Current Science
CS 5 (6), 5251-5271 (2025)
Received 15 Oct 2025 | **Accepted** 30 October 2025 | **Online Available** 04 November 2025
https://doi.org/10.52845/CS/2025-5-6-2

Current Science
Current Science

OPEN ACCESS **ISSN(O)2795-8639**

Original Article

YOLO-RicePest: High-Precision Rice Pest Detection Method Based on Improved YOLOv5s

Xiuhui Liao[1*]

[1]College of information and Intelligence, Hunan Agricultural University, Changsha 410128, China

***Corresponding Author: Xiuhui Liao**

Abstract:
Rice yields are threatened by diverse pests, demanding accurate field detection. Yet small object scale, high inter-class similarity, and cluttered backgrounds make recognition difficult. We present YOLO-RicePest, an improved YOLOv5s framework. Specifically, the efficient multi-scale attention (EMA) module is integrated after each C3 block in the backbone to strengthen the representation of spatial cues such as texture and shape. This enhances the model's capacity to differentiate visually similar pest categories. In the neck, a content-aware reassembly of features (CARAFE) module is incorporated to adaptively generate reassembly kernels conditioned on input feature content, thereby retaining fine-grained details critical for small-object detection. Furthermore, we design a novel bounding-box regression loss, Inner-Focaler-EIoU, which combines linear interval map-ping with an auxiliary bounding-box mechanism. This loss alleviates the deficiencies of CIoU in handling small and occluded objects and facilitates faster convergence. To validate the proposed method, we construct a rice pest dataset (RPD) comprising 2,550 images across 18 common pest species, captured using an automated trapping system. Experimental results demonstrate that YOLO-RicePest achieves a mAP50 of 87.9%, substantially surpassing the baseline YOLOv5s model (82.4%), thereby confirming its effectiveness and robustness for rice pest detection.
Keywords: Rice Pest Detection, Object Detection, Attention Mechanism, YOLOv5s.

1. Introduction

Agriculture, as the foundational industry upon which humanity depends for survival, retains an irreplaceable importance even amid today's rapid technological advancements—particularly against the backdrop of a growing global population and increasingly severe food security challenges [1]. Rice stands as one of the world's most vital food crops, with its yield having a direct impact on national economies and public welfare. However, rice cultivation frequently faces attacks from multiple pests, severely compromising both yield and quality, and even causing regional crop failures. Timely and accurate identification and control of rice field pests are crucial for ensuring food security and promoting sustainable agricultural development [2].

Currently, the identification and classification of rice pests still primarily rely on professional pest control personnel conducting field surveys and making judgments based on experience [3]. This approach is not only labor-intensive and inefficient but also limited in accuracy by individual expertise, making it difficult to meet the demands of modern agriculture for large-scale, real-time, and high-precision monitoring [4]. Particularly in vast rural areas, farmers generally lack systematic knowledge of pest identification, resulting in relatively weak capabilities for early prevention and control of pests.

Traditional image-based pest detection methods typically involve multiple steps, including image acquisition, augmentation, segmentation, feature extraction, and classification [5]. Image augmentation enhances image quality by applying various transformations to the original image, thereby facilitating more effective extraction of regions of interest. However, such methods heavily rely on manually designed features, resulting in poor generalizability, low efficiency, and difficulty in capturing complex and variable pest features. In recent years, deep learning technology has advanced rapidly, garnering widespread attention across various fields with continuously expanding applications. Particularly in pest detection tasks, deep learning has demonstrated significant advantages, becoming a research hotspot [6]. Extensive studies have focused on effectively applying it to crop pest detection tasks.

Currently, deep learning-based object detection methods primarily fall into two categories: single-stage detection methods and two-stage detection methods. Their detection process typically involves two core steps: feature extraction and classification/localization. Single-stage detection methods feature relatively simple structures, typically employing regression strategies to simultaneously perform object localization and classification across the entire image. Representative examples include You Only Look Once (YOLO) and Single Shot MultiBox Detector (SSD). In contrast, two-stage detection methods divide the detection task into two phases: first, generating candidate object regions through selective search or regional proposal networks (RPNs), followed by classification and bounding box regression on these regions. Among these, the R-CNN family (e.g., Faster R-CNN [7]) excels in detection accuracy but suffers from slow inference speeds, making it unsuitable for applications demanding high real-time performance. In contrast, single-stage detection methods like the YOLO series maintain high detection accuracy while offering significantly faster inference. This makes them more practical for scenarios requiring rapid response, such as agricultural pest detection [8,9].

Against this backdrop, this paper proposes YOLO-RicePest, a rice pest detection network based on an improved YOLOv5s. Its main contributions are as follows:

1. Introducing the EMA [10] module after each C3 module in the backbone network enhances the network's attention to spatial information (e.g., texture, shape) with only a slight increase in parameters, thereby improving its ability to distinguish pests with similar body colors.

2. Replacing the upsampling method in the neck network with the CARAFE [11] module, which adaptively generates recomposed convolutional kernels based on input feature content, thereby preserving more details of small targets. Additionally, this module possesses a larger receptive field, effectively integrating local and contextual information to enhance the model's detection performance for rice pests.

3. Introduce a novel bounding box loss, Inner-Focaler-EIoU, which employs linear interval mapping and auxiliary bounding box mechanisms. This approach not only mitigates the limitations of Complete IoU (CIoU) loss when handling small-sized pests and occlusion issues in rice fields but also accelerates model convergence.

4. A rice pest dataset (RPD) comprising 2,550 high-quality images of 18 common pests was constructed using automated trapping devices. Experimental results demonstrate that YOLO-RicePest achieves an mAP50 of 87.9% in natural environments, outperforming the original YOLOv5s (82.4%), thereby validating the effectiveness and practicality of the proposed method.

2 Related Work

2.1 Traditional Image Processing-Based Pest Detection Methods

Before the widespread adoption of deep learning, pest detection primarily relied on traditional image processing techniques. Early studies predominantly employed manually designed feature extraction methods, such as Gabor wavelet transform, scale-invariant feature transform (SIFT), and histogram of oriented gradients (HOG). Zhang et al. [12] combined Gabor wavelets with image segmentation theory to effectively address the challenge of separating similar-appearing objects from backgrounds in images of stored-grain pests. Lecun et al. [13] employed a multi-resolution segmentation approach to achieve the identification and counting of pests like the rice leaf roller. For feature extraction, researchers often integrated morphological, textural, and local features for multidimensional analysis. Wen et al. [14] further improved classification accuracy for fruit tree pests by fusing local and global image features.

Although these methods have achieved certain successes in specific scenarios, they still face significant limitations in practical applications. On one hand, traditional features exhibit poor adaptability to lighting variations, object occlusions, and pose changes, making them ill-suited for complex agricultural environments. On the other hand, manually designed low-level features lack the capacity to capture high-level semantic information about pests (e.g., body structure or behavioral patterns), resulting in limited generalization capabilities of traditional classifiers (e.g., SVM, KNN) when encountering new pest species or morphological variations.

2.2 Deep Learning-Based Pest Detection Methods

Deep learning-based object detection methods have been widely applied in agricultural pest identification. Early studies often built upon Faster R-CNN, enhancing detection performance for field pests by improving feature extraction networks or region proposal strategies. Guan et al. [15] proposed the GC-Faster RCNN model, incorporating a hybrid attention mechanism, achieving notable results in handling multi-scale variations and highly similar pest interferences. Wang et al. [16] introduced enhanced feature pyramid networks (FPN) [17] to boost small-object feature extraction capabilities, and combined it with ROI Align to significantly improve detection accuracy for small-scale agricultural pests. Furthermore, Mask R-CNN [18], which integrates detection and segmentation capabilities, has been widely applied to pest detection tasks. Its segmentation branch precisely locates pest edges, significantly improving recognition accuracy for complex-shaped pests. Wang et al. [19] proposed an enhanced Mask R-CNN integrating a convolutional block attention module (CBAM) [20], FPN, and a dual-channel downsampling module to improve detection accuracy for small-scale pests.

In recent years, compared to two-stage detectors (such as Faster R-CNN), the YOLO series has gradually become the mainstream direction in pest detection research due to its efficient single-stage architecture and excellent real-time performance. Song et al. [21] replaced the backbone network of YOLOv4 [22] with DenseNet and introduced the Focal Loss function. Experimental results demonstrate that this approach significantly improves detection accuracy for complex pest categories while mitigating sample imbalance issues. Similarly, Liu et al. [23] enhanced tomato pest recognition accuracy by incorporating a ternary attention mechanism based on YOLOv4. Tan et al. [24] proposed an improved approach combining point-line distance intersection over union (PLDIoU) loss and convolutional attention mechanism (CBAM) on a dataset fused with Mosaic-9 and Mixup data augmentation strategies, effectively boosting detection accuracy and efficiency for dense small passion fruit pests. Gao et al. [25] replaced the original PAN with FC-FPN in YOLOv5's neck structure. This adaptive fusion of multi-scale information addresses the challenge of highly variable pest targets. Mo et al. [26] proposed YOLONDD, which combines the novel NMIoU loss function with the dynamic attention head (DyHead). It demonstrated superior accuracy and robustness in multi-scale detection tasks for tomato leaf pests, outperforming multiple mainstream object detection algorithms.

3 Materials and Methods

3.1 Data Collection

This study selected Zhuge Village and Paotuan Village in Jingzhou County, Huaihua City, Hunan Province, China (a typical southern hilly rice-growing region) as sample sites. This area primarily cultivates rice, featuring a humid climate and complex ecological environment, making it a high-risk breeding ground for pests such as rice planthoppers and rice stem borers. Data collection relied on a team-developed rice field insect trapping device (equipped with high-definition cameras) and was conducted from May 20 to August 30, 2022 (the critical rice growth period). The data encompassed diverse weather and day/night lighting conditions, with most captures concentrated during the early morning and dusk hours when pests are most active.

Figure 1 illustrates the overall technical workflow of this study. The process comprises three stages: First, the self-developed device captured extensive pest imagery from rice fields. Raw images underwent processing—including denoising, cropping, and augmentation-to construct a high-quality, large-scale rice pest dataset. Second, the processed image data were fed into both the original YOLOv5s model and the improved model for training, yielding experimental results before and after enhancement. Finally, differences in detection accuracy, recall, and other metrics between the pre- and post-improvement models were analyzed to validate the proposed method's effectiveness.

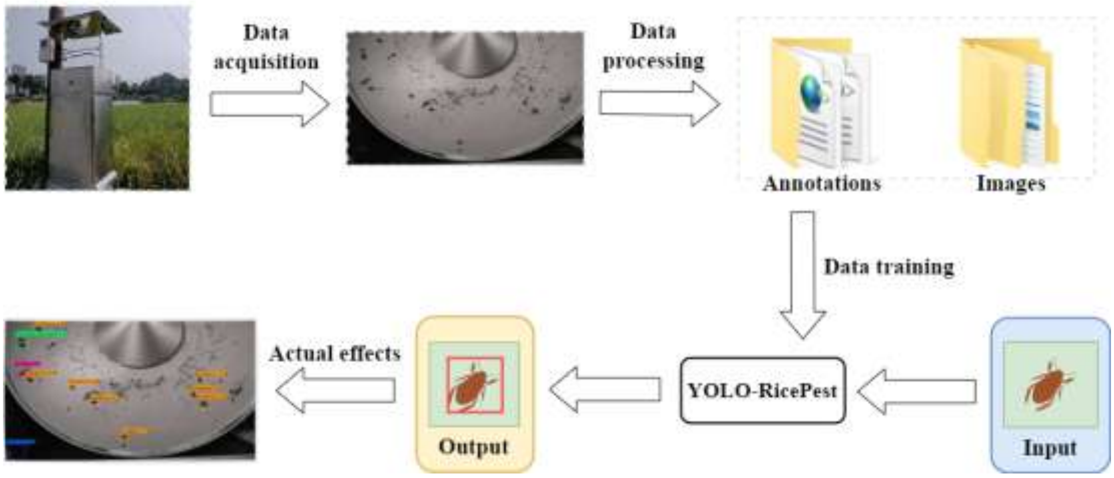

Figure 1 Technology roadmap for rice pest detection task

3.2 RPD Dataset

To construct a high-quality rice pest detection dataset, we screened the original images based on shooting periods and pest activity patterns, removing invalid samples with insufficient lighting, blurred images, or high content redundancy to enhance overall data quality. Subsequently, the retained images underwent manual annotation (as shown in **Figure 2**). Specifically, LabelImg software was used to load samples, where tightly enclosing bounding boxes were drawn around each pest, and corresponding pest categories were assigned to each annotated box.

Figure 2 Visual interface of LabelImg software

To enhance the model's generalization capabilities and adapt to complex, variable field conditions, data augmentation techniques such as brightness enhancement, flipping, and rotation were applied to expand the sample size, as illustrated in **Figure 3**.

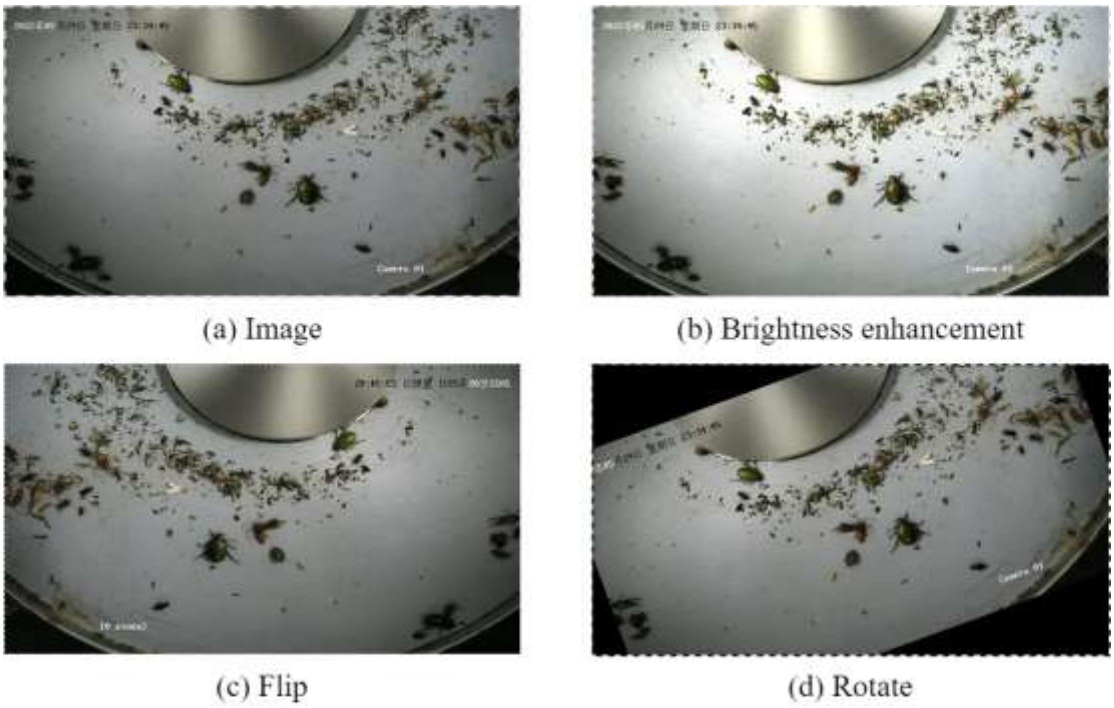

Figure 3 Examples of rice pest image using various data augmentation methods

The final RPD dataset comprises 2,550 images, covering 18 categories and totaling 66,145 rice pests. The detailed distribution of each pest category after data augmentation is shown in **Table 1**. Finally, the dataset was divided into training, validation, and test sets at an 8:1:1 ratio, providing a robust data foundation for model training and performance evaluation.

Table 1 The RPD dataset on the detailed distribution of various pest categories

Category name	Number	Category name	Number
Carabidae	17705	Rove Beetle	1775
NarangaaenescensMoore	1195	Cnaphalocrocis	8215
Chilosuppressalis	6000	Mayfly	650
Amatidae	60	Tropido thoraxelegans	590
Linnaeus	390	Lepidoptera	1085
Gryllotalpidae	110	Dytiscidae	1160
Coleoptera	14125	Hydrophilidae	8375
Melolonthidae	2545	Spilosoma menthastri	1170
Cricket	380	Leaf hopper	615

3.3 Methods

3.3.1 Basis for Selecting the YOLOv5 Model and Architectural Analysis

In the field of object detection, numerous outstanding algorithmic models are available for selection. Choosing an appropriate model is crucial for detecting rice pest images. Considering the unique application scenarios and data characteristics of rice pest images, model selection requires analysis from the following three aspects:

1. First, rice pest detection primarily serves agricultural pest control. Given the vast diversity and sheer volume of pests encountered in real-world farming, the selected model must deliver high detection accuracy while maintaining rapid processing speed to meet large-scale image handling and response demands.

2. Second, rice pest detection faces complex agricultural challenges. On one hand, pests exhibit unpredictable activity patterns, complicating real-time image acquisition; On the other hand, the significant variation in species and body size of pests inhabiting rice fields imposes higher demands on the model's robustness and generalization capabilities. Therefore, achieving precise detection of pest categories requires the model to exhibit excellent stability and adaptability.

3. Finally, considering the future development of intelligent pest detection systems in actual agricultural environments, the model should also possess good portability to enable efficient operation on resource-constrained platforms such as edge computing devices.

Given the YOLO series models' strong balance between detection speed and accuracy, particularly in real-time object detection tasks, this study selected the YOLOv5 model—known for its high stability and excellent generalization capabilities—as the base network for rice pest detection. The YOLOv5 model comprises four components: Input, Backbone, Neck, and Prediction.

The input section includes Mosaic data augmentation, image resizing, and adaptive anchor box calculation strategies. Mosaic augmentation enhances sample diversity by randomly scaling, cropping, and stitching four images, while strengthening batch normalization effects. Image resizing uniformly scales images of varying dimensions, effectively reducing redundant computations and accelerating detection speed. Adaptive anchor calculation evaluates the deviation between predicted and ground truth boxes, then updates anchor parameters in reverse to obtain more suitable anchors and improve detection accuracy.

Backbone consists of Conv, C3, and SPPF modules for extracting deep image features. The C3 module integrates ResNet's residual architecture with CSPNet's [27] cross-stage partial connection concept, enhancing feature extraction while significantly reducing computational load. The SPPF module introduces multi-scale max pooling to expand the model's receptive field and strengthen nonlinear expression capabilities [28].

The neck section employs FPN and PAN structures to achieve multi-scale feature fusion. FPN enhances the model's detection capability for small objects, while PAN further strengthens feature representation by improving information transfer between different scales. Their combination effectively boosts the model's perception of multi-scale objects in complex scenes [29].

In the prediction stage, the model employs CIoU to measure the discrepancy between predicted and ground-truth bounding boxes, thereby improving localization accuracy. Simultaneously, it incorporates the Non-Maximum Suppression (NMS) mechanism [30] to eliminate redundant boxes, enhancing detection performance for multiple and occluded objects. The IoU definition is shown in **Equation 1**:

$$IoU(B,B_{gt}) = \frac{|B \cap B_{gt}|}{|B \cup B_{gt}|} \quad (1)$$

where B and B_{gt} represent two bounding boxes corresponding to a specific target. IoU is defined as the intersection area between box B and box B_{gt} divided by their union area. A higher IoU value indicates greater overlap between the two boxes and more accurate model predictions; conversely, a lower IoU value signifies poorer localization accuracy of the model.

CIoU builds upon IoU by incorporating center point distance penalties and aspect ratio penalties. This enables faster convergence to more accurate prediction boxes, thereby enhancing the model's regression performance. CIoU is defined as shown in **Equations 2** and **3**:

$$CIoU(B,B_{gt}) = IoU(B,B_{gt}) - \frac{\rho^2(B,B_{gt})}{c^2} - \alpha v \quad (2)$$

$$\nu = \frac{4}{\pi}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2 \quad (3)$$

where $\rho^2(B,B_{gt})$ is the squared Euclidean distance between the center points of the B box and the B_{gt} box. c is the diagonal length of the minimum bounding rectangle that simultaneously encloses both the B box and the B_{gt} box. ν is a similarity value measuring the width-to-height ratio between the B box and the B_{gt} box. α is a weight used to balance the similarity of the width-to-height ratios between the two boxes. w_{gt} and h_{gt} are the width and height of the B_{gt} box, respectively, while w and h are the width and height of the B box, respectively.

3.3.2 The Overview of YOLO-RicePest

To address the frequent false negatives, false positives, and overall low recognition accuracy encountered in practical rice pest detection due to small object sizes and occlusion issues, this paper proposes the improved model YOLO-RicePest (as shown in **Figure 4**) based on YOLOv5s. First, the EMA module is embedded after each C3 module in the backbone network to guide the network in automatically adjusting channel feature weights, thereby enhancing focus on key pest features. Second, recognizing that the nearest-neighbor interpolation upsampling method in the original Neck network can lead to loss of small object information, we replace it with the CARAFE module. By expanding the receptive field and integrating content-aware reorganization strategies, CARAFE achieves comprehensive fusion of local details and contextual information, enhancing the retention capability for small objects. Simultaneously, we introduce Inner-Focaler-EIoU as a novel bounding box loss function to more effectively address the performance bottlenecks of CIoU loss when handling small-to-medium objects and occlusion scenarios, thereby improving detection accuracy. Furthermore, we replaced the Leaky ReLU activation function with SiLU. By integrating the advantages of linear transformation and the Sigmoid function, SiLU not only mitigates the "neuron death" issue but also enhances the model's expressive power and generalization capabilities in complex scenarios.

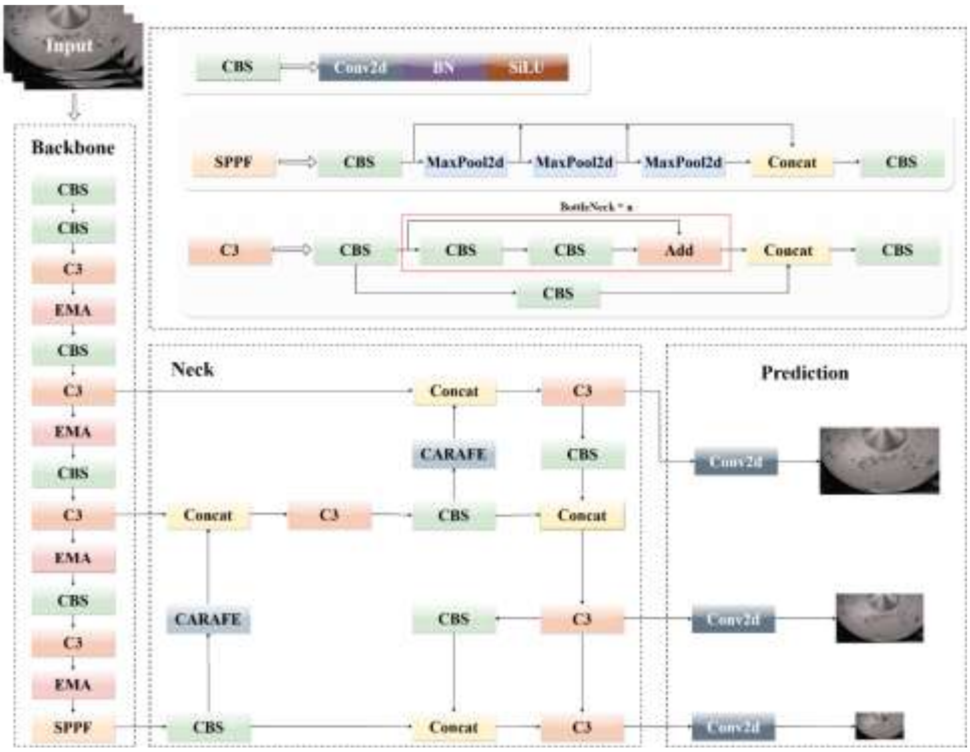

Figure 4 Architecture of YOLO-RicePest

3.3.3 Efficient Multi-scale Attention Module

During the forward propagation process of the model, shallow feature maps primarily focus on low-level information such as texture, contours, and edges. When the body color of a particular rice pest closely resembles that of other pests or the background, subtle differences in color distribution alone are insufficient for precise differentiation. These low-level features become critical for distinguishing between pest species. Therefore, we introduced an EMA module after the four C3 modules in the backbone network. This enhances the network's ability to differentiate between pests with similar colors while maintaining computational efficiency. The detailed structure of the EMA module is shown in **Figure 5**.

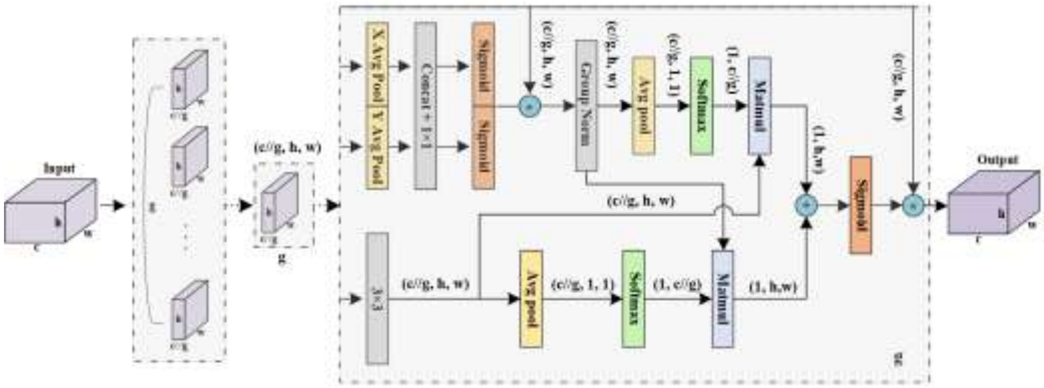

Figure 5 Structural diagram of the EMA module

The EMA module is a lightweight, high-performance attention mechanism comprising three components: feature reconstruction grouping, multi-scale parallel processing paths, and cross-spatial information interaction. By remapping partial channel dimensions to batch dimensions and dividing the entire feature map into multiple subgroups along the channel dimension, this module ensures balanced spatial semantic representation within each subgroup, thereby enhancing the model's perception of information across different scales. Specifically, input features are first divided into g subgroups. Within each subgroup, processing occurs through two distinct branches: a lightweight 1×1 convolution branch and a standard 3×3 convolution branch. The 1×1 branch performs one-dimensional global pooling along both horizontal and vertical directions to capture long-range spatial dependencies while preserving position-dependent structural information. Unlike the channel compression mechanism in the coordinate attention (CA) [31] module, the EMA module bypasses the channel dimension reduction stage. Instead, it directly concatenates the pooling results from both directions and fuses them via 1×1 convolution. The output is then decomposed into feature maps for each direction, and spatial attention weights are generated using a Sigmoid activation function to achieve weighted enhancement of the original input.

Concurrently, the 3×3 branch employs standard convolutions to extract local contextual features, further compensating for the limitations of single-scale processing. Features generated by both branches undergo global average pooling and are normalized via Softmax to dynamically adjust response strengths across channels. Finally, the module further integrates information through cross-dimensional interaction mechanisms to capture pixel-level pairwise relationships. Compared to CBAM and CA, the EMA module enhances the model's ability to identify pests by enabling it to capture multi-scale feature representations of objects and increase focus on object regions, without significantly increasing computational overhead.

3.3.4 Content-Aware ReAssembly of Features

Upsampling, as one of the key operations in deep convolutional neural networks, is widely applied in tasks such as object detection and semantic segmentation. YOLOv5s defaults to nearest-neighbor interpolation for upsampling. While this method is computationally fast and low-overhead, it considers only the pixel values closest to the target pixel, ignoring the influence of other neighboring pixels. This fails to effectively preserve the original image's detailed information, often leading to severe loss of image details and consequently affecting the model's detection accuracy for small-scale pests. In contrast, while methods like transposed convolution and pixel shuffle partially address the lack of detail awareness, they introduce significant parameter overhead and pose higher training difficulties.

To address this, we introduce CARAFE, a more efficient and lightweight approach with the structure shown in Figure 6. This module features a larger receptive field, enabling effective integration of local and contextual information. Simultaneously, it adaptively generates reassembled convolutional kernels based on input feature content, achieving more precise upsampling while maintaining low computational overhead. CARAFE consists of two core components: the kernel prediction module and the content-aware reassembly module. Specifically, the kernel prediction module performs channel compression on the input feature map via 1×1 convolutions to reduce computational load for subsequent operations. Subsequently, a k×k convolutional layer applies content-aware encoding to the compressed features, predicting an upsampling kernel for each position. This expands the receptive field and enhances utilization of broader contextual information. Furthermore, each upsampling kernel undergoes normalization via a Softmax function to ensure its validity and rationality.

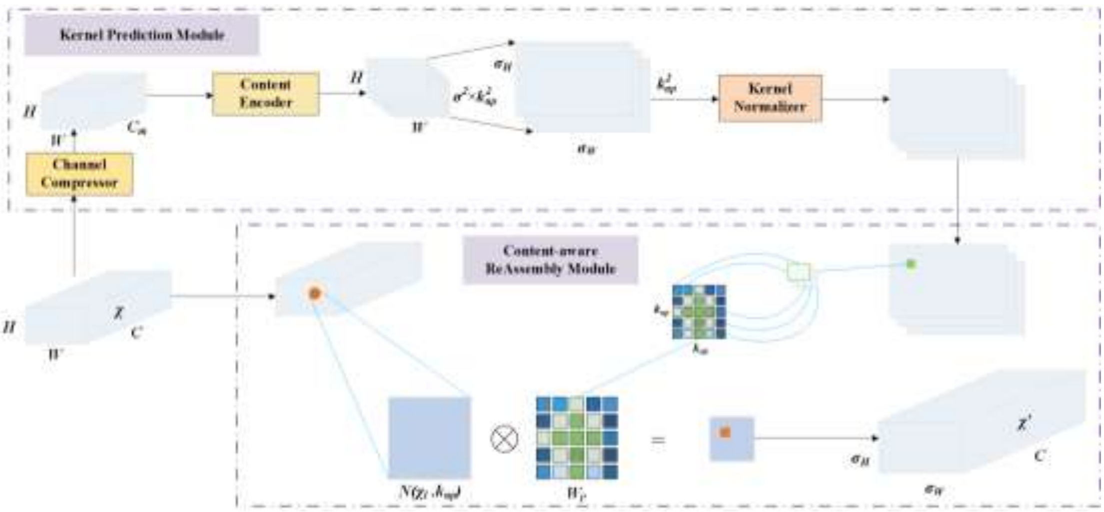

Figure 6 Structural diagram of the CARAFE module

The content-aware reassembly module generates the final upsampled result by reverse-mapping each pixel of the output feature map to the k×k region centered at that pixel in the original feature map, then performing a dot product with the upsampling kernel predicted at the corresponding position. Benefiting from its content-aware mechanism and larger receptive field, CARAFE effectively captures target feature information within local regions, thereby enhancing feature reconstruction accuracy. Consequently, we replaced the default upsampling method in YOLOv5s with CARAFE. This replacement maintains high computational efficiency while strengthening the model's feature expression capabilities and detection performance in complex scenes.

3.3.5 Inner-Focaler-EIoU

The accuracy of object detection models depends not only on their ability to distinguish between different object categories but also heavily relies on the precise prediction of bounding box locations. YOLOv5s employs the CIoU loss function to optimize bounding box regression, a method that simultaneously considers factors such as IoU overlap, center point distance, and aspect ratio. However, when handling scenarios with significant sample difficulty variations—such as small object detection or occlusion—CIoU tends to assign greater training weights to easier regression samples. This limits the model's overall regression performance in complex scenarios.

To address this issue, this paper proposes a loss function named Inner-Focaler-EIoU. This function extends Efficient IoU (EIoU) to replace the original CIoU loss, aiming to enhance YOLOv5s' bounding box prediction accuracy in complex scenarios. EIoU decomposes the aspect ratio loss term into separate differences between predicted and ground truth boxes in width and height, simplifying computations to accelerate model convergence. It is defined as shown in **Equation 4**:

$$EIoU(B,B_{gt}) = IoU(B,B_{gt}) - \frac{\rho^2(B,B_{gt})}{(w^c)^2+(h^c)^2} - \frac{\rho^2(w,w_{gt})}{(w^c)^2} - \frac{\rho^2(h,h_{gt})}{(h^c)^2} \quad (4)$$

where w_c and h_c denote the width and height of the minimum bounding rectangles for the B and B_{gt} boxes, respectively.

Focaler-EIoU further incorporates linear interval mapping [32] to adjust EIoU values. Specifically, by setting upper and lower thresholds, samples with IoU values within this range are assigned higher gradient weights, while samples with excessively low or high IoU values receive lower response weights. This method effectively enhances the model's learning capability for challenging samples, defined as shown in **Equation 5**:

$$Focaler-EIoU = \begin{cases} 0, & EIoU < d \\ \dfrac{EIoU-d}{u-d}, & d \leqslant EIoU \leqslant u \\ 1, & EIoU \geqslant d \end{cases} \quad (5)$$

where d and u denote the lower and upper thresholds, typically satisfying $0 \leq d < u < 1$.

Furthermore, inspired by the concept of Inner-IoU [33], we introduce an auxiliary bounding box mechanism on top of Focaler-EIoU. These auxiliary boxes are scaled-down versions of the original bounding boxes. By performing finer-grained IoU evaluations within the box interior, this mechanism significantly improves the model's localization accuracy for highly overlapping targets.

In summary, Inner-Focaler-EIoU not only inherits the advantages of EIoU but also effectively mitigates the limitations of CIoU loss in handling small object pests and occlusion issues in rice through the introduction of linear interval mapping and auxiliary bounding box mechanisms. This enhances the model's detection performance for rice pests.

4 Experimental Results and Analyses

4.1 Experimental Setup and Evaluation Metrics

This paper employs the YOLO-RicePest model for rice pest detection tasks. The optimizer selected is stochastic gradient descent (SGD), with a maximum learning rate set to 1e-2 and a minimum learning rate set to 1e-4. The cosine annealing learning rate scheduler [34] is utilized for learning rate adjustment. The model is initialized by loading the pretrained weights yolov5s.pt. The lower bound threshold d and upper bound threshold u for the bounding box loss function are set to 0.00 and 0.95, respectively. The specific experimental environment is detailed in **Table 2**.

Table 2 Experimental running software configuration and hardware configuration

Parameter	**Configuration**
CPU	Intel i9-13900K
GPU	NVIDIA GTX 4090
Memory	64 GB
Operating System	Windows 10
CUDA	CUDA 12.1
Pytorch	Pytorch 2.3.0

To comprehensively evaluate the detection performance of the improved YOLOv5s network, this paper selects Precision (P), Recall (R), F1-score, average precision (AP), and mean average precision (mAP) as evaluation metrics.

P denotes the proportion of correctly identified positive samples relative to the total number of predicted positive samples, defined as shown in **Equation 6**:

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

where true positive (TP) denotes the number of samples correctly classified as positive, and false positive (FP) denotes the number of samples incorrectly classified as positive.

R denotes the proportion of correctly identified positive samples relative to the total number of positive samples, defined as shown in **Equation 7**:

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

where false negative (FN) represents the number of samples incorrectly identified as negative.

AP is a key metric for evaluating an object detection model's performance on a single category. It reflects the balance between precision and recall at different confidence thresholds by calculating the area under the PR curve. AP ranges from 0 to 1, with higher values indicating superior detection performance in that category. Its specific definition is shown in **Equation 8**:

$$AP = \int_0^1 P(R)d(R) \approx \sum_{k=1}^{N} P(k)\Delta R(k) \tag{8}$$

where P(R) denotes the corresponding function form of the PR curve. The right side of the formula approximates the integral under the PR curve by performing a weighted sum of precision P(k) and the increment ΔR(k) of recall, thereby yielding the AP value.

mAP is the average of AP across all classes, comprehensively evaluating a model's overall performance in multi-class detection tasks. mAP holistically reflects a model's detection capability across the entire dataset, making it a widely adopted core metric for multi-class object detection. Values closer to 1 indicate stronger overall model performance. mAP is defined as shown in **Equation 9**:

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \qquad (9)$$

where n is the number of classes, and AP_k is the AP value for the kth class.

4.2 YOLO-RicePest Model Comparison Results and Performance Analysis

The comparative experimental results on the self-built pest dataset RPD are shown in **Table 3**. The YOLO-RicePest model outperforms the baseline YOLOv5s model across all metrics. Specifically, YOLO-RicePest achieved a precision of 93.4%, representing a relative improvement of 4.2% compared to YOLOv5s' 89.6%. Recall increased to 80.5%, a relative improvement of 9.8% over YOLOv5s' 73.3%. mAP50 increased from 82.4% to 87.9%, a relative gain of 6.7%; mAP50-95 also significantly improved to 64.5%, representing an 8.0% relative increase over YOLOv5s' 59.7%. With only a 0.14M increase in parameters, YOLO-RicePest demonstrates high accuracy and practicality for rice pest detection tasks.

Table 3 Performance comparison between YOLO-RicePest and YOLOv5s models

Model	**P/%**	**R/%**	**mAP50/%**	**mAP50-95/%**	**Params/M**	**GFLOPs**	**FPS**
YOLOv5s	89.6	73.3	82.4	59.7	**6.73**	**15.9**	**142.5**
YOLO-RicePest	**93.4**	**80.5**	**87.9**	**64.5**	6.87	16.5	121.4

Note: the best results are highlighted in bold.

To further highlight the improved model's advantages, we plotted PR curves for YOLOv5s and YOLO-RicePest under identical experimental conditions and selected three representative test images for visual comparison. **Figure 7** shows the PR curve comparison between the two models. It can be observed that the curve of YOLO-RicePest shifts upward and to the right overall and is smoother, indicating that it maintains higher precision in the high recall region while significantly reducing false positive rates. Regarding detection performance across categories, AP values improved for most pest categories. For categories with ample samples, such as Coleoptera (14,125 samples) and Hydrophilidae (8,375 samples), AP increased from 0.895 and 0.940 to 0.900 and 0.946, respectively, demonstrating the model's optimization for mainstream categories. For categories with smaller sample sizes, the improvement was even more pronounced. For example, Amatidae (60 samples) saw its AP rise from 0.229 to 0.421, demonstrating the enhanced model's stronger generalization capability in scenarios with limited data. Simultaneously, for key pest categories with larger sample sizes, such as Chilosuppressalis (6,000 samples) and Cnaphalocrocis (8,215 samples), although the AP improvement was modest (rising from 0.957 to 0.962 and from 0.973 to 0.974, respectively), it maintained stable high accuracy, further demonstrating the robustness of the model enhancement.

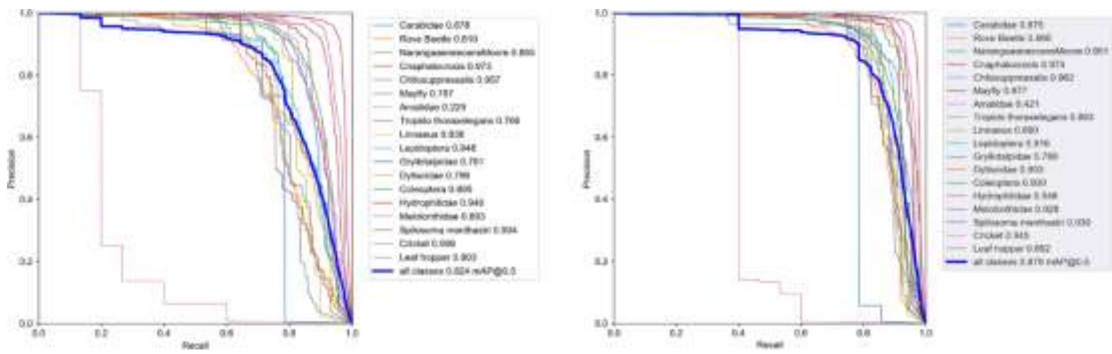

Figure 7 Comparison of P-R curves between YOLOv5s (left) and YOLO-RicePest (right) models

Figure 8 displays the inference results of both models on test set images. The first column shows the original test images, while the second and third columns correspond to the detection results of YOLOv5s and YOLO-RicePest, respectively. As seen in the first test image, YOLO-RicePest demonstrates superior performance in detecting small-scale pests, thanks to the cross-spatial attention mechanism of the EMA module and the content-aware feature reorganization of the CARAFE module. For instance, the Spilosoma menthastri in the upper left corner and the Chilosuppressalis at the bottom are accurately identified and annotated in YOLO-RicePest's detection results (see red dashed boxes in the third column).

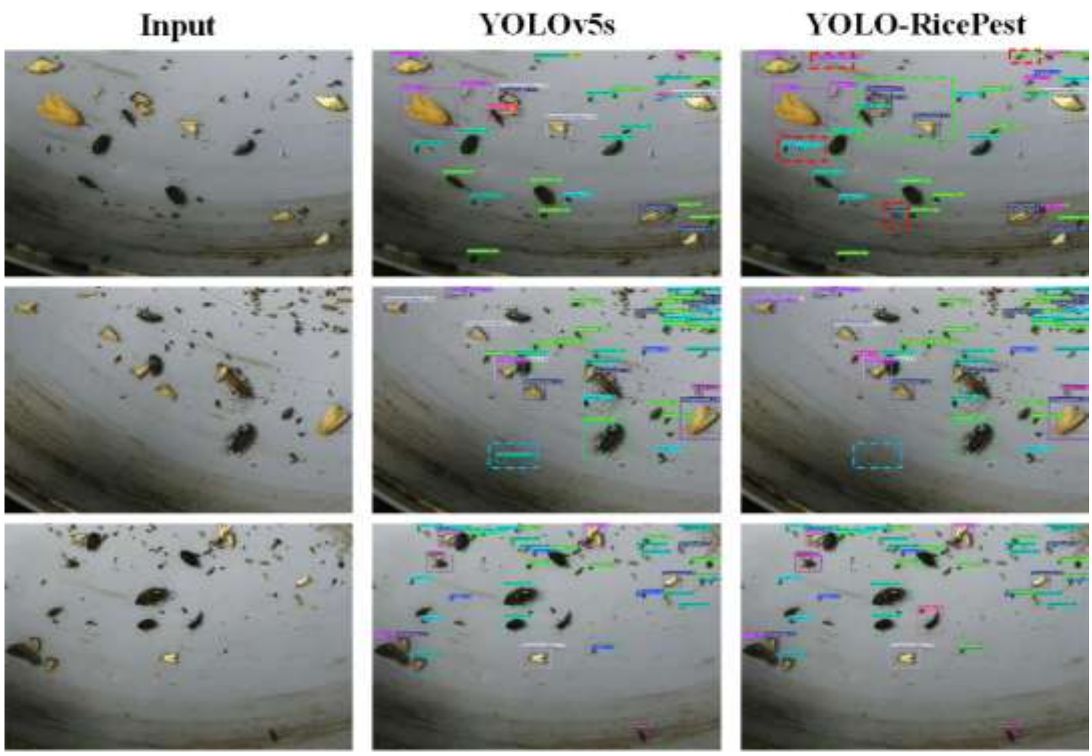

Figure 8 Comparison of test image detection results between YOLO-RicePest and YOLOv5s models

Furthermore, the Inner-Focaler-EIoU loss function optimizes overlapping object bounding boxes, enabling the model to achieve more precise localization in high-overlap scenarios. For instance, the object in the green dashed box, initially misclassified as Narangaa enescens Moore by YOLOv5s, is correctly detected as

Chilosuppressalis in YOLO-RicePest with a more reasonable bounding box convergence. The EMA module simultaneously enhances the model's focus on object regions, effectively reducing false detections caused by background interference, as shown in the blue dashed box in the second test image.

Combined experimental data and visualization results demonstrate that YOLO-RicePest maintains stable detection performance for small-sample objects, high-overlap scenarios, and complex backgrounds while improving accuracy and recall. These experiments fully validate the model's effectiveness and robustness, indicating its significant practical value for rice pest detection tasks. This provides a reliable basis for subsequent practical deployment and further optimization.

4.3 Ablation Experiment

We divided the ablation experiments into three parts, sequentially examining the impact of the Inner-Focaler-EIoU, EMA module, and CARAFE module on the overall network performance. To ensure experimental fairness, when conducting ablation on a specific innovation, the remaining YOLO-RicePest architecture remained unchanged, thereby guaranteeing the accuracy and reliability of the experimental results.

4.3.1 Ablation Experiments on Different Bounding Box Regression Losses

Bounding box regression loss is a critical factor in object detection performance. YOLOv5s defaults to CIoU, which optimizes regression by constraining overlap, center offset, and aspect ratio. However, in scenarios involving small objects or occlusions, it tends to assign excessive weight to easily regressible samples, thereby weakening learning efficiency for challenging samples. To address this, we propose the Inner-Focaler-EIoU loss. This loss decouples width-height constraints based on EIoU to accelerate convergence, introduces a gradient reweighting strategy to boost the weight of moderately difficult samples, and incorporates an auxiliary bounding box mechanism for precise evaluation of internal overlap. This approach enhances localization accuracy for highly overlapping objects.

To validate the method's effectiveness, we conducted ablation experiments comparing CIoU, EIoU, Inner-EIoU, Focaler-EIoU, and Inner-Focaler-EIoU. Results are shown in **Table 4**. The table demonstrates that replacing CIoU with EIoU improves Precision and Recall from 92.3% and 77.2% to 92.8% and 79.0%, respectively, while boosting mAP50 and mAP50-95 to 86.9% and 62.4%. This indicates that decoupling width-height constraints enhances the model's overall regression performance. Further introduction of auxiliary bounding boxes and gradient reweighting strategies further improved Precision to 93.7% and 92.8%, Recall to 79.9% and 79.2%, with mAP50 reaching 87.7% and 87.5%, while mAP50-95 improved to 64.3% and 63.6%, respectively. This demonstrates that both strategies effectively enhance the model's ability to localize challenging samples. Ultimately, the Inner-Focaler-EIoU approach, which integrates the strengths of both strategies, achieved the best performance: Precision, Recall, mAP50, and mAP50-95 reached 93.4%, 80.5%, 87.9%, and 64.5%, respectively, outperforming other configurations in both accuracy and generalization capability.

Table 4 Ablation study results of YOLO-RicePest under different bounding box regression losses

Supervision Setting	**P/%**	**R/%**	**mAP50/%**	**mAP50-95/%**
CIoU	92.3	77.2	86.1	61.0
EIoU	92.8	79.0	86.9	62.4

Inner-EIoU	**93.7**	79.9	87.7	64.3
Focaler-EIoU	92.8	79.2	87.5	63.6
Inner-Focaler-EIoU (Ours)	93.4	**80.5**	**87.9**	**64.5**

Note: the best results are highlighted in bold.

Notably, Inner-Focaler-EIoU rapidly reduces regression error during early training while maintaining smooth convergence in the middle and late stages, avoiding gradient oscillations. This approach promotes more balanced attention between easy-to-regress and hard-to-regress samples, reducing missed detections and enhancing overall stability and robustness.

4.3.2 Ablation Experiments on Introducing Different Attention Modules into the Backbone Network

Attention mechanisms enhance detection performance by dynamically adjusting feature map weights, enabling the model to focus on critical regions. To analyze the impact of introducing different attention mechanisms into the backbone network on YOLO-RicePest, we conducted comparative experiments with squeeze-and-excitation (SE) [35], CA, CBAM, and EMA modules for comparative experiments, with results shown in **Table 5**. The SE module employs only channel attention, compressing feature maps via global average pooling before sequentially passing through a fully connected layer and Sigmoid activation function to relearn channel weights, thereby enhancing the model's focus on critical feature channels. The CA module further decomposes channel attention into two one-dimensional feature encoding processes along horizontal and vertical directions, capturing long-range dependencies in feature maps while preserving positional information. Building upon this, the CBAM module introduces spatial attention, utilizing large-scale convolutions to enhance the model's perception of object regions, enabling simultaneous focus on important features across both channel and spatial dimensions.

Table 5 Ablation study results of YOLO-RicePest with different attention modules

Attention Module	**P/%**	**R/%**	**mAP50/%**	**mAP50-95/%**
NoAttention	92.0	77.7	86.2	61.2
SE	92.1	78.7	86.4	61.7
CA	92.0	79.1	87.1	63.1
CBAM	93.3	79.9	87.6	64.0
EMA (Ours)	**93.4**	**80.5**	**87.9**	**64.5**

Note: the best results are highlighted in bold.

As shown in the table, incorporating attention modules improves model performance to varying degrees. Specifically, CA and CBAM elevate mAP50 to 87.1% and 87.6%, respectively, demonstrating that jointly modeling channel and spatial attention effectively enhances feature representation. However, while CA captures long-range dependencies, it lacks explicit spatial attention modeling, resulting in insufficient focus on object locations. CBAM simultaneously models both channel and spatial attention but relies on large convolutional kernels for spatial feature extraction, incurring high computational overhead. It can only capture local correlations and struggles to balance distant information, leading to slightly inferior performance in complex scenes. In contrast, the EMA module employs efficient cross-channel interactions and feature reconstruction grouping mechanisms. This approach preserves long-range dependencies while significantly reducing computational complexity, enabling the model to better capture feature representations of multi-scale pest objects. Ultimately, YOLO-RicePest with the EMA module achieved optimal results with Precision, Recall, mAP50, and mAP50-95 scores of 93.4%, 80.5%, 87.9%, and 64.5%, respectively. This demonstrates that the EMA module effectively enhances detection accuracy and is more suitable for rice pest detection tasks.

4.3.3 Ablation Studies Under Different Upsampling Strategies

YOLOv5s defaults to nearest neighbor interpolation for upscaling. While this method offers fast computation and low computational overhead, it relies solely on the pixel value nearest to the target pixel, disregarding contributions from other neighboring pixels. This approach risks losing feature map details, thereby impairing the model's ability to detect small-scale pests. To investigate the impact of different upscaling strategies on performance, we replaced the default upscaling method with bilinear interpolation, transposed convolution, pixel shuffle, and the CARAFE module, respectively, while keeping the rest of the network architecture unchanged. The specific experimental results are shown in **Table 6**.

Table 6 Ablation study results of YOLO-RicePest with different upsampling strategies

Attention Module	**P/%**	**R/%**	**mAP50/%**	**mAP50-95/%**
NoAttention	92.0	77.7	86.2	61.2
SE	92.1	78.7	86.4	61.7
CA	92.0	79.1	87.1	63.1
CBAM	93.3	79.9	87.6	64.0
EMA (Ours)	**93.4**	**80.5**	**87.9**	**64.5**

Note: the best results are highlighted in bold.

As shown in the table, bilinear interpolation mitigates detail loss to some extent by performing weighted averaging of neighboring pixels, improving mAP50 and mAP50-95 to 87.3% and 63.8%, respectively. In contrast, transposed convolution, despite learning upscaling weights, tends to introduce checkerboard artifacts, resulting in performance inferior to nearest-neighbor interpolation. Pixel shuffle achieves learnable upscaling through channel rearrangement, enhancing detail perception while maintaining computational efficiency. It achieves mAP50 and mAP50-95 scores of 87.4% and 64.1%, respectively, outperforming the three aforementioned methods. The CARAFE module employed in this paper dynamically generates upsampling convolutional kernels using content-aware mechanisms and a large receptive field. It significantly enhances spatial detail recovery while maintaining controllable computational overhead, achieving optimal results. This demonstrates that the CARAFE module is more suitable for rice pest detection tasks, effectively improving the model's feature representation capability and small object detection accuracy in complex scenes.

5 Discussion

Agriculture, as the foundational industry ensuring human survival, directly impacts food security through its yield and quality. During rice cultivation, infestations by various pests significantly affect crop yield and quality, potentially leading to regional crop failures. Traditional manual field inspections are not only time-consuming and labor-intensive but also limited in accuracy by individual experience, failing to meet modern agriculture's demand for large-scale, high-precision, real-time monitoring.

Therefore, developing efficient, intelligent rice pest detection methods is crucial for enhancing control efficiency and safeguarding food security.

In recent years, deep learning has demonstrated significant advantages in image recognition, with object detection technology gaining particular attention in agricultural pest identification. Deep learning-based object detection methods primarily fall into two categories: single-stage and two-stage detection algorithms. Single-stage methods like YOLO and SSD feature simple architectures, simultaneously performing object localization and classification through regression strategies, enabling faster inference speeds. Two-stage methods like Faster R-CNN first generate candidate regions before performing classification and regression, offering higher accuracy but slower speed. In agricultural pest detection scenarios demanding high real-time performance, single-stage detection algorithms hold greater application value.

Given this, we propose the YOLO-RicePest model, optimized for rice pest detection based on an enhanced YOLOv5s architecture. This approach introduces an EMA module into the backbone network, replaces the neck network sampling with a CARAFE module, and adopts an Inner-Focaler-EIoU bounding box loss. These modifications significantly enhance detection performance for small objects and highly similar pests while maintaining fast inference speeds. On the RPD dataset, YOLO-RicePest demonstrates clear advantages over the original YOLOv5s: Precision increased from 89.6% to 93.4%, Recall rose from 73.3% to 80.5%, mAP50 climbed from 82.4% to 87.9%, and mAP50-95 improved from 59.7% to 64.5%. These results fully validate the model's practicality and superiority in natural environments.

The performance gains of YOLO-RicePest primarily stem from three key optimizations. First, introducing an EMA module after each C3 block in the backbone network enhances spatial information, enabling more precise differentiation of texture-similar pests. Second, the neck network employs a CARAFE module for upsampling, which adaptively generates convolutional kernels based on input features. This approach preserves small-object details while integrating local and contextual information, enhancing the model's ability to detect pests in complex environments. Finally, the Inner-Focaler-EIoU bounding box loss is introduced. Through linear interval mapping and auxiliary bounding box mechanisms, it mitigates the limitations of CIoU in handling small-to-medium objects and occlusion scenarios while accelerating model convergence.

Furthermore, to support model training, we constructed the RPD dataset based on automated trapping devices, encompassing 18 common rice pests with 2,550 high-quality images. Experimental results on this dataset demonstrate that YOLO-RicePest outperforms the original YOLOv5s in accuracy, recall, and mean average precision, highlighting the effectiveness and practicality of the proposed method for real-world agricultural pest detection tasks.

Although YOLO-RicePest performs excellently in natural environments, several areas warrant improvement. For instance, detection accuracy for extremely small objects and densely occluded regions remains suboptimal. Furthermore, computational complexity and deployment efficiency require optimization for resource-constrained scenarios. Future research will focus on exploring multi-scale feature fusion, lightweight network design, and adaptive loss functions to further enhance the performance and applicability of rice pest detection.

6 Conclusions

This paper proposes YOLO-RicePest, a rice pest detection network based on an enhanced YOLOv5s architecture, designed for automatic identification of multiple pests in rice fields. By introducing an EMA module after each C3 module in the backbone network, the network's ability to focus on texture and shape features is enhanced, effectively improving the discrimination of pests with similar body colors. In the neck network, the CARAFE module replaces traditional sampling operations by adaptively generating convolutional kernels based on input features. This approach better preserves small object details while integrating local and contextual information. Additionally, the Inner-Focaler-EIoU bounding box loss mitigates detection accuracy impacts from small/medium objects and occlusions through linear interval mapping and auxiliary bounding box mechanisms, simultaneously accelerating model convergence.

Leveraging automated trapping devices, we constructed the high-quality Rice Pest Detection (RPD) dataset covering 18 common rice pests and conducted experiments on this dataset. Compared to the original YOLOv5s, YOLO-RicePest achieves a 3.8 percentage point increase in Precision, a 7.2 percentage point increase in Recall, a 5.5 percentage point improvement in mAP50, and a 4.8 percentage point improvement in mAP50-95. These results fully validate the effectiveness and practicality of the proposed method for rice pest detection tasks. Furthermore, we analyzed the limitations of YOLO-RicePest and proposed future improvement directions: (1) Further optimize the multi-scale feature fusion strategy to enhance detection accuracy for extremely small objects and densely occluded regions; (2) Explore lightweight network designs and adaptive loss weighting mechanisms to improve the model's deployment efficiency and adaptability in resource-constrained scenarios.

Conflict of Interest Statement: The author declare no conflict of interest.

Funding Statement: No external funding received.

References

1. çalves, J.; Silva, E.; Faria, P.; Nogueira, T.; Ferreira, A.; Carlos, C.; Rosado, L. Edge-compatible deep learning models for detection of pest outbreaks in viticulture. Agronomy 2022,12,3052. https://doi.org/10.3390/agronomy12123052
2. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multir esolution gray-scale and rotation invariant text ure classification with local binary patterns. I EEE Transactions on pattern analysis and mac hine intelligence 2002, 24, 971–987. https://doi.org/10.1109/TPAMI.2002.1017623
3. OIV. State of the world vitivinicultural sector in 2020. STATE OF THE WORLD VITIVINICULTURAL SECTOR IN 202020 21.
4. Kaya, Y.; Kayci, L.; Tekin, R.; Faruk Ertugrul ,˘ Ö. Evaluation of texture features for automa tic detecting butterfly species using extreme le arning machine. Journal of Experimental & T heoretical Artificial Intelligence 2014, 26, 267 –281. https://doi.org/10.1080/0952813X.2013. 861875
5. Kang, S.H.; Cho, J.H.; Lee, S.H. Identification of butterfly based on their shapes when viewed from different angles using an artificial neural network. Journal of Asia-Pacific Entomology 2014, 17, 143–149. https://doi.org/10.1016/j.aspen.2013.12.004
6. Archana, R.; Jeevaraj, P.E. Deep learning models for digital image processing: a review. Artificial Intelligence Review 2024, 57, 11. https://doi.org/10.1007/s10462-023-10631-z
7. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 2015, 28.
8. Wang, L.; Shoulin, Y.; Alyami, H.; Laghari, A.A.; Rashid, M.; Almotiri, J.; Alyamani, H.J.; Alturise, F. A novel deep learning-based single shot multibox detector model for object detection in optical remote sensing images,

2024.
9. Badgujar, C.M.; Poulose, A.; Gan, H. Agricult ural object detection with You Only Look On ce (YOLO) Algorithm: A bibliometric and sys tematic literature review. Computers and Elect ronics in Agriculture 2024, 223, 109090. https ://doi.org/10.1016/j.compag.2024.109090
10. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, Greece, 4–10 June 2023. https://doi.org/10.1109/ICASSP49 357.2023.10096516
11. Wang, J., Chen, K., Xu, R., Liu, Z., Loy, C. C ., Lin, D. Carafe: Content-aware reassembly o f features. In Proceedings of the IEEE/CVF In ternational Conference on Computer Vision (I CCV), Seoul, Korea, 27 October-2 November 2019. https://doi.org/10.1109/ICCV.2019.003 10
12. Fan, L. The research on automatic identification of butterfly species based on the digital image. Beijing Forestry University 2015.
13. Zhou, A.M.; Ma, P.P.; Xi, T.Y.; Wang, J.; Feng, J.; Shao, Z.; Tao, Y.; Yao, Q. Automatic identification of butterfly specimen images at the family level based on deep learning method. Acta Entomol. Sin 2017, 60, 1339–1348.
14. Xie, J.; Hou, Q.; Shi, Y.; Peng, L.; Jing, L.; Zhuang, F.; Zhang, J.; Tang, X.; Xu, S. The automatic identification of butterfly species. arXiv preprint arXiv:1803.06626 2018.
15. Guan, B., Wu, Y., Zhu, J., Kong, J., Dong, W. GC-Faster RCNN: The object detection algorithm for agricultural pests based on improved hybrid attention mechanism. Plants 2025,14,1106. https://doi.org/10.3390/plants 14071106
16. Wang, Z., Qiao, L., Wang, M. Agricultural pest detection algorithm based on improved faster RCNN. In International Conference on Computer Vision and Pattern Analysis (ICCPA), Guangzhou China, 19-21 November 2021.
17. Lin, T. Y., Dollár, P., Girshick, R., He, K., Ha riharan, B., Belongie, S. Feature pyramid net works for object detection. In Proceedings of t he IEEE conference on computer vision and p attern recognition (CVPR), HONOLULU, 21-26 July 2017. https://doi.org/10.1109/CVPR.2 017.106
18. He, K., Gkioxari, G., Dollár, P., Girshick, R. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (ICCV), Venice, Italy, 22-29, October 2017. https://doi.org/10.1109/ICCV.2017.322
19. Wang, X., Hu, C., Wang, X., Zha, H., Chen, X., Yuan, S., Ye, Z. Research on multi class p ests identification and detection based on fusi on attention mechanism with Mask-RCNN-C BAM. Frontiers in Agronomy 2025, 7, 15784 12. https://doi.org/10.3389/fagro.2025.157841 2
20. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8-14 September, 2018. https://doi.org/10.10 07/978-3-030-01234-2_1
21. Song, L., Liu, M., Liu, S., Wang, H., Luo, J. Pest species identification algorithm based on improved YOLOv4 network. Signal, Image and Video Processing 2023, 17, 3127-3134. https://doi.org/10.1007/s11760-023-02534-x
22. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004. 10934 2020.
23. Liu, J., Wang, X., Miao, W., Liu, G. Tomato p est recognition algorithm based on improved YOLOv4. Frontiers in Plant Science 2022, 13, 814681. https://doi.org/10.3389/fpls.2022.814 681
24. Tan, Z., Ye, D., Wang, J., Wang, W. P4CN-

YOLOv5s: a passion fruit pests detection method based on lightweight-improved YOLOv5s. Frontiers in Plant Science 2025, 16,1612642. https://doi.org/10.3389/fpls.2025.1612642
25. Gao, Z., Li, Z., Zhang, C., Wang, Y., Su, J. Double Self-Attention Based Fully Connected Feature Pyramid Network for Field Crop Pest Detection, 2024. https://doi.org/10.21203/rs.3.rs-5068163/v1
26. Mo, L., Xie, R., Ye, F., Wang, G., Wu, P., Yi, X. Enhanced Tomato Pest Detection via Leaf Imagery with a New Loss Function. Agronomy 2024, 14, 1197. https://doi.org/10.3390/agronomy14061197
27. Wang, C. Y., Liao, H. Y. M., Wu, Y. H., Chen, P. Y., Hsieh, J. W., Yeh, I. H. CSPNet: A new backbone that can enhance learning capability of CNN. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (CVPR), Virtual, 14-19, June 2020. https://doi.org/10.1109/CVPRW50498.2020.00203
28. Lou, H.; Liu, X.; Bi, L.; Liu, H.; Guo, J. BD-YOLO: detection algorithm for high-resolution remote sensing images. Physica Scripta 2024, 99, 066003. https://doi.org/10.1088/1402-4896/ad418f
29. Ke, J.; He, L.; Han, B.; Li, J.; Gao, X. ProFPN: Progressive feature pyramid network with soft proposal assignment for object detection. Knowledge-Based Systems 2024, 299,112078. https://doi.org/10.1016/j.knosys.2024.112078
30. Zhang, T.; Chen, C.; Liu, Y.; Geng, X.; Aly, M.M.S.; Lin, J. PSRR-MaxpoolNMS++: Fast Non-Maximum Suppression with Discretization and Pooling. IEEE Transactions on Pattern Analysis and Machine Intelligence 2024. https://doi.org/10.1109/TPAMI.2024.3485898
31. Hou, Q., Zhou, D., Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), Virtual, 19-25, June 2021. https://doi.org/10.1109/CVPR46437.2021.01350
32. Zhang, H., Zhang, S. Focaler-iou: More focused intersection over union loss. arXiv preprint arXiv:2401.10525, 2024.
33. Zhang, H., Xu, C., Zhang, S. Inner-IoU: more effective intersection over union loss with auxiliary bounding box. arXiv preprint arXiv:2311.02877, 2023.
34. Loshchilov, I., Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983, 2016.
35. Hu, J., Shen, L., Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR), Salt Lake City, 18-22 June 2018. https://doi.org/10.1109/CVPR.2018.00745