insects

MDPI

Article

Forest Pest Detection Algorithm Based on Improved YOLOv11 with Intraspecific Classification Strategy

Xiuhao Liang [1,2,3], Jun Xiang [1,2,4], Sheng Qin [5,6], Wangjiao Liao [1,2], Junxiu Liu [5,6], Kanyakorn Piraonapicha [7], Nguyen Thanh Tuan [8], Wei Wei [1,2,*] and Dongxia Zou [1,2,*]

1 Guangxi Forestry Research Institute, Nanning 530002, China
2 Guangxi Laboratory of Forestry, Guangxi Key Laboratory of Superior Timber Trees Resource Cultivation, Guangxi Key Laboratory of Special Non-Wood Forests Cultivation and Utilization, Nanning 530002, China
3 School of Technology, Beijing Forestry University, Beijing 100083, China
4 School of Low-Altitude Economy, Central South University of Forestry and Technology, Changsha 410004, China
5 Guangxi Key Lab of Brain-Inspired Computing and Intelligent Chips, School of Electronic and Information Engineering, Guangxi Normal University, Guilin 541004, China
6 Key Laboratory of Nonlinear Circuits and Optical Communications (Guangxi Normal University), Education Department of Guangxi Zhuang Autonomous Region, Guilin 541004, China
7 Queen Sirikit Botanic Garden, The Botanical Garden Organization, Chiang Mai 50180, Thailand
8 Vietnam National University of Forestry in Dong Nai, Dong Nai 810000, Vietnam
* Correspondence: weiwei@gxlky.com.cn (W.W.); zoudongxia@gxlky.com.cn (D.Z.)

Simple Summary

Forest pests cause serious damage to trees, but identifying them from images is difficult because different pest species often look very similar, and even the same pest can look very different at different life stages, such as larva, nymph, and adult. This confusion makes it hard for detection systems to tell pests apart accurately. In this study, a pest image dataset covering different life stages of the same species is built and an improved YOLOv11 detection model is developed. The model is designed to better capture the subtle visual differences between similar pest species and between different growth stages of the same pest. Experimental results show that the proposed model achieved a detection precision of 97.19%, a recall of 97.20%, and an overall accuracy ($AP_{50–90}$) of 88.25%, outperforming existing methods. In practical use, this model can accurately identify both the pest type and its occurrence stage, helping forest managers apply targeted control measures, avoid excessive pesticide use, and reduce control costs and environmental pollution.

Abstract

Forest pest detection has challenges such as the high similarity in appearance among different species, which leads to feature confusion. To address this issue, this paper proposes an improved YOLOv11 forest pest detection model with a fine-grained intraspecific classification strategy. The bottleneck of identification caused by the diversity of internal morphologies is effectively alleviated via constructing a dataset with adults, larvae, and nymphs, i.e., using intraspecific classification strategy. This model incorporates a CBA (Convolution-BatchNorm-AReLU) module, Scale-based Dynamic (SD) loss, and an Adaptively Spatial Feature Fusion (ASFF) module. The CBA module integrates an attention mechanism to dynamically adjust the gradient, which alleviates vanishing gradients and accelerates the convergence of model parameters. SD loss adaptively balances localization and scale loss based on a target scale, which reduces IoU fluctuations and enhances cross-scale robustness. Resolution conflicts and optimizing semantic alignment are alleviated by introducing the ASFF module to achieve learnable spatial multi-scale feature fusion. Experimental results show that the improved YOLOv11 model outperforms existing methods,

Check for updates

Academic Editor: Denis J. Wright

Received: 12 September 2026
Revised: 13 October 2026
Accepted: 20 October 2026
Published: 26 October 2026

Copyright: © 2026 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.

with Precision, Recall, $AP_{50–90}$, and F1 score reaching 97.19%, 97.20%, 88.25%, and 97.17%, respectively. In practical applications, the proposed model can accurately capture subtle differences in the microscopic features of the same pest during different life cycles. This research project can accurately identify the types and occurrence stages of pests, match targeted control measures, avoid the use of excessive pesticides, and significantly reduce the cost of control and environmental pollution.

Keywords: fine-grained classification; attention mechanism; object detection; deep learning

1. Introduction

According to statistics from the Food and Agriculture Organization, although global forest resources are vast in total volume, pest and disease infestations cause an annual loss of 15% to 25% in timber stock [1]. In severely affected regions, this has even led to crises involving widespread tree mortality and ecological degradation. Traditional methods for identifying forest pests and diseases primarily rely on manual observation and empirical judgment. These approaches are defined by their low efficiency, poor timeliness, high labor intensity, significant lag and subjective errors, making them ill-suited to meet the demands of smart management in forestry [2].

Artificial intelligence methods demonstrate significant potential in the field of agriculture and forestry image identification, providing new methods for the intelligent and precise prevention of pests [3,4]. Some work combines neural networks and other methods to process agricultural and forestry images; for example, the saliency method [5], machine learning [6], image processing methods [7], self-learning salient feature maps [8], and support vector machines [9]. However, due to the complex morphology of pests and the large variation in target scales, it is difficult to ensure accurate identification of features, resulting in relatively weak accuracy and generalization ability. To overcome these limitations, object detection methods are introduced to identify pests. The performance of real-time monitoring for pests using Faster R-CNN, SSD, YOLOv3, and Cascade R-CNN was discussed in [10]. PestNet was proposed to detect and classify large-scale multi-species pests [11]. On this basis, single-stage detectors represented by YOLO, owing to their favorable balance between speed and accuracy, have become the mainstream backbone for pest-detection improvements. A review of existing works shows that their improvement strategies can be categorized into five principal directions, each with distinct advantages and limitations. (1) Backbone reconstruction: replacing the native backbone with stacked residual structures [12], Transformer hybrid modules [13], or re-parameterized networks [14] enhances multi-scale feature extraction, but usually at the cost of increased parameters and computational overhead, thereby raising deployment costs. (2) Neck feature fusion: improving multi-scale feature aggregation facilitates information flow and benefits small-object detection; however, additional fusion paths and operators increase architectural complexity and may introduce redundant connections. (3) Attention modules: attention designs such as global context-aware modules [13] strengthen key-region features and suppress background interference with negligible parameter overhead, yet attention weights are sensitive to the training data distribution and may overfit on classes with scarce samples. (4) Loss function optimization: introducing losses such as NWD [15] to refine bounding-box regression for small objects is cost-effective and plug-and-play, but its benefits are constrained by target scale distribution and are limited for extremely small objects. (5) Lightweight design: lightweight convolutional modules [12,15,16], pruning, and knowledge distillation directly compress models for edge deployment, but generally incur accuracy degradation, requiring a careful trade-off between accuracy and efficiency. Two representative cases fall within

this taxonomy. CPD-YOLO [17] targets cotton pest detection by combining neck redesign (R&B-FPN), a dynamic detection head (DyHead), and Inner-Shape IoU loss, improving F1 score and mAP by 7.44% and 7.08%, respectively, while the combined modifications substantially increase structural complexity. YOLOv8-DDS [18] introduces a dynamic sample operator, combines deformable ConvNets v2 with C2f structure, and adopts seam carving technology to reconstruct the detection head, which outperforms the baseline model by 2.4% in Precision, 5.6% in Recall, and 2.2% in mAP50. Overall, existing studies mostly pursue one or a few directions at the architectural level, whereas systematically alleviating feature confusion among visually similar species from the perspective of data supervision remains largely unexplored.

Recently, some researchers have developed methods to address the problems caused by the highly similar external features of different pest species. The highly similar physical appearances, colors and backgrounds of certain pest categories make identification extremely difficult. To solve this problem, a one-stage anchor-free detection network was proposed to mitigate the disturbance from the similar physical appearance of pests [19]. Some diseases and pests have regional characteristics, which, due to the high similarity of their features, make it difficult to classify different insect species on a large scale [20]. Based on the YOLO model, the approach of [21] introduces EIOU to address the issue caused by high similarity among many pests, and the problem caused by extremely similar colors and backgrounds of pests and their surroundings. The approach of [22] employs the Xception model as a classifier for further identification. It is difficult for human experts to identify various types of insects due to subtle differences in appearance among various types of insects [23]. For the same insect, there are significant intra-species differences and huge interspecies similarities during egg, larvae, pupae, and adult stages, which makes the identification of pests challenging. A multi-view complementary feature fusion network was proposed [23], which combines Grad-CAM with attention selection mechanisms to locate the important areas of insect images and precisely integrate these important areas to reconstruct attention areas. Based on the above literature reports, pests generally have small morphological differences among different types, extremely similar physical appearances, colors, and backgrounds, with strong regional characteristics. The egg, larvae, pupae and adult stages of the same insect are different and have different characteristics. Although these methods are aimed at addressing these issues, there is still confusion regarding the characteristics of different periods within the same category. Because characteristics of pests vary at different stages, such as adult insects, nymphs, and larvae, they are clearly inconsistent. The core principle of deep learning is to summarize and generalize the regularity of the unified characteristics of a certain thing. Therefore, the problem caused by the significant intra-species differences and huge intraspecific similarities is still a challenge.

In conclusion, most of the improvements have focused on the network structure level. However, from the perspective of data supervision, through fine-grained intra-species classification strategies, there is still a lack of research on alleviating the confusion of similar species' characteristics; few studies have jointly designed the three structural directions of gradient dynamic adjustment, scale self-adaptive regression, and integration of learnable spatial features. Therefore, an improved YOLOv11 with intraspecific classification is proposed for forestry pest detection, which employs an Attention-based Rectified Linear Unit (AReLU) activation function, Scale-based Dynamic (SD) loss function and Adaptively Spatial Feature Fusion (ASFF) module. The AReLU is introduced for gradient focusing to mitigate vanishing gradients. SD loss is used to achieve scale-aware balancing, which can reduce loss oscillation. ASFF is used to resolve feature conflicts, since it can achieve adaptive fusion. The major contributions of this paper are as follows. (1) Improved YOLOv11 with intraspecific classification significantly enhances the stability, accuracy, and cross-scale

generalization of the forestry pest detection model. (2) The dataset for inter-species classification setup improves the model accuracy by subdividing the adult, nymph, and larval stages of the pests. (3) To mitigate vanishing gradients and accelerate convergence, a CBA (Convolution-BatchNorm-AReLU) module is proposed, which uses the AReLU activation function to integrate attention into activation to dynamically regulate gradients. (4) To address the fluctuations in Intersection-over-Union (IoU) caused by the diversity of target scales, SD loss is introduced to employ a scale-aware mechanism to adaptively balance localization and scale losses, reducing IoU fluctuation and enhancing cross-scale robustness by prioritizing position for small objects and scale for large ones. (5) To solve the feature conflicts caused by scale differences, the ASFF module is introduced to enable learnable spatial multi-scale fusion, adaptively weighting hierarchical features to resolve resolution conflicts and optimize semantic alignment. Improved YOLOv11 possesses outstanding adaptability to complex environments and multi-scale feature fusion capabilities, enabling the forestry pest detection model to more accurately capture the subtle differences in the microscopic features of different life stages (adults, larvae, nymphs) of pests within the same species. It achieves high-precision identification and classification of forestry pests, aiming to alleviate problems such as gradient disappearance, multi-scale target detection conflicts, and missed detections or false detections caused by the diversity of internal morphologies of pests in traditional methods.

The rest of the paper is organized as follows: The Materials and Methods are presented in Section 2. The Experiments and Results are presented in Section 3. A conclusion is provided in Section 4.

2. Materials and Methods

2.1. Dataset Construction

The data in this work were collected in the southern areas of Guangxi, Chiang Rai, Chiang Mai, and Nan Province in Thailand, and Lạng Sơn, Nghệ An, and Quảng Trị provinces in Vietnam. The host tree species are Camellia oleifera, Eucalyptus and Illicium verum, etc. The images are captured in the wild using a combination of single-lens reflex cameras and macro lenses (such as Canon 5D Mark series + EF 100 mm f2.8L IS (Canon Inc.; Shimomaruko, Ota-ku, Tokyo, Japan), Olympus E-M5 Mark II + ED 60 mm F2.8 Macro (Olympus Corporation; Nishi-Shinjuku, Shinjuku-ku, Tokyo, Japan)) or ultra-wide macro and telephoto macro lenses (such as Huawei (Huawei Technologies Co., Ltd.; Bantian, Longgang District, Shenzhen, Guangdong, China), OnePlus (OnePlus Technology (Shenzhen) Co., Ltd.; Shenzhen, Guangdong, China), OPPO (OPPO Guangdong Mobile Communications Co., Ltd.; Chang'an, Dongguan, Guangdong, China), and iPhone (Apple Inc.; Apple Park, Cupertino, California, USA)). Classification is conducted by 4 entomologists from Guangxi Forestry Research Institute, including one researcher, two professor-level engineers and one senior engineer. After preprocessing (including cleaning, deduplication, etc.) and annotation, an image dataset of forestry pests is constructed, including 77 species. It is a field ecological dataset, containing species such as *Porthesia scintillans* (Walker), *Eucorysses grandis* (Thunberg), *Poecilocoris latus* (Dallas) (*P. latus*), *P. latus nymph*, *Lawana imitata* Melichar, etc. Due to the presence of adult, nymph and larval stages in the dataset, there are significant differences in their morphological and color features (i.e., differences in morphological features among different life stages within a species), such as in *Hyposidra talaca* (Walker) (*H. talaca*), *Spodoptera litura* (Fabricius), *Biston suppressaria* (Guenée) (*B. suppressaria*), *P. latus*, *Salurnis marginella* (Guérin-Méneville) (*S. marginella*), *Tessaratoma papillosa* Drury and *Bothrogonia ferruginea* (Fabricius) (*B. ferruginea*). As shown in Figure 1, the adult and nymph stages of *B. ferruginea* and *Salurnis marginella* have some similarities in morphological features, but most features are different; the adult and larval stages of *B.*

suppressaria have huge differences; the three stages (early, middle, and mature) of *H. talaca* larva also have significant distinctions. Therefore, the pest class is set to 88 by separating the different stages of the pests with an intraspecific classification strategy, such as adult insects, nymphs, and larvae, namely the ForestryPest84 dataset. During the dataset construction process, these images are divided into training, validation, and test sets in an 8:1:1 ratio to evaluate the generalization ability of the model. Subsequently, the data is preprocessed, annotated, and compiled into a comprehensive dataset suitable for training and testing purposes. The species composition of ForestryPest84 is shown in Table 1. The images of forestry pests in the dataset contain a series of complex scenes, including single targets, multiple targets (the same species and different species), small targets, and complex backgrounds, to evaluate the detection performance of the proposed algorithm under different complexities, scales, and target sizes, as shown in Figure 2. Figure 3 shows the visualization analysis of the dataset annotation files. It can be clearly seen from the class number distribution histogram in Figure 3A that there is an uneven distribution of classes. It can be clearly seen from Figure 3B that there is an overlap of labels, with many small targets and significant scale variations.

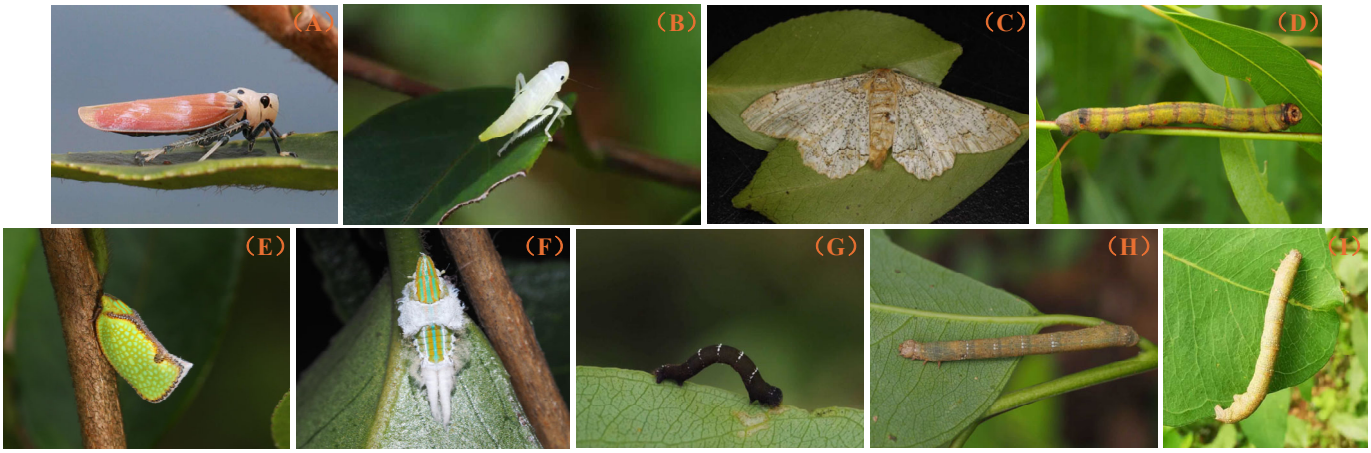

Figure 1. The intraspecific differences in the features of different life stages (adults, nymphs, and larvae) in forestry pest images. (**A**) *B. ferruginea* adult; (**B**) *B. ferruginea* nymph; (**C**) *B. suppressaria* adult; (**D**) *B. suppressaria* larva; (**E**) *S. marginella* adult; (**F**) *S. marginella* nymph; (**G**) *H. talaca* larva (early instar); (**H**) *H. talaca* larva (middle instar); (**I**) *H. talaca* larva (late instar).

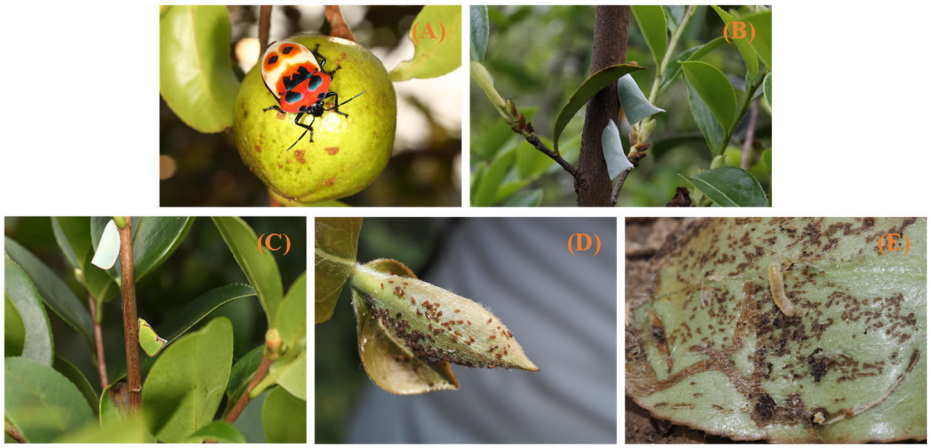

Figure 2. Some examples of the forestry pest identification dataset. (**A**) single targets; (**B**) multiple targets (belongs to the same species); (**C**) multiple targets (belongs to different species); (**D**) small targets; (**E**) complex backgrounds.

Table 1. The species composition of ForestryPest84.

Eucorysses grandis (Thunberg)	*Hyposidra talaca* (*Walker*) middle instar	*Poecilocoris latus* Dallas	*Coreidae*	*Salurnis marginella* (Guérin-Méneville) nymph
Miltochrista striata (Bremer et Grey)	*Hyposidra talaca* (Walker) late instar	*Poecilocoris latus* Dallas nymph	*Toxoptera aurantii* (Boyer de Fonscolombe)	*Mahasena colona* Sonan
Ricania speculum (Walker)	*Empoasca* (Matsumurasca) onukii Matsuda	*Aleurotrachelus camelliae* Kuwana	*Eumeta minuscula* (Butler)	*Cantao ocellatus* (Thunberg)
Amata divisa (Walker)	*Geometridae*	*Colasposoma dauricum auripenne* (Motschulsky)	*Ectropis obliqua* (Prout)	*Myllocerinus ochrolineatus* Voss
Amata sperbius (Fabricius)	*Euproctis varians* (Walker)	*Chalioides kondonis* Mats	*Eterusia aedea* (Linnaeus)	*Campsiomorpha napoana*
Tortricidae	*Euproctis varians* (Walker)	*Leptocentrus leucaspis* (Walker)	*Caloptilia theivora* (Walsingham)	*Satapa ferruginea* (Walker)
Porthesia scintillans (Walker)	*Euproctis flava* (Bremer)	*Lawana imitata* Melichar	*Dasychira baibarana* Matsumura	*Mesophalera stigmata* (Butler) larva
Pleonomus canaliculatus Fal-dermann	*Spodoptera litura* (Fabricius) larva	*Porthesia similis* (Fueszly)	*Spodoptera frugiperda*	*Graphium doson postianum*
Ricania cacaonis Chou et Lu	*Spodoptera litura* (Fabricius)	*Arbela baibarana* Matsumura	*Tessaratoma papillosa* Drury	*Erthesina fullo* (Thunberg)
Bradybaena similaris (Ferussac)	*Mictis tenebrosa*	*Euricania ocellus* (Walker)	*Hypomeces squamosus* Fabricius	*Ctenuchidae*
Zeuzera coffeae Nietner	*Sympiezomias citri* Chao	*Geisha distinctissima* (Walker)	*Altica cirsicola* Ohno	*Mythimna seperata* (Walker)
Eilema moorei (Leech)	*Archips machlopis* (Meyrick)	*Dysdercus cingulatus* (Fabricius)	*Fruticicolidae*	*Bothrogonia ferruginea* (Fabricius)
Calliteara thwaitesi Moore	*Orgyia postica* Walker	*Darna furva*	*Acanthoecia larminati* Heylaerts	*Bothrogonia ferruginea* (Fabricius) nymph
Hyposidra talaca (Walker)	*Paratrachelophrous nodicornis* Voss	*Chrysocoris stollii* (Wolff)	*Fulgoridae* nymph	*Smaragdina migrifrons* (Hope)
Ascotis selenariaSchiffermuller et Denis	*Thalassodes quadraria* Guenée	*Epicauta ruficeps*	*Pentatomidae stinkbug* nymph	*Pyrops candelaria*
Parasa repanda Walker	*Biston suppressaria* (Guenée) larvae	*Pyrrhocoridae*	*Pseudopsyra bilobata* Karny	*Ceroplastes floridensis* Comstock
Hyposidra talaca (Walker) early instar	*Biston suppressaria* (Guenée)	*Mogannia hebes* (Walker)	*Salurnis marginella* (Guérin-Méneville)	

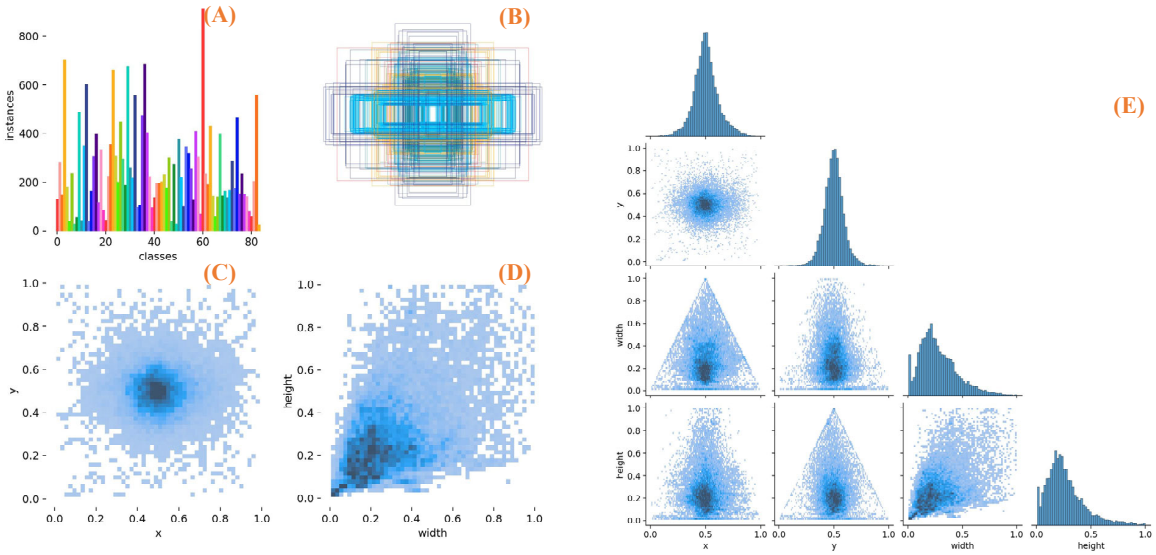

Figure 3. Visualization of training set annotation files. (**A**) Class number; (**B**) length and width of label frame; (**C**) distribution of central points; (**D**) width and height distribution; (**E**) label distribution details.

2.2. Improved YOLOv11 Model Structure

Figure 4 shows the architecture of the improved YOLOv11 model that integrates the AReLU activation function, SD loss, and ASFF module. This architecture consists mainly of the backbone network, the neck network, and the head network. The backbone network extracts features from the input image, the neck network fuses the hierarchical features extracted by the backbone network, and the head network uses the fused feature maps to predict the detection results. The basic principle of the improved YOLOv11 model is to enhance the original pest images using Mosaic and Mixup to enhance the generalization ability of the model. The enhanced images are fed into the backbone network to extract features corresponding to the three levels. Finally, the detection head that integrates the ASFF module predicts the type and corresponding position of the forestry pest targets. Its main improvements are as follows:

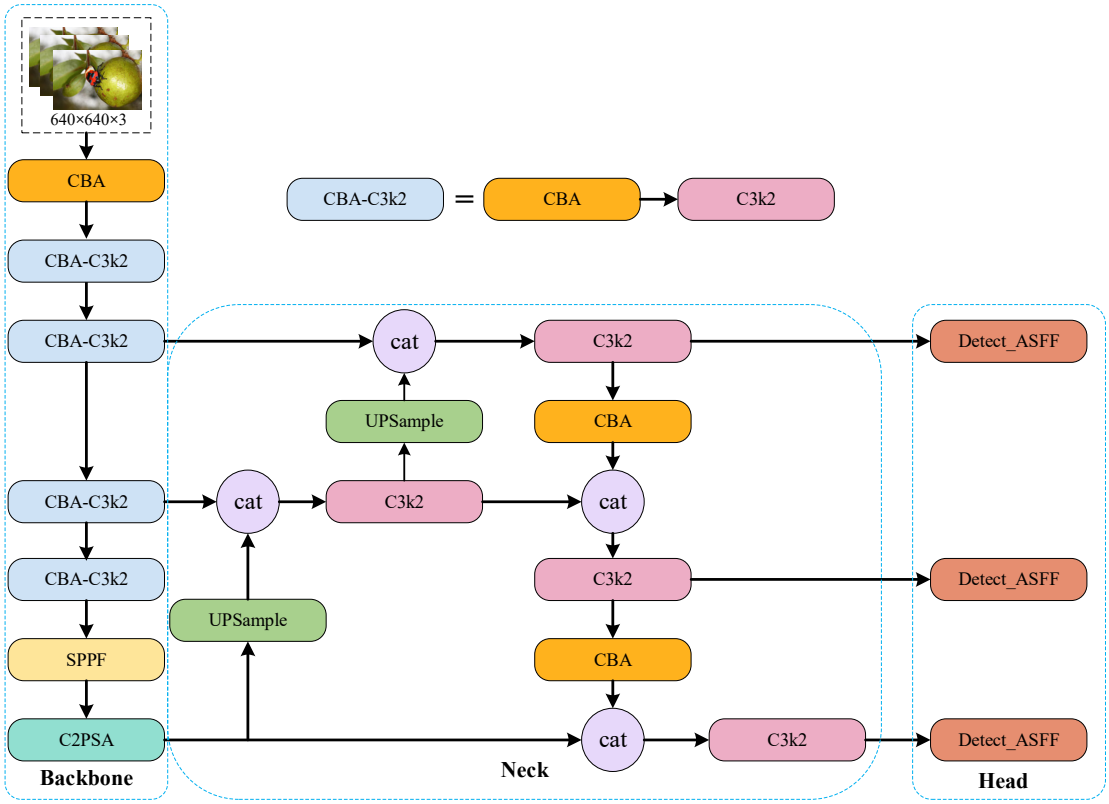

Figure 4. The network architecture of improved YOLOv11.

(1) To mitigate vanishing gradients and accelerate convergence, a CBA module is proposed to integrate attention into activation to dynamically regulate gradients. The AReLU activation function is introduced in the CBA module, which incorporates an attention mechanism into the element-wise activation pathway to dynamically regulate gradient flow and learn element-level residuals. This preserves effective activation signals during forward propagation while significantly amplifying the magnitude of gradients back-propagated from deeper to shallower layers, particularly achieving "gradient focusing" within activated regions. Consequently, it effectively mitigates the vanishing gradient problem, thereby enhancing training stability and convergence speed.

(2) To address the fluctuations in Intersection-over-Union (IoU) caused by the diversity of target scales, the SD loss function is introduced to leverage a target scale-aware mechanism to dynamically adjust the relative weights between localization loss and scale loss, thus strengthening the model's bounding box regression capability across varying scales. Existing detection frameworks typically employ composite loss functions with fixed

weighting, causing small objects to exhibit severe fluctuations due to the high sensitivity of IoU to minor perturbations, while large objects may overlook scale deviations. SD loss incorporates target scale as a prior guide to construct a continuously differentiable dynamic modulation function that automatically balances the optimization focus between localization accuracy and scale consistency, emphasizing positional sensitivity for small objects and prioritizing scale matching for large objects. This mechanism reduces reliance on unstable IoU values, significantly diminishes loss oscillation during training, and improves the model's robustness and generalization performance in cross-scale scenarios.

(3) To address the feature conflicts caused by scale differences, the ASFF module is introduced to implement a learnable multi-scale feature fusion strategy along the spatial dimension, fully exploiting and coordinating complementary information from multi-level features. Unlike traditional hand-crafted fusion methods (such as summation or concatenation), ASFF adaptively learns fusion weights from different hierarchical feature maps at each spatial location, enabling the network to select the most discriminative scale responses based on local semantic content. This mechanism effectively alleviates feature conflicts arising from resolution discrepancies and inconsistent receptive fields, preventing reduced training efficiency caused by erroneous information alignment.

2.3. CBA Module

Figure 5 shows a structural comparison between the CBS (Convolution-BatchNorm-SiLu) module in the original YOLOv11 [24] and the CBA module in improved YOLOv11. The sub-plot at the top of Figure 5 shows the structure of the CBS module, which consists of a convolutional layer, a Batch Normalization (BN) layer, and an activation function with Sigmoid-weighted Linear Unit (SiLU). In contrast, the CBA module replaces the SiLU activation function with an AReLU activation function [25]. The SiLU function introduces nonlinear factors, enabling the network to learn and fit more complex functional relationships. However, it is relatively weak in handling the problem of gradient vanishing and its stability is greatly affected by the input values and network parameters. AReLU, combined with the attention module and the ReLU module, learns element-wise residuals relative to the ReLU activation part, which helps alleviate the problem of gradient vanishing. The AReLU function is defined by

$$F(x_i, \alpha, \beta) = \begin{cases} C(\alpha)x_i, & x_i < 0 \\ (1 + \sigma(\beta))x_i, & x_i \geq 0 \end{cases}, \quad (1)$$

where $x_i \in X$ represents the input of the current neural network layer, α and β are learnable parameters that determine the different levels of attention given to positive and negative elements, function $C(\cdot)$ restricts the input variables to the range $[0.01, 0.99]$, and σ represents the Sigmoid function. This combination is based on the learned scale parameters α and β, which respectively amplify the positive elements and suppress the negative elements. Therefore, the model learns the residual elements of the activated elements, which helps to alleviate the problem of gradient disappearance.

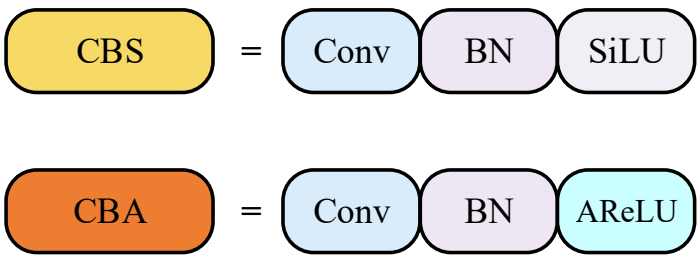

Figure 5. The structure of both the CBS and CBA modules.

2.4. SD Loss

In the identification of forestry pests, there are issues such as complex environments, large variations in target scale, and a predominance of small targets. However, in the object detection task, loss based on the IoU ratio has a fluctuation problem. The IoU loss of bounding box (BBox) labels can fluctuate up to 86% at the boundary, and the stability of IoU loss for smaller targets is even worse, seriously affecting the stability of the model and the regression effect [26]. The existing loss functions do not fully consider the sensitivity differences in scale and position under different target scales. The instability of small targets is more obvious, and the position loss of mask labels is difficult to converge when dealing with missed detections, easily leading to false alarms. Additionally, as shown in Figure 3B, the ForestryPest84 dataset is a multi-scale dataset. Therefore, SD loss [26] is introduced in this work to replace the Complete IoU (CIoU) loss function. The influence coefficients of scale loss and position loss are dynamically adjusted based on the target scale to reduce the fluctuation of IoU labels.

The overall loss in the YOLOv11 algorithm [24] is defined by

$$Loss_{Total} = \lambda_1 L_{CIoU} + \lambda_2 L_{DFL} + \lambda_3 L_{BCE}, \tag{2}$$

where L_{CIoU} represents the position loss, which is the bounding box regression loss based on CIoU; L_{DFL} represents the confidence loss, which is the Distribution Focal loss, converting the bounding box regression into a classification problem to improve the positioning accuracy. L_{BCE} represents the classification loss, which is the binary cross-entropy loss and supports multi-label classification; λ_1, λ_2 and λ_3 respectively represent the weight values assigned to these parameters. The position loss L_{CIoU} is replaced by SD loss, and its equation definition is as follows:

$$Loss_{Total} = \lambda_1 \mathcal{L}_{SD} + \lambda_2 L_{DFL} + \lambda_3 L_{BCE}, \tag{3}$$

where $\mathcal{L}_{SD}$ represents SD loss, and its formula is

$$L_{SD} = \beta_{L_{BS}} \times L_{BS} + \beta_{L_{BL}} \times \mathcal{L}_{BL}, \tag{4}$$

among them, $\beta_{L_{BS}}$ and $\beta_{L_{BL}}$ are the influencing factors of L_{BS} and L_{BL} respectively. The calculation formulas for L_{BS} and L_{BL} are

$$L_{BS} = 1 - IoU + \alpha v, \tag{5}$$

$$L_{BL} = \frac{\rho^2\left(b_p, b_{gt}\right)}{c^2}, \tag{6}$$

where IoU represents the intersection of the union of the predicted true value and the actual true value BBox. αv measures the consistency of the aspect ratio of the BBox, $\rho(\cdot)$ is the Euclidean distance, b_p and b_{gt} are the centroids of the predicted BBox and the target BBox, respectively, and c is the diagonal length of the 2 BBoxes, while $\beta_{L_{BS}} = 1 - \delta - \beta_B$ and $\beta_{L_{BL}} = 1 + \delta - \beta_B$, where δ is a tunable parameter. For the BBox labels of this task, the influence coefficient is determined by β_B, and its formula is

$$\beta_B = min\left(\frac{B_{gt}}{B_{gtmax}} \times R_{OC} \times \delta, \delta\right), \tag{7}$$

where B_{gt} is the current target box area, $B_{gtmax} = 81$ in the formula is the maximum size of the IRST as stipulated by the Society of Photo-Optical Instrumentation Engineers in the United States, and R_{OC} represents the size ratio between the original image and the current feature map. When the area of the target BBox is greater than 81, L_{SD} degenerates into L_{CIoU}.

2.5. ASFF

In single-stage object detection, the inconsistency among features of different scales limits the improvement of its performance. As shown in Figure 2B, the ForestryPest84 dataset is a multi-scale dataset. To mitigate this problem, an ASFF module [27] is introduced in this work to be integrated into the detection head, which learns the spatial fusion weights of different-scale feature maps adaptively and effectively filters conflicting information during training, enhancing the scale invariance of the features. As shown in Figure 6, first, the resolution feature representation of the l layer ($l \in \{1,2,3\}$) is denoted as x^l. For a certain layer l, the features x^n of other layers n ($n \neq l$) are adjusted to have the same shape as x^l. In YOLO, there are three feature layers with different resolutions and channel numbers, so the upsampling and downsampling strategies for each scale are modified accordingly. For upsampling, one 1 × 1 convolution layer is first used to compress the feature channels in the l layer, and then the resolution is enhanced using interpolation. For 1/2 downsampling, a 3 × 3 convolution (with a stride of 2) is used to simultaneously modify the number of channels and the resolution. For 1/4 downsampling, a 1 × 1 max-pooling layer with a stride of 2 is added before the 2 × 2 convolution.

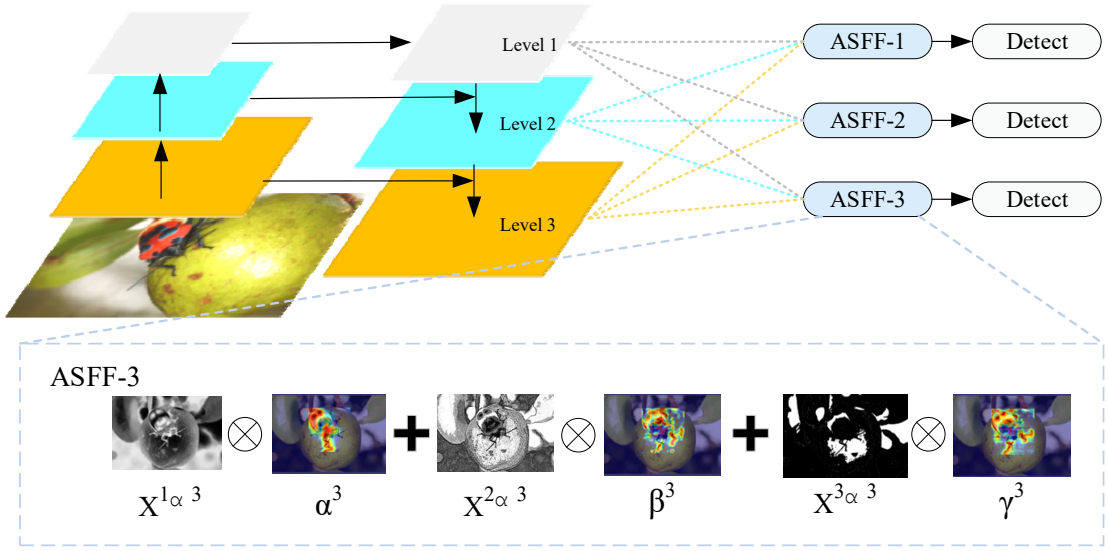

Figure 6. Illustration of the adaptive spatial feature fusion mechanism.

Suppose the adjusted feature vector is $x_{ij}^{n \to l}$ and the fusion formula is

$$y_{ij}^l = \alpha_{ij}^l \cdot x_{ij}^{1 \to l} + \beta_{ij}^l \cdot x_{ij}^{2 \to l} + \gamma_{ij}^l \cdot x_{ij}^{3 \to l}, \quad (8)$$

where y_{ij}^l denotes the vector (i,j) of the output feature map y^l between channels α_{ij}^l, β_{ij}^l, and γ_{ij}^l are the spatial importance weights of the three different levels of feature mappings learned adaptively by the network. Note that α_{ij}^l, β_{ij}^l, and γ_{ij}^l can be simple scalar variables, shared across all channels, and satisfy $\alpha_{ij}^l + \beta_{ij}^l + \gamma_{ij}^l = 1$, with $\alpha_{ij}^l, \beta_{ij}^l, \gamma_{ij}^l \in [0,1]$. The weight maps λ_α^l, λ_β^l and λ_γ^l are calculated from $x^{1 \to l}$, $x^{2 \to l}$ and $x^{3 \to l}$ via 1 × 1 convolution layers, and they are learned through standard backpropagation.

3. Experiments and Results

3.1. Experimental Setup and Evaluation Indicators

All experiments were run in the same environment with the hardware configuration of an i9-13900K CPU, NVIDIA RTX 4090 and the code framework of PyTorch 2.6.0. To quantitatively evaluate the detection methods for forestry pests, Precision (P), Recall (R), F1 score (F1) and Average Precision (AP) are applied as the evaluation metrics for the

performance of the pest identification model. In all experiments, for the improved YOLOv11 and comparison methods, the number of training epochs is set to 50, the initial learning rate is 0.01, the momentum parameter of Stochastic Gradient Descent (SGD) is set to 0.937, and the batch size is set to 16.

3.2. Ablation Experiment

To validate the effect of the improved YOLOv11 model with SD loss, AReLU and ASFF, the control variates method is employed to uniformly apply the ForestryPest84 dataset from this research to evaluate the improved YOLOv11 and the baseline model YOLOv11n. Table 2 shows the results using different configurations. Baseline + SD loss is better than Baseline in all evaluation metrics. Baseline + SD loss outperforms Baseline in the evaluation metric P, while other metrics slightly decline. This indicates that SD loss has an improvement on P as SD loss effectively reduces the impact of IoU fluctuations on the model by dynamically adjusting the loss coefficient. The model's regression ability is enhanced on targets of different scales, and the target identification stability and detection accuracy are improved. Baseline + SD loss + AReLU outperforms Baseline + SD loss in terms of evaluation metrics P, R and F1 by 0.24%, 0.07% and 0.15%, respectively, while $AP_{50–90}$ slightly decreases. The results show that using the AReLU activation function to learn element-level residuals via the attention module can effectively amplify the gradients propagated from the downstream for the activated input, thereby alleviating the problem of gradient disappearance. Furthermore, the evaluation metrics P, R and F1 are improved. At the same time, it can flexibly adjust the activation effect when dealing with complex and diverse data. Baseline + SD loss + AReLU + ASFF outperforms Baseline + SD loss + AReLU in all evaluations except for P. This indicates that ASFF enables the model to more reasonably utilize multi-scale feature information, avoiding the problem of training efficiency decline caused by feature conflicts. Furthermore, the model's detection accuracy for different-scale targets is improved in complex scenarios, enhancing the overall performance and robustness of the model. Obviously, the overall performance of the improved YOLOv11 model is superior to that of YOLOv11. By merely adding 1.88 M parameters, P has increased by 0.38%, R has risen by 0.01%, $AP_{50–90}$ has improved by 0.29%, and F1 has increased by 0.20%. From the overall effect, the improved YOLOv11 model that integrates the SD loss, AReLU, and ASFF modules has significant improvements in terms of the evaluation metrics of P, $AP_{50–90}$, and F1.

Table 2. P, R, $AP_{50–90}$, F1 (%) and parameters of the model in the ablation experiment. Baseline denotes the YOLOv11n model.

Method	P	R	$AP_{50–90}$	F1	Parameters/M
Baseline (YOLOv11)	96.81	97.19	87.96	97.00	2.58
Baseline + SD loss	97.19	96.51	87.80	96.85	2.63
Baseline + SD loss + AReLU	97.43	96.58	87.74	97.00	2.63
Baseline + SD loss + AReLU + ASFF (Improved YOLOv11)	97.19	97.20	88.25	97.19	4.46

The Precision-Recall curves of YOLOv11 and improved YOLOv11 models are shown in Figure 7. The gray curve represents the detection accuracy of different classes of forestry pests, while the blue curve represents overall average accuracy. From the gray curve, it can be seen that improved YOLOv11 is overall superior to YOLOv11. From the blue curve, it is difficult to compare the differences in the enclosed areas. Thus, the comparison is made based on the values on the right side of the image, i.e., improved YOLOv11 (98.40%) is superior to YOLOv11 (98.30%). The detection results shown in the upper subfigure of Figure 8 only have one pest target (*Darna furva*); obviously, improved YOLOv11 (96.00%) is superior to YOLOv11 (93.00%). The middle subfigure of Figure 8 has two pests (*Orgyia*

postica (*O. postica*)). The improved YOLOv11 (87.00% and 92.00%) is superior to YOLOv11 (86.00% and 90.00%). The lower subfigure of Figure 8 has six pests (*P. latus*) and the environment is relatively complex. Improved YOLOv11 still outperforms YOLOv11, while YOLOv11 identified seven targets and the recognition accuracy of *P. latus* is relatively low. In conclusion, the detection performance of the improved YOLOv11 model is superior to that of the YOLOv11 model.

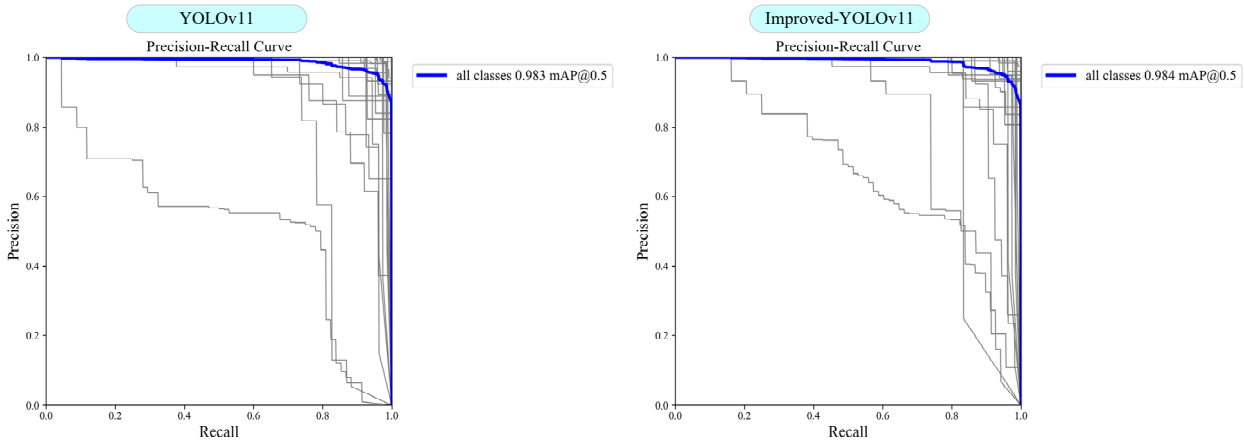

Figure 7. The Precision-Recall curves of YOLOv11 and improved YOLOv11.

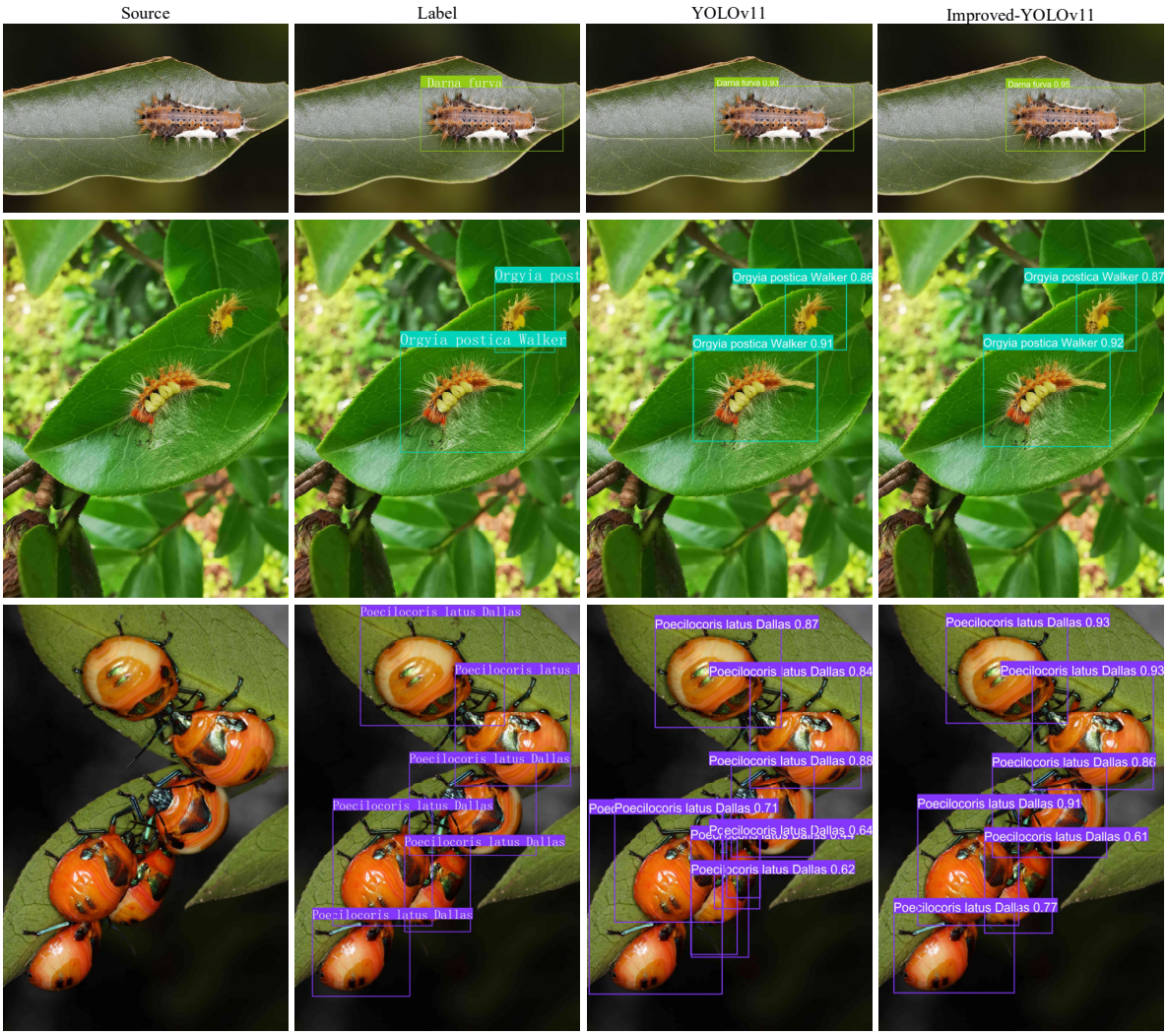

Figure 8. Detection results of the improved YOLOv11 model and Baseline.

3.3. Intraspecific Experiment

This section presents experiments verifying differences in morphological features among different life stages within the same species. Considering the significant variations in color and morphological features of pests across multiple growth stages, which lead to suboptimal pest identification performance in practical applications, this subsection merges adults, larvae, and nymphs of the same pest species into a single class based on the ForestryPest84 dataset (thereby constructing the ForestryPest77 dataset) to conduct experiments validating intra-species feature differences in forestry pests. As shown in Table 3, the YOLOv11 and improved YOLOv11 models trained on the ForestryPest84 dataset outperform those trained on ForestryPest77 across all evaluation metrics, including P, R, $AP_{50–90}$, and F1 score. As illustrated in Figure 9, both the YOLOv11 and improved YOLOv11 models trained on ForestryPest84 demonstrate superior identification performance compared to those trained on ForestryPest77, particularly for *B. ferruginea* and *B. suppressaria*. The experimental results in this subsection indicate that categorizing adults, larvae, and nymphs of the same pest species as separate classes can significantly enhance model performance. Table 3 also compares the model complexity and inference efficiency of YOLOv11 and improved YOLOv11 on ForestryPest77 (77 classes) and ForestryPest84 (84 classes). It can be seen that the number of parameters (Parameters) has a weak positive correlation with the number of dataset classes: as the number of classes increases from 77 to 84, the number of parameters of both models increases slightly. Specifically, YOLOv11 increases from 2,611,961 to 2,627,724 (approximately +0.6%), and improved YOLOv11 increases from 4,396,797 to 4,464,528 (approximately +1.5%). This is mainly due to the linear expansion of the output dimensions of the detection head classification branch with the increase in the number of classes, while the parameters of the model backbone and feature fusion part remain unchanged, thus limiting the increase in total parameters to within 2%. At the same time, the inference delay (Latency) is almost unaffected: the delay of YOLOv11 decreases from 2.5 ms to 2.7 ms, and that of improved YOLOv11 decreases from 2.4 ms to 2.5 ms, with the change range being within 0.2 ms. It is worth noting that although improved YOLOv11 introduces modules such as CBA and ASFF, resulting in its parameter count being significantly higher than that of the original YOLOv11 (approximately +68%), due to its faster convergence and more efficient computation, the delay on the 84-class dataset (2.5 ms) is actually lower than that of YOLOv11 on the same dataset (2.7 ms). The above results indicate that the category expansion brought by the intra-species classification strategy only causes a minor increase in the parameters of the detection head and has almost no impact on the model's inference speed. This shows that fine-grained intra-species classification does not increase the actual deployment burden while significantly improving the recognition granularity, and is feasible for real-time deployment on resource-constrained edge devices in forests.

Table 3. Comparison of detection performance of the YOLOv11 and improved YOLOv11 models trained on the ForestryPest84 and ForestryPest77 dataset (%).

Method	Dataset	P	R	$AP_{50–90}$	F1 Score	Parameters	Latency/ms
YOLOv11	ForestryPest77	95.97	96.37	86.63	96.17	2,611,961	2.5
YOLOv11	ForestryPest84	96.81	97.19	87.96	97.00	2,627,724	2.7
Improved YOLOv11	ForestryPest77	95.82	97.08	86.92	96.45	4,396,797	2.4
Improved YOLOv11	ForestryPest84	97.19	97.20	88.25	97.19	4,464,528	2.5

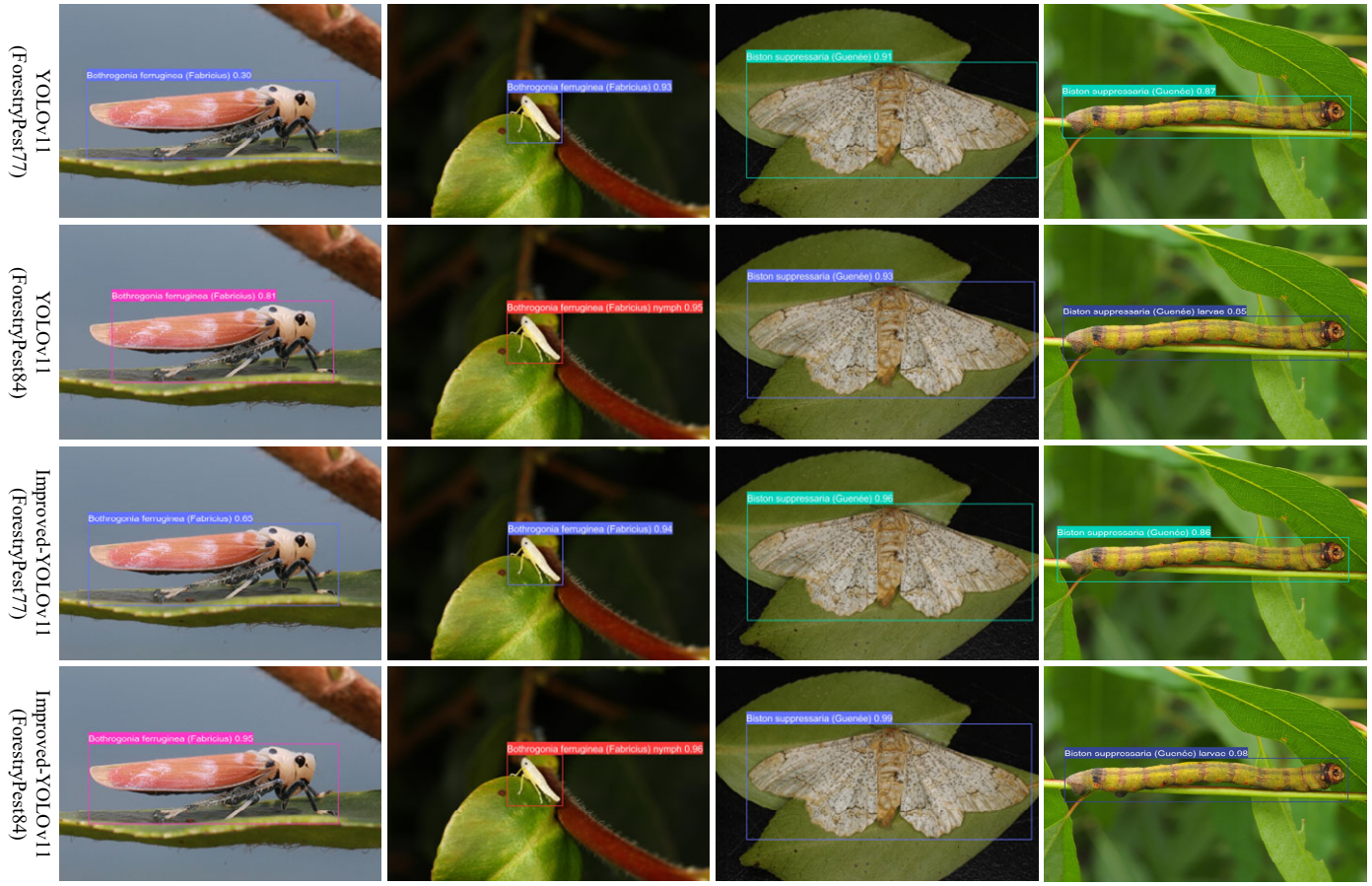

Figure 9. Example of experimental results showing the differences in features among different stages within the same pest species (YOLOv11 and improved YOLOv11).

3.4. Quantitative Analysis

To comprehensively analyze the performance of the detection model in locating and identifying forestry pests, this subsection presents a quantitative analysis. Specifically, the proposed improved YOLOv11 and the comparison methods (YOLOv3 [28], YOLOv5, YOLOv7 [29], YOLOv8 [30] and YOLOv10 [31]) are evaluated on the ForestryPest84 dataset. As shown in Table 4, improved YOLOv11 outperformed YOLOv3, YOLOv5, YOLOv8, and YOLOv10 in evaluation metrics P, R, $AP_{50–90}$, and F1, with values of 97.19%, 97.20%, 88.25%, and 97.17% respectively. It is inferior to YOLOv7. In terms of parameters, improved YOLOv11 had fewer parameters than YOLOv3 and YOLOv7 and more parameters than YOLOv5, YOLOv8, and YOLOv10, with 4.46 M. In terms of detection time (detection time = preprocessing time + inference time + post-processing time), improved YOLOv11 outperformed YOLOv3, YOLOv5, YOLOv7, and YOLOv10, but is inferior to YOLOv8 and YOLOv10, with a latency of 2.5 ms. Although improved YOLOv11 is inferior to YOLOv7 in the evaluation metrics R, $AP_{50–90}$, and F1, considering the actual application, where model parameters and detection time are taken into account to reduce computing power usage and achieve real-time recognition, and taking into account all evaluation metrics, the proposed improved YOLOv11 model is the best. In summary, improved YOLOv11 has made a trade-off between accuracy, model parameters, and latency. Considering multiple indicators and the results of ablation experiments, the improved YOLOv11 model is more suitable for forestry pest detection and identification.

Table 4. Comparison of detection performance of different models (%).

Method	Dataset	P	R	$AP_{50–90}$	F1	Parameters/M	Latency/ms
YOLOv3	ForestryPest84	96.81	96.55	84.77	96.68	61.51	11.1
YOLOv5	ForestryPest84	95.66	96.17	80.92	95.91	1.76	5.7
YOLOv7	ForestryPest84	97.12	98.06	89.31	97.59	36.93	8.6
YOLOv8	ForestryPest84	96.93	96.64	87.49	96.78	3.01	1.9
YOLOv10	ForestryPest84	90.70	90.00	78.20	90.35	2.70	2.3
YOLOv11	ForestryPest84	96.81	97.19	87.96	97.00	2.63	2.7
Improved YOLOv11	ForestryPest84	97.19	97.20	88.25	97.17	4.46	2.5

3.5. Qualitative Analysis

To further evaluate the performance of the improved YOLOv11 model from a visual perspective, this subsection presents qualitative analysis results comparing improved YOLOv11 with the baseline methods, as shown in Figures 10–12. In Figure 10, improved YOLOv11 performs well, outperforming YOLOv3 and YOLOv5. In Figure 11, where the image contains two pests (*O. postica*), improved YOLOv11 outperforms YOLOv3, YOLOv5, and YOLOv10. In Figure 12, with six pests (*P. latus*) and a complex environment, improved YOLOv11 achieved the best performance, with no misdetections or missed detections. As a whole, its detection accuracy surpassed YOLOv7. Meanwhile, YOLOv3 missed three targets, YOLOv5 missed one target, YOLOv8 detected one extra target, and YOLOv10 missed four targets. In summary, improved YOLOv11 demonstrates a more marked advantage in identifying forestry pests within complex scenes involving multiple targets.

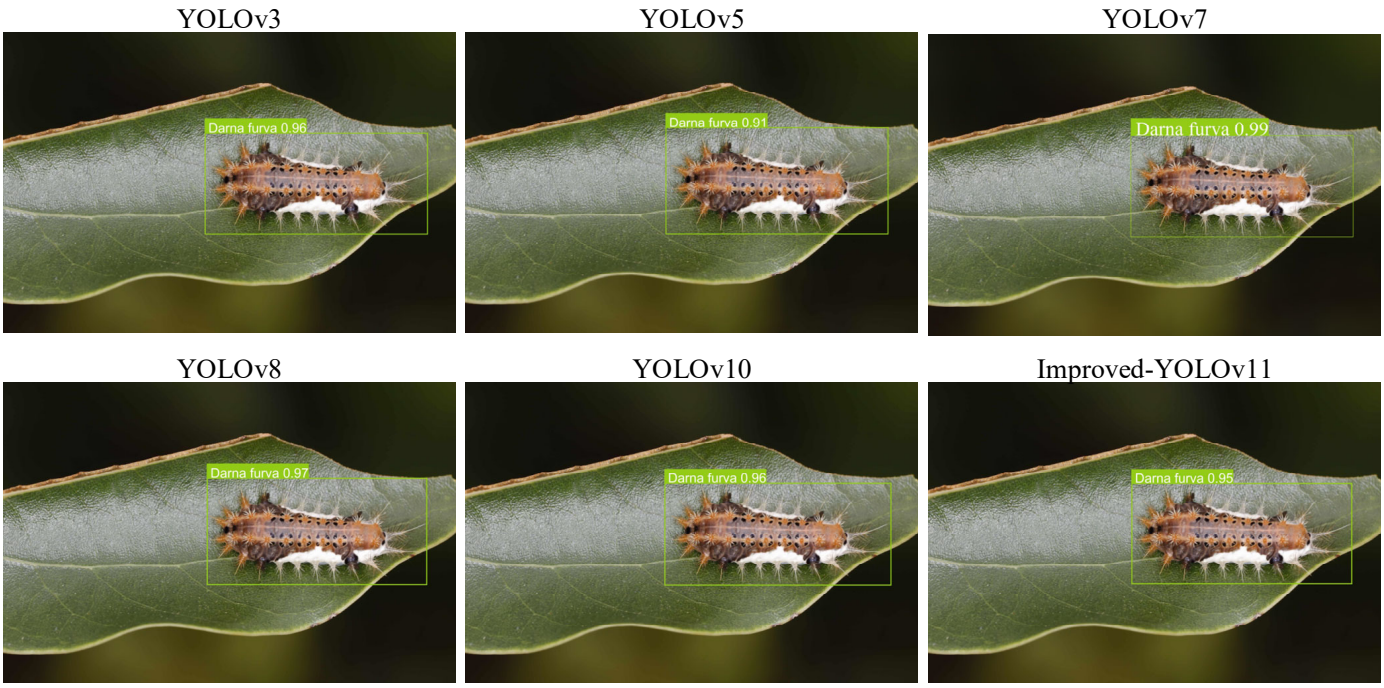

Figure 10. Comparison of individual object detection results from different models.

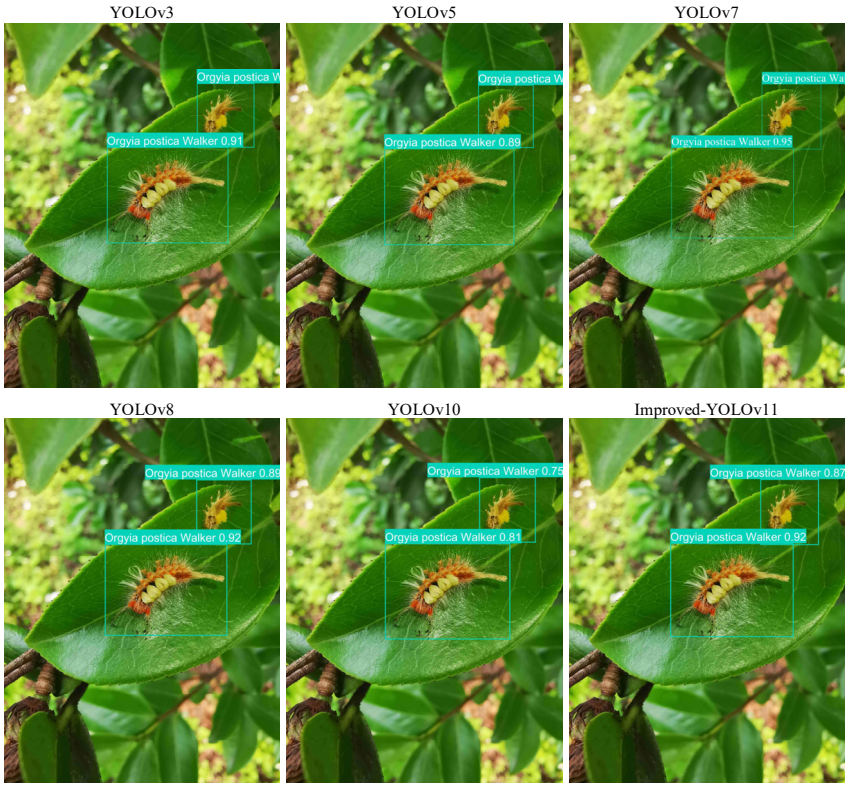

Figure 11. Comparison of two object detection results from different models.

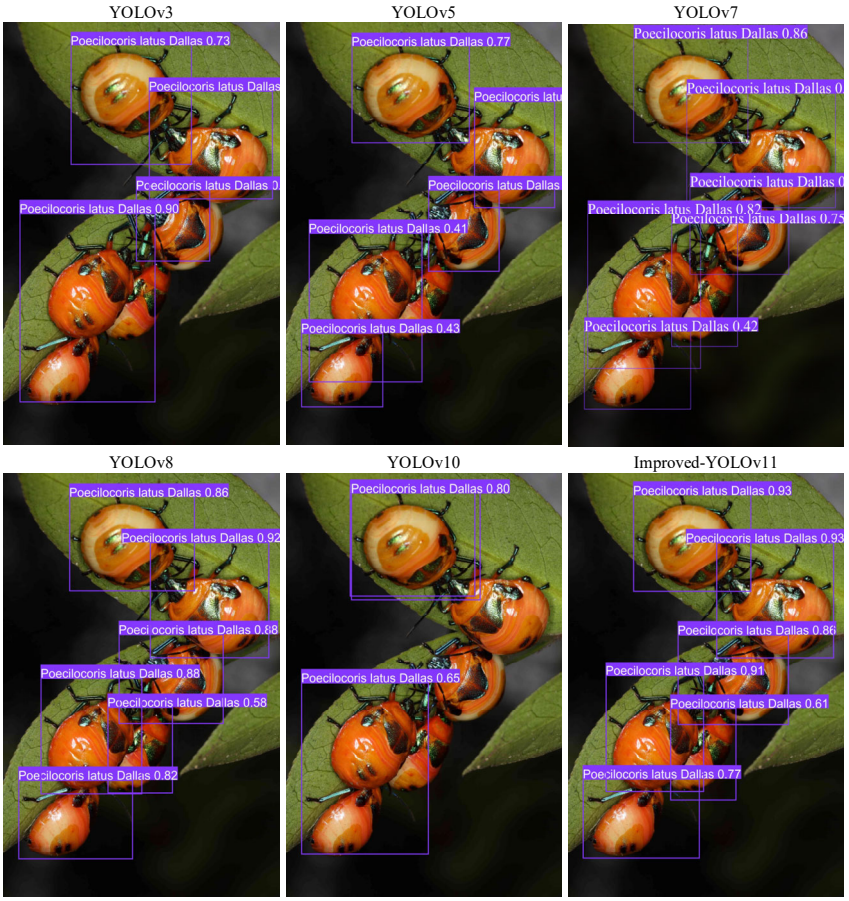

Figure 12. Comparison of multiple object detection results from different models.

3.6. Comparison of Different Models with Respect to the IP102 Dataset

To further verify the advantages of the proposed improved YOLOv11 model in detecting insect images, training is carried out on the IP102 dataset [32] with 50 epochs, and compared with State-of-the-Art (SOTA) methods (FRCNN, FPN, SD300, RefineDet, YOLOv3, C3M-YOLO and Paddy-YOLO [32–34]) applied to the IP102 dataset. The IP102 dataset was established by Wu et al. [32] as a benchmark for large-scale pest detection in agriculture, offering a robust resource for identification tasks. This publicly accessible repository comprises 102 unique pest classes, supported by a collection of roughly 22,284 images. Notably, the dataset provides systematic coverage of the entire pest life cycle, from larval to adult stages. It can be accessed via the GitHub website (https://github.com/xpwu95/IP102 (accessed on 3 March 2026)). The results of the improved YOLOv11 and SOTA methods are shown in Table 5. Compared with the SOTA methods, improved YOLOv11 achieves a P of 60.18%, an R of 59.16%, an AP_{50} of 59.03%, an $AP_{50–90}$ of 38.10%, and an F1 score of 59.67%. Among them, P, R, AP_{50}, and F1 outperform all compared methods, while $AP_{50–90}$ is comparable to that of Paddy-YOLO (38.50% vs. 38.10%). It should be emphasized that the intraspecific classification strategy is not applied to IP102, since this dataset provides only species-level labels and its field-captured images do not allow reliable differentiation of developmental stages. Therefore, the improvements on IP102 arise solely from architecture-level designs, namely the CBA module, SD loss, and the ASFF module, indicating that these components generalize well across datasets. Nevertheless, the absolute performance on IP102 remains moderate, and the gap with our custom dataset is analyzed as follows.

Table 5. Comparison of detection performance of the SOTA methods (%).

Method	P	R	AP_{50}	$AP_{50–90}$	F1
FRCNN [32]	21.05	-	47.87	-	-
FPN [32]	28.10	-	54.93	-	-
SD300 [32]	21.49	-	47.21	-	-
RefineDet [32]	22.84	-	49.01	-	-
YOLOv3 [32]	25.67	-	50.64	-	-
C3M-YOLO [33]	57.40	-	57.20	34.90	-
Paddy-YOLO [34]	52.50	52.30	54.30	38.50	52.4
Improved YOLOv11	60.18	59.16	59.03	38.10	59.67

Analysis of the performance gap on IP102. The absolute performance of improved YOLOv11 on IP102 is notably lower than that on our custom dataset, which can be attributed to four factors. (i) Category scale and imbalance: IP102 covers 102 pest classes with only about 22,284 images in total, averaging roughly 218 images per class, and follows a long-tailed distribution in which the majority of classes suffer from severe sample scarcity, limiting the learning of discriminative fine-grained features. (ii) Higher inter-class similarity: IP102 contains a large number of closely related species with highly similar appearances (e.g., moths and their sibling species), making misclassification far more likely than on our custom dataset. (iii) Complex natural backgrounds: IP102 images are captured in real field environments with occlusion, illumination variations, and interference from host plants, whereas a considerable proportion of our custom dataset was collected under relatively standardized imaging conditions. (iv) Absence of the intraspecific classification strategy: On our custom dataset, the intraspecific strategy splits each species by developmental stage, providing fine-grained supervision that effectively alleviates feature confusion among similar categories; this mechanism is unavailable on IP102, which accounts for a substantial part of the performance gap. Overall, the IP102 experiment serves as a cross-dataset generalization validation rather than a proof of universal generalization capability; the

generalization of the proposed model to broader and more complex pest datasets warrants further investigation on diverse benchmarks in future work.

These results indicate that the improved YOLOv11 model, through its CBA module, SD loss, and ASFF module, improves detection accuracy on both the custom dataset with the intraspecific classification strategy and the public IP102 dataset without it, demonstrating the effectiveness and transferability of the proposed designs. The moderate absolute performance on IP102 suggests that further gains may be expected when fine-grained strategies are applicable, and that the model's generalization to larger and more complex datasets remains to be fully validated.

3.7. Discussion

Regarding the cost and deployment analysis of intraspecific classification strategies, it should be noted that the intraspecific classification strategy introduces additional costs in the annotation process: annotators need to determine the developmental stage of each individual beyond the species label. However, this incremental burden is limited. Firstly, the morphological features used for determining the developmental stage (such as body size, the development degree of wing buds, and body segment structure) are macroscopic and intuitive, and annotators with basic plant protection knowledge can complete this task after a short training period, without relying on insect classification experts to identify samples one by one. Secondly, the number of categories has only increased from 77 in ForestryPest77 to 84 (an increase of approximately 9.1%), indicating that intraspecific splitting only involves some species categories, rather than reconstructing all the data. More importantly, the stage labels are a one-time investment, and the stage-level recognition ability obtained from them enables the control measures to precisely correspond to the sensitive developmental periods of the pests, and the long-term application shows that the benefits are much higher than the costs. At the deployment level, a small increase in the number of categories only linearly expands the output dimension of the detection head, and the impact on model parameters and inference speed can be ignored. At the same time, images of different developmental stages of the same species complement each other as supervisory signals, enhancing the model's ability to represent the morphological diversity of the species, which is one of the sources of the improvement in accuracy and recall rate in this paper. This strategy is not the optimal choice in all scenarios: for tasks that only require species-level incidence surveys, the marginal benefit of stage-level annotation is limited, and in this case, the coarse-grained solution is more cost-effective. In practical applications, one should weigh the identification granularity, annotation cost, and pest control benefits when choosing between different options.

4. Conclusions

In this paper, an improved YOLOv11 model is proposed for forestry pest detection. Specifically, SD loss, AReLU and ASFF are introduced. SD loss dynamically adjusts the influence coefficients of scale loss and position loss based on target scale, reducing the impact of IoU fluctuations on the model. The AReLU activation function learns residuals at the element level through an attention mechanism. It effectively amplifies gradients propagated from downstream layers for activated inputs, thereby mitigating the vanishing gradient problem. The ASFF module adapts by learning fusion weights for feature maps at different scales. This enhances the utilization of multi-scale feature information and prevents training efficiency degradation caused by feature conflicts. Furthermore, considering the differences in features among various life stages within the same species, the adult, larval and nymphal stages of pests of the same species are each classified into one class, and the differences in features among different life stages within the species are analyzed. Experimental results

demonstrate that the proposed improved YOLOv11 model achieves P, R, $AP_{50–90}$, and F1 of 97.19%, 97.20%, 88.25%, and 97.17%, respectively, effectively mitigating gradient vanishing and multi-scale interference in complex environments and the differences in features among different stages within the same pest species. This research enables the identification of forestry pests, safeguarding forest ecosystems and ensuring the healthy, rapid development of the forestry industry. The proposed model can accurately identify types of pests and their occurrence stages, providing support for targeted and precise pest control. It should be noted that the additional annotation cost brought about by intraspecific classification is one-time, and the impact of category expansion on the model's inference cost can be ignored. The intraspecific classification strategy of this study relies on accurate labeling of developmental stages. For transitional individuals with blurred stage boundaries (such as late-stage larvae), consistency of labeling still requires expert experience to ensure identification; in the future, methods for automatic verification of uncertain labels based on uncertainty can be explored to further reduce this reliance.

Author Contributions: Conceptualization, X.L. and W.W.; methodology, X.L.; validation, J.X., S.Q., J.L., D.Z. and W.W.; formal analysis, W.W.; resources, K.P. and N.T.T.; data curation, D.Z. and W.L.; writing—original draft preparation, X.L.; writing—review and editing, X.L., J.X., D.Z., S.Q., W.L., J.L. and W.W.; supervision, W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Guangxi Science and Technology Major Project (Department of Science and Technology of Guangxi Zhuang Autonomous Region) under grant GuikeAA24263019, and the Research team project of Guangxi Forestry Research Institute (Guangxi Forestry Research Institute) under grant 2025RT07; in part by the Guangxi Public-Private Partnership Science and Technology Project (Department of Science and Technology of Guangxi Zhuang Autonomous Region) under grant GuikeJB2504240004, Guangxi Science and Technology Major Project (Department of Science and Technology of Guangxi Zhuang Autonomous Region) under grant GuikeAA24263008 and GuikeAA24263014, and Central Government Finance Promotion Project Major Project (Department of Forestry of Guangxi Zhuang Autonomous Region) under grant [2026]TG03.

Data Availability Statement: The source code and data during this study are not publicly available due to the IP management policy of the institution but are available from the corresponding author on reasonable request.

Acknowledgments: We thank Kanyakorn Piraonapicha at Queen Sirikit Botanic Garden (Thailand) and Nguyen Thanh Tuan at Vietnam National University of Forestry in Dong Nai (Vietnam) for collecting the pest images.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fei, S.; Morin, R.S.; Oswalt, C.M.; Liebhold, A.M. Biomass Losses Resulting from Insect and Disease Invasions in US Forests. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 17371–17376. [CrossRef]
2. Liang, X.; Yang, L.; Liao, W.; Huang, L.; Chen, J.; Yang, W.; Meng, F.; Huang, C.; Wei, W.; Wang, G. Ecological Identification of *Camellia* spp. Pests Based on Convolutional Neural Networks. *Guangxi For. Sci.* **2023**, *52*, 361–366. [CrossRef]
3. Liang, X.; Xiang, J.; Qin, S.; Xiao, Y.; Chen, L.; Zou, D.; Ma, H.; Huang, D.; Huang, Y.; Wei, W. Small Target Detection Algorithm Based on SAHI-Improved-YOLOv8 for UAV Imagery: A Case Study of Tree Pit Detection. *Smart Agric. Technol.* **2025**, *12*, 101181. [CrossRef]
4. Xiang, J.; Liang, X.; Jiang, J.; Zou, D. GIBN-Net for Automated Windthrown Tree Detection in Eucalyptus Plantations Using UAV Imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *146*, 105121. [CrossRef]
5. Nanni, L.; Maguolo, G.; Pancino, F. Insect Pest Image Detection and Recognition Based on Bio-Inspired Methods. *Ecol. Inform.* **2020**, *57*, 101089. [CrossRef]
6. Huang, M.L.; Chuang, T.C.; Liao, Y.C. Application of Transfer Learning and Image Augmentation Technology for Tomato Pest Identification. *Sustain. Comput. Inform. Syst.* **2022**, *33*, 100646. [CrossRef]

7. Mishra, M.; Singh, P.K.; Brahmachari, A.; Debnath, N.C.; Choudhury, P. A Robust Pest Identification System Using Morphological Analysis in Neural Networks. *Period. Eng. Nat. Sci.* **2019**, *7*, 483–495. [CrossRef]
8. Li, W.; Chen, P.; Wang, B.; Xie, C. Automatic Localization and Count of Agricultural Crop Pests Based on an Improved Deep Learning Pipeline. *Sci. Rep.* **2019**, *9*, 7024. [CrossRef]
9. Zhong, Y.; Gao, J.; Lei, Q.; Zhou, Y. A Vision-Based Counting and Recognition System for Flying Insects in Intelligent Agriculture. *Sensors* **2018**, *18*, 1489. [CrossRef]
10. Wang, Q.J.; Zhang, S.Y.; Dong, S.F.; Zhang, G.C.; Yang, J.; Li, R.; Wang, H.Q. Pest24: A Large-Scale Very Small Object Data Set of Agricultural Pests for Multi-Target Detection. *Comput. Electron. Agric.* **2020**, *175*, 105585. [CrossRef]
11. Liu, L.; Wang, R.; Xie, C.; Yang, P.; Wang, F.; Sudirman, S.; Liu, W. PestNet: An End-to-End Deep Learning Approach for Large-Scale Multi-Class Pest Detection and Classification. *IEEE Access* **2019**, *7*, 45301–45312. [CrossRef]
12. Xiang, Q.; Huang, X.; Huang, Z.; Chen, X.; Cheng, J.; Tang, X. Yolo-Pest: An Insect Pest Object Detection Algorithm via CAC3 Module. *Sensors* **2023**, *23*, 3221. [CrossRef]
13. Mu, Y.; Luo, L.; Ni, L.; Sun, Y.; Gong, H. An Efficient and Accurate Pest Detection Model Based on Improved YOLOv5s Architecture for Complex Agricultural Environments. *Appl. Soft Comput.* **2026**, *187*, 114329. [CrossRef]
14. Chen, G.; Wang, J. TSRA-Net: A Pest Detection Method Based on Token Statistics and Reparameterized Adaptive Networks. *Expert Syst. Appl.* **2026**, *304*, 130659. [CrossRef]
15. Yang, X.; Hu, H.; Chen, X.; Li, W.; Zhou, Z.; Li, W. Lightweight Detection and Recognition Model for Small Target Pests on Sticky Traps in Multi-Source Scenarios. *Smart Agric.* **2025**, *7*, 111–123. [CrossRef]
16. Dai, Q.; Liang, S.; Li, Z.; Lyu, S.; Xue, X.; Song, S.; Huang, Y.; Zhang, S.; Fu, J. YOLOv11-RDTNet: A Lightweight Model for Citrus Pest and Disease Identification Based on an Improved YOLOv11n. *Agronomy* **2025**, *15*, 1252. [CrossRef]
17. Yu, G.; Ma, B.; Zhang, R.; Xu, Y.; Lian, Y.; Dong, F. CPD-YOLO: A Cross-Platform Detection Method for Cotton Pests and Diseases Using UAV and Smartphone Imaging. *Ind. Crops Prod.* **2025**, *234*, 121515. [CrossRef]
18. Junjie, H.; Zheng, M.; Yuzhu, W.; Yujian, B.; Yizhe, W. YOLOv8-DDS: A Lightweight Model Based on Pruning and Distillation for Early Detection of Root Mold in Barley Seedling. *Inf. Process. Agric.* **2025**, *12*, 581–594. [CrossRef]
19. Wang, R.; Dong, S.; Jiao, L.; Du, J.; Huang, Z.; Zheng, S.; Kang, C. OSAF-Net: A One-Stage Anchor-Free Detector for Small-Target Crop Pest Detection. *Appl. Intell.* **2023**, *53*, 24895–24907. [CrossRef]
20. Prathima, K.; Kanchan, R.G.; Arekal, S.; Shalini, A.N.; Mishra, G. Agricultural Pests and Disease Detection. In *Proceedings of the 2021 International Conference on Forensics, Analytics, Big Data, Security*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6.
21. Lou, L.; Liu, J.; Yang, Z.; Zhou, X.; Yin, Z. Agricultural Pest Detection Based on Improved Yolov5. In *Proceedings of the ACM International Conference Proceeding Series*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 7–12.
22. Kuzuhara, H.; Takimoto, H.; Sato, Y.; Kanagawa, A. Insect Pest Detection and Identification Method Based on Deep Learning for Realizing a Pest Control System. In *Proceedings of the 2020 59th Annual Conference of the Society of Instrument and Control Engineers of Japan, SICE*; IEEE: Piscataway, NJ, USA, 2020; Volume 2020, pp. 709–714.
23. An, J.; Du, Y.; Hong, P.; Zhang, L.; Weng, X. Insect Recognition Based on Complementary Features from Multiple Views. *Sci. Rep.* **2023**, *13*, 2966. [CrossRef]
24. Khanam, R.; Hussain, M. YOLOv11: An Overview of the Key Architectural Enhancements. *arXiv* **2024**, arXiv:2410.17725.
25. Chen, D.; Li, J.; Xu, K. AReLU: Attention-Based Rectified Linear Unit. In Proceedings of the ICLR 2021, Virtual, 3–7 May 2021; pp. 1–18.
26. Yang, J.; Liu, S.; Wu, J.; Su, X.; Hai, N.; Huang, X. Pinwheel-Shaped Convolution and Scale-Based Dynamic Loss for Infrared Small Target Detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI Press: Washington, DC, USA, 2025; Volume 39, pp. 9202–9210. [CrossRef]
27. Liu, S.; Huang, D.; Wang, Y. Learning Spatial Fusion for Single-Shot Object Detection. *arXiv* **2019**. [CrossRef]
28. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE Computer Society: Washington, DC, USA, 2018; pp. 1–6.
29. Wang, C.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. *arXiv* **2022**. [CrossRef]
30. Reis, D.; Kupec, J.; Hong, J.; Daoudi, A. Real-Time Flying Object Detection with YOLOv8. *arXiv* **2023**. [CrossRef]
31. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv* **2024**. [CrossRef]
32. Wu, X.; Zhan, C.; Lai, Y.K.; Cheng, M.M.; Yang, J. IP102: A Large-Scale Benchmark Dataset for Insect Pest Recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2019; Volume 1, pp. 8779–8788.

33. Zhang, L.; Zhao, C.; Feng, Y.; Li, D. Pests Identification of IP102 by YOLOv5 Embedded with the Novel Lightweight Module. *Agronomy* **2023**, *13*, 1583. [CrossRef]
34. Zheng, Y.; Zheng, W.; Du, X. Paddy-YOLO: An Accurate Method for Rice Pest Detection. *Comput. Electron. Agric.* **2025**, *238*, 110777. [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.