AN EFFICIENT RICE PEST DETECTION METHOD BASED ON YOLOV11N
/
基于 YOLOv11n 的高效水稻害虫检测方法

Xiaoke WANG, Zhichao ZHAO, Qiuyang HU, Jiyu LAI, Tiefeng WU*
School of Information and Control Engineering, Qingdao University of Technology, Qingdao 266520 / China
Tel: +86-15653723323; *E-mail:* 791479558@qq.com
DOI: https://doi.org/10.35633/inmateh-79-52

Keywords: *Rice pest detection, YOLOv11n, Detail-Preserving Contextual Fusion, Multi-cognitive Visual Adapter, Edge-device deployment*

ABSTRACT
Accurate and efficient rice-pest detection is essential for field scouting, early warning and precision control. This study presents a lightweight YOLOv11n-based framework for accurate edge deployment. It integrates a Multi-cognitive Visual Adapter (Mona), a Detail-Preserving Contextual Fusion module (DPCF) and Wise-IoU-based non-maximum suppression (Wise-IoU-NMS). Mona improves multi-scale feature representation with limited overhead; DPCF preserves fine-grained pest details during contextual fusion; and Wise-IoU-NMS stabilizes localization of overlapping and densely distributed targets. On a public dataset of 5,212 images from six near-balanced rice-pest classes, divided 8:1:1 for training, validation and testing, the proposed model achieved 95.4% mAP@50, 94.8% precision and 94.7% recall with 2.65 M parameters and 6.5 GFLOPs. The F1-confidence curve identified 0.476 as the alarm threshold. Deployment on Jetson Nano using ONNX and frame-level logging demonstrated a practical workflow for real-time monitoring and selective spraying.

摘要
水稻害虫的快速、准确检测对于田间巡检、虫情预警和精准防控具有重要意义。本文提出一种面向边缘端高精度部署的轻量化 YOLOv11n 检测框架，融合多认知视觉适配器（Multi-cognitive Visual Adapter, Mona）、细节保留型上下文融合模块（Detail-Preserving Contextual Fusion, DPCF）和基于 Wise-IoU 的非极大值抑制策略（Wise-IoU-NMS）。Mona 以较低开销增强多尺度特征表达，DPCF 在上下文融合中保留细粒度虫体信息，Wise-IoU-NMS 则提高重叠及密集目标的定位稳定性。在包含 5,212 张图像、6 类近均衡水稻害虫并按 8:1:1 划分训练集、验证集和测试集的公开数据集上，所提出模型取得 95.4% 的 mAP@50、94.8% 的精确率和 94.7% 的召回率，参数量为 2.65 M，计算量为 6.5 GFLOPs。F1-confidence 曲线确定 0.476 为报警阈值；基于 ONNX 的 Jetson Nano 部署及逐帧记录验证了该方法用于实时监测和选择性施药的可行流程。

INTRODUCTION
Rice is one of the most important staple crops worldwide, and pest outbreaks during the growing season directly affect yield, grain quality and pesticide-use decisions. Field scouting still relies heavily on visual inspection and manual counting, which are labor-intensive, subjective and difficult to scale across large planting areas. A practical rice-pest detection system therefore needs not only high recognition accuracy, but also real-time inference, stable performance under occlusion and illumination variation, and compatibility with low-power edge devices used in field monitoring or selective spraying.

Deep-learning-based object detection has become an important tool for agricultural pest monitoring. Two-stage detectors such as Faster R-CNN generally provide strong localization ability but require relatively high computational cost, whereas one-stage detectors represented by the YOLO series offer a better accuracy-speed trade-off for real-time applications. Recent studies on rice or field pest identification, including MobileNet-CA-YOLO, AgriPest-YOLO and transformer-based rice-pest detectors, have improved detection accuracy by introducing lightweight backbones, attention mechanisms or multi-scale feature fusion (*Jia et al., 2023; Zhang et al., 2023; Wang et al., 2025*). Complementary international studies have broadened both the sensing scenarios and the task definitions used for rice-pest identification. Hassan et al. combined remote-sensing imagery with a modified YO-CNN to identify major rice pests and demonstrated the potential of wide-area automated scouting *(Hassan et al., 2023)*.

Saufi and Suharjito developed a trap-image CNN for white rice stem borer detection and population counting, but reported that inference still depended on server resources rather than direct processing on the field device *(Saufi and Suharjito, 2024)*. Quach et al. released an expert-reviewed rice-pest dataset assembled from heterogeneous sources, highlighting the importance of data diversity and standardized preprocessing for smart-farming systems *(Quach et al., 2024)*. In parallel, RP-DETR used an end-to-end Transformer architecture to strengthen long-range feature interaction and multi-scale fusion for rice-pest recognition, thereby providing a complementary alternative to convolution-dominated detectors *(Wang et al., 2025)*.

Taken together, these model families address different operational needs. Image-level CNN systems are relatively simple and effective in controlled trap or aerial scenes, but they do not always localize several neighboring pests in a single field image and may require remote-server support *(Hassan et al., 2023; Saufi and Suharjito, 2024)*. Lightweight YOLO detectors generally offer a favorable accuracy-speed balance for edge deployment *(Jia et al., 2023; Zhang et al., 2023; Xiong et al., 2024)*; however, performance remains sensitive to the loss of fine-grained cues during downsampling and to small, dense or overlapping target layouts *(Sunkara and Luo, 2023; Wang et al., 2025)*. Transformer-based detectors capture longer-range dependencies and provide an end-to-end alternative for multi-scale rice-pest detection *(Wang et al., 2025)*. Moreover, headline accuracy values should not be compared directly across studies because pest categories, acquisition platforms, annotation protocols and test sets differ *(Quach et al., 2024; Xiong et al., 2024)*. Collectively, these observations indicate that field-oriented rice-pest detection should be treated as a coupled problem of fine-grained visual representation, reliable localization in dense or occluded scenes, and resource-efficient inference, rather than as an isolated improvement to one network component. Accordingly, the present study develops an integrated lightweight detection framework that coordinates feature representation, multi-scale information interaction and prediction refinement within a unified pipeline. The framework is intended to retain weak boundary and texture cues of very small pests, use broader contextual information to distinguish visually similar categories, reduce missed and duplicate detections in overlapping infestations, and maintain an accuracy-latency balance suitable for edge deployment. Its practical relevance is assessed through class-wise PR/F1 analysis, confidence-threshold selection and Jetson Nano validation at an average inference speed of 36 FPS.

However, several research gaps remain. First, small rice pests are often densely clustered or partially occluded, resulting in increased missed detections and duplicate bounding boxes. Second, many studies focus primarily on benchmark accuracy while providing limited analysis of class-specific errors, confidence-threshold selection, and practical alarm criteria. Third, the deployment performance of improved models on resource-constrained edge devices has often not been sufficiently validated. To address these gaps, this study developed a lightweight YOLOv11n-based detector integrating Mona, DPCF, and Wise-IoU-NMS. The proposed model was evaluated through ablation experiments, cross-model comparisons, class-wise precision–recall and F1-score analyses, and a deployment protocol designed for the Jetson Nano.

MATERIALS AND METHODS

DATA SOURCES AND PREPROCESSING

Because publicly available rice-pest datasets with uniform annotation standards remain limited, this study constructed a rice-pest image dataset from open-source images and expert-reviewed annotations. The dataset contains six target categories: green leafhopper, brown planthopper, rice bug, leaf folder, stem borer and whorl maggot. Representative images from the dataset are shown in Fig. 1.

Fig. 1 - Representative images from the dataset

The dataset used in this study was sourced from publicly available pest images collected from multiple online platforms (Fig. 1).

Before inclusion, the images were manually screened to remove duplicates, irrelevant samples and low-quality records, and the copyright and usage conditions of the source images were checked. Each retained image was annotated with pest category labels and bounding boxes under the guidance of agricultural experts and authoritative entomological references, thereby improving annotation reliability and domain consistency.

A total of 5,212 high-resolution images were collected, covering six rice-pest categories and diverse visual conditions, including occlusion, deformation, complex backgrounds and illumination variation. The dataset was divided into training, validation and test sets at a ratio of 8:1:1, with 4,170, 521 and 521 images, respectively. The six pest categories were arranged in a near-balanced manner: each class contained 868-869 images, corresponding to 16.65-16.67% of the dataset. This distribution reduces class-prior bias and allows the precision, recall and F1-score curves to be interpreted as class-comparable evidence rather than being dominated by one abundant pest category. To improve model generalization and reduce imbalance-related bias, data augmentation was applied, including random rotation, brightness adjustment, horizontal flip, translation, local occlusion and Gaussian noise.

Table 1

Class-wise composition of the rice-pest dataset

Pest class	Chinese name	Total images	Proportion (%)	Train/Val/Test
green-leafhopper	绿叶蝉	869	16.67	695 / 87 / 87
brown-planthopper	褐飞虱	869	16.67	695 / 87 / 87
rice-bug	稻蝽	869	16.67	695 / 87 / 87
leaf-folder	卷叶螟	869	16.67	695 / 87 / 87
stem-borer	螟虫	868	16.65	695 / 87 / 86
whorl-maggot	稻瘿蚊	868	16.65	695 / 86 / 87
Total	-	5,212	100.00	4,170 / 521 / 521

Note: *The dataset was organized as a near-balanced six-class rice-pest dataset. The train/validation/test split follows the 8:1:1 protocol.*

EXPERIMENTAL ENVIRONMENT AND PARAMETER CONFIGURATION

Experimental Environment Configuration

Model training was conducted in Python 3.8 using the PyTorch deep-learning framework. The training workstation ran Ubuntu 22.04.3 with CUDA 12.2 acceleration and used an Intel(R) Core(TM) i5-8300H CPU @ 2.30 GHz and an NVIDIA A100 GPU with 40 GB memory. This environment was used for model development, parameter optimization and benchmark comparison, whereas the edge-device validation was conducted separately on Jetson Nano as described below.

Experimental Parameter Configuration

YOLOv11n was used as the baseline model. All models were trained with an input resolution of 640 x 640 pixels. Stochastic Gradient Descent (SGD) was used for optimization with an initial learning rate of 0.01, momentum of 0.937 and cosine annealing for learning-rate scheduling. The batch size was set to 16 and the maximum number of epochs was 300. The same training and testing configuration was maintained across ablation and comparison experiments to ensure fair evaluation of the proposed modules.

EDGE-DEVICE VALIDATION PROTOCOL

To address the requirement for real-life validation, the final YOLOv11n-Mona-DPCF-Wise-IoU model was prepared for deployment on an NVIDIA Jetson Nano B01 (4 GB). The edge-device workflow included model export to ONNX, graph simplification, TensorRT FP32 engine conversion and local inference on 640 x 640 input frames. The deployment interface integrates image, video and camera-stream input, real-time visualization of bounding boxes and confidence scores, and result logging for subsequent pest occurrence analysis.

In practical use, detections with confidence scores above 0.476, the threshold selected from the F1-confidence curve, are treated as valid pest alarms. Each alarm record includes the pest category, confidence score, bounding-box coordinates and frame information. This design allows the system to support field scouting, multi-class pest monitoring, pest-density recording and future integration with selective spraying equipment.

ALGORITHM IMPROVEMENT

YOLOv11n Algorithm

This study focuses on YOLOv11n as the baseline detector because of its lightweight structure and real-time potential. The YOLOv11n network consists of four major components: the input layer, backbone, neck feature-fusion network and detection head. The overall network structure is shown in Fig. 2.

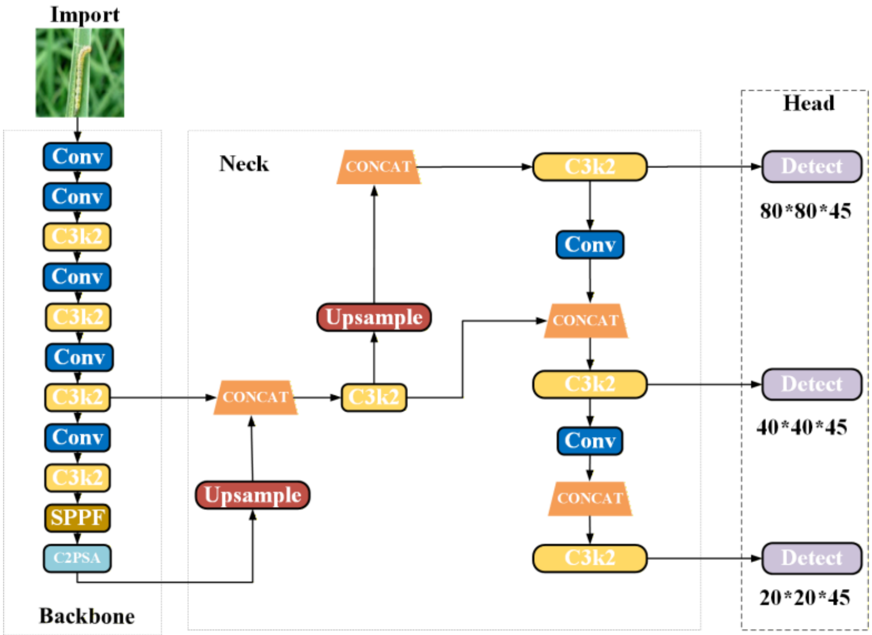


Fig. 2 - Architecture of the YOLOv11n network model

Multi-cognitive Visual Adapter

To address efficient and accurate pest detection under complex visual conditions, Mona (Multi-cognitive Visual Adapter) is introduced as a lightweight feature adaptation module (Fig. 3). Mona is designed to strengthen visual knowledge transfer without substantially increasing trainable parameters. It contains scaled layer normalization and a multi-branch depthwise convolutional filtering module. The scaled LayerNorm uses learnable weights to balance normalized and original features, improving the stability of transferred representations. The convolutional branches use 3 x 3, 5 x 5 and 7 x 7 depthwise kernels to capture pest features at different receptive fields, followed by pointwise convolution and GeLU activation. Residual connections maintain gradient flow and semantic consistency. This design improves fine-grained pest feature representation while preserving the lightweight nature of YOLOv11n.

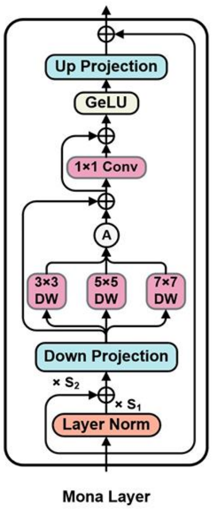


Fig. 3 - Architecture of the Multi-cognitive Visual Adapter

Detail-Preserving Contextual Fusion

To enhance feature fusion for rice-pest detection, the DPCF module is incorporated into the neck architecture (Fig. 4). DPCF addresses the limitations of static fusion schemes such as simple concatenation or summation, which may not sufficiently adapt to multi-scale pests, background interference and partial occlusion. Instead, DPCF performs content-adaptive fusion of hierarchical feature maps, enabling the model to preserve local details while incorporating broader contextual information.

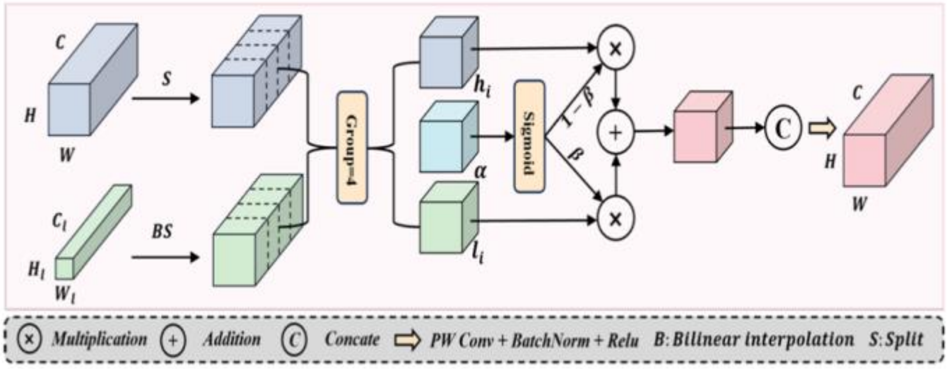


Fig. 4 - Architecture of the Detail-Preserving Contextual Fusion module

Structurally, DPCF is composed of two key components: Depthwise Separable Pyramid Convolution (DSPC) and Channel-wise Dynamic Fusion (CDF). DSPC applies a set of depthwise separable convolutions with multiple receptive fields (3×3, 5×5, 7×7) to enhance contextual representation at each scale. The outputs are then passed through the CDF module, which generates dynamic fusion weights across feature channels using attention mechanisms. These weights allow the model to selectively emphasize semantically relevant features while suppressing noisy or redundant ones. Residual connections are employed to preserve gradient stability and retain original semantic pathways.

Compared to conventional PANet-style feature aggregation, DPCF dynamically learns spatial-semantic relationships, yielding a more adaptive fusion process that is sensitive to local structure and global context. This is particularly advantageous for pest detection, where the targets are often small, densely distributed, and partially occluded. By modeling scale-aware and position-sensitive features, DPCF significantly improves the model's ability to discriminate subtle pest contours from noisy backgrounds.

Wise-IoU-NMS Non-Maximum Suppression Algorithm

The original YOLOv11n adopts conventional non-maximum suppression (NMS). When the Intersection over Union (IoU) between a candidate box and the highest-confidence box exceeds a predefined threshold, the candidate box is removed or heavily suppressed. This strategy may discard valid boxes in overlapping pest scenes and can therefore reduce recall. To alleviate this problem, this study adopts a Wise-IoU-based NMS strategy. The method adjusts suppression weights according to overlap and localization quality, so that boxes are retained or suppressed more adaptively rather than being removed only by a fixed IoU threshold. This improves localization stability for dense and partially occluded pest targets.

In this strategy, M denotes the detection box with the highest confidence score, B_i denotes the i-th candidate box, epsilon denotes the NMS threshold, s_i denotes the classification confidence score and IoU(B_i, M) denotes the overlap between the candidate box and M. The adjusted confidence score is used to determine whether the candidate box should be retained. This design reduces false suppression of valid pest boxes and mitigates duplicate detections caused by multiple high-confidence boxes corresponding to the same target.

For bounding-box regression, h and w denote the height and width of the minimum enclosing box, respectively, while x and y denote the center coordinates of the predicted and ground-truth boxes. The Wise-IoU formulation dynamically adjusts localization weights according to box quality, allowing the model to focus more effectively on difficult samples. The schematic representation is shown in Fig. 5.

Compared with traditional NMS, Wise-IoU-NMS considers overlap, center distance and localization quality jointly. This mechanism is particularly useful for small rice pests that appear close to one another or are partially occluded by leaves and stems. By reducing unnecessary suppression and duplicate alarm boxes, Wise-IoU-NMS improves practical detection reliability in complex field images.

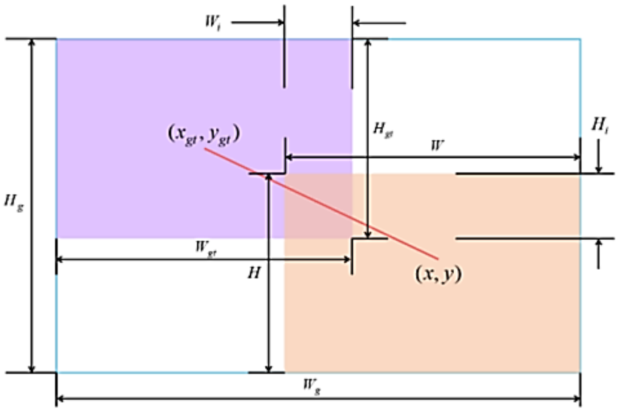


Fig. 5 - The architecture of the Wise-IoU

Overall Improved Network Architecture

This study introduces three key improvements to the YOLOv11n algorithm: Mona for lightweight feature adaptation, DPCF for detail-preserving contextual fusion and Wise-IoU-NMS for refined box suppression. The refined network architecture, denoted as YOLOv11n-Mona-DPCF-Wise-IoU, is depicted in Fig. 6.

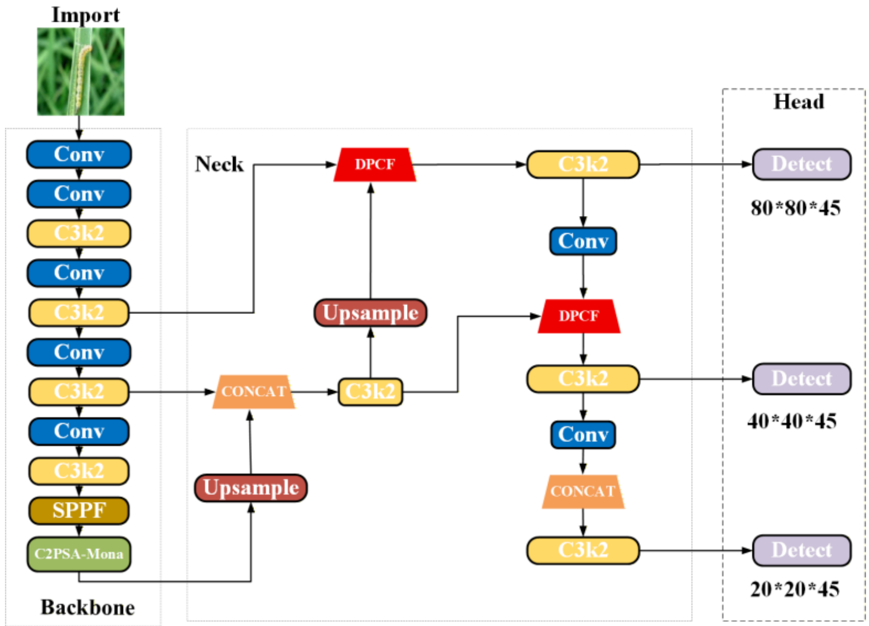


Fig. 6 - Improved YOLOv11n structure

RESULTS

A. Ablation Experiment

To evaluate the contribution of each module to detection accuracy and model complexity, a series of ablation experiments were conducted under identical training and testing conditions. Mona, DPCF and Wise-IoU-NMS were incrementally added to the YOLOv11n baseline. The results on the test set are summarized in Table 2.

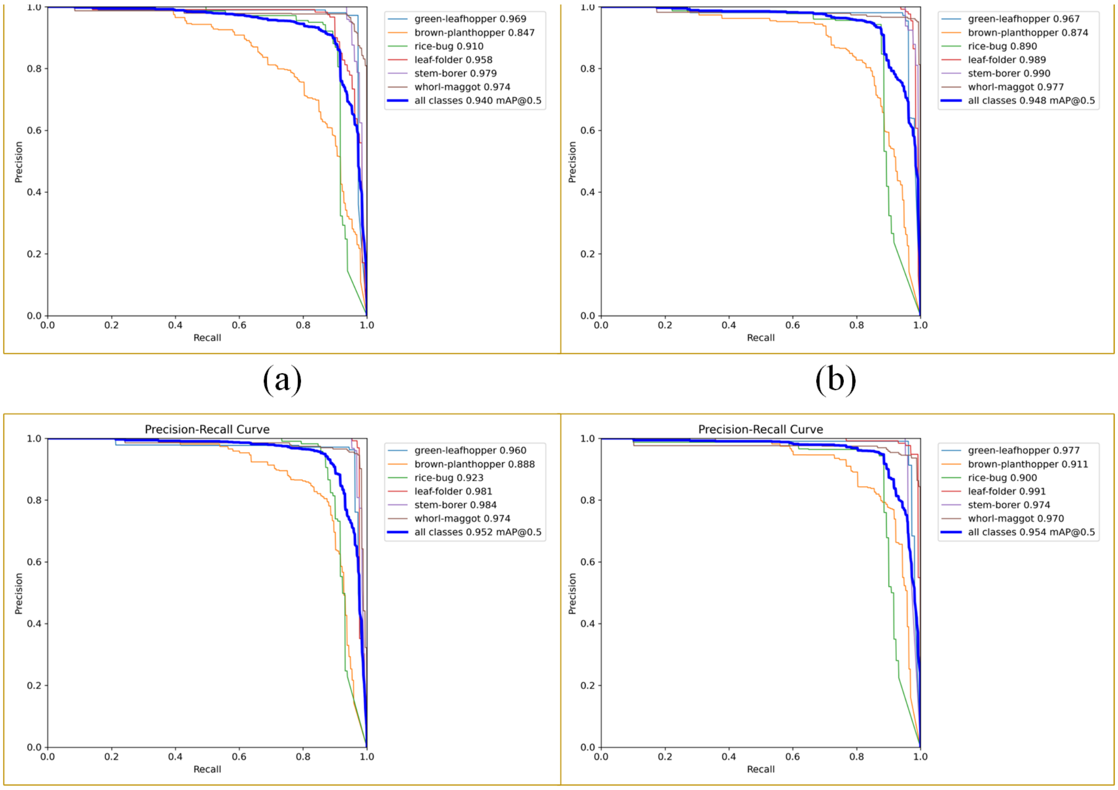
Table 2

Ablation results of the proposed model on the test set

Model	Params (M)	GFLOPs	mAP@50
YOLOv11n	2.58	6.3	0.94
YOLOv11n-Mona	2.63	6.4	0.948
YOLOv11n-Mona-DPCF	2.65	6.5	0.952
YOLOv11n-Mona-DPCF-Wise-IoU	2.65	6.5	0.954

The ablation results demonstrate the effectiveness of the proposed components. The baseline YOLOv11n achieved 94.0% mAP@50 with 2.58 M parameters and 6.3 GFLOPs. After introducing Mona, mAP@50 increased to 94.8%, while the parameter count and computational cost only rose slightly to 2.63 M and 6.4 GFLOPs. Adding DPCF further improved mAP@50 to 95.2%, indicating that detail-preserving contextual fusion strengthened multi-scale pest representation. Finally, the complete YOLOv11n-Mona-DPCF-Wise-IoU model achieved the best mAP@50 of 95.4% with 2.65 M parameters and 6.5 GFLOPs. These results show that the proposed modules improve accuracy with limited additional cost, which is important for practical edge deployment.

Figure 7 illustrates the class-wise Precision-Recall (PR) curves of four YOLOv11n variants, corresponding to the baseline and each stage of the proposed incremental enhancements. Subfigure (a) represents the baseline YOLOv11n, which achieved an mAP@0.5 of 0.940. Although the baseline performed well for stem borer (0.979) and whorl maggot (0.974), it showed weaker performance for brown planthopper (0.847), indicating class-specific sensitivity. After Mona was introduced, mAP@0.5 increased to 0.948, with improved precision-recall balance for visually similar pest classes. Incorporating DPCF further improved mAP@0.5 to 0.952 by strengthening multi-scale feature aggregation. The complete YOLOv11n-Mona-DPCF-Wise-IoU model achieved the highest mAP@0.5 of 0.954 and produced more stable PR curves across categories. This confirms that the proposed modules progressively improve class-wise detection reliability, as summarized in Table 2.



(a) (b) (c) (d)

Fig. 7 - Precision-Recall (PR) curves

Figure 8 presents the F1-confidence curves of four sequentially enhanced YOLOv11n models, providing insight into classification robustness and the selection of practical alarm thresholds.

The baseline YOLOv11n reached a peak macro-average F1 score of 0.91 at a confidence threshold of 0.472, but its F1 values were lower for brown planthopper and rice bug, reflecting the difficulty caused by class imbalance and inter-class similarity. After Mona was introduced, the peak F1 score increased to 0.93 at a threshold of 0.437. Adding DPCF further stabilized the curves, with a macro F1 score of 0.93 at 0.452. The complete YOLOv11n-Mona-DPCF-Wise-IoU model maintained the highest F1 score of 0.93 and achieved it at a confidence threshold of 0.476. Therefore, 0.476 was selected as the recommended operating threshold for practical pest alerts in the deployment experiment. Detections with confidence scores above this threshold were treated as valid alarm events, while lower-confidence predictions were retained only for diagnostic review.

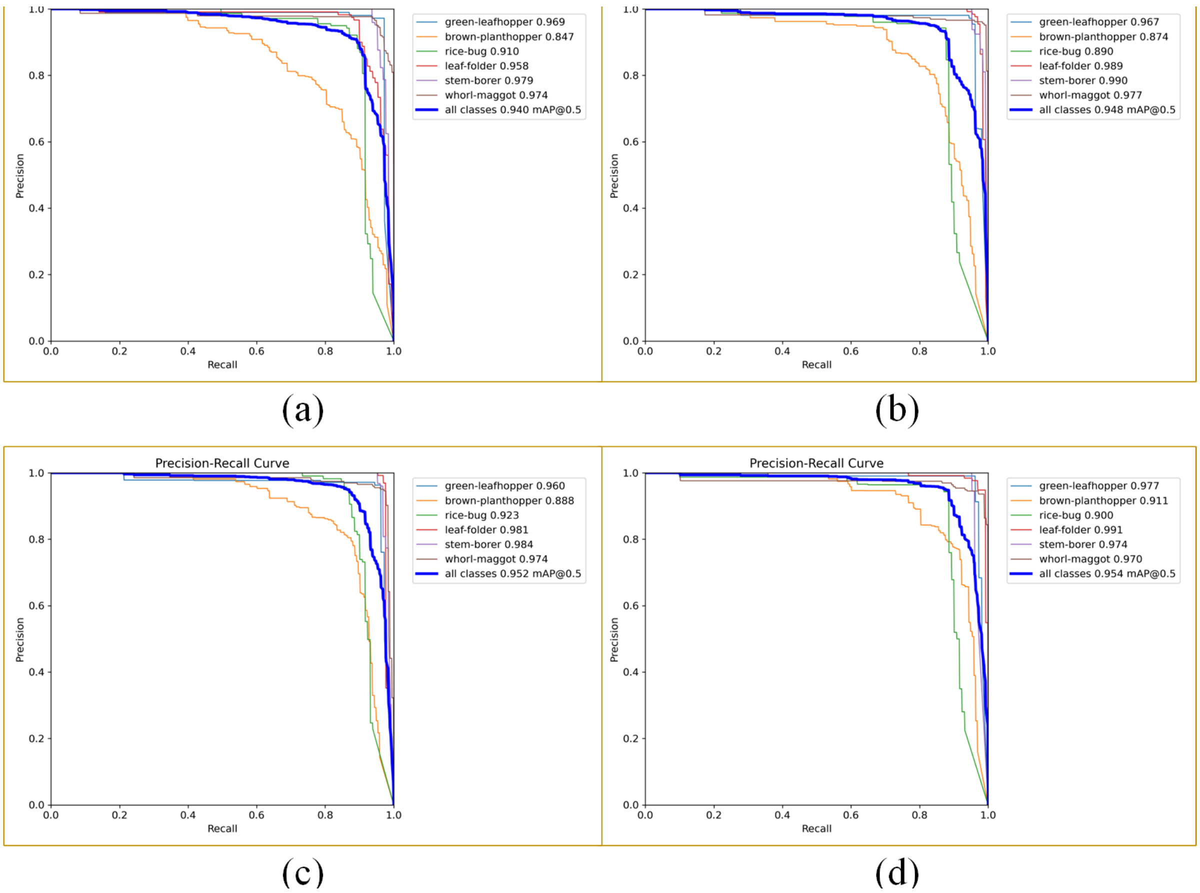


Fig. 8 - F1-Confidence curves

B. Comparative Experiment

To comprehensively evaluate the performance of the enhanced model, this study conducted comparative experiments between the proposed algorithm, YOLOv11n-Mona-DPCF-Wise-IoU, and several other prominent object detection algorithms, including SSD, Faster R-CNN, YOLOv10n, YOLOv12n, TPH-YOLOv5, RT-DETR, and DEIM. The results of these comparative experiments are summarized in Table 3.

Table 3

Comparison results with mainstream object-detection models

Model	Params (M)	GFLOPs	P	R	mAP@50	FPS
SSD	34.3	45.8	0.863	0.807	0.829	78
Faster R-CNN	69.2	217.8	0.894	0.917	0.858	84
YOLOv10n	2.69	8.2	0.911	0.904	0.933	106
YOLOv12n	2.50	5.8	0.931	0.914	0.939	117
TPH-YOLOv5	45.37	260.2	0.928	0.909	0.935	101
RT-DETR	31.99	103.3	0.933	0.921	0.937	95
DEIM	4.0	7.0	0.921	0.916	0.924	105
YOLOv11n	2.58	6.3	0.935	0.925	0.940	112
YOLOv11n-Mona-DPCF-Wise-IoU	2.65	6.5	0.948	0.947	0.954	109

The comparative results in Table 3 show that classical detectors such as SSD and Faster R-CNN obtained lower mAP@50 values of 82.9% and 85.8%, respectively, while requiring higher computational costs. YOLOv10n and YOLOv12n achieved 93.3% and 93.9% mAP@50 with high inference speeds, but their precision and recall were lower than those of the proposed method. The baseline YOLOv11n achieved 94.0% mAP@50 with 2.58 M parameters and 6.3 GFLOPs, demonstrating a strong lightweight baseline for rice-pest detection.

The proposed YOLOv11n-Mona-DPCF-Wise-IoU model achieved the best overall performance, with 95.4% mAP@50, 94.8% precision and 94.7% recall. It retained a compact model size of 2.65 M parameters and 6.5 GFLOPs and maintained high inference speed in the workstation test. Compared with the baseline, the proposed method improved mAP@50 by 1.4 percentage points while adding only 0.07 M parameters and 0.2 GFLOPs. These results indicate that the proposed architecture provides a more favorable balance between accuracy, robustness and computational efficiency than the compared detectors.

Figure 9 shows the mAP@50 convergence curves of the baseline YOLOv11n and the proposed YOLOv11n-Mona-DPCF-Wise-IoU model over 200 training epochs. The improved model converged faster in the early training stage and maintained a stable performance advantage throughout training, ultimately approaching an mAP@50 of 0.954. This indicates that Mona, DPCF and Wise-IoU-NMS not only improve detection accuracy but also enhance training stability.

Figure 10 visually compares the detection performance of YOLOv10n, YOLOv12n, YOLOv11n and the proposed YOLOv11n-Mona-DPCF-Wise-IoU model on three representative pest images. Across the samples, the improved model produced tighter bounding boxes and higher confidence scores.

In the first image, the proposed model localized the pest with a confidence score of 0.84, higher than the compared models (0.71-0.79). In the second image, which contained overlapping pests, the proposed model detected two targets with confidence scores of 0.87 and 0.83, whereas other models missed one target or produced lower-confidence detections. In the third image, which had high brightness and a complex background, the proposed model still detected the target with a confidence score of 0.88. These four valid detections all exceeded the recommended alarm threshold of 0.476, supporting the feasibility of threshold-based pest warning in practical monitoring.

Overall, the results demonstrate that Mona, DPCF and Wise-IoU-NMS improve both localization precision and detection stability, particularly under small-target, occlusion and illumination-challenging conditions. The remaining practical requirement is to verify whether this accuracy-efficiency balance can be maintained on low-power edge hardware; therefore, a Jetson Nano deployment validation was added in the following section.

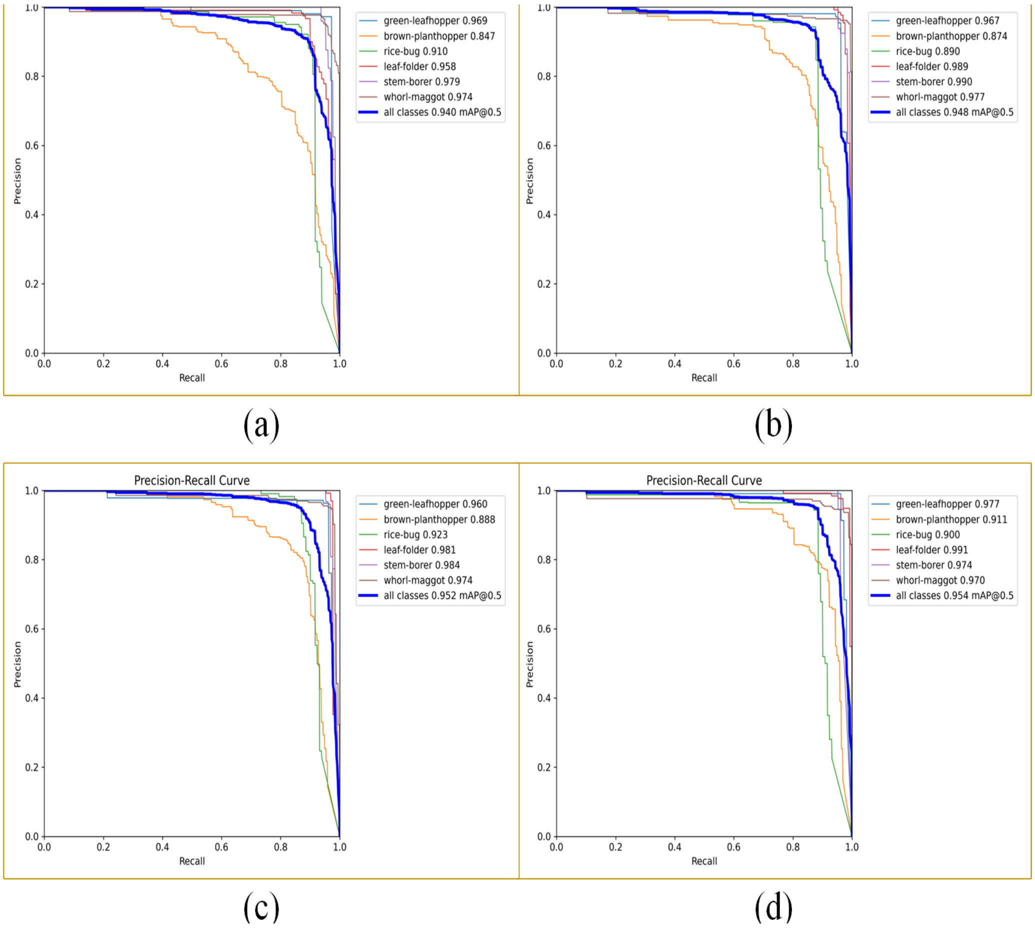


Fig. 9 - mAP@50 convergence curves

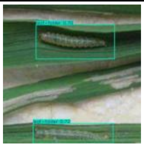
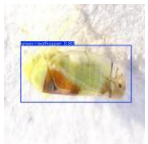
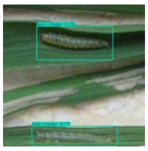
Image	1	2	3
YOLOv10n			
YOLOv12n			
YOLOv11n			
YOLOv11n-Mona-DPCF-Wise-IoU			

Fig. 10 - Comparison of model detection performance

C. Edge-Device Deployment and Practical Application Validation

To further clarify the engineering application scenario requested by the reviewer, the proposed model was deployed on Jetson Nano for rice-field pest monitoring (Fig. 11). The interface provides functions for opening images, videos and camera streams, starting or stopping detection and saving detection results. During operation, the model identifies six rice-pest categories, namely green leafhopper, brown planthopper, rice bug, leaf folder, stem borer and whorl maggot, and displays the predicted class, confidence score and bounding box in real time.

The deployment result demonstrates that the lightweight YOLOv11n-based model can be transferred from workstation evaluation to an embedded monitoring terminal. In field or greenhouse patrol scenarios, the confidence threshold can be used as an alarm rule: targets above the threshold are recorded as pest occurrences, while repeated frame records can be aggregated to estimate pest density. When connected to a fixed camera, mobile inspection platform or selective sprayer, the system can provide timely multi-class pest warnings and location-aware decision support for precision pest management.

Fig. 11 - Jetson Nano deployment interface of the rice-pest detection system

CONCLUSIONS

This study developed a lightweight rice-pest detector by integrating Mona, DPCF and Wise-IoU-NMS with YOLOv11n. The model achieved 95.4% mAP@50, 94.8% precision and 94.7% recall with 2.65 M parameters and 6.5 GFLOPs. Ablation and comparative experiments confirmed improved accuracy with limited computational overhead, particularly for small, occluded and complex-background targets. The Jetson Nano workflow further demonstrated the feasibility of embedded monitoring and threshold-based alarms. Future work should quantify long-term field robustness and end-to-end latency under varying environmental conditions.

ACKNOWLEDGEMENT

This study was funded by the Key Research and Development Program of Shandong Province (No. 2024RZB0202) and the Shandong Provincial Natural Science Foundation (No. ZR2025QC600).

REFERENCES

[1] Ahmad, I., Yang, Y., Yue, Y., Ye, C., Hassan, M., Cheng, X., Wu, Y., Zhang, Y. (2022). Deep Learning Based Detector YOLOv5 for Identifying Insect Pests (基于深度学习的 YOLOv5 检测器用于识别昆虫害虫). *Appl. Sci.*, Vol.12, 10167.

[2] Böckmann, E., Pfaff, A., Schirrmann, M., Pflanz, M. (2021). Rapid and low-cost insect detection for analysing species trapped on yellow sticky traps (用于分析黄色粘虫板上诱集物种的快速低成本昆虫检测方法). *Sci. Rep.*, Vol.11, 12854.

[3] Deserno, M., Briassouli, A. (2021). Faster R-CNN and EfficientNet for Accurate Insect Identification in a Relabeled Yellow Sticky Traps Dataset (基于 Faster R-CNN 和 EfficientNet 的重标注黄色粘虫板数据集昆虫 精准识别方法). In Proceedings of the 2021 IEEE International Workshop on Metrology for Agriculture and Forestry (MetroAgriFor), Trento-Bolzano, Italy, 3 November 2021; pp.209–214.

[4] Di, Y., Phung, S., Van Den Berg, J., Clissold, J., Bouzerdoum, A. (2023). TP-YOLO: A Lightweight Attention-Based Architecture for Tiny Pest Detection (TP-YOLO:一种用于微小害虫检测的轻量级注意力架构). In *Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP)*, Kuala Lumpur, Malaysia, 8 October 2023; pp.3394–3398.

[5] Gutierrez, A., Ansuategi, A., Susperregi, L., Tubío, C., Rankic, I., Lenža, L. (2019). A Benchmarking of Learning Strategies for Pest Detection and Identification on Tomato Plants for Autonomous Scouting Robots Using Internal Databases (基于内部数据库的自主侦察机器人番茄植株害虫检测与识别学习策略基准测试). *Sensors*, 2019, 5219471.

[6] Guan, S., Lin, Y., Lin, G., Su, P., Huang, S., Meng, X., Liu, P., Yan, J. (2024). Real-Time Detection and Counting of Wheat Spikes Based on Improved YOLOv10 (基于改进 YOLOv10 的小麦穗实时检测与计数). *Agronomy*, Vol.14, 1936.

[7] Huang, J., Huang, Y., Huang, H., Zhu, W., Zhang, J., Zhou, X. (2022). An Improved YOLOX Algorithm for Forest Insect Pest Detection (一种用于森林害虫检测的改进 YOLOX 算法). *Comput. Intell. Neurosci.*, 2022, 5787554. [PubMed]

[8] Jia, L., Wang, T., Chen, Y., Zang, Y., Li, X., Shi, H., Gao, L. (2023). MobileNet-CA-YOLO: An Improved YOLOv7 Based on the MobileNetV3 and Attention Mechanism for Rice Pests and Diseases Detection (MobileNet-CA-YOLO：基于 MobileNetV3 和注意力机制的改进 YOLOv7 用于水稻病虫害检测). *Agriculture*, Vol.13, 1285.

[9] Khalid, S., Oqaibi, H., Aqib, M., Hafeez, Y. (2023). Small Pests Detection in Field Crops Using Deep Learning Object Detection (基于深度学习目标检测的田间作物微小害虫检测). *Sustainability*, Vol.15, 6815.

[10] Liu, J., Wang, X. (2020). Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network (基于改进 Yolo V3 卷积神经网络的番茄病虫害检测). *Front. Plant Sci.*, Vol.11, 898.

[11] Pinto-Zevallos, D., Vänninen, I. (2013). Yellow sticky traps for decision-making in whitefly management: What has been achieved? (用于粉虱管理决策的黄色粘虫板：取得了哪些成果？). *Crop Prot.*, Vol.47, pp.74–84.

[12] Ramalingam, B., Mohan, R., Pookkuttath, S., Gómez, B., Borusu, C., Teng, T., Tamilselvam, Y. (2020). Remote Insects Trap Monitoring System Using Deep Learning Framework and IoT (基于深度学习框架和物联网的远程昆虫诱捕监测系统). *Sensors*, Vol.20, 5280. [PubMed]

[13] Sunkara, R., Luo, T. (2023). No More Strided Convolutions or Pooling: A New CNN Building Block for Low-Resolution Images and Small Objects (摒弃步幅卷积或池化：一种用于低分辨率图像和小目标的新型 CNN 构建模块). Springer, Berlin/Heidelberg, Vol.13715, pp.443–459.

[14] Tan, L., Liu, S., Gao, J., Liu, X., Chu, L., Jiang, H. (2024). Enhanced Self-Checkout System for Retail Based on Improved YOLOv10 (基于改进 YOLOv10 的零售增强自助结账系统). J. Imaging, Vol.10, 248.

[15] Wen, C., Chen, H., Ma, Z., Zhang, T., Yang, C., Su, H., Chen, H. (2022). Pest-YOLO: A model for large-scale multi-class dense and tiny pest detection and counting (Pest-YOLO：一种用于大规模多类别密集微小害虫检测与计数的模型). *Front. Plant Sci.*, Vol.13, 973985.

[16] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G. (2024). YOLOv10: Real-Time End-to-End Object Detection (YOLOv10：实时端到端目标检测). arXiv, 2024, arXiv:2405.14458.

[17] Xia, D., Chen, P., Wang, B., Zhang, J., Xie, C. (2018). Insect Detection and Classification Based on an Improved Convolutional Neural Network (基于改进卷积神经网络的昆虫检测与分类). *Sensors*, Vol.18, 4169.

[18] Yang, S., Zhou, G., Feng, Y., Zhang, J., Jia, Z. (2024). SRNet-YOLO: A model for detecting tiny and very tiny pests… (SRNet-YOLO：一种用于检测微小及极微小害虫的模型…). *Front. Plant Sci.*, Vol.15, 1416940.

[19] Zhang, W., Huang, H., Sun, Y., Wu, X. (2023). AgriPest-YOLO: A rapid light-trap agricultural pest detection method based on deep learning (AgriPest-YOLO：一种基于深度学习的快速灯光诱集农业害虫检测方法). *Front. Plant Sci.*, Vol.13, 1079384.

[20] Zhang, L., Zhao, C., Feng, Y., Li, D. (2023). Pests Identification of IP102 by YOLOv5 Embedded with the Novel Lightweight Module (嵌入新型轻量级模块的 YOLOv5 用于 IP102 害虫识别). *Agronomy*, Vol.13, 1583.