

ITC 1/54 Information Technology and Control Vol. 54 / No. 1/ 2025 pp. 164-174 DOI 10.5755/j01.itc.54.1.39160	Improved YOLOv8n Based Lotus Seedpod Detection Algorithm	
	Received 2024/10/18	Accepted after revision 2025/01/25
	HOW TO CITE: Zhang, K., Xu, H., Hu, M., Tang, T., Ye. B. (2025). Improved YOLOv8n Based Lotus Seedpod Detection Algorithm. <i>Information Technology and Control</i> , 54(1), 164-174. https://doi.org/10.5755/j01.itc.54.1.39160	

Improved YOLOv8n Based Lotus Seedpod Detection Algorithm

Kun Zhang, Huimin Xu, Miao Hu, Tao Tang, Bingliang Ye

Faculty of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China

Corresponding author: zist_ybl@zstu.edu.cn

Due to the influence of the lotus seedpod's shape, appearance, color, and growth environment, lotus seedpod detection faces challenges such as low efficiency, low accuracy, and issues with false negatives and false positives. To address these problems, an improved lotus pod detection algorithm, FSM-YOLOv8, is proposed based on YOLOv8n. First, the C2f-Faster module reduces the number of model parameters while ensuring the structural feature extraction capability of the YOLOv8n network. Then, the SimAM attention mechanism is applied to the model feature extraction module, which enhances the multi-scale and spatial feature extraction capability of the model. Finally, MPDIoU is used as the boundary loss function to effectively solve the problem of low detection rate caused by the spatial overlap and occlusion of the lotus seedpods and lotus leaves. The research results indicate that the improved FSM-YOLOv8 achieves detection precision, recall rate, and mAP@0.5 of 85.5%, 84.7%, and 88.7%, respectively, on the lotus seedpod detection dataset. Compared to the YOLOv8n model, this represents improvements of 1.1%, 0.8%, and 0.9%, with a 13.4% reduction in model parameters. The proposed algorithm enables rapid lotus seedpod detection in complex environments, meeting the recognition requirements for lotus seedpod harvesting robots during the picking process.

KEYWORDS: Lotus seedpod; YOLOv8; Detection; FasterNet; SimAM; MPDIoU.

1. Introduction

Lotus seedpod is an aquatic cash crop, rich in nutrients and of high economic value. After thousands of years of careful cultivation and continuous development in China, it has been widely planted in various places, and the production and export of lotus seed and lotus seedpod have ranked among the world's top.

Due to its huge market demand, the economic benefits brought by lotus seedpod planting are very considerable, showing vigorous development potential [20, 24]. The harvesting of lotus seedpods is mainly concentrated from June to October, with production peaking in July and August. In order to obtain the

most nutritious lotus seedpod, it is necessary to pick them when they are in full growth [23]. However, at present, China's lotus seedpod picking is still mainly dependent on artificial completion, according to the Hangzhou West Lake fruit lotus cost statistics, the cost of artificial picking occupies more than 60% of the production cost. The manual picking method is inefficient and difficult to ensure the seasonality and freshness of lotus seedpods, which cannot meet the market demand. Therefore, it is crucial to realise the automation of lotus seedpod harvesting, which can not only solve all the problems of manual harvesting, but also significantly reduce the company's production cost, improve the harvesting efficiency and quality, and further enhance the competitiveness in the lotus seedpod market.

Detection, localization and tracking of lotus seedpods is the main research focus of the vision system of lotus seedpod picking robots. The difficulty in testing the lotus seedpod lies in its unique visual characteristics: the front of the lotus leaf is bright green and the back is slightly lighter in color, while the fresh lotus seedpod has a soft green color until the mature stage when it becomes lighter in color and similar to the color of the lotus leaf. The growing environment of the lotus seedpod is complex and changing. Due to the influence of the wind, the lotus seedpod swings irregularly, making it difficult to distinguish between the lotus leaf and the lotus seedpod. In particular, the corner of the lotus leaf is also similar in shape to the lotus seedpod when folded, and these factors combine to make detecting the height of the lotus seedpod a major challenge (Figure 1).

Target detection and tracking has opened up endless possibilities for modern agriculture, especially in target detection to achieve in-depth and extensive application, significantly making up for the shortcomings of traditional detection methods. Most of these cutting-edge technologies are built on the powerful cornerstone of Convolutional Neural Networks (CNNs) and can be clearly divided into two camps based on their different processing strategies: two-stage detectors, which are oriented to high-precision and nuanced target recognition; The other category, one-stage detectors, focuses on efficiency and getting the job done faster. In agricultural applications, these two types of detectors have each demonstrated outstanding performance and benefits.

Figure 1

The lotus seedpod and its growth environment



Two-stage methods, such as the R-CNN family [4, 5, 6, 19] and SPPNet [7], first locate the region of interest in the image and then detect objects from that region. Single-stage methods such as SSD [22], YOLO [16, 17, 18, 1, 10, 12, 21], RetinaNet [14] and CenterNet [25] directly predict the target location and category, eliminating the region proposal step, which has greatly improved the detection efficiency, although the detection accuracy is slightly reduced. Jawaharlal Nehru et al. [11] improved the YOLO algorithm by improving the clustering algorithm, large data set weighting, multi-scale training, and optimising non-maximisation to improve the model efficiency and verify its feasibility in target detection. However, the data body is single and the practicality is not strong. Li et al. [13] introduced GSCov module and V-GSCSP module to improve the accuracy and speed as well as the computational requirement. Yang et al. [9] enhanced the receptive field of YOLOv8 by introducing an improved version of the DBlock convolution and DWR module. By combining a multi-branch CA attention mechanism and shared convolution, they optimized feature extraction and representation, significantly reducing the model parameters and improving the accuracy of object detection.

As deep learning algorithms continue to improve, fruit and vegetable detection has been fruitful, but environmental factors in the field are much more complex than in laboratory conditions. In a lotus seedpod

planting environment, the growth position of the lotus leaves exhibits a range of levels, and the growth position of the lotus seedpod displays considerable variation in height. This results in a substantial amount of shade and overlap of the lotus seedpod, coupled with the ever-changing intensity of the light. Collectively, these factors present a significant challenge to the accuracy of identification and localization of the lotus seedpod. In addition to meeting the requisite specifications for mobile deployment, the Lotus seedpod picking device must also account for the real-time detection accuracy and speed. In this paper, the FSM-YOLOv8 network modelling algorithm is designed to address these challenges. The algorithm mitigates the issues of high computational complexity inherent to the model, as well as the considerable number of parameters, and offers a swift and precise lotus seedpod recognition model for lotus seedpod picking robots with constrained computational resources. This enables them to adeptly navigate the multifarious challenges encountered in the actual picking environment.

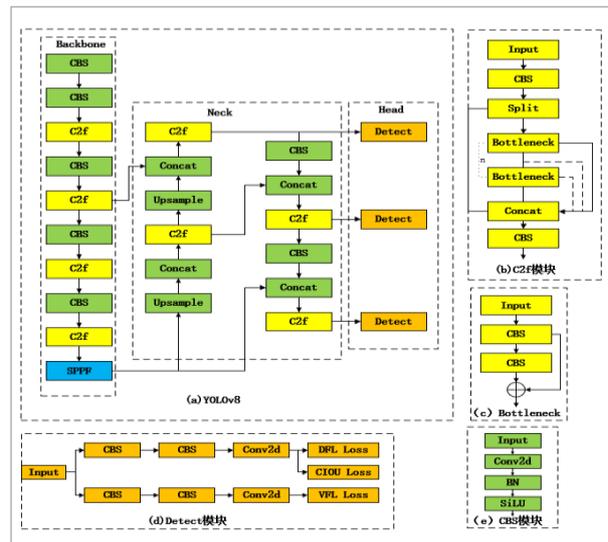
2. YOLOv8 Network Modeling and Improvements

2.1. YOLOv8n Network Model

The YOLO algorithm exhibits superior detection speed and more suitable deployment conditions than the two-stage target detection algorithm. YOLOv8 separates the classification and localization branches without sharing parameters and modifies the detection head of the two tasks to a decoupled head structure, thereby alleviating the intrinsic conflict between feature data. The implementation of the anchorless frame concept in label assignment and loss function optimization facilitates enhanced consistency between classification and regression. The integration of DFL Loss and CIou Loss further enhances the model's performance. The comprehensive structure of YOLOv8 is illustrated in Figure 2.

The YOLOv8n network model has undergone structural fine-tuning. The backbone of the network model employs the C2f module, derived from the CSPNet concept, which incorporates the ELAN design concept of YOLOv7 [21] and the C3 module of YOLOv5. This results in an efficient C2f module. The model

Figure 2
YOLOv8n network architecture



concludes with the introduction of the SPPF module, which captures and fuses feature information at varying scales through the application of maxpool operations in parallel. The neck network plays a top-down role in the model, integrating the extracted features and later passing them to the recognition head. The model employs a cutting-edge feature fusion method, namely PAN-FPN, which effectively integrates the strengths of path aggregation networks and feature pyramid networks. The PAN model is capable of incorporating high-resolution feature details on low-resolution feature maps, thereby enhancing the network's perceptual capabilities. In contrast, FPN is capable of achieving effective feature fusion at varying scales through the use of top-down and bottom-up pathways. The combination of the two enables the PAN-FPN to achieve multi-scale perception at multiple layers and feature fusion over a wider range of scales, Improved detection capabilities. With regard to the output, YOLOv8n makes use of YOLOX's head separation technique to decouple the classification and detection processes [3]. This technique mainly involves loss calculation and target detection frame screening, which directly improves the accuracy and efficiency of centroid and bounding box detection.

This paper addresses the lotus seedpod harvesting requirements by optimizing the YOLOv8n algorithm

and proposing the FSM-YOLOv8 model. By introducing Fasterblock to enhance speed and integrating SimAM to construct the C2f-FS backbone module, the model strengthens feature extraction and localization capabilities. Additionally, the MPDIoU loss function is adopted to improve small object detection accuracy while significantly addressing occlusion and overlap issues.

2.2. Improved Modeling of YOLOv8n Networks

Since the C2f module in the YOLOv8n model suffers from parameter redundancy, large model size, and unsatisfactory performance on small target detection tasks, these factors together limit its effective use on mobile platforms. To overcome this limitation, we innovatively introduce the FSM-YOLOv8 network model, which is clearly shown in Figure 3. The core design goal of this model is to realize the dual leaps of model lightweighting and small target detection accuracy improvement, so as to perfectly meet the deployment requirements of mobile devices. By carefully tuning the model architecture and parameter configurations, FSM-YOLOv8 not only maintains its efficient detection capability, but also drastically reduces the complexity and size of the model, ensuring

its smooth operation in resource-constrained mobile environments and demonstrating excellent performance and adaptability.

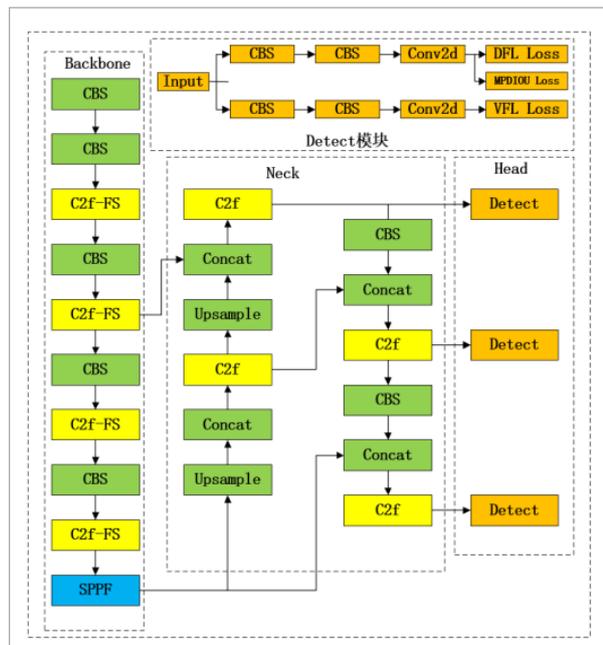
In Figure 3, the structure of the FSM-YOLOv8 network model highlights a comprehensive optimization strategy. Through the optimization of both the architecture and the loss function, FSM-YOLOv8 significantly reduces computational costs while enhancing the ability to capture fine-grained features, thereby improving the model's detection accuracy and stability in complex field environments. Tailored to the practical needs of lotus seedpod detection, the model demonstrates exceptional robustness and adaptability, offering an efficient and reliable technical solution for field crop target detection.

2.2.1. C2f-Faster Module

To enhance the real-time monitoring capability of the lotus seedpod harvesting equipment, we have incorporated the efficient FasterNet network architecture. FasterNet [2] is a lightweight neural network designed to improve the efficiency of target detection, and its unique feature is the innovative introduction of a new convolutional module called Partial Convolution, as clearly shown in Figure 4. The overall structure of FasterNet is well-designed and divided into four hierarchical levels, each with a built-in Fasterblock module, and no additional embedding or merging layers are added between these levels to maintain architectural simplicity. The main function of the end of the network is feature transformation and classification, which is responsible for transforming the extracted high-level features into effective representations that can be used for classification and recognition, so as to successfully complete the task of real-time detection of lotus seed pods. The PConv layer is immediately followed by two 1×1 convolutional layers, which can fully exploit the feature information of each channel, and such a layout strategy allows the network to automatically focus on the central key areas of the input feature map and explore them in depth during the feature extraction process, thus capturing more feature information. central key areas during the feature extraction process, thus capturing richer and more accurate feature representations.

One of the highlights of this structure is its excellent ability to deal with complex scenes (e.g. occluded regions in a lotus seedpod) while ensuring that the

Figure 3
Network structure diagram of FSM-YOLOv8



clarity and integrity of feature information is maintained. However, in the pursuit of building high-performance neural networks, while normalization layers and activation layers are indispensable elements, normalization layers are highly sensitive to the distribution of input data. They automatically adjust the activation of output data, reducing the model's dependence on initial weights, ensuring the stability of mean and variance, and effectively alleviating the internal covariate shift phenomenon during training, thereby accelerating the convergence speed. Although they can stabilize the training process, excessive application may inadvertently limit the diversity of target feature expressions, thereby weakening the model's detection accuracy, and introducing unnecessary computational overhead, leading to a decrease in response speed. Based on this insight, FasterNet adopts a cautious but efficient strategy in its architectural design: the BN layer corrects the bias of the feature values in the right direction, which improves the speed of model inference. The non-linear property introduced by the activation function is a richer feature.

The redesigned C2f-Faster module is illustrated in Figure 5. It features a significant structural advancement, whereby all bottleneck units within the conventional C2f module have been replaced with FasterBlock units. First of all, PConv (Partial Convolution) reduces the amount of computation and memory access by performing convolution operations only on a part of the input feature map. When the input data passes through the PConv layer, redundant feature information can be removed, which reduces the demand for computing power while increasing the computing speed. Subsequently, the feature map that has undergone preliminary feature extraction will flow through two 1×1 convolution layers in sequence. These two convolution layers are responsible for further refining the features, eliminating irrelevant information, and retaining only the effective features that are crucial for subsequent processing. Ultimately, through this series of efficient operations, the C2f-Faster module not only significantly reduces the number of 3×3 standard convolutions in the original C2f module, but also ensures that the model is able to fully and deeply exploit the information in each channel, while maintaining the diversity and richness of the feature representation (Figure 6).

Figure 4
PConv structure

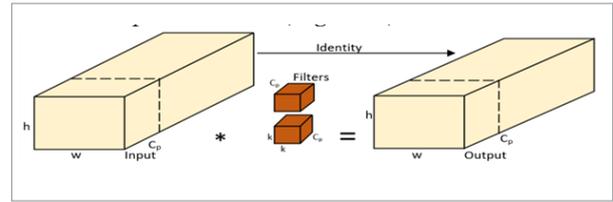


Figure 5
Faster Net Structure

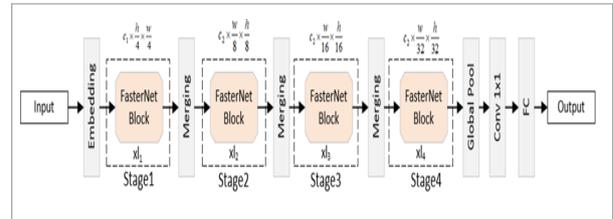
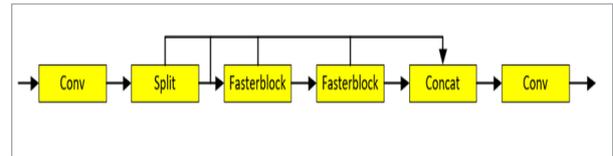


Figure 6
C2f-Faster structure



2.2.2. SimAM Attention Mechanism Module

SimAM [8] can infer three-dimensional attention weights for feature maps, taking into account the information of both channel dimension and spatial dimension. Compared with the traditional methods that only consider channel attention or spatial attention, it can capture the relationships between features more comprehensively, focus better on the key object features in the image, help the model learn more discriminative features, and thus improve the performance of the model in various visual tasks. When processing the feature map, the SimAM attention mechanism employs a minimum energy e_i^* based evaluation method. It is used to accurately measure and highlight the importance of each target neuron by analysing the similarity between the neuron and its surrounding features, which can be applied to the model to accurately extract the feature information of the detection frame and the target. The lower the minimum energy, the greater the difference

between the neurons, and the more important the neurons are in the overall extraction of information, and therefore can also be indicated by the importance of the neurons, wherein e_i^* is defined as follows:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda}. \quad (1)$$

In Equation (1), where λ is the canonical term; t for a channel-targeted neuron, $\hat{\mu}$ is the average of all input characteristics on a channel, and is the variance of all input features on a channel; the expressions for $\hat{\sigma}^2$ and $\hat{\mu}$ are as follows:

$$\hat{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \hat{\mu})^2 \quad (2)$$

$$\hat{\mu} = \frac{1}{M} \sum_{i=1}^M x_i. \quad (3)$$

In Equations (2)-(3), M All neurons; x_i : neurons on specific channels; X' : Extended feature information for feature enhancement.

$$X' = \text{sigmoid}\left(\frac{1}{E}\right) \Theta X. \quad (4)$$

In Equation (4), multiplication by elements and E denotes the set of all minimum energies e_i^* . The sigmoid function limits the E -value between 0 and 1 to avoid vanishing gradients.

The SimAM attention mechanism is introduced at the backend, seamlessly integrated with FasterBlock to create the innovative FSBlock. This is then incorporated into the C2f module, resulting in the upgraded C2f-FSM (as shown in Figures 7-8).

Figure 7
Sblock structure

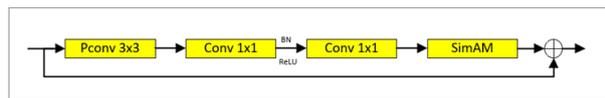
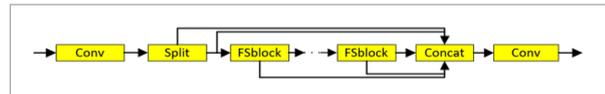


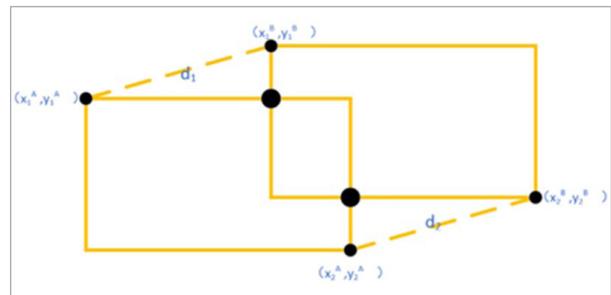
Figure 8
C2f-FS network structure



2.2.3. MPDIoU Improvements

In the original YOLOv8 loss function, the overlap area between target bounding boxes determines their similarity. This approach can lead to lower IoU values when handling partially occluded targets, which negatively affects detection accuracy. In contrast, MPDIoU [15] addresses occlusion by focusing on the minimum distance between bounding boxes, enabling more accurate target localization and improving positioning precision. Furthermore, MPDIoU refines the spatial distribution information of targets by considering the geometric relationships between bounding boxes, such as the distance between their centers and differences in aspect ratio. When targets are partially occluded, the original IoU method struggles to handle overlapping regions, resulting in decreased detection accuracy. However, by incorporating geometric information, MPDIoU can more accurately estimate the boundaries of occluded targets, thereby reducing the impact of occlusion on localization. In scenarios involving small target occlusion, the IoU calculation in the original loss function may not be sensitive enough to the occluded regions, leading to significant localization errors. In contrast, MPDIoU not only considers the overlapping area but also incorporates the geometric features of the bounding boxes, allowing the model to maintain high localization accuracy and robustness, even in cases of small targets or severe occlusion. When dealing with occluded regions, the original YOLOv8 loss function suffers from reduced IoU values, which can cause regression targets to shift and impact detection results. However, MPDIoU, through its minimum point distance metric, effectively mitigates the negative impact of occlusion on the regression process, ensuring more accurate regression and ultimately improving detection stability and reliability in occluded environments. The structure is shown in Figure 9.

Figure 9
MPDIoU



The precise formula is as follows:

$$d_1^2 = (x_1^B - x_1^A)^2 + (y_1^B - y_1^A)^2 \quad (5)$$

$$d_2^2 = (x_2^B - x_2^A)^2 + (y_2^B - y_2^A)^2 \quad (6)$$

$$MPDIOU = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \quad (7)$$

where h is the width and height of the input image; (x_1^A, y_1^A) and (x_2^A, y_2^A) are the coordinates of the upper left and lower right points of the real box; (x_1^B, y_1^B) and (x_2^B, y_2^B) are the coordinates of the upper left and lower right points of the prediction box; d_1^2 and d_2^2 are the square of the Euclidean distance of the true frame and the Euclidean distance of the predicted frame.

2.3. Experimental Environment and Dataset

The lotus seedpod image dataset was sourced from the lotus pond cultivation base in the He Tang planting area, located in Qianmu, Linping District, Hangzhou, Zhejiang Province, China. Image collection was conducted between July and October 2023, yielding a total of 3,000 images. Natural scenes are often characterized by intricate and diverse conditions, such as intense direct sunlight, soft diffused lighting, various shooting angles (including top-down, upward, and side views), and differing shooting distances.

To enhance the model's generalization ability, reduce the risk of overfitting, and ensure diversity in the dataset, the original images were subjected to brightness adjustments, noise addition, mirror flipping,

translation, and rotation. This augmented the dataset to 6,000 images. These augmentation techniques generate new images with different characteristics while preserving the original semantic content. The training, validation, and test set proportions were set at 8:1:1. (Figure 10).

Figure 10

Lotus seedpod data set image



2.4. Evaluation Metrics

To comprehensively evaluate the performance of the detection model, we employed widely recognized metrics such as model parameters, precision, recall, and mean average precision (mAP). Model parameters reflect the complexity of the model, serving as a direct indicator of its structure. Precision is a standard value used to assess the model's reliability and usability, representing the proportion of true positive detections among all positive detections. Recall, on the other hand, refers to the proportion of actual targets that are correctly detected by the model. These metrics provide a well-rounded assessment of the model's effectiveness and efficiency.

2.5. Model Selection and Field Experiments

YOLOv8n optimizes feature fusion and expression capabilities while maintaining its lightweight design, making it more adaptable in complex scenarios compared to YOLOv5n. The inference speed and

Table 1

Experimental environment and parameter settings

Project	Version environment	Parameter name	Parameter setting
GPU	NVIDIA RTX 3080	Batch Size	8
CUDA	Cuda12.1	Image Size	640
Python	Python3.9	lr0	0.01
PyTorch	Pytorch2.0.1	Optimizer	SGD
Operating system	Windows10	Epoch	300

computational efficiency have been significantly improved, demonstrating exceptional performance, particularly in tasks requiring high real-time processing. Additionally, YOLOv8n enhances deployment flexibility in embedded systems and edge computing environments, effectively reducing hardware resource consumption while improving computational stability.

This study developed a lotus seedpod harvesting experimental system at Zhejiang Sci-Tech University, and the performance of the detection model was tested (experimental scene shown in Figure 11). The system utilizes a three-coordinate structure, consisting mainly of a PC, STM-32 controller, Intel RealSense L515 depth camera, and an end effector. The L515 depth camera is mounted along the X-axis to capture lotus seedpod images from a top-down view, enabling visual detection in complex scenarios. The Jetson Xavier NX serves as the main controller, responsible for detecting harvestable lotus seedpods and calculating their 3D coordinates. The STM-32 controller works in coordination with the three-coordinate platform to control the end effector, executing the harvesting task.

Figure 11

Lotus Seedpod Harvesting Experimental Equipment



3. Analysis of Results

3.1. Ablation Experiment

To test the effectiveness of the improved detection algorithm, five sets of experiments are conducted,

with Experiment 1 being the base experiment YOLOv8n, the original model. Experiment 2 Replacing Bottleneck in C2f with Fasterblock on the original YOLOv8n model. Experiment 3 Adding the SimAM attention mechanism to the end of the Fasterblock module. Experiment 4 introduces the use of MPDIoU as a replacement for CIoU Loss. Experiment 5 models the FSM-YOLOv8 algorithm, which has been designed within the context of this paper.

Table 2

Results of ablation experiments

Model	P _{recision} /%	R _{ecall} /%	mAP@0.5/%	P _{arameters} /10 ⁶
YOLOv8n	84.4	83.9	87.8	3.01
+C2f-FasterNet	83.8	83.6	87.4	2.61
+C2f-FS	84.9	84.1	87.9	2.61
+MPDIoU	84.6	84.3	87.8	3.01
Ours method	85.5	84.7	88.7	2.61

As shown in Table 2, the C2f-FasterNet module optimizes inference efficiency and computation speed, but in target detection tasks with complex backgrounds, accuracy and recall decrease due to information loss, reduced feature extraction capability, and the trade-off introduced by the lightweight design. Therefore, by integrating SimAM after the C2f-Faster module, the model can more effectively focus on important features of the target by optimizing the attention distribution of feature channels. This enhancement improves the model's ability to detect small targets, complex scenes, and occluded targets, resulting in an improvement in both accuracy and recall. By replacing the conventional loss function with the MPDIoU loss function, which is more suitable for small targets, and training the model for 300 iterations, FSM-YOLOv8 achieves improvements of 1.1%, 0.8%, and 0.9% in Precision, Recall, and mAP@0.5, respectively, on the lotus seedpod detection dataset. Furthermore, FSM-YOLOv8 reduces model size while improving detection accuracy.

Table 3

Comparative tests

Model	P _{recision} /%	R _{ecall} /%	mAP@0.5/%	P _{arameters} /10 ⁶
YOLOX-tiny	79.4	80.6	83.1	5.1
YOLOv5n	73.2	69.5	70.7	1.77
YOLOv5s	84.7	84.4	84.5	7.03
YOLOv6	82.9	81.2	83.7	4.29
YOLOv8n	84.4	83.9	87.8	3.01
Ours method	85.5	84.7	88.7	2.61

3.2. Comparative Experiment

In the same experimental setting, the FSM-YOLOv8 lotus seedpod detection algorithm was compared with detection models such as YOLOX-tiny, YOLOv5n, YOLOv5s, YOLOv6, and YOLOv8n to further validate the model's superiority. The results of the comparison experiment are presented in Table 3.

As shown in the table above, for lotus seedpod detection, the optimized FSM-YOLOv8 algorithm achieves a precision of 85.5%, a recall of 84.7%, and a mean average precision (mAP@0.5) of 88.7%. This performance significantly surpasses that of mainstream object detection algorithms, including YOLOX-tiny, YOLOv5n, YOLOv6, YOLOv7-tiny, and YOLOv8n. Compared to YOLOX-tiny, YOLOv5n, YOLOv6, and YOLOv8n models, FSM-YOLOv8 requires only 2.61MB of parameters. This not only improves training speed but also enhances overall performance.

Figure 12 compares the Recall, Precision, and mAP@0.5 curves between YOLOv8n and FSM-YOLOv8. As can be seen from the figure, FSM-YOLOv8 outperforms YOLOv8n in terms of Recall, Precision, and mAP metrics. Figure 13 presents the detection results of FSM-YOLOv8 applied to the original images of the lotus seedpod dataset. It is observable from Figure 13 that in rural settings, FSM-YOLOv8 achieves higher detection scores for small-sized targets, thereby mitigating the issue of missed detections to a certain extent.

Figure 12

Comparison of YOLOv8n and FSM-YOLOv8 curves

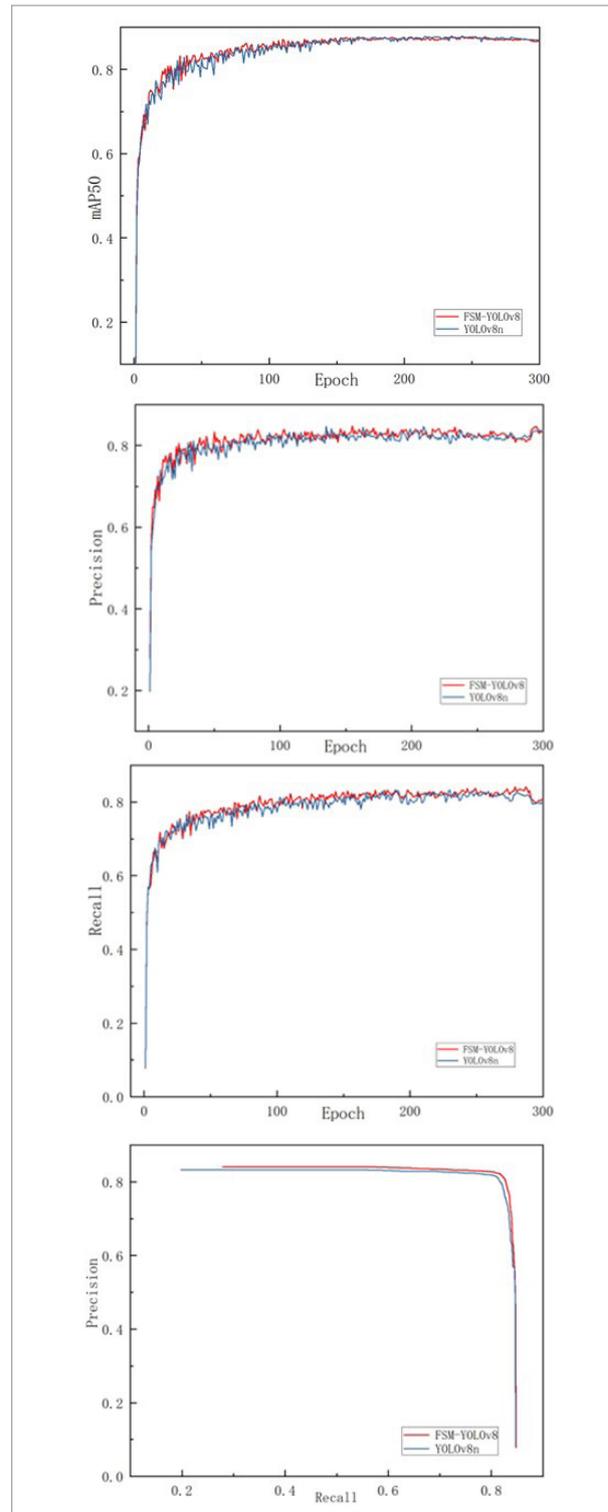


Figure 13

FSM-YOLOv8 detection effect



5. Conclusions

This study proposes the FSM-YOLOv8 optimization algorithm for lotus seedpod detection tasks.

References

1. Bochkovskiy, A., Wang, C. Y., Liao, H. Y. M. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv e-prints, 2020. DOI: 10.48550/arXiv.2004.10934.
2. Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C.-H. Run, Don't Walk: Chasing Higher FLOPs for Faster Neural Networks // 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023. <https://doi.org/10.1109/CVPR52729.2023.01157>
3. Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J. YOLOX: Exceeding YOLO Series in 2021. 2021. DOI: 10.48550/arXiv.2107.08430.

The model achieves deeper feature exploration in processing multi-channel information, enriching the diversity of feature extraction while improving efficiency and effectively reducing computational burden. With the introduction of a parameter-free attention mechanism, the model can accurately capture key features of the target bounding box. Through local self-similarity calculation, it effectively suppresses interfering features and concentrates computational resources on extracting critical information. To further optimize detection performance, the replacement of the loss function simplifies the computation process while comprehensively improving regression accuracy and detection precision. Compared to existing methods, FSM-YOLOv8 performs better in reducing false positives and false negatives, with relatively lower hardware resource demands. This approach holds potential for intelligent lotus seedpod harvesting tasks and provides a valuable reference for optimizing agricultural object detection technologies.

Data Sharing Agreement

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

Funding

The author(s) received no financial support for the research.

4. Girshick, R. Fast R-CNN // Proceedings of ICCV, 2015. <https://doi.org/10.1109/ICCV.2015.169>
5. Girshick, R., Donahue, J., Darrell, T., Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. IEEE Computer Society, 2014. <https://doi.org/10.1109/CVPR.2014.81>
6. He, K., Gkioxari, G., Piotr, D., Girshick, R. Mask R-CNN. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017. <https://doi.org/10.1109/TPAMI.2018.2844175>
7. He, K., Zhang, X., Ren, S., Sun, J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9), 1904-1916. https://doi.org/10.1007/978-3-319-10578-9_23
8. Yang, L., Zhang, R. Y., Li, L., Xie, X. SimAM: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. PMLR, 2021.
9. Yang, P., Shi, T., Yuan, Y., Jiang, H. Fish Catch Sorting and Detection Model Improved Based on YOLOv8 Model. Information Technology and Control, 2024, 53(4), 1291-1310. <https://doi.org/10.5755/j01.itc.53.4.38761>
10. YOLOv5. Available online: <https://github.com/ultralytics/yolov5> (accessed on 12 April 2021).
11. Jawaharlalnehru, A., Sambandham, T., Sekar, V., Ravikumar, D., Loganathan, V., Kannadasan, R., Khanm A. A., Wechtaisong, C., Haq, M. A., Alhussen, A., Alzamil, Z. S. Target Object Detection from Unmanned Aerial Vehicle (UAV) Images Based on Improved YOLO Algorithm. Electronics (2079-9292), 2022, 11(15). <https://doi.org/10.3390/electronics11152343>
12. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., Wei, X. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. arXiv preprint arXiv:2209.02976, 2022. DOI: 10.48550/arXiv.2209.02976.
13. Li, H., Li, J., Wei, H., Liu, Z., Zhan, Z., Ren, Q. Slim-Neck by GSConv: A Lightweight-Design for Real-Time Detector Architectures. Journal of Real-Time Image Processing, 2024, 21(3). <https://doi.org/10.1007/s11554-024-01436-6>
14. Lin, T. Y., Goyal, P., Girshick, R., He, K., Dollar, P. Focal Loss for Dense Object Detection. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99), 2999-3007. <https://doi.org/10.1109/TPAMI.2018.2858826>
15. Ma, S., Xu, Y. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression. 2023. DOI: 10.48550/arXiv.2307.07662.
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection // Computer Vision & Pattern Recognition. IEEE, 2016. <https://doi.org/10.1109/CVPR.2016.91>
17. Redmon, J., Farhadi, A. YOLO9000: Better, Faster, Stronger. IEEE, 2017: 6517-6525. <https://doi.org/10.1109/CVPR.2017.690>
18. Redmon, J., Farhadi, A. YOLOv3: An Incremental Improvement. arXiv e-prints, 2018. DOI: 10.48550/arXiv.1804.02767.
19. Ren, S., He, K., Girshick, R., Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6), 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
20. Tang, T., Wang, X., Ma, Z., Hong, W., Yu, G., Ye, B. Lightweight Detection Method for Lotus Seedpod in Natural Environment. International Journal of Agricultural and Biological Engineering, 2024, 16(6), 197-206. <https://doi.org/10.25165/ijabe.20231606.8281>
21. Wang, C. Y., Bochkovskiy, A., Liao, H. Y. M. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. arXiv e-prints, 2022. <https://doi.org/10.1109/CVPR52729.2023.00721>
22. Wei, L., Dragomir, A., Dumitru, E., et al. SSD: Single Shot MultiBox Detector. Springer, Cham, 2016. https://doi.org/10.1007/978-3-319-46448-0_2
23. Zhang, Y., Zeng, H., Wang, Y., Zeng, S., Zheng, B. Structural Characteristics and Crystalline Properties of Lotus Seed Resistant Starch and Its Prebiotic Effects. Food Chemistry, 2014, 155(Jul. 15), 311-318. <https://doi.org/10.1021/cr500072j>
24. Zhu, F. Structures, Properties, and Applications of Lotus Starches. Food Hydrocolloids, 2017, 63(Feb.), 332-348. <https://doi.org/10.1016/j.foodhyd.2016.08.034>
25. Zhu, X., Wang, Q., Zhang, B., Sun, Z. An Improved Feature Enhancement CenterNet Model for Small Object Defect Detection on Metal Surfaces. Advanced Theory & Simulations, 2024, 7(8). <https://doi.org/10.1002/adts.202301230>

