Conv	[128, 256, 3, 1]
11	−1	0	Space to Depth	[1]
12	−1	656,896	C2f	[1024, 256, 1, True]
13	−1	164,608	SPPF	[256, 256, 5]
14	−1	249,728	PSA	[256, 256]
15	−1	0	Upsample	[None, 2, ‘nearest’]
16	[−1, 10]	0	Concat	[1]
17	−1	164,608	C2f	[512, 128, 1]
18	−1	0	Upsample	[None, 2, ‘nearest’]
19	[−1, 7]	0	Concat	[1]
20	−1	41,344	C2f	[256, 64, 1]
21	−1	4258	CBAM	[64, 7]
22	−1	36,992	Conv	[64, 64, 3, 2]
23	[−1, 17]	0	Concat	[1]
24	−1	123,648	C2f	[192, 128, 1]
25	−1	18,048	SCDown	[128, 128, 3, 2]
26	[−1, 14]	0	Concat	[1]
27	−1	282,624	C2fCIB	[384, 256, 1, True, True]
28	[21, 24, 27]	862,888	v10Detect	[4, [64, 128, 256]]

2.3.1. Space-to-Depth Convolution Optimization

In the pest detection task, although the traditional YOLOv10 network has high real-time performance and detection ability, it still has some significant defects in dealing with small targets, complex backgrounds, and low-resolution images. After research, these defects are mainly due to the stride convolution and pooling operations [19] in its infrastructure. These operations inevitably lose fine-grained information during the downsampling process, resulting in a significant decrease in the detection accuracy of small insects. Because of the dense distribution of a large number of insects, the detection ability of the YOLOv10 network will also be greatly affected, and it is very easy to have missed detection and false detection problems.

To address the premature loss of small-target features caused by conventional downsampling methods, this study applies Space-to-Depth Convolution to structurally optimize the Backbone, ensuring complete feature transmission and effectively enhancing small-target detection accuracy.

As illustrated in Figure 4, Space-to-Depth primarily comprises two core components: SPD (Space-to-Depth) and Non-strided Convolution. The primary function of the SPD layer is to split the input feature map into multiple sub-feature maps based on a specified stride and concatenate these sub-feature maps along the channel dimension, forming a new feature map. This operation effectively encodes spatial information into additional channel dimensions, thereby reducing the spatial size while preserving all information. As shown in Equation (1), where $X(S \times S \times C_1)$ denotes the input feature map dimensions and $f_{i,j}$ represents the SPD layer’s splitting formula, $S \times S$ indicates the spatial resolution of the feature map and C_1 denotes the number of channels. During the splitting process, the SPD layer partitions the input feature map along rows and columns according to the specified stride. For example, when the stride length is 2, four sub-feature maps of size $\frac{S}{2} \times \frac{S}{2} \times C_1$

will be generated. During concatenation along the channel dimension, the new feature map size after concatenation is as described in Equation (1):

$$\left\{ \begin{array}{l} f_{0,0} = X[0 : S : scale, 0 : S : scale] \\ f_{1,0} = X[1 : S : scale, 0 : S : scale] \\ \vdots \\ f_{scale-1,0} = X[scale - 1 : S : scale, 0 : S : scale] \\ f_{0,1} = X[0 : S : scale, 1 : S : scale] \\ f_{scale-1,1} = X[scale - 1 : S : scale, 1 : S : scale] \\ f_{0,scale-1} = X[0 : S : scale, scale - 1 : S : scale] \\ f_{scale-1,scale-1} = X[scale - 1 : S : scale, scale - 1 : S : scale] \\ X' = \left(\frac{S}{scale}, \frac{S}{scale}, scale^2 C_1 \right) \end{array} \right. \quad (1)$$

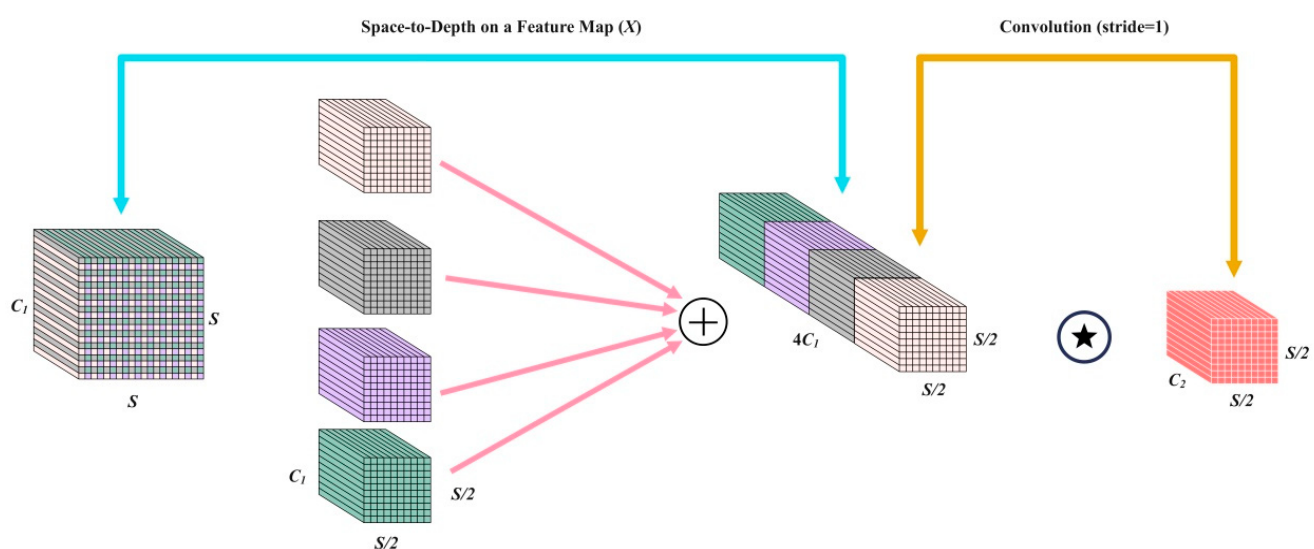


Figure 4. Space-to-Depth Convolution. (★): It means output after convolution with a stride of 1).

After passing through the SPD layer, the feature map's channel number increases significantly. To further reduce the model's computational cost while extracting discriminative features, a Non-strided Convolution layer is introduced after the SPD layer for channel dimension reduction. As shown in Equation (2), X'' represents the output feature map after channel reduction. To precisely retain discriminative features and enhance the model's perceptual ability, the channel reduction process incorporates a Non-strided Convolution with C_2 filters, ensuring efficient feature extraction while maintaining key information:

$$X'' = \left(\frac{S}{scale}, \frac{S}{scale}, C_2 \right) \quad (2)$$

2.3.2. Convolutional Block Attention Module Optimization

Research findings indicate that while the original YOLOv10 [20,21] network demonstrates strong object detection capabilities, it still faces notable limitations when detecting small targets and incomplete (partially occluded or fragmented) insects. These challenges arise because the downsampling operation in the YOLOv10 network excessively compresses small-target features, reducing detection accuracy. Additionally, the network relies heavily on global features for target recognition.

To address these issues and enhance the YOLOv10 network's detection accuracy for small targets and incomplete insects, as well as improve feature extraction capability and

robustness, this study incorporates the Convolutional Block Attention Module into the network's small-object detection layer.

As shown in Figure 5, as a lightweight attention mechanism, CBAM (Convolutional Block Attention Module) [22] is composed of the Channel Attention Module and Spatial Attention Module. It can significantly enhance the feature extraction ability of the model by adaptively adjusting the attention distribution of the model in the channel dimension and the spatial dimension.

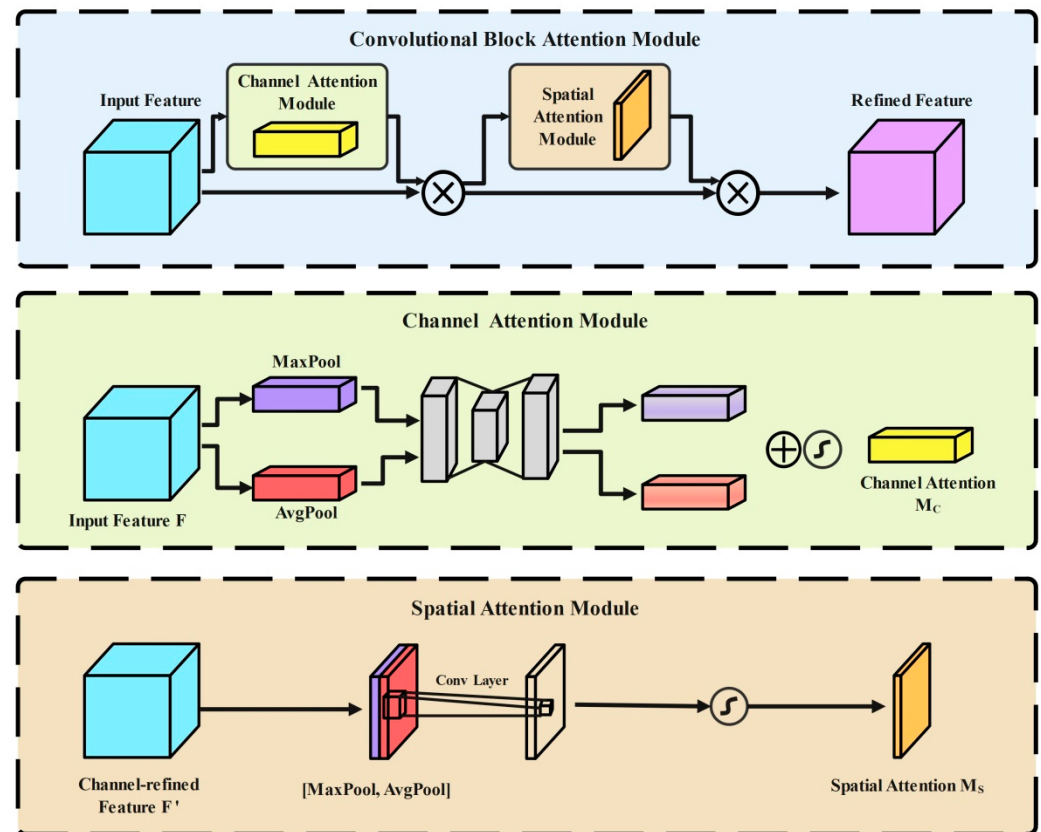


Figure 5. Convolutional Block Attention Module.

In the CBAM, the CAM (Channel Attention Module) extracts both global and local information from the feature map along the channel dimension using global average pooling and max pooling operations. The resulting pooled features are then passed through a MLP (multi-layer perceptron) [23] for further processing. As shown in Equation (3), F represents the input feature map and σ denotes the sigmoid activation function. The final channel attention map generated by CAM is applied to adjust the feature responses of each channel, effectively enhancing channels that carry critical pest-related features such as insect texture and color.

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (3)$$

The SAM (Spatial Attention Module) generates two spatial feature maps of size $1 \times H \times W$ by performing max pooling and average pooling along the channel dimension. These two feature maps are then concatenated along the channel axis and processed through a convolutional layer to extract spatial attention features. As shown in Equation (4), $f^{7 \times 7}$ represents a 7×7 convolution operation. This process allows the SAM to effectively focus

attention on regions where insects are located, thereby improving the localization accuracy of small and incomplete insects:

$$M_s(F) = \sigma\left(f^{7 \times 7}([AvgPool(F); MaxPool(F)])\right) \quad (4)$$

2.3.3. Loss Function Optimization

In the YOLOv10 network, the Bounding Box Regression Loss [24] primarily relies on the relative positions and shapes between predicted and ground-truth bounding boxes while often overlooking the geometric properties of the bounding boxes themselves. This limitation leads to reduced regression accuracy when dealing with insects with significant shape variations or size changes, ultimately affecting the model's detection performance.

To address this issue, improve detection accuracy for small insects, enhance model robustness, and accelerate convergence, this study introduces Shape Weights and Scale Adjustment Factors to optimize the loss function of the YOLOv10 network. This method incorporates the geometric shape, aspect ratio, and scale change of the model bounding box into the model training process to more accurately describe the matching degree between the prediction box and the real box and enhance the detection performance of the model for complex shape targets. Consequently, the model's detection performance for targets with complex shapes is enhanced. As shown in Equation (5), the calculation of the Shape-Sensitive Distance [25,26] in the improved loss function is weighted based on aspect ratio differences along the horizontal and vertical directions. In this context, x_c and y_c represent the center coordinates of the predicted bounding box, while x_c^{gt} and y_c^{gt} denote the center coordinates of the ground-truth box. The diagonal length of the minimum enclosing box is represented by c , while H_h and W_w indicate the shape weights in the horizontal and vertical directions, respectively:

$$distance^s = \frac{H_h \times (x_c - x_c^{gt})^2}{c^2} + \frac{W_w \times (y_c - y_c^{gt})^2}{c^2} \quad (5)$$

$$W_w = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (6)$$

$$H_h = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (7)$$

For the calculation of the Shape Error Term, as shown in Equation (8), the shape error Ω^s represents the cumulative shape error of the bounding box along the horizontal and vertical directions. The term θ is used to amplify the impact of the error, while ω_w and ω_h denote the shape error coefficients in the horizontal and vertical directions, respectively. The optimized loss function is computed as described in Equation (9), where IoU represents the Intersection over Union between the predicted and ground-truth bounding boxes. This enhanced formulation incorporates both geometric shape and scale considerations, allowing the model to perform more accurate bounding box regression, especially for insects with varying shapes and sizes:

$$\Omega^s = \sum_{t=w,h} (1 - e^{-\omega_t})^\theta, \theta = 4 \quad (8)$$

$$\begin{cases} \omega_w = H_h \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ \omega_h = W_w \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (9)$$

$$L_{S-IoU} = 1 - IoU + distance^s + 0.5 \times \Omega^s \quad (10)$$

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \quad (11)$$

2.4. Model Evaluation Metrics and Training Configuration

To further evaluate the performance of the improved YOLOv10 network in tea plantation pest detection tasks, this study introduces Precision, Recall, F1-score (balanced score), and mAP as performance evaluation metrics. Precision represents the proportion of correctly detected pests among all targets predicted as a specific pest class by the model. Recall indicates the proportion of actual pest targets successfully detected by the model among all ground-truth pests. The F1 score is the harmonic mean of accuracy and recall, serving as a comprehensive measure of the model's detection capability. As shown in Equations (12)–(16), T_P represents the number of correctly detected pests, F_P represents the number of incorrectly detected pests, and F_N represents the number of pests missed by the model. Additionally, AP (Average Precision) [27] refers to the average accuracy of a certain category under different IoU thresholds, which is a comprehensive index to evaluate the positioning accuracy and prediction accuracy. The size of the AP is determined by the accuracy and recall rate of the model. It is the area under a certain type of PR curve in all predicted pictures (the horizontal axis is Recall, and the vertical axis is Precision), and mAP is the average of all types of AP. r_i are the Recall values corresponding to the first interpolation points of the Precision interpolation segments, arranged in ascending order:

$$Precision = \frac{T_P}{T_P + F_P} \quad (12)$$

$$Recall = \frac{T_P}{T_P + F_N} \quad (13)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (14)$$

$$AP = \sum_{i=1}^{n-1} (r_{i+1} - r_i) P_{inter}(r_i + 1) \quad (15)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (16)$$

To evaluate the performance of the improved YOLOv10 network in tea plantation pest detection, this study conducted three sets of comparative experiments using four object detection networks: I-YOLOv10-SC (where “I” indicates IoU optimization, “S” represents Space-to-Depth Convolution optimization, and “C” stands for Convolutional Block Attention Module optimization), the original YOLOv10, Faster R-CNN, and SSD. Model training and testing were performed on the same dataset under identical hardware and software configurations to ensure scientific rigor and reliability of the test results.

The operating system used in this study was Windows 10, with model training conducted in GPU mode. The main machine configuration included a 12th Gen Intel(R) Core(TM) i5-12600KF 3.70 GHz processor, a 1TB hard drive, and a Colorful NVIDIA GeForce RTX 4060Ti Ultra W OC 16 G graphics card, running NVIDIA-SMI 561.09 with CUDA version 12.6. The manufacturer of the equipment is Wuhan Qicaihong Company, Wuhan, Hubei Province, China. The network development environment was Python 3.9 and PyCharm 2024. During training, the batch size was uniformly set to 64, and the number of epochs was 500.

To further improve model performance and stability, hyperparameter tuning strategies were applied to both the original YOLOv10 and the I-YOLOv10-SC network. In the process of model training, to effectively prevent overfitting or underfitting, the weight

decay parameter was set to 0.0005, controlling model complexity, suppressing unnecessary parameter growth, and enhancing generalization. In order to accelerate the convergence speed and stability of the model, the preheating period is set to 3 to ensure that the model can converge steadily at the beginning of training and avoid performance oscillation caused by too fast update. In addition, in order to further optimize the convergence process of the model, the preheating initial momentum and the preheating initial bias learning rate are set to 0.8 and 0.1, respectively.

Regarding the loss function design, the bounding box loss gain was set to 7.5 to strengthen bounding box regression optimization and improve target localization accuracy [28]. The classification loss gain was set to 0.5 to balance the model's target class recognition with other loss terms. Finally, the keypoint object loss gain was set to 1.0, ensuring a reasonable contribution from the keypoint prediction task to the overall loss function.

3. Results

3.1. Model Result Analysis

The loss function is mainly used to measure the difference between the predicted results and the real results of the model in the tea garden pest detection task. The smaller the value, the higher the matching degree between the detection results of the model and the real position of the pest target, and the better the detection performance of the model. As shown in Figure 6, the loss function of the I-YOLOv10-SC network decreases rapidly during the initial training phase. By the 40th epoch, the loss reduction rate slows significantly, 20 epochs earlier than the original YOLOv10 network. After 200 epochs, the loss function stabilizes, achieving convergence 100 epochs earlier than the original model. The final stabilized loss values for the bounding box, classification, and keypoint detection in the training set are below 0.65, 0.4, and 1.15, respectively. Compared to the original YOLOv10 network, the bounding box loss decreased by 18.75%, the classification loss dropped by 27.27%, and the keypoint loss reduced by 8%. Moreover, the model's oscillation during training was significantly mitigated, indicating improved stability and convergence efficiency.

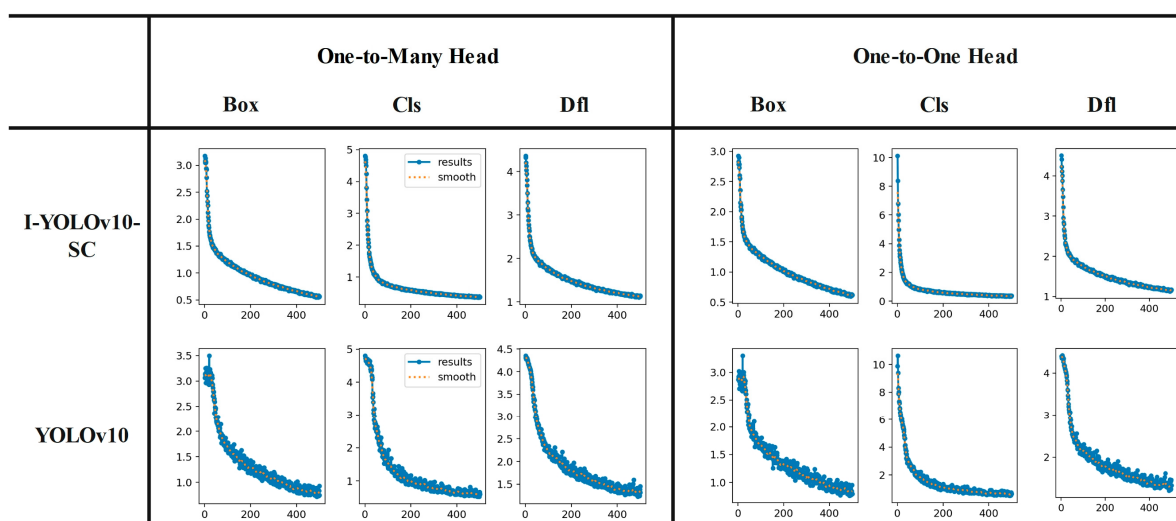


Figure 6. Loss function variation curve.

As shown in Figure 7, the I-YOLOv10-SC network achieved a Precision of 97.36%, a Recall of 97.11%, and an F1-score of 97.23%. Compared to the original YOLOv10 network, these metrics improved by 5.88, 6.67, and 6.27 percentage points, respectively. Compared with the original network, the enhanced I-YOLOv10-SC network demonstrates more accurate bounding box localization and stronger target classification capabilities, providing

a solid technical foundation for intelligent pest monitoring and precise pest control in tea plantations.

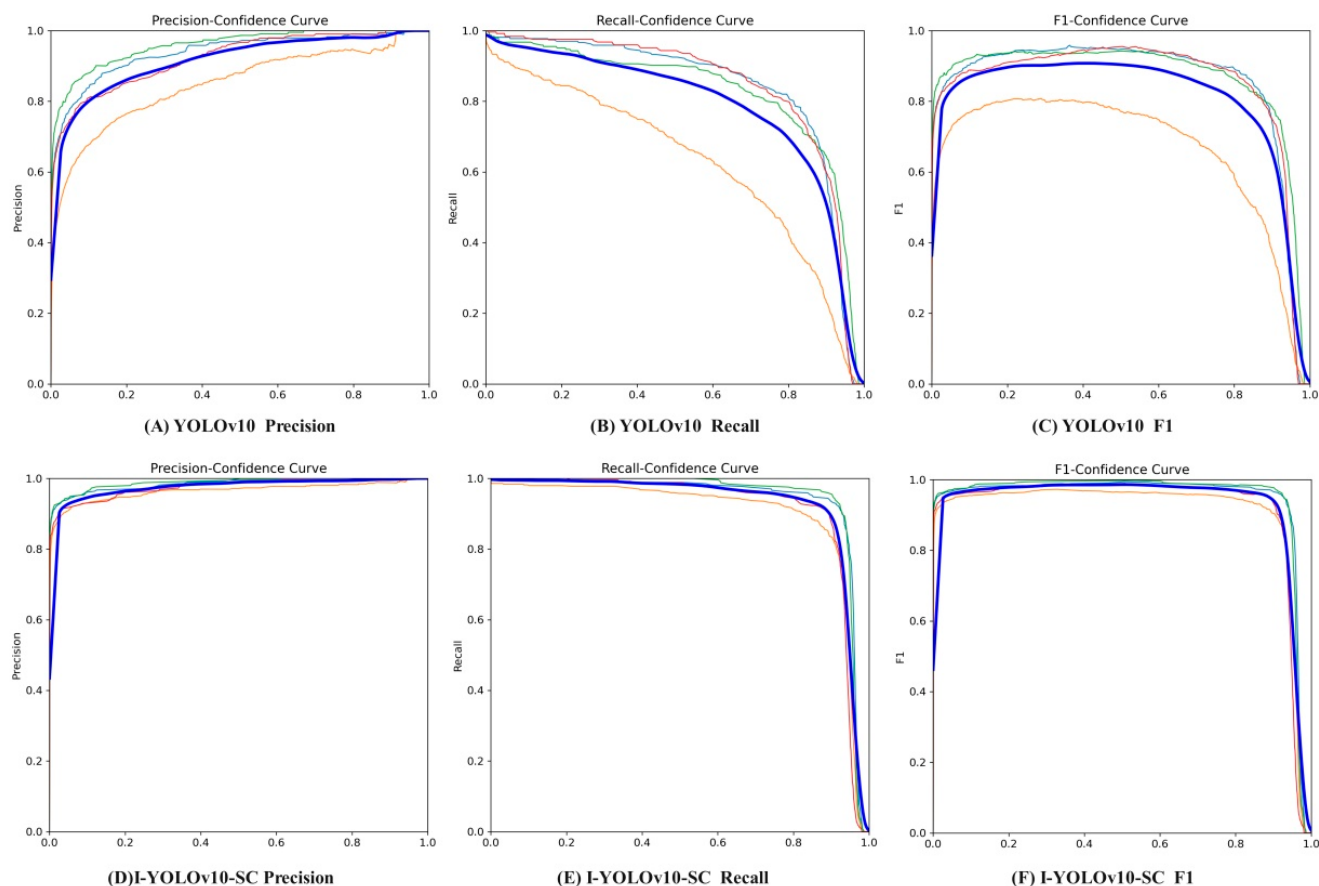


Figure 7. Performance metrics curve. (Note: Light blue is *T. aurantii*, green is *X. fornicatus*, red is *A. apicalis*, yellow is *E. pirisuga*, and dark blue is all categories.).

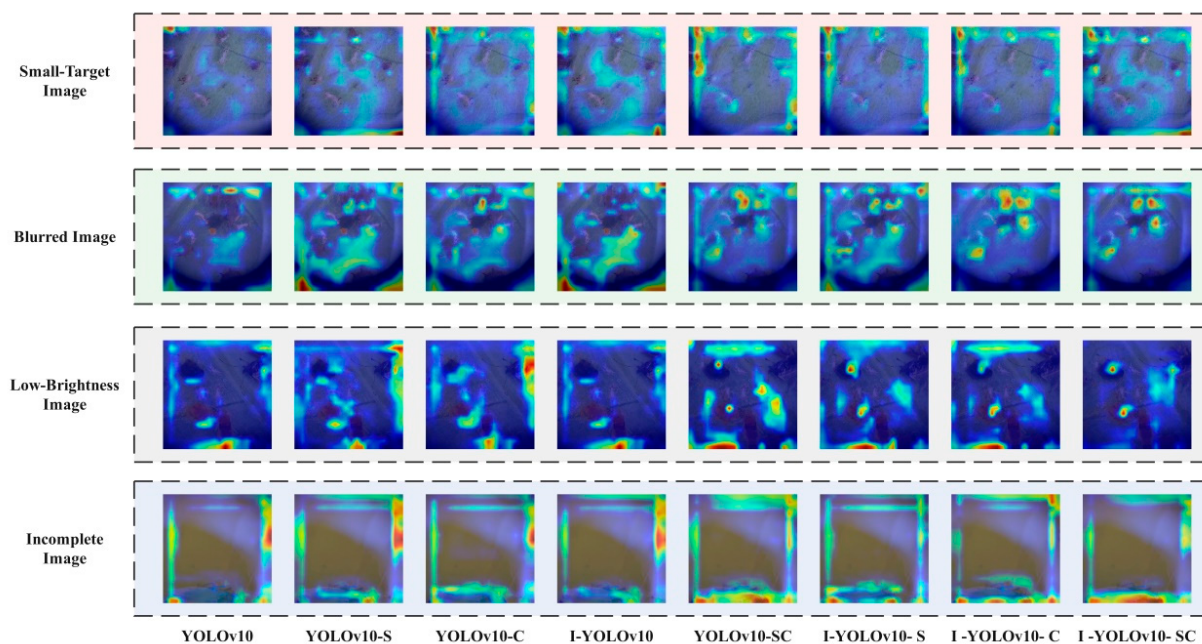
3.2. Ablation Study

To verify the effectiveness of the I-YOLOv10-SC network in tea plantation pest detection tasks and evaluate the performance improvements brought by Space-to-Depth Convolution, the Convolutional Block Attention Module [29], and the loss function optimization strategy, an ablation study was conducted using the tea plantation dataset from Yunnan Agricultural University. To ensure scientific reliability, all experiments were conducted under the same dataset, hardware environment, and hyperparameter settings. Each model was trained three times, and the highest performance metrics were recorded. As shown in Table 2, the integration of Space-to-Depth Convolution improved Recall by 1.53% and mAP by 1.31%, enhancing the spatial resolution of small-object features while reducing detail loss during feature fusion. The Convolutional Block Attention Module increased Precision by 0.2% and mAP by 1.11%, showing significant effectiveness in optimizing target localization and detection accuracy. The loss function optimization strategy improves the Recall by 1.03% and the mAP by 0.91%. The bounding box regression accuracy is improved and the positioning error is significantly reduced. After the overall improvement, compared with the original YOLOv8 network, the accuracy, recall, and mAP of the I-YOLOv10-SC network were increased by 5.88%, 6.67%, and 4.26%, respectively.

Table 2. Ablation study results.

Model	Precision (%)	Recall (%)	mAP (%)	Layers	Parameters	Gradients
YOLOv10	91.48	90.44	94.51	402	2,498,168	2,498,152
YOLOv10-S	91.43	91.97	95.82	379	3,313,656	3,313,640
YOLOv10-C	91.68	90.01	95.62	410	2,502,426	2,502,410
I-YOLOv10	91.42	91.47	95.42	402	2,498,168	2,498,152
YOLOv10-SC	91.73	93.44	96.91	387	3,317,914	3,317,898
I-YOLOv10-S	90.77	92.91	96.47	379	3,313,656	3,313,640
I-YOLOv10-C	90.16	92.15	96.28	410	2,502,426	2,502,410
I-YOLOv10-SC	97.36	97.11	98.77	387	3,317,914	3,317,898

To further evaluate the performance and limitations of the I-YOLOv10-SC network in tea plantation pest detection tasks, this study applies GradCAM (Gradient-weighted Class Activation Mapping) for visual analysis of the model's attention regions. GradCAM is a deep learning interpretability technique primarily used for visualizing convolutional neural networks. Its core concept involves combining gradient information with feature maps in the classification network to generate class activation maps of the input image. In this study, GradCAM calculates the gradient information between the model's prediction and convolutional feature maps through a backpropagation mechanism. This process determines the importance of each position in the feature map for the final classification decision. The heatmaps highlight regions of interest where the model concentrates its attention, with brighter areas indicating a stronger focus on pest-related features. As illustrated in Figure 8, the GradCAM heatmaps highlight that the I-YOLOv10-SC network demonstrates a stronger target focus in pest detection compared to the original YOLOv10 network. The enhanced heatmaps indicate more precise localization and attention to pest-related features, reflecting significant improvements in the network's detection accuracy and interpretability.

**Figure 8.** GradCAM heatmaps from the ablation study.

3.3. Model Comparison Experiments

In tea plantation pest detection tasks, model performance directly impacts detection accuracy and stability in real-world applications. To comprehensively evaluate the ef-

fectiveness of the improved I-YOLOv10-SC network, this study conducted performance comparison experiments using the I-YOLOv10-SC network, the original YOLOv10, Faster R-CNN, and SSD. As shown in Table 3, the I-YOLOv10-SC network outperformed all baseline models across key evaluation metrics. Compared to the original YOLOv10 network, Faster R-CNN [30], and SSD [31], the I-YOLOv10-SC network achieved Precision improvements of 5.88%, 26.47%, and 14.11%, respectively. Its Recall increased by 6.67%, 19.80%, and 21.82%, while its F1 score rose by 6.27%, 23.27%, and 18.16%. Additionally, the mAP values increased by 4.26%, 19.71%, and 13.46%.

Table 3. Model comparison experiment results.

Model	Precision (%)	Recall (%)	F1 (%)	mAP (%)
I-YOLOv10-SC	97.36	97.11	97.23	98.77
YOLOv10	91.48	90.44	90.96	94.51
Faster-RCNN	70.89	77.31	73.96	79.06
SSD	83.25	75.29	79.07	85.31

To further verify the robustness and adaptability of the I-YOLOv10-SC network in tea plantation pest detection tasks, this study conducted comparative analyses of detection performance in complex scenarios, including incomplete insects, small insect targets, low-brightness images, and blurry images. Selected comparison results are shown in Figure 9. To ensure objective evaluation, an external validation dataset collected from the tea plantation behind Yunnan Agricultural University was used for testing. The results indicate that compared to the original YOLOv10 network, the I-YOLOv10-SC network demonstrated significantly improved detection capabilities for small targets and complex scenes, effectively reducing false positives and missed detections. These findings confirm that the I-YOLOv10-SC network provides strong technical support for intelligent pest detection and precise pest management in tea plantations.

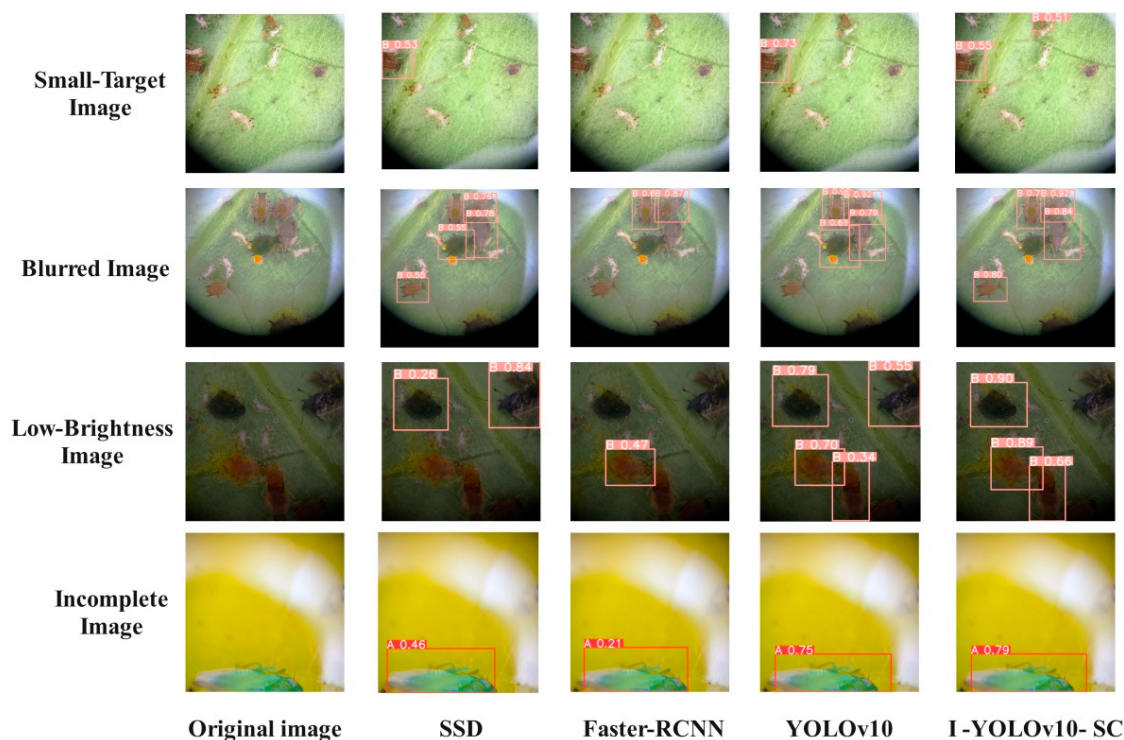


Figure 9. External validation comparison.

4. Discussion

The I-YOLOv10-SC model proposed in this study is a deep learning model designed for the detection of small and incomplete insect targets in tea gardens, optimized to improve precise detection performance. The results indicate that compared to the original YOLOv10, as well as networks such as Faster R-CNN and SSD, I-YOLOv10-SC shows significant improvements in both detection accuracy and model efficiency. Compared with the Hypertuned-YOLO based on EigenCAM developed by Stefano Frizzo Stefenon et al., the F1 score is increased by 10.53% and the mAP is increased by 6.57% [32]. Compared with the YOLOu-Quasi-ProtoPNet network based on DenseNet-161, the F1 score is increased by 2.06%. These results fully demonstrate that I-YOLOv10-SC not only has stronger target detection ability, but also outperforms most existing advanced models in terms of performance indicators [33].

One of the core innovations of this model is the introduction of Space-to-Depth convolution, which successfully addresses the common problem of feature loss during downsampling operations in traditional networks. This improvement allows I-YOLOv10-SC to better preserve the fine-grained features of small insect targets, enhancing its ability to detect pests that are small in size and set against complex backgrounds.

Additionally, the inclusion of the Convolutional Block Attention Module (CBAM) further enhances detection accuracy, especially when facing small or partially occluded insects. CBAM selectively focuses attention on the most critical features, helping the model concentrate on key pest-related attributes, thus improving its ability to handle partial occlusion and fragmented appearances of insects. By integrating Shape Weights and Scale Adjustment Factors into the loss function, the model's detection performance is significantly improved. These adjustments allow the model to better match the predicted bounding boxes with the actual pest locations, reducing false positives and improving localization accuracy.

The I-YOLOv10-SC model has significant potential for intelligent pest monitoring in tea gardens. By precisely identifying and locating pests, the model can monitor the species and number of pests in real-time, providing accurate decision support to farmers and reducing the overuse of pesticides. Farmers can intervene locally in specific areas based on the actual detection results, not only reducing pesticide use but also minimizing pollution to soil and water sources. Furthermore, accurate pest monitoring helps protect the ecological environment and reduce harm to beneficial organisms, promoting the sustainability of agricultural production. This study lays a solid foundation for the advancement of precision pest control technologies and provides strong support for the development of smart agricultural technologies, contributing to the green transformation of agriculture and the achievement of sustainable development goals.

5. Conclusions

The I-YOLOv10-SC network was developed using a collaborative optimization method, greatly enhancing the accuracy and generalization for detecting small insects. This provides a new solution for the challenges of detecting small objects in object detection tasks. By incorporating Space-to-Depth Convolution into the YOLOv10 backbone, the model reduces detail loss in faraway and low-resolution images, improving its ability to detect small objects and accurately predict bounding boxes. The CBAM enhances the detection of small objects, helping the network better locate and identify small and incomplete insects. It also introduces Shape Weights and Scale Adjustment Factors in the loss function, which improves the accuracy of bounding box predictions and speeds up model training. Experimental data reveal that the I-YOLOv10-SC network stabilizes after only 200 training epochs, approximately 100 epochs earlier than the original network. Its bounding box loss, classification loss, and keypoint loss stabilize below 0.65, 0.4, and 1.15, respectively, repre-

senting reductions of 18.75%, 27.27%, and 8%. The model also demonstrates significantly reduced oscillation, indicating better training stability.

Ablation studies further validate the effectiveness of each proposed improvement. Space-to-Depth Convolution increases Recall by 1.53% and mAP by 1.31%, while CBAM boosts Precision by 0.2% and mAP by 1.11%. The loss function optimization strategy raises Recall by 1.03% and mAP by 0.91%. As a result, the overall I-YOLOv10-SC network outperforms the original YOLOv10 model with Precision, Recall, and mAP improvements of 5.88%, 6.67%, and 4.26%, respectively, with only a minimal parameter and gradient increase of less than 1 M.

Comparative experiments demonstrate that I-YOLOv10-SC surpasses the original YOLOv10 network, Faster R-CNN, and SSD in key metrics. Precision improved by 5.88%, 26.47%, and 14.11%, respectively, Recall by 6.67%, 19.80%, and 21.82%, F1-score by 6.27%, 23.27%, and 18.16%, and mAP by 4.26%, 19.71%, and 13.46%. The enhanced YOLOv10 network significantly strengthens robustness in small-object detection and adaptability to complex environments, reducing false positives and missed detections. These improvements provide effective technical support for intelligent pest monitoring and precision pest control in tea plantations [34], laying a solid foundation for future applications in smart agriculture [35,36].

Author Contributions: Conceptualization, writing—original draft preparation, W.Y.; methodology, L.L.; software, J.X.; formal analysis, T.S.; investigation, X.W., Q.W. and J.H.; conceptualization, writing—review and editing, funding acquisition, B.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the study on the screening mechanism of phenotypic plasticity characteristics of large-leaf tea plants in Yunnan driven by AI based on data fusion (202301AS070083); Yunnan Menghai County Smart Tea Industry Science and Technology Mission (202304BI090013) and the National Natural Science Foundation (32060702).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Our code has been compressed and sent to the editorial department.

Conflicts of Interest: Author Tingting Sun was employed by the company China Tea (Yunnan) Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Li, X.Y.; Fang, T.; Gao, T.; Gui, H.; Chen, Y. Widespread presence of gut bacterium *Glutamicibacter ectropis* sp. nov. confers enhanced resistance to the pesticide bifenthrin in tea pests. *Sci. Total Environ.* **2024**, *955*, 176784. [[CrossRef](#)] [[PubMed](#)]
2. Chen, Z.M.; Luo, Z.X. Management of Insect Pests on Tea Plantations: Safety, Sustainability, and Efficiency. *Annu. Rev. Entomol.* **2024**. [[CrossRef](#)] [[PubMed](#)]
3. Zhang, Q.; Zhang, Y.; Wang, Y.H. Transcriptomic Analysis of the Effect of Pruning on Growth, Quality, and Yield of Wuyi Rock Tea. *Plants* **2023**, *12*, 3625. [[CrossRef](#)] [[PubMed](#)]
4. Yoo, J.; Lee, J.; Jeung, S.; Jung, S.; Kim, M. Development of a Deep Learning-Based Flooding Region Segmentation Model for Recognizing Urban Flooding Situations. *Sustainability* **2024**, *16*, 11041. [[CrossRef](#)]
5. Liu, K.C.; Liu, P.Z.; Gao, S.S. Research on the Trusted Traceability Model of Taishan Tea Products Based on Blockchain. *Appl. Sci.* **2024**, *14*, 10630. [[CrossRef](#)]
6. Li, H.X.; Yuan, W.X.; Xia, Y.X.; Wang, Z.J.; He, J.J.; Wang, Q.M.; Zhang, S.H.; Li, L.M.; Yang, F.; Wang, B.J. YOLOv8n-WSE-Pest: A Lightweight Deep Learning Model Based on YOLOv8n for Pest Identification in Tea Gardens. *Appl. Sci.* **2024**, *14*, 8748. [[CrossRef](#)]
7. Li, H.J.; Wang, Y.L.; Wang, Y.T.; Chen, J.J. A multi-memory-augmented network with a curvy metric method for video anomaly detection. *Neural Netw.* **2025**, *184*, 106972. [[CrossRef](#)] [[PubMed](#)]
8. Yu, Y.; Zhou, Q.; Wang, H.; Lv, K.; Zhang, L.J.; Li, J.; Li, D.M. LP-YOLO: A Lightweight Object Detection Network Regarding Insect Pests for Mobile Terminal Devices Based on Improved YOLOv8. *Agriculture* **2024**, *14*, 1420. [[CrossRef](#)]

9. Lima, B.P.D.M.; Borges, L.D.A.B.; Hirose, E.; Borges, D.L. A lightweight and enhanced model for detecting the Neotropical brown stink bug, *Euschistus heros* (Hemiptera: Pentatomidae) based on YOLOv8 for soybean fields. *Ecol. Inform.* **2024**, *80*, 102543. [\[CrossRef\]](#)
10. He, J.J.; Zhang, S.H.; Yang, C.H.; Wang, H.Q.; Gao, J.; Huang, W.; Wang, Q.M.; Wang, X.H.; Yuan, W.X.; Wu, Y.M.; et al. Pest Recognition in Microstates State: An Improvement of YOLOv7 Based On Spatial and Channel Reconstruction Convolution for Feature Redundancy and Vision Transformer with Bi-Level Routing Attention. *Front. Plant Sci.* **2024**, *15*, 1327237. [\[CrossRef\]](#)
11. Wen, R.Y.; Yao, Y.; Li, Z.J.; Liu, Q.Y.; Wang, Y.J.; Chen, Y.Z. LESM-YOLO: An Improved Aircraft Ducts Defect Detection Model. *Sensors* **2024**, *24*, 4331. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Tan, L.; Wu, H.; Xu, Z.F.; Xia, J.M. Multi-object garbage image detection algorithm based on SP-SSD. *Expert Syst. Appl.* **2025**, *263*, 125773. [\[CrossRef\]](#)
13. Lye, R.; Min, H.; Dowling, J.; Obertová, Z.; Estai, M.; Bachtar, N.A.; Franklin, D. Deep learning versus human assessors: Forensic sex estimation from three-dimensional computed tomography scans. *Sci. Rep.* **2024**, *14*, 30136. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Wang, Z.J.; Zhang, S.H.; Chen, L.J.; Wu, W.D.; Wang, H.Q.; Liu, X.H.; Fan, Z.P.; Wang, B.J. Microscopic Insect Pest Detection in Tea Plantations: Improved YOLOv8 Model Based on Deep Learning. *Agriculture* **2024**, *14*, 1739. [\[CrossRef\]](#)
15. Fotouhi, F.; Menke, K.; Prestholt, A.; Gupta, A.; Carroll, M.E.; Yang, H.-J.; Skidmore, E.J.; O'Neal, M.; Merchant, N.; Das, S.K.; et al. Persistent monitoring of insect-pests on sticky traps through hierarchical transfer learning and slicing-aided hyper inference. *Front. Plant Sci.* **2024**, *15*, 1484587. [\[CrossRef\]](#)
16. Yin, Y.F.; Zhang, S.; Zhang, Y.C.; Zhang, Y.; Xiang, S.L. Aircraft trajectory prediction in terminal airspace with intentions derived from local history. *Neurocomputing* **2025**, *615*, 128843. [\[CrossRef\]](#)
17. Zhao, D.; Cheng, Y.L.; Mao, S.Z. Improved Algorithm for Vehicle Bottom Safety Detection Based on YOLOv8n: PSP-YOLO. *Appl. Sci.* **2024**, *14*, 11257. [\[CrossRef\]](#)
18. Wang, B.; Huang, G.Z.; Li, H.X.; Chen, X.L.; Zhang, L.; Gao, X.H. Hybrid CBAM-EfficientNetV2 Fire Image Recognition Method with Label Smoothing in Detecting Tiny Targets. *Mach. Intell. Res.* **2024**, *21*, 1145–1161. [\[CrossRef\]](#)
19. Zhu, Y.T.; Peng, M.F.; Wang, X.Y.; Huang, X.J.; Xia, M.; Shen, X.T.; Jiang, W.W. LGCE-Net: A local and global contextual encoding network for effective and efficient medical image segmentation. *Appl. Intell.* **2024**, *55*, 66. [\[CrossRef\]](#)
20. Xue, L.Y.; Zhang, W.J.; Lu, L.Z.; Chen, Y.S.; Li, K.B. Unsupervised Domain Adaptation for Simultaneous Segmentation and Classification of the Retinal Arteries and Veins. *Int. J. Imaging Syst. Technol.* **2024**, *34*, e23151. [\[CrossRef\]](#)
21. Sharma, A.; Kumar, V.; Longchamps, L. Comparative performance of YOLOv8, YOLOv9, YOLOv10, YOLOv11 and Faster R-CNN models for detection of multiple weed species. *Smart Agric. Technol.* **2024**, *9*, 100648. [\[CrossRef\]](#)
22. Li, Y.; Guo, Z.H.; Sun, Y.; Chen, X.A.; Cao, Y.L. Weed Detection Algorithms in Rice Fields Based on Improved YOLOv10n. *Agriculture* **2024**, *14*, 2066. [\[CrossRef\]](#)
23. Deng, J.X.; Liu, J.B.; Ma, X.Q.; Qin, X.Z.; Jia, Z.H. Local Feature Enhancement for Nested Entity Recognition Using a Convolutional Block Attention Module. *Appl. Sci.* **2023**, *13*, 9200. [\[CrossRef\]](#)
24. Ahmed, A.; Wong, M.S.F.; Ilyas, U.S.; Serene, S.M.L.; Mustafa, A.; Aymn, A. Predictive analytics of oil-based non-newtonian nanofluid's viscosity with multi-layer perceptron neural networks. *Phys. Scr.* **2025**, *100*, 016004. [\[CrossRef\]](#)
25. Yu, M.; Li, Y.X.; Li, Z.L.; Yan, P.; Li, X.T.; Tian, Q.; Xie, B.L. Dense detection algorithm for ceramic tile defects based on improved YOLOv8. *J. Intell. Manuf.* **2024**, 1–16, prepubl. [\[CrossRef\]](#)
26. Emre, B.; Mattia, R.A. A database of calculated solution parameters for the AlphaFold predicted protein structures. *Sci. Rep.* **2022**, *12*, 7349.
27. Jiang, J.F.; Strother, C.M. Interactive decomposition and mapping of saccular cerebral aneurysms using harmonic functions: Its first application with “patient-specific” computational fluid dynamics (CFD) simulations. *IEEE Trans. Med. Imaging* **2013**, *32*, 153–164. [\[CrossRef\]](#)
28. Jing, L.; Yin, Y.H.; Li, L.H.; Li, L.H.; Wang, Z.H.; Zhou, Y.F. Small Object Detection in Traffic Scenes Based on Attention Feature Fusion. *Sensors* **2021**, *21*, 3031. [\[CrossRef\]](#)
29. Chen, G.Q.; Cui, B.; Chen, Y.M.; Hong, X.B.; Zhang, Y.J. Research on small target detection algorithm for water surface in complex weather environment. *J. Phys. Conf. Ser.* **2024**, *2897*, 012043. [\[CrossRef\]](#)
30. Zhang, S.Y.; Wang, W.M.; Wang, Z.B.; Li, H.L.; Li, R.C.; Zhang, S.X. Extreme R-CNN: Few-Shot Object Detection via Sample Synthesis and Knowledge Distillation. *Sensors* **2024**, *24*, 7833. [\[CrossRef\]](#)
31. Mo, H.H.; Wei, L.J. Tomato yellow leaf curl virus detection based on cross-domain shared attention and enhanced BiFPN. *Ecol. Inform.* **2025**, *85*, 102912. [\[CrossRef\]](#)
32. Stefenon, F.S.; Seman, O.L.; Klaar, R.C.A.; Ovejero, R.G.; Leithardt, V.R.Q. Hypertuned-YOLO for interpretable distribution power grid fault location based on EigenCAM. *Ain Shams Eng. J.* **2024**, *15*, 102722. [\[CrossRef\]](#)
33. Stefenon, F.S.; Singh, G.; Souza, B.J.; Freire, R.Z.; Yow, K.C. Optimized hybrid YOLOu-Quasi-ProtoPNet for insulators classification. *IET Gener. Transm. Distrib.* **2023**, *17*, 3501–3511. [\[CrossRef\]](#)

34. Li, H.F.; Kong, M.; Shi, Y. Tea Bud Detection Model in a Real Picking Environment Based on an Improved YOLOv5. *Biomimetics* **2024**, *9*, 692. [[CrossRef](#)] [[PubMed](#)]
35. Guan, X.L.; Wan, H.; He, Z.X.; Jiang, R.; Ou, Y.Z.; Chen, Y.L.; Gu, H.N.; Zhou, Z.Y. Pomelo-Net: A lightweight semantic segmentation model for key elements segmentation in honey pomelo orchard for automated navigation. *Comput. Electron. Agric.* **2025**, *229*, 109760. [[CrossRef](#)]
36. Barve, N.A.; Lajurkar, R.M.; Kharbade, B.S.; Bagde, A.S.; Waghmare, S.J.; Karande, R.A.; Sathe, S.N. Advancing Precision Agriculture: The Role of UAVs and Drones in Sustainable Farming. *Asian Res. J. Agric.* **2024**, *17*, 987–992. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.