Proceedings of Machine Learning Research vol 278:1–8, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

YOLOv5n-MobileNetv4: A Lightweight Crop Pest Detection Algorithm

Qin Yang QINYANG_8866@163.COM
Leshan Vocational and Technical College, Leshan 614000, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Pests significantly impact crop quality and yield, making their management crucial for agriculture. The real-time and accurate identification of agricultural pests is essential for implementing effective pest control measures, ensuring timely intervention, and minimizing crop losses. Existing pest detection systems face challenges such as low accuracy and excessive parameters, which hinder their efficiency and practicality in real-world applications. Therefore, this paper proposed a real-time pest detector for embedded devices by combining the state-of-the-art mobile device-based MobileNetv4 and lightweight You Only Look Once (YOLO) v5n object detection algorithms, achieving high efficiency and performance. The proposed YOLOv5n-MobileNetv4 model replaced the YOLOv5n backbone with the MobileNetv4 backbone, effectively reducing parameters while maintaining high detection accuracy. Experimental results showed that the improved model achieved 82.1% mAP50 at 87.7 frames per second (FPS). It achieved a 36.2% reduction in parameters and a 31.1% increase in speed, with a slight accuracy drop.

Keywords: Pest Detection, YOLOv5n, MobileNetv4, Embedded Devices.

1. Introduction

Crop pests pose a significant challenge to agricultural development, directly impacting crop output. Traditional pest detection relies on expert on-site inspections, which are time-consuming and ineffective for timely intervention (Jiang, 2023). The pest control industry faces challenges such as insufficient monitoring and early warning systems, improper pesticide use, and delays in effectiveness. However, with the modernization of agriculture, deep learning has emerged as a breakthrough in target detection, offering innovative solutions for efficient crop pest monitoring (Ali et al., 2023).

Traditional deep learning models usually require numerous computing resources for training and inference, which may not be feasible for some resource-limited devices, such as embedded devices used in agricultural fields (Kuzuhara et al., 2020). The deep learning-based YOLOv5 object detection algorithm excels in crop pest detection with advantages such as real-time performance, high accuracy, effective small-object detection, and a lightweight design ideal for embedded devices and drone deployment. Furthermore, YOLOv5n is the fastest detection network in the YOLOv5 series. Therefore, a cost-effective real-time crop pest detection algorithm based on YOLOv5n for agricultural IoT edge devices was proposed to balance accuracy and speed.

2. Related Works

Based on the idea of meta-learning, Guo and Shang (2024) designed a VGG-based prototype network called VGG-meta learning (VGG-ML) for identifying plant pest species in small samples in 2024. VGG16 was used to extract pest features, and the model was trained with a heterogeneous approach to improve accuracy and recognize new pest categories, even with small sample sizes. The proposed network achieved an accuracy of 81.85% on a test set of eight common rice pests.

In 2023, Ding et al. (2023) designed and built a crop pest monitoring system based on the Internet of Agriculture (IOA) technology and the YOLOX-s target detection algorithm. The system's overall framework comprised four layers: the information perception layer, the information transmission layer, the information service layer, and the system application layer. The proposed method achieved 84.98 mAP50 on the Locust and Longicorn test dataset.

In 2023, Han et al. (2023) proposed a crop disease detection model using infrared thermal imaging and an improved YOLOv5, with CSP-Darknet as the backbone. The model replaced stride 2 convolution with the SPD-Conv module to improve accuracy while maintaining parameter size, and outputs three feature layers of different scales. The SE mechanism was also introduced to enhance feature extraction by recalibrating channel features. The proposed model achieved a mAP50 of 85.7%.

Liu et al. (2023) proposed the MSDB-ResNet model for rice pest and disease recognition, based on an improved ResNet with a multi-scale dual-branch structure in 2023. The model introduced the ConvNeXt residual block to optimize the calculation ratio of the residual block and built a dual-branch structure. By adjusting the convolution kernel size in each branch, it effectively extracted disease features of different sizes from input images.

In 2023, Xu et al. (2023) proposed a solution combining data augmentation and Real-ESRGAN super-resolution enhancement to create a high-quality wheat pest dataset, IP-AugESRWheat, addressing issues like imbalanced categories, small scale, and low resolution . They also introduced the lightweight and efficient ECA-EffV2 model to improve the extraction of wheat pest features, enhancing the model's performance. The proposed algorithm achieved 84.8% accuracy.

Although the above methods achieve high detection accuracy for pest and disease recognition, they do not address the challenges faced by resource-constrained edge devices on the Internet of Things (IoT). Therefore, reducing model parameters and model complexity is a good solution.

3. Methodology

This paper built a crop pest detector using the YOLOv5n model, which had the smallest complexity in the YOLOv5 series. To further reduce the model's complexity while maintaining performance, the MobileNetv4 backbone was used to replace the original YOLOv5n backbone. This modification achieved a balance between detection accuracy and processing speed, making it suitable for resource-constrained environments.

3.1. YOLOv5n Architecture

YOLOv5n presents a highly efficient and compact architecture that optimizes detection performance, as shown in Fig. 1. The model's architecture is composed of three primary components: the backbone, responsible for extracting features from the input data; the neck, which performs feature fusion to combine multi-scale features for improved representation; and the head, which is responsible for making the final object predictions using feature maps sizes of 80×80, 40×40, and 20×20. This modular design enables the model to efficiently process images and deliver accurate detection results.

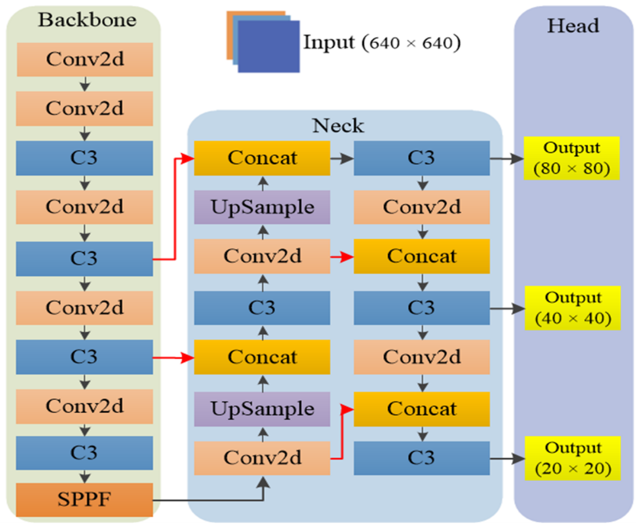


Figure 1: YOLOv5 structure.

3.2. MobileNetv4 Network

MobileNetv4 is an efficient CNN architecture developed by Google, specifically designed to enhance deep learning applications on mobile and edge computing devices. Building on the previous versions, MobileNetv4 places greater emphasis on optimizing computational resources while still improving performance, particularly in processing speed and model size. Key advancements include the introduction of the Universal Inverted Bottleneck (UIB) search block and Mobile MQA, which help optimize the model's efficiency through a refined neural architecture search method. Table 1 displays the overall architecture of the MobileNetv4 backbone.

Table 1: Overall architecture of the MobileNetv4 backbone.

Input	Operator	Output
$640 \times 640 \times 3$	Conv	$320 \times 320 \times 32$
$320 \times 320 \times 32$	Conv $\times$ 2	$160 \times 160 \times 16$
$160 \times 160 \times 16$	Conv	$80 \times 80 \times 48$
$80 \times 80 \times 48$	Conv	$80 \times 80 \times 32$
$80 \times 80 \times 32$	Universal_Inverted_Residual $\times$ 6	$40 \times 40 \times 48$
$40 \times 40 \times 48$	Universal_Inverted_Residual $\times$ 6	$20 \times 20 \times 64$
$20 \times 20 \times 64$	Conv	$20 \times 20 \times 480$

Where the shape of the feature map denotes length $\times$ width $\times$ channel. When the input shape is $640 \times 640 \times 3$, the MobileNetv4 backbone outputs three feature maps with shapes of $80 \times 80 \times 32$, $40 \times 40 \times 48$, and $20 \times 20 \times 480$ to predict small, medium, and large objects. First, the feature map with a shape of $80 \times 80 \times 32$ is obtained through four convolutional layers. Second, the feature map with a shape of $40 \times 40 \times 48$ is obtained through six Universal_Inverted_Residual layers. Similarly, the feature map with a shape of $20 \times 20 \times 64$ is obtained through six Universal_Inverted_Residual layers. Finally, the feature map with a shape of $20 \times 20 \times 480$ is obtained through a convolutional layer.

3.3. Backbone Improvement

To maintain architectural compatibility with the original model, the MobileNetv4 backbone replaced the YOLOv5n backbone by generating three feature maps with dimensions of 80 × 80, 40 × 40, and 20 × 20. This dimensional consistency was critical for maintaining the hierarchical feature representation capabilities of the network. Fig. 2 shows the integration of the MobileNetv4 backbone, preserving the model's detection framework while enhancing computational efficiency.

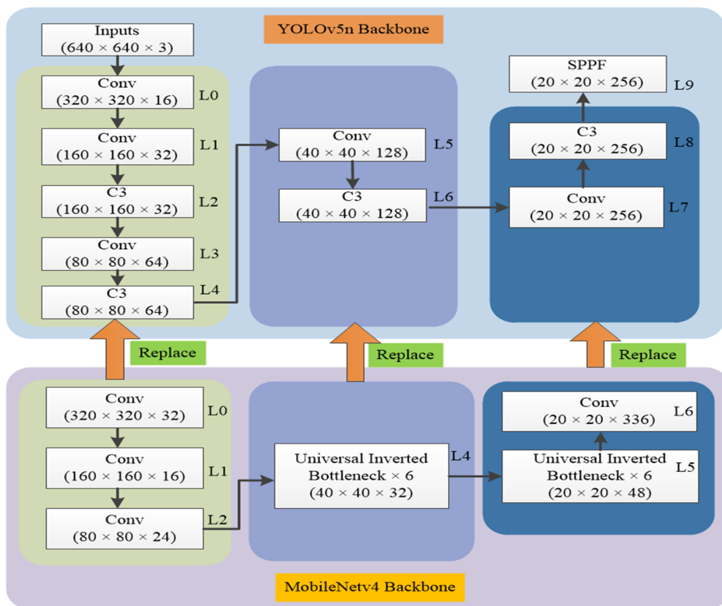


Figure 2: Backbone improvement.

The hierarchical feature extraction process was restructured through systematic layer substitution: MobileNetv4's initial layers (0-2) supplanted YOLOv5s layers 0-4 for generating 80 × 80 feature maps, while layer 4 of MobileNetv4 replaced YOLOv5s layers 5-6 to produce 40 × 40 feature maps. The final transformation involved MobileNetv4 layers 5-6 superseding YOLOv5s layers 7-8 for 20 × 20 feature map generation.

4. Results

The customized pest dataset used in this study consisted of 2,029 images sourced from the IP102 pest dataset. Of these, 1,623 were for training, and 406 were designated for validation. The GPU used for training was the Nvidia RTX 4060. The training epochs, image size, and batch size were set to 200, 640, 32. The model's performance was evaluated using frames per second (FPS) to measure detection speed and mean average precision at IoU threshold 0.5 (mAP_0.5) to assess detection accuracy.

4.1. Training Curves

Fig. 3 shows the training curves of the YOLOv5n model. The box_loss, obj_loss, and cls_loss gradually dropped during the training and validation period. The recall, mAP_0.5, and mAP_0.5:0.95

metrics gradually increased during 200 training epochs. In contrast, the precision metric decreased in the initial 40 epochs and increased gradually in the last 160 epochs.

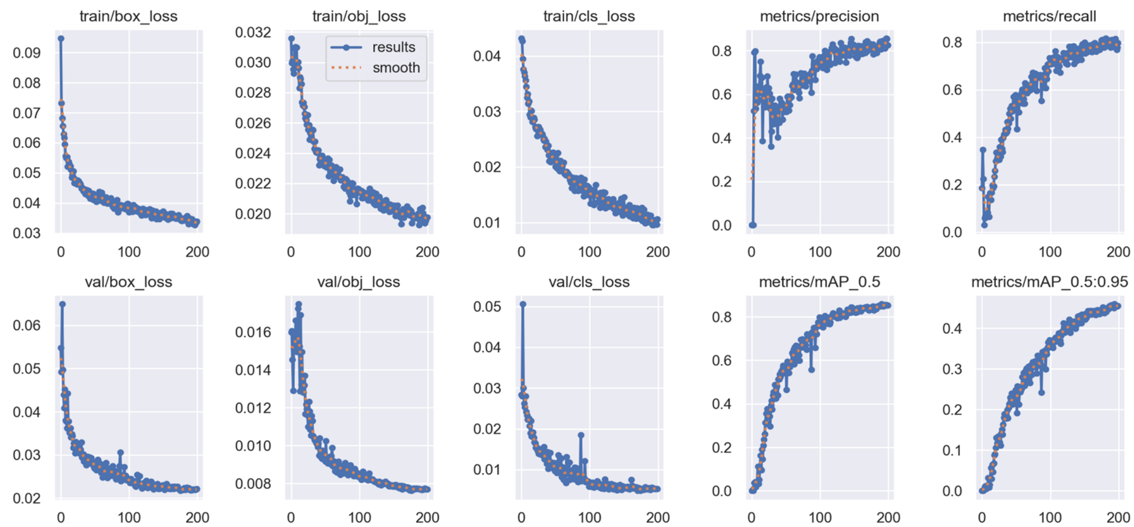


Figure 3: Training Curves of the YOLOv5n model.

Fig. 4 shows the training curves of the YOLOv5n-MobileNetv4 model. The training curve of the YOLOv5n-MobileNetv4 model was similar to that of the YOLOv5m model.

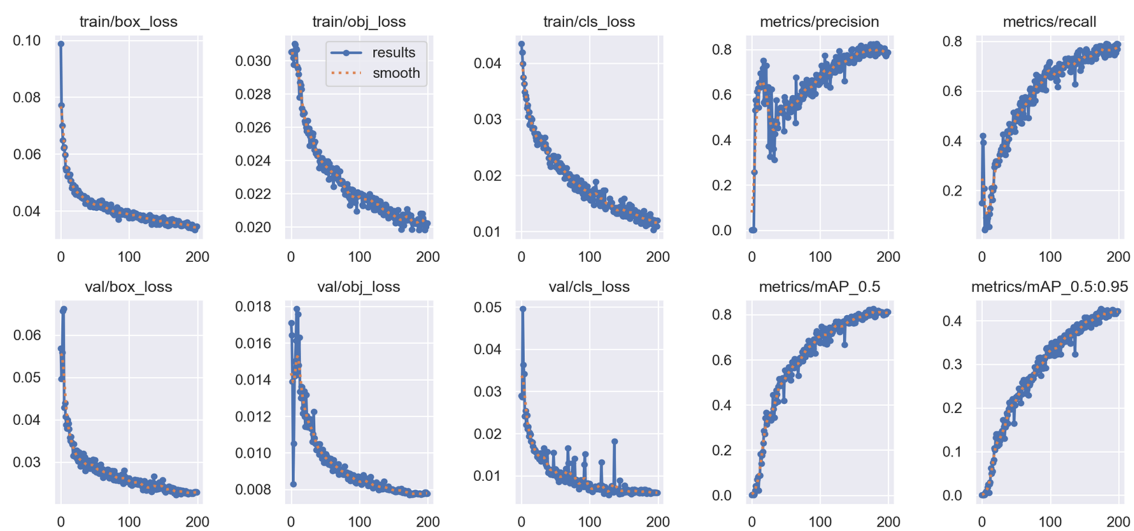


Figure 4: Training Curves of the YOLOv5n-MobileNetv4 model.

4.2. Validation Results

A validation set of 400 images was used to verify the performance of the detection model. The validation results of the YOLOv5n model are shown in Table 2.

Table 2: Validation results of the YOLOv5n model.

Class	Precision (%)	Recall (%)	mAP_0.5 (%)
all	83.5	80.6	85.6
grub	92.4	83.9	91.4
mole cricket	93.8	90.9	96.3
wireworm	75.7	71.6	76.0
black cutworm	72.2	76.1	78.6

The YOLOv5n model achieved 85.6% mAP_0.5, with precision and recall values of 83.5% and 80.6%, respectively. The model performed well for specific pests, such as mole cricket with 96.3% mAP_0.5 and grub with 91.4% mAP_0.5, but showed lower performance for wireworm with 76.0% mAP_0.5 and black cutworm with 78.6% mAP_0.5. The validation results of the YOLOv5n-MobileNetv4 are displayed in Table 3.

Table 3: Validation results of the YOLOv5n-MobileNetv4 model.

Class	Precision (%)	Recall (%)	mAP_0.5 (%)
all	82.3	78.2	82.1
grub	91.5	81.9	90.5
mole cricket	93.1	92.0	96.2
wireworm	68.7	71.6	69.8
black cutworm	75.8	67.2	72.1

The YOLOv5n-MobileNetv4 model achieved a mAP_0.5 of 82.1%, with precision and recall values of 82.3% and 78.2%, respectively. The "mole_cricket" class achieved 96.2% mAP_0.5, and the "grub" achieved 90.5% mAP_0.5. However, the model had a lower performance for wireworm with 69.8% mAP_0.5 and black cutworm with 72.1% mAP_0.5.

4.3. Performance Comparison

Table 4 compares the performance of the YOLOv5n and YOLOv5n-MobileNetv4 models.

Table 4: Performance comparison of two models.

Model	Precision (%)	Recall (%)	mAP_0.5 (%)	Parameters	FPS
YOLOv5n	83.5	80.6	85.6	1,764,577	125
YOLOv5n-MobileNetv4	82.3	78.2	82.1 (-3.5%)	1,126,649 (-36.2%)	163.9 (+31.1%)

The YOLOv5n model achieved a higher precision of 83.5% and recall of 80.6% with a mAP_0.5 of 85.6% but had more parameters of 1,764,577 and a lower FPS of 125. In contrast, the YOLOv5n-MobileNetv4 model had slightly lower performance in terms of precision, recall, and mAP_0.5 of 82.3%, 78.2%, and 82.1%, respectively, but had significantly fewer parameters of 1,126,649 and a higher FPS of 163.9, showing a 31.1% speed improvement.

4.4. Performance Comparison of Other Lightweight Algorithms

Table 5 compares the performance of the three lightweight models.

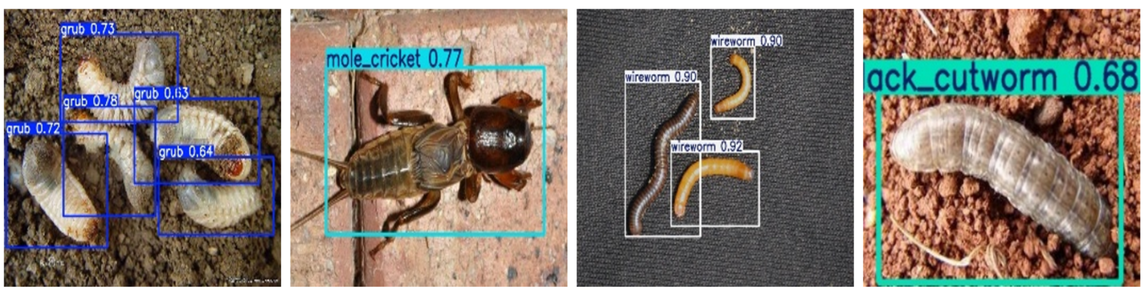
Table 5: Performance comparison of two models.

Model	Precision (%)	Recall (%)	mAP_0.5 (%)	Parameters	FPS
YOLOv5n-MobileNetv4	82.3	78.2	82.1	1,126,649	163.9
YOLOv6n	87.3	84.2	87.6	4,155,420	63.2
YOLOv8n	87.8	87.5	88.4	2,583,712	130.5

The YOLOv8n model achieved the highest accuracy of 88.4% with 2,583,712 parameters, and the YOLOv6n model achieved an accuracy of 87.6% with 4,155,420 parameters. Although the accuracy of the YOLOv8n and YOLOv6n models is higher than that of the YOLO5n-MobileNetv4 model, their parameters are about two and three times more. Therefore, the YOLOv5n-MobileNetv4 model balanced accuracy and speed.

4.5. Detection Results

Fig. 5 shows the detection results of YOLOv5n-MobileNetv4 model.



(a) Grub (b) Mole_cricket (c) Wireworm (d) Black Cutworm

Figure 5: Detection results of the YOLOv5n-MobileNetv4 model.

The YOLOv5n-MobileNetv4 model correctly detected four pests: Grub, Mole Cricket, Wireworm, and Black Cutworm in sample images. The "grub" class achieved a confidence value of average 0.65, the "black_cutworm" class achieved a comparable confidence value of 0.68. The "mole_cricket" class achieved a confidence value of 0.77, while the "wireworm" class achieved a high confidence value of over 0.90.

5. Conclusions

This paper presented a lightweight and efficient crop pest detection algorithm, YOLOv5n-MobileNetv4, designed for real-time applications in resource-constrained environments like embedded devices and IoT systems. By replacing the YOLOv5n backbone with MobileNetv4, the proposed model achieved a significant reduction in parameters. The experimental results demonstrated that proposed model maintained high detection accuracy of 82.1% while significantly improving processing speed, with a 31.1% increase in FPS and a 36.2% reduction in parameters.

Acknowledgments

This research was funded by the Science and Technology Bureau of Leshan, Leshan, China (Grant number 22ZDYJ0154).

References

F. Ali, H. Qayyum, and M. J. Iqbal. Faster-pestnet: a lightweight deep learning framework for crop pest detection and classification. *IEEE Access*, 11:104016–104027, 2023. doi: 10.1109/ACCESS.2023.3317506.

Y. Ding, W. Wang, J. Yu, Y. Wang, and M. Jin. Research on crop pest monitoring system based on ioa technology and yolox-s algorithm. *Internet of Things Technologies*, 13(7):26–30, 2023. doi: 10.16667/j.issn.2095-1302.2023.07.007.

X. Guo and H. Shang. Plant pest identification prototype network vgg-ml based on meta learning. *Journal of Nanjing Agricultural University*, 47(2):392–401, 2024.

X. Han, Y. Xu, R. Feng, T. Liu, J. Bai, and Y. Lan. Early identification of crop diseases based on infrared thermography and improved yolo v5. *Transactions of the Chinese Society of Agricultural Machinery*, 54(12):300–307, 2023.

Q. Jiang. Research on application of image pattern recognition in agricultural disease monitoring and prediction. *Journal of Agricultural Catastrophology*, 13(8):74–76, 2023.

H. Kuzuhara, H. Takimoto, Y. Sato, and A. Kanagawa. Insect pest detection and identification method based on deep learning for realizing a pest control system. In *59th Annual Conference of the Society of Instrument and Control Engineers of Japan*, pages 709–714. IEEE Inc, Chiang Mai, 2020.

Y. Liu, K. Hu, J. Nie, and T. Xie. Rice disease and pest identification based on msdb-resnet. *Journal of South China Agricultural University*, 44(6):978–985, 2023.

X. Xu, L. Jia, H. Li, X. Jia, B. Zhang, J. Zhou, and H. Pu. Application of deep learning to wheat pest classification methods. *Journal of Sichuan Agricultural University*, 41(6):1079–1089, 2023. doi: 10.16036/j.issn.1000-2650.202307279.