Automation and Machine Learning (2026)
Clausius Scientific Press, Canada
DOI: 10.23977/autml.2026.070114
ISSN 2516-5003 Vol. 7 Num. 1

Research on Farmland Pest Image Detection and Recognition Based on SE-YOLOv5

Wangxia Xia, Jianzhong Sun

School of Information and Intelligent Engineering, University of Sanya, Sanya, 572022, Hainan, China

Keywords: Farmland Pest Detection; YOLOv5; SE Attention Mechanism; Small Target Detection

Abstract: Aiming at the problems of low efficiency and high missed detection rate in traditional manual farmland pest detection, as well as the defects of the original YOLOv5 algorithm in detecting small-target pests and pests in complex backgrounds, an improved YOLOv5 algorithm integrated with the SE attention mechanism (SE-YOLOv5) is proposed to realize accurate and real-time detection of farmland pests. Taking six common rice pests as the research objects, a pest dataset containing 4656 images was constructed, and the robustness of the dataset was improved through preprocessing steps such as label cleaning, size normalization and data augmentation. The SE attention mechanism was embedded between the C3 module and the SPPF layer of the YOLOv5s backbone network to strengthen the model's adaptive weight assignment for pest feature channels, highlight effective features and suppress redundant background features. The experimental results show that the mAP@50 of the SE-YOLOv5 algorithm reaches 97.0%, an increase of 2.0 percentage points compared with the original YOLOv5s. The inference speed for a single image is 6.0 ms, and the detection standard deviation and coefficient of variation are reduced to 0.031 and 0.032, respectively. The detection accuracy of small-target pests such as brown planthoppers is significantly improved. This algorithm performs excellently in detection accuracy, real-time performance and stability, providing an effective technical solution for the intelligent monitoring of farmland pests.

1. Introduction

Farmland pests are one of the core factors restricting the yield and quality of crops. Failure to monitor and control them in a timely manner will lead to severe economic losses and threaten food security. China has a large planting area of food crops such as rice and wheat, with a wide variety of pests that spread rapidly. Traditional detection methods rely on manual field inspections, which are greatly affected by personnel experience, subjective judgment and field environment. They have the problems of low detection efficiency, high missed detection and misjudgment rates, and poor timeliness, which are difficult to meet the large-scale and intelligent prevention and control needs of modern agriculture.

With the rapid development of computer vision and deep learning technologies, agricultural pest and disease image recognition technology based on target detection algorithms has become a research hotspot. Deep learning target detection algorithms are divided into two-stage (e.g., Faster R-CNN) and one-stage (e.g., YOLO series, SSD) categories. Among them, the YOLO series algorithms are more suitable for the real-time detection scenario of farmland pests due to their advantages of end-to-end detection, fast inference speed and model lightweight. In recent years, deep learning models based on attention mechanisms have achieved remarkable progress in the fine-grained classification of plant diseases and pests [1]. As a classic lightweight algorithm in this series, YOLOv5 has the characteristics of high detection accuracy and low deployment difficulty, and is widely used in agricultural pest and disease detection. However, in actual farmland scenarios, pest targets have problems such as small size, high color similarity with the background, and diverse forms. The original YOLOv5 has insufficient feature extraction ability for small-target pests, which is prone to missed detection and low detection accuracy.

To address the above problems, this paper proposes an improved YOLOv5 algorithm integrated with the SE attention mechanism. By embedding the SE attention mechanism in the backbone network, the model's ability to learn key features of pests is enhanced, and the detection accuracy of small-target pests and pests in complex backgrounds is improved. At the same time, a standardized rice pest dataset is constructed, and comparative experiments are designed to verify the performance of the improved algorithm, providing technical support for the intelligent monitoring of farmland pests.

2. Relevant Theories and Technical Foundations

2.1 Core Structure of YOLOv5 Algorithm

This paper selects YOLOv5s_v6.0 as the base model, which has fewer parameters and fast inference speed, suitable for lightweight detection scenarios. Its network structure is mainly divided into three parts: backbone network, neck network and head network:

Backbone network: Composed of Conv layer, C3 layer and SPPF layer, it is responsible for image feature extraction. The Conv layer (convolution + batch normalization + SiLU activation) realizes basic feature extraction; the C3 layer is based on the CSPNet structure, which deepens the network depth through residual connection and improves the feature extraction ability; the SPPF layer integrates local and global features through multi-scale spatial pyramid pooling, enhancing the model's adaptability to targets of different sizes.

Neck network: Adopts the FPN+PAN structure to realize multi-scale feature fusion and transmission, combining high-level semantic features with low-level detailed features to improve the target detection ability.

Head network: Responsible for target classification and bounding box regression, outputting the target category probability and position coordinates to complete the final detection.

2.2 SE Attention Mechanism

Proposed by Hu Jie et al., the core of the SE attention mechanism is to obtain the importance of feature channels through learning, assign high weights to effective feature channels and low weights to redundant feature channels, so as to realize adaptive optimization of feature channels [2]. This mechanism has a simple structure and low parameter overhead, does not change the size of the feature map of the original network, and can be flexibly embedded in convolutional neural networks. Attention mechanisms have been proven to effectively improve the feature expression ability in small-scale target detection tasks [3], and its workflow is divided into three steps:

(1) Squeeze: Perform global average pooling on the input feature map, convert the two-dimensional feature map into a one-dimensional feature vector, and obtain the global receptive field

of each channel.

(2) Excitation: Construct a nonlinear mapping between channels through two fully connected layers, learn the channel weight coefficients, and normalize the weights to the interval [0,1] through the Sigmoid activation function.

(3) Scale: Multiply the weight coefficients with the original feature map channel by channel, adjust the feature channels by weighting, highlight effective features and suppress redundant features.

The SE attention mechanism can effectively strengthen the model's sensitivity to key features, and is especially suitable for detection tasks with inconspicuous target features and complex backgrounds, providing an effective idea for solving the problem of small-target detection of farmland pests.

3. Construction and Preprocessing of Farmland Pest Dataset

3.1 Research Objects and Data Sources

Taking six common pests of rice, the main food crop in southern China, as the research objects, namely brown planthopper, green leafhopper, leaf folder, rice worm, borer and rotifer. Among them, the brown planthopper is a typical small-target pest with high color similarity to the rice background, leading to high detection difficulty. The experimental dataset is sourced from public dataset websites, and a total of 4656 original pest images are collected, covering pest samples under different light conditions, shooting angles and field backgrounds, providing a rich foundation for model training.

3.2 Data Preprocessing

The original images have problems such as inconsistent labels, differences in image quality and noise interference, which require standardized preprocessing:

(1) Label cleaning and reconstruction: The public datasets have the problems of inconsistent labels for the same pest and non-uniform label formats. The LabelImg annotation tool is used to re-annotate pest targets in YOLO format, and generate txt format label files containing pest categories and normalized coordinates of bounding boxes.

(2) Image normalization and denoising: All images are uniformly adjusted to 640×640 pixels to match the input size of YOLOv5s; the Gaussian filtering algorithm is used to remove image noise, improve image clarity and reduce the interference of noise on feature extraction.

(3) Data augmentation: To solve the problems of low proportion of small-target samples and insufficient dataset diversity, three methods are adopted for data augmentation: data stitching (randomly stitching different pest images to simulate the coexistence scenario of multiple pests), scale scaling (scaling images at a random ratio of [0.5,2.0]), and geometric transformation (random clockwise rotation by 90°), so as to improve the generalization ability of the model.

Finally, the processed dataset is divided into training set, validation set and test set at a ratio of 75%:20%:5%. The training set is used for model parameter learning, the validation set for hyperparameter adjustment, and the test set is an independent dataset without manual labels, which simulates the actual application scenario to evaluate the model performance.

3.3 Experimental Platform and Evaluation Metrics

The experiment builds a platform based on the Windows 11 operating system, with hardware configuration of Intel i5-13600KF CPU, 32G memory, NVIDIA RTX 4060Ti 16GB GPU, parallel computing environment of CUDA 11.6, programming language of Python 3.9 and deep learning framework of PyTorch.

Classic target detection evaluation metrics are selected to assess the model performance:

Intersection over Union (IoU) measures the overlap degree between the predicted bounding box and the real bounding box with a set threshold of 0.5; mean Average Precision (mAP@50) is the core metric reflecting the overall detection accuracy of the model; inference speed is the detection time (ms) for a single image, reflecting real-time performance; standard deviation and coefficient of variation are also introduced to evaluate the detection stability of the model for different pests.

4. Design of SE-YOLOv5 Algorithm and Model Training

4.1 Embedding Strategy of SE Attention Mechanism

The core reason for the low detection accuracy of the original YOLOv5s for small-target pests is that it is difficult to distinguish the effective pest features from redundant background features in the feature extraction process. Therefore, the SE attention mechanism is embedded between the C3 module and the SPPF layer of the backbone network, which is the late stage of feature extraction in the backbone network where high-level semantic features are extracted. Embedding the SE attention mechanism at this position can achieve:

(1) Channel weighting optimization of high-level pest features, highlighting the key features of small-target pests and suppressing redundant field background features.

(2) Better integration of global and local features, improving the model's perception and positioning accuracy of targets.

(3) Allowing the network to dynamically adjust the weights of feature channels according to the input image, enhancing the adaptability to different field scenarios.

Similar studies have embedded the CBAM attention mechanism in YOLOv5s for pest detection, with the mAP increased by 12.9 percentage points [4]; the introduction of attention modules in multi-scale feature fusion has also been verified effective [5]. After the embedding of the SE attention mechanism, the feature transfer process and feature map size of the original network are unchanged, ensuring the lightweight and inference speed of the model. The network structure of the improved SE-YOLOv5 is shown in Figure 1.

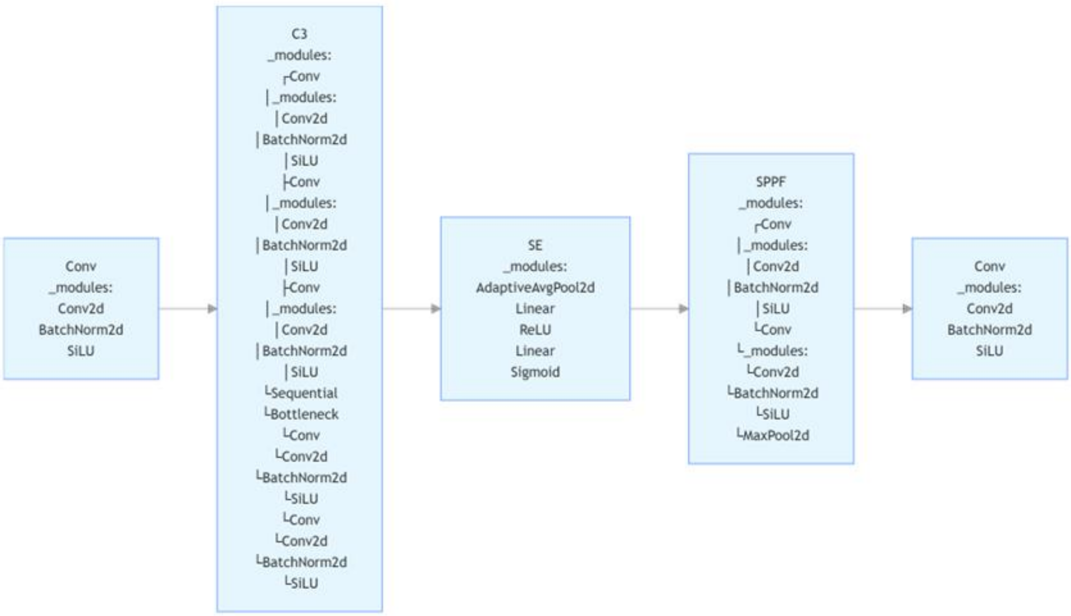

Figure 1: Network structure of SE-YOLOv5

4.2 Hyperparameter Setting for Model Training

To improve the training efficiency and generalization ability of the model, transfer learning is adopted with the official pre-trained weight of yolov5s.pt as the initial weight. The hyperparameter settings for model training are as follows: input size of 640×640, optimizer of SGD, learning rate of 0.01, momentum of 0.937, batch size of 24, initial training epochs of 50. Monitored by the validation set, the model did not converge completely, and the epochs were finally adjusted to 100. In the training process, CIoU loss is used as the bounding box regression loss, and BCEWithLogits loss as the classification and confidence loss. The model parameters are continuously optimized through backpropagation.

5. Experimental Results and Analysis

5.1 Single Experimental Results of SE-YOLOv5 Model

After 100 epochs of training, the loss function value and validation set performance metrics of the SE-YOLOv5 model tend to converge. During the training process, box_loss, obj_loss and cls_loss gradually decrease and stabilize, and mAP@50 gradually rises to 97.0% and remains stable, indicating that the model has no overfitting or underfitting phenomenon.

The detection accuracy of the model for six rice pests is shown in Table 1. The overall mAP@50 reaches 97.0%, and the detection accuracy of all pests exceeds 90%, among which the detection accuracy of green leafhopper and borer reaches 99.0%; the detection accuracy of brown planthopper is increased to 90.9%, a significant improvement compared with the original YOLOv5, solving the problem of low detection accuracy of the original model for small-target pests. The inference speed of the model for a single image is 6.0 ms, meeting the engineering requirements of real-time farmland pest detection.

Table 1: Detection results of SE-YOLOv5 model for different pests (mAP@50)

Category	**mAP50**
Overall	0.970
Brown planthopper	0.909
Green leafhopper	0.990
Leaf folder	0.979
Rice worm	0.976
Borer	0.990
Rotifer	0.976

5.2 Comparative Experimental Results and Analysis

To verify the superiority of the SE-YOLOv5 algorithm, a comparative experiment is designed to compare the performance of the original YOLOv5 and SE-YOLOv5. The performance comparison results of each model are shown in Table 2.

Table 2: Performance comparison of different models for farmland pest detection

Model method	mAP@50/%	Inference speed/ms	Standard deviation	Coefficient of variation
YOLOv5	95.0	6.5	0.059	0.063
SE-YOLOv5	97.0	6.0	0.031	0.032

It can be seen from Table 2 that the performance of the improved model is superior to the original YOLOv5 in all aspects:

(1) Detection accuracy: The mAP@50 reaches 97.0%, an increase of 2.0 percentage points compared with the original YOLOv5, indicating that the SE attention mechanism has a more significant optimization effect on pest feature channels.

(2) Real-time performance: The inference speed for a single image is 6.0 ms, the fastest among all comparative models, indicating that the embedding of the SE attention mechanism does not increase the model computation, but instead improves the inference efficiency through feature optimization.

(3) Detection stability: The standard deviation of 0.031 and coefficient of variation of 0.032 are the lowest values, indicating that the model has the smallest difference in detection accuracy for six pests, more balanced detection, and significantly improved adaptability to small-target pests and pests in complex backgrounds.

5.3 Detection Effect in Actual Scenarios

Rice pest images taken in the field are selected for testing, including actual scenarios such as different light conditions, coexistence of multiple pests and complex backgrounds. The results show that SE-YOLOv5 can quickly and accurately identify pest targets in the images, accurately frame the pest positions, and effectively detect small-target pests such as brown planthoppers and green leafhoppers with similar colors to rice leaves, without obvious missed detection or misjudgment; while the original YOLOv5 has partial missed detection of small-target pests in actual scenarios, and the prediction accuracy of bounding boxes for pests in complex backgrounds is low. This indicates that the SE-YOLOv5 algorithm has a good detection effect in actual farmland scenarios and can meet the actual application requirements.

6. Conclusions and Prospects

6.1 Conclusions

Aiming at the actual demand of farmland pest detection, this paper proposes an improved YOLOv5 algorithm integrated with the SE attention mechanism, and carries out experimental research on six common rice pests. The main conclusions are as follows:

(1) The constructed rice pest dataset has good diversity and robustness after preprocessing and data augmentation, including 4656 standardized images, providing a reliable sample basis for the training of farmland pest detection models.

(2) Embedding the SE attention mechanism between the C3 module and the SPPF layer of the YOLOv5s backbone network can effectively strengthen the model's ability to learn key pest features, highlight the features of small-target pests and suppress redundant background features, without changing the original network structure and ensuring the model lightweight.

(3) Experimental results show that the SE-YOLOv5 algorithm has significantly improved detection stability, and the detection accuracy of small-target pests such as brown planthoppers is

obviously improved.

(4) Actual scenario tests show that the algorithm can effectively cope with farmland scenarios with different light conditions and complex backgrounds, with high detection accuracy and good real-time performance, providing an efficient technical solution for the intelligent monitoring of farmland pests.

6.2 Prospects

The SE-YOLOv5 algorithm proposed in this paper has achieved good results in farmland pest detection, and can be further optimized from the following aspects in the future:

(1) Expanding multi-crop pest detection: At present, the research is only carried out on rice pests. In the future, pest samples of wheat, corn, vegetables and other crops can be expanded to optimize the model generalization ability and realize the unified detection of multiple crops and multiple pests.

(2) Fusion of multi-attention mechanisms: The SE attention mechanism only optimizes feature channels without considering spatial position information. In the future, it can be combined with attention mechanisms such as CA and CBAM that integrate spatial-channel information to further improve the feature extraction ability [4].

(3) Lightweight deployment of the model: Combined with technologies such as model pruning, quantization and knowledge distillation, the SE-YOLOv5 is optimized for lightweight, reducing the hardware deployment requirements and realizing the edge-side deployment on embedded devices such as UAVs and field monitoring cameras.

(4) Multi-modal data fusion: Combine image data with field environment data (temperature, humidity) and pest occurrence law data to build a multi-modal pest detection model, improve the accuracy and prediction ability of pest detection, and provide support for the early prevention and control of pests.

Acknowledgement

This Project supported by the Education Department of Hainan province (Project number: Hnky2025-38)

References

[1] Wu Yunzhi, Hu Nan, Iftikhar Ahmad, et al. Fine-Grained Classification of Plant Diseases and Pests Based on Transfer Learning[J]. Journal of Suzhou University, 2024, (9): 29-33.

[2] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(08): 2011-2023.

[3] Sheng Wenshun, Yu Xiongfeng, Lin Jiayan, et al. Small-Scale Object Detection Algorithm Fusing Attention and Feature Pyramid[J]. Computer Engineering, 2024, (1): 242-250.

[4] Zheng Junjian, Lan Yubin, Xiong Wanjie, et al. Pest Identification Method for Pakchoi Based on Improved YOLOv5s Model[J]. Transactions of the Chinese Society of Agricultural Engineering, 2024, 40(13): 124-133.

[5] Qu Haicheng, Wang Meng, Chai Rui. Efficient Vehicle Detection in Remote Sensing Images Based on Bidirectional Multi-Scale Feature Fusion[J]. Computer Engineering and Applications, 2024, (12): 346-356.