A LIGHTWEIGHT WEED IDENTIFICATION METHOD FOR CHINESE CABBAGE FIELDS BASED ON THE IMPROVED YOLOv11

基于改进 YOLOv11 的轻量化白菜地杂草识别方法

Shunshun JI[1)], Jiajun SUN[1)], Chao ZHANG[1,*)]
[1)] College of Information Science and Engineering, Shandong Agricultural University, TaiAn, Shandong/China
Tel: +86-15621550101; E-mail: zhangch@sdau.edu.cn
DOI: https://doi.org/10.35633/inmateh-76-73

Keywords: *YOLO-LMSW, Object detection, Weed recognition, Lightweight*

ABSTRACT

Accurate and real-time weed identification was a key technology for ensuring the efficient operation of field weeding robots. It played a vital role in reducing pesticide usage, minimizing environmental pollution, and protecting the agricultural ecosystem. In response to these demands, this paper proposed a lightweight and high-precision weed recognition method, YOLO-LMSW, tailored for Chinese cabbage field scenarios. Firstly, a lightweight multi-scale convolutional module (LMSConv) was designed, upon which the LMSC3k2 structure was constructed to reduce computational complexity and enhance multi-scale feature extraction capabilities. In terms of network architecture, a lightweight backbone network, LMSNet, was built to significantly reduce model parameters while maintaining detection accuracy. Additionally, a detection head, LMSHead, was designed to further optimize the model structure. To improve localization accuracy and convergence speed, the Wise-IoU (WIoU) loss function was introduced. Experimental results demonstrated that, compared to YOLOv11n, YOLO-LMSW achieved improvements of 1.2%, 1.0%, and 0.6% in precision, recall, and mAP50 respectively, while reducing the Params and FLOPs by 34.6% and 36.5%, respectively. Furthermore, the model was deployed on a Jetson Orin Nano (4GB) embedded device, where it achieved an average inference time of 117 ms per image. These results demonstrated its application potential in the actual deployment of field weeding robots.

摘要

精准且实时的杂草识别是保障田间除草机器人高效运行的关键技术，对于减少农药施用、降低环境污染以及保护农业生态环境具有重要意义。针对上述需求，本文提出了一种适用于白菜地场景的轻量化高精度杂草识别方法 YOLO-LMSW。本方法首先设计了一种轻量化多尺度卷积模块（LMSConv），并在此基础上构建了 LMSC3k2，用以降低计算复杂度并增强多尺度特征提取能力。在网络架构方面，构建了轻量化主干网络 LMSNet，以在保持检测精度的同时显著减少模型参数量。此外，设计了检测头 LMSHead，进一步优化模型结构。为提升定位精度与收敛速度，引入了 Wise-IoU（WIoU）作为损失函数。试验结果表明，相较于 YOLOv11n，YOLO-LMSW 在精确率、召回率与 mAP50 上分别提升了 1.2%、1.0%和 0.6%，参数量与每秒浮点运算数分别减少了 34.6%和 36.5%。此外，该模型被部署于 Jetson Orin Nano（4GB）嵌入式设备上，单张图像的平均推理时间为 117 毫秒。这些结果展示了其在田间除草机器人实际部署中的应用潜力。

INTRODUCTION

Chinese cabbage was an essential component of the daily diet and held a significant position in the agricultural industry *(Hasan et al., 2023)*. However, the presence of weeds during its growth period severely affected both yield and quality *(Khan et al., 2021; Xiang et al., 2024)*. Weeds demonstrated high resilience and adaptability to agricultural environments. Their root systems were often more developed than those of crops, allowing them to compete more effectively for nutrients *(Darbyshire et al., 2024)*. Moreover, weeds reproduced rapidly, and without timely intervention, could cause substantial yield losses. Manual weeding was labor-intensive and inefficient, while chemical herbicides introduced harmful residues, making them unsuitable for the production of organic vegetables. Furthermore, chemical usage contributed to environmental pollution. As a result, achieving green, efficient, and intelligent weeding became a critical need *(Praveenraj et al., 2024; Zhang et al., 2025)*. Computer vision-based automatic weed recognition provided accurate spatial information about weed distribution, enabling more precise and automated weeding operations.

At that time, researchers both domestically and internationally conducted extensive studies on weed recognition, encompassing traditional and deep learning-based approaches *(Chen et al., 2022)*. Traditional methods typically relied on color, shape, and texture features, which were extracted using conventional image processing techniques. These features were then fed into classical classifiers such as support vector machines (SVMs), achieving a certain degree of success *(Hua et al., 2022; Sabzi et al., 2020)*. However, such methods often suffered from poor robustness under complex field conditions. With the advancement of artificial intelligence, deep learning techniques were increasingly applied in the field of weed recognition *(Zhang et al., 2024)*. *Shang et al., (2022),* improved the Faster R-CNN algorithm by incorporating an enhanced double-threshold non-maximum suppression algorithm and employing transfer learning. The improved model achieved a precision of 96.58%, a recall of 94.82%, and an F1 score of 95.06%. *Cai et al., (2023),* proposed a semi-supervised semantic segmentation network for weed recognition in pineapple fields and evaluated the model's effectiveness by placing the attention mechanism at different network layers.

The methods proposed by the aforementioned scholars demonstrated promising recognition performance; however, many of their models were overly complex, making them unsuitable for real-time operation on weeding robots. As a result, single-stage object detectors such as the YOLO series gained popularity due to their high inference speed and accuracy *(Yan et al., 2025)*. *Ji et al., (2024),* conducted lightweighting research on the YOLOv5 algorithm. In their study, the lightweight network PP-LCNet was employed to redesign the feature extraction backbone, thereby reducing the model's parameter count. They introduced the Ghost convolution module to lightweight the feature fusion network and incorporated a standardized attention module at the end of the fusion stage. As a result, the memory usage of the weed recognition model was reduced to 6.23 MB. *Shao et al., (2023),* proposed GTCBS-YOLOv5s, a lightweight model for weed recognition in rice fields. This model achieved both reduced complexity and improved accuracy, attaining an average precision of 91.1% and a recognition speed of 85.7 FPS. *Hu et al., (2024),* designed an efficient deep learning model for lettuce field weed detection and severity classification. They reduced the size of the YOLOv7 network using a scaling factor ($\tau = 0.5$), and further enhanced the model by integrating ECA and CA attention mechanisms. They also introduced ELAN-B3 and DownC modules to construct the Multimodule-YOLOv7-L framework, which was applied in a real-time mechanical–laser collaborative weeding robot.

Despite these advancements, three major challenges remained unresolved: (1) the high computational complexity of deep networks, (2) low detection accuracy for small or occluded weed targets, and (3) limited feasibility of real-time deployment on low-power embedded devices. To address these issues, this study proposed a lightweight weed detection algorithm for Chinese cabbage fields, named YOLO-LMSW, based on YOLOv11n. The proposed method introduced a novel lightweight backbone network (LMSNet) and a simplified detection head (LMSHead), which together significantly reduced computational cost while preserving high detection accuracy. In addition, the Wise-IoU (WIoU) loss function was employed to enhance localization performance and accelerate convergence. The effectiveness of each module was validated through ablation and comparison experiments. Furthermore, the model was deployed on a Jetson Orin Nano embedded device, where it achieved an average inference time of 117 ms per image, confirming its potential for real-time, field-level application in intelligent weeding systems.

MATERIALS AND METHODS

Data Collection and Preprocessing

The dataset used in this study was collected from Shandong Xiangniao Ecological Farm, located in Qihe County, Dezhou City, Shandong Province (116°46′E, 36°39′N), as shown in Figure 1(a). The images were captured using a Redmi K60 smartphone during the months of September and October. A total of 850 images were collected, covering Chinese cabbage and several representative weed species, mainly including Portulaca oleracea, Eleusine indica, Cirsium setosum, Chenopodium album, and Amaranthus retroflexus. The images were captured under diverse lighting conditions, including strong sunlight and shadows. The images were annotated using the open-source tool Labelme, with labels divided into two categories: crop and weed. Representative sample images were presented in Figure 1(b). The dataset was subsequently divided into training, validation, and test sets in a 7:2:1 ratio. Considering the limited number of original samples, which could have led to model overfitting, four methods were used to enhance the dataset: image flipping, rotation, noise addition, and brightness adjustment. After augmentation, the total number of images increased to 3400.

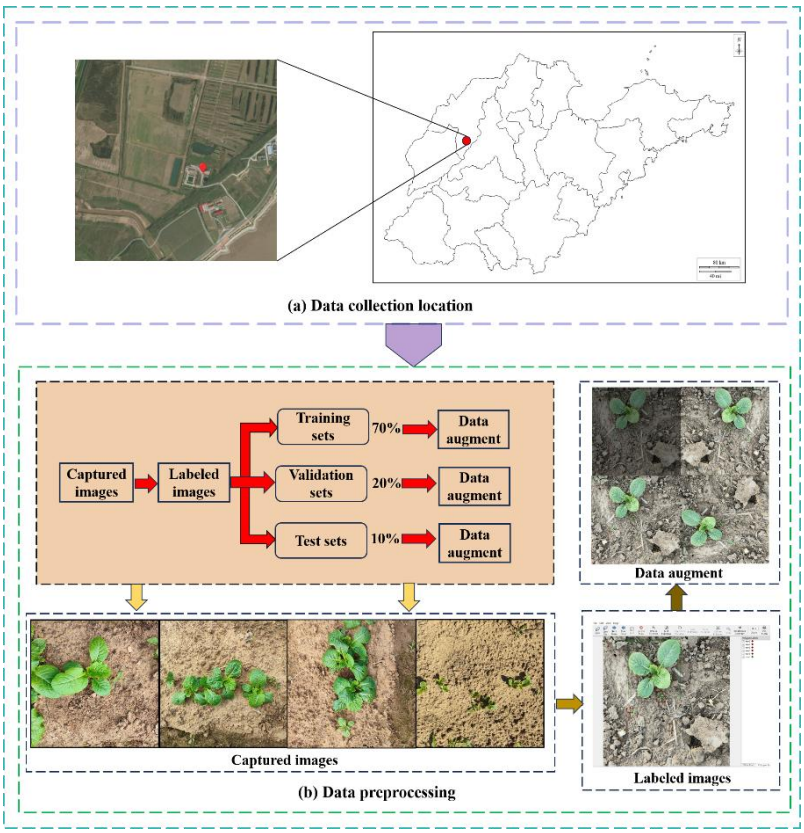

Fig. 1 - Data acquisition and preprocessing flow chart

YOLO-LMSW

YOLOv11 was one of the classic algorithms in the YOLO (You Only Look Once) series *(Redmon et al., 2016)*. As a single-stage object detection algorithm, it balanced accuracy and real-time performance. The overall architecture of YOLOv11 comprised three main components: the backbone, neck, and detection head. The backbone was responsible for extracting fundamental features from the input image, the neck facilitated multi-scale feature fusion, and the head performed the final tasks of object localization and classification. However, the conventional YOLOv11 model entailed a relatively large computational burden, making it difficult to meet the real-time inference requirements of embedded devices.

To address this, and with the aim of improving both detection accuracy and inference speed, this study proposed a lightweight optimization of YOLOv11n. Firstly, a lightweight convolutional module with multi-scale feature extraction capability, LMSConv, was designed, upon which the LMSC3k2 structure was built to form the lightweight backbone network, LMSNet. Subsequently, a lightweight detection head, LMSHead, was proposed to further reduce the model size while maintaining accuracy. Finally, the loss function was replaced with WIoU to enhance the model's regression performance and accelerate convergence during training. The final YOLO-LMSW network architecture, illustrated in Figure 2, achieved significantly improved real-time performance and deployment efficiency, while preserving high detection accuracy.

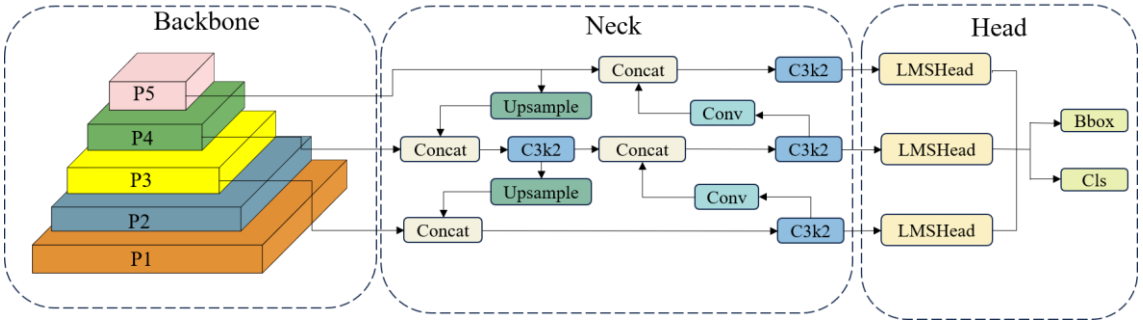

Fig. 2 - YOLO-LMSW Network Architecture

LMSNet

To address the high computational cost of the conventional YOLOv11 backbone and its performance bottlenecks on edge devices—particularly in the context of Chinese cabbage and weed targets that exhibited similar color and significant scale variation—this study proposed a structurally optimized, lightweight multi-scale backbone network, termed LMSNet. The architecture of LMSNet was illustrated in Figure 3.

Firstly, a convolutional module named LMSConv was designed, which combined lightweight characteristics with effective multi-scale feature extraction capabilities. This module replaced part of the Standard Convolution (SC) operations to reduce model complexity. Building upon LMSConv, the LMSC3k2 module was constructed to further enhance the representation of multi-scale features. In the LMSNet backbone, the P1 and P2 layers retained SC operations to ensure sufficient extraction of high-resolution features. In contrast, the P3, P4, and P5 layers incorporated LMSConv and LMSC3k2 modules to achieve more efficient multi-scale feature extraction while alleviating the computational burden. Additionally, the original SPPF and C2PSA modules from YOLOv11 were preserved to strengthen feature extraction and fusion capabilities, ensuring that the lightweight design did not compromise the robustness and accuracy of the model in weed recognition tasks.

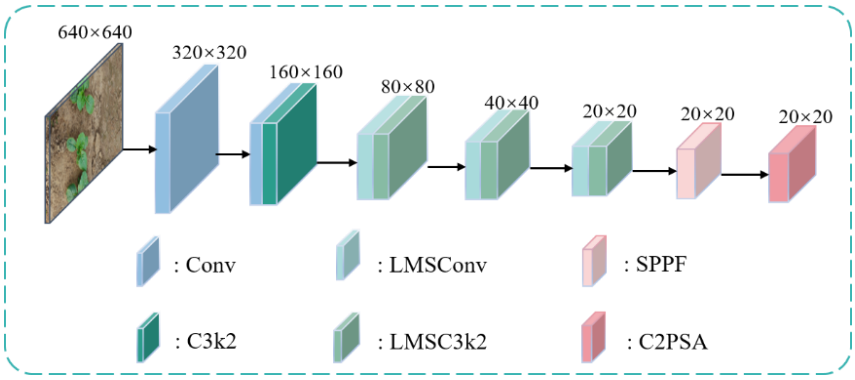

Fig. 3 - LMSNet structure

While SC performed well in feature extraction, their high computational complexity made them less suitable for lightweight models on edge devices. Moreover, due to the significant scale variation between Chinese cabbage and weeds in field environments, single-scale convolution operations struggled to effectively capture features of multi-scale targets.

To address these issues, this study drew inspiration from the ShuffleNet *(Ma et al., 2018)* and Xception *(Chollet, 2017)* architectures and proposed a lightweight multi-scale convolutional module, termed LMSConv. This module significantly reduced the number of parameters and computational load while maintaining model accuracy. Specifically, the input feature map was first split along the channel dimension into two parts. One part was processed with a 3×3 Depthwise Convolution (DWConv) to capture fine-grained details and enhance the perception of small-scale targets; the other part was processed with a 5×5 DWConv to obtain a larger receptive field, thereby improving contextual modelling of large-scale objects. As a typical lightweight convolution operator, DWConv substantially reduced the computational cost and number of parameters compared to SC, making it well-suited for mobile and embedded scenarios. The structure was illustrated in Figure 4(a). Following the application of different-scale DWConv, the resulting feature maps were concatenated along the channel dimension to restore the original number of channels. A 1×1 SC was then applied for channel fusion, enabling cross-channel multi-scale feature interaction and further enhancing feature representation. The complete structure of LMSConv was depicted in Figure 4(b).

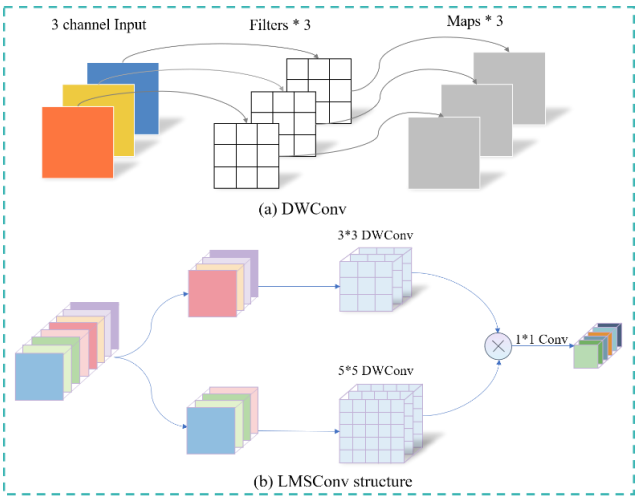

Fig. 4 - The structures of DWConv and LMSConv

Compared to SC, LMSConv significantly reduced computational complexity while enhancing feature representation through the introduction of multi-scale convolutional branches. To quantitatively assess the difference in computational cost between LMSConv and SC, the ratio of their FLOPs was calculated in this study.

$$r_{LMSConv} = \frac{C_{in} \times C_{out} \times H \times W \times 3 \times 3}{\frac{C_{in}}{2} \times 3 \times 3 + \frac{C_{in}}{2} \times 5 \times 5 + C_{in} \times C_{out}} = \frac{9C_{out}}{17 + C_{out}} \quad (1)$$

As shown in formula (1), the gap in FLOPs between LMSConv and SC widened progressively with an increase in the number of output channels, further validating the proposed module's advantage in achieving lightweight computation.

Building upon LMSConv, the Bottleneck structure within the original C3k2 module was modified by replacing its two SC with LMSConv, resulting in a new module termed Bottleneck_LMS. This module retained the original structure's feature extraction capability while significantly reducing the computational load. Subsequently, a Multidimensional Collaborative Attention *(Yu et al., 2023)* (MCA) module was introduced at the output of the C3k2 module to enhance the global modelling capacity of feature representations. MCA integrated information across the channel, height, and width dimensions, thereby improving the network's ability to perceive fine-grained targets and enhancing recognition accuracy. The structure of the MCA module was shown in Figure 5(a).

By placing the MCA attention module after the 1×1 SC within the C3k2 module, the fused channel features were further refined and enhanced, effectively improving the model's ability to perceive global contextual information. The final constructed LMSC3k2 module, illustrated in Figure 5(b), maintained strong feature extraction capabilities while significantly reducing the number of parameters, thereby achieving an improved balance between lightweight design and representational power.

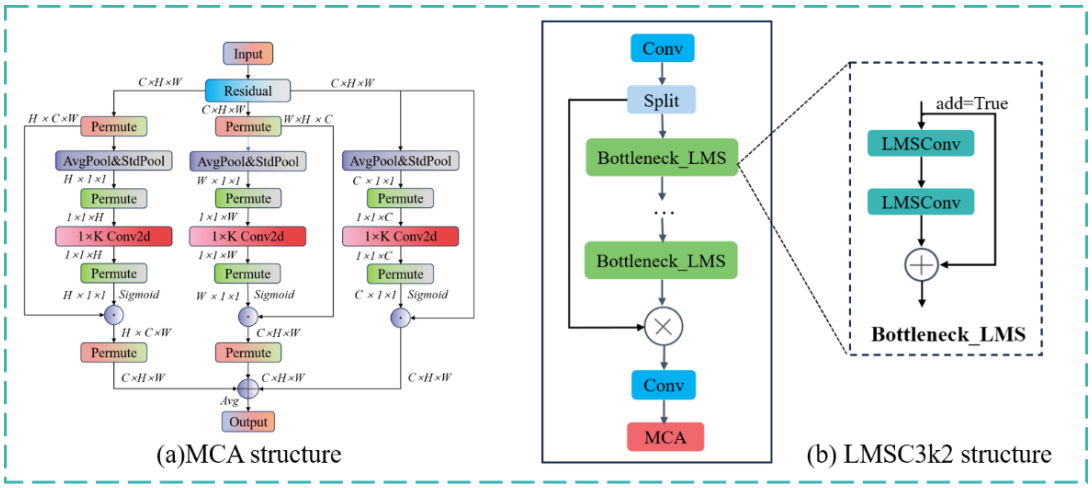

Fig. 5 - The structure of MCA and LMSC3k2

LMSHead

The traditional YOLOv11 detection head consisted of two branches: one branch comprising two 3×3 SC layers and the other employing two 3×3 Depthwise Separable Convolution (DSConv) layers, each of which was processed and then output via a 1×1 Conv2d module. However, this structure resulted in high computational complexity and a large parameter count.

To address this issue, this research redesigned the detection head and proposed LMSHead. LMSConv was introduced into LMSHead, and the original parallel structure was changed to a sequential structure, thus significantly reducing the number of parameters and computational complexity. Specifically, a 3×3 LMSConv was used to replace the original two 3×3 SC layers and two 3×3 DSConv layers. The feature maps transmitted by the neck network first passed through the LMSConv, followed by a 1×1 Conv2d layer, which separately output the regression and classification results. This design simplified the original four convolution operations into one, reducing the number of parameters in the detection head by approximately 72% without sacrificing accuracy. The LMSHead structure was illustrated in Figure 6.

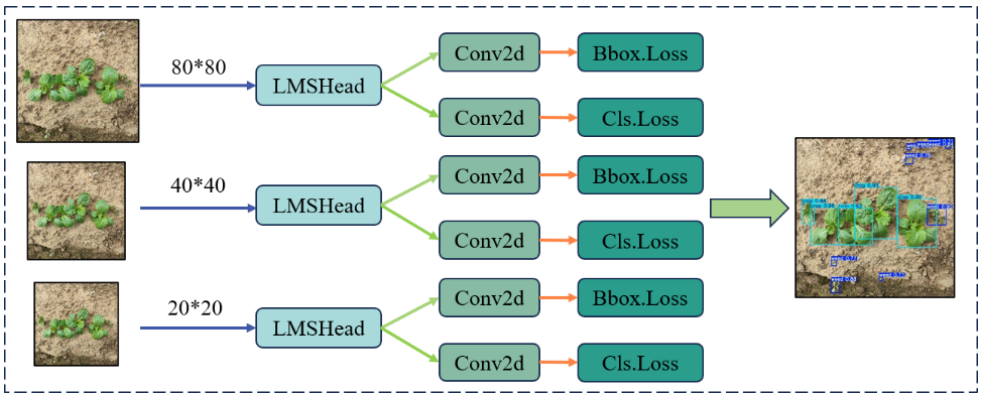

Fig. 6 - LMSHead structure

WIoU

YOLOv11 employed Complete Intersection over Union (CIoU) as its bounding box regression loss function *(Zheng et al., 2022)*. CIoU enhanced the evaluation of predicted boxes by incorporating penalties for the distance between center points, aspect ratio discrepancies, and the IoU itself, thereby providing a more comprehensive assessment of bounding box quality. However, in real-world field environments, Chinese cabbage and weeds often differed significantly in scale, and a substantial proportion of weed instances appeared as small objects. These small targets were prone to being confused with the background, making them difficult to annotate and detect accurately, and thus more likely to be treated as low-quality samples. The use of CIoU exacerbated the penalization of such samples, which in turn impaired the model's recognition capability. To address this issue, this study introduced the WIoU loss *(Tong et al., 2023)*. WIoU incorporated a dynamic, non-monotonic focusing mechanism along with gradient gain to mitigate the negative effects of penalizing low-quality samples, thereby enhancing the model's overall detection performance.

The calculation method of WIoU was presented in formula (2):

$$L_{WIoU} = rR_{WIoU}L_{IoU} \tag{2}$$

$$R_{WIoU} = exp\left(\frac{(a-a_{gt})^2 + (b-b_{gt})^2}{(W_{min}^2 + H_{min}^2)}\right) \tag{3}$$

where: L_{IoU} denoted the IoU-based bounding box loss, R_{WIoU} represented distance attention, a and b represented the horizontal and vertical coordinates of the predicted box center point, respectively, and a_{at} and b_{at} represented the horizontal and vertical coordinates of the real box center point, respectively. W_{min} and H_{min} indicated the width and height of the minimum bounding rectangle between the predicted box and the real box, respectively.

The quality of anchor boxes was characterized by the outlier degree β, defined as:

$$\beta = \frac{L_{IoU}^*}{L_{IoU}}, r = \frac{\beta}{\delta\alpha^{\beta-\delta}} \tag{4}$$

where: * denoted the separation operation. r was the gradient gain, and α and δ were hyperparameters.

Training Platform and Parameters

The model was executed on an experimental platform with a 64-bit Windows 10 operating system and was equipped with CUDA version 12.1. The development environment was PyCharm, running Python version 3.8 and PyTorch version 2.4.1, with an NVIDIA Tesla T4 GPU. The training configuration included 200 epochs, a batch size of 32, an image size of 640 × 640, an initial learning rate of 0.01, and a momentum factor of 0.937. Stochastic Gradient Descent (SGD) was employed as the optimizer.

Performance evaluation indicators

This study used Precision (P), Recall (R), and mean average precision (mAP) as evaluation metrics for model recognition accuracy. The IoU threshold for mAP was set to 0.5. The mAP50, commonly referred to as mAP50, was adopted as a standard metric to facilitate comparison with other models and to provide a more comprehensive assessment of model performance. Floating Point Operations per Second (FLOPs) and the number of Parameters (Params) were used to evaluate model lightweighting. The formulas for P, R, and mAP were defined as follows:

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$mAP = \frac{1}{k}\sum_{i=1}^{k} AP_i \tag{7}$$

where: TP stood for true case, FP for false positive case, FN for false negative case. k denoted the total number of detected object categories.

RESULTS

Ablation experiments

To assess the specific contribution of each proposed module to the model's performance, a series of ablation experiments were conducted, and the results were presented in Table 1. A tick mark "√" indicated that the corresponding module was incorporated in a given experimental configuration. YOLOv11n was used as the baseline model in the first group of experiments. In the second and third groups, the LMSNet backbone and the LMSHead detection head were introduced individually. The results showed that both modules significantly reduced the Params and FLOPs, while improving detection accuracy. In the fourth experiment, the WIoU loss function was introduced independently. Although this module had a minimal impact on computational cost, it contributed to a certain improvement in detection accuracy, verifying its effectiveness in enhancing localization performance. The fifth, sixth and seventh groups of experiments combined two of the three modules respectively, and the final comprehensive performance was all better than that of the benchmark model. In the eighth experiment, all three modules—LMSNet, LMSHead, and WIoU—were integrated. This configuration achieved improvements of 1.2%, 1.0%, and 0.6% in P, R, and mAP50, respectively, while reducing the Params and FLOPs by 34.6% and 36.5%. These results fully validated the effectiveness and practicality of the proposed method in enhancing detection accuracy and accelerating inference speed.

Table 1

Ablation experiments

LMSNet	LMSHead	WIoU	P(%)	R(%)	mAP50(%)	Params(M)	FLOPs(G)
			93.8	91.3	96.4	2.6	6.3
√			94.1	91.4	96.8	2.0	5.4
	√		93.3	92.3	96.6	2.3	4.9
		√	94.7	90.7	96.7	2.6	6.3
√	√		94.0	91.1	96.5	1.7	4.0
√		√	93.1	93.0	96.8	2.0	5.4
	√	√	93.4	91.9	96.7	2.3	4.9
√	√	√	95.0	92.3	97.0	1.7	4.0

As shown in Figure 7, the loss of YOLO-LMSW on both the training and validation sets consistently remained lower than that of YOLOv11n, indicating improved localization accuracy and a faster convergence rate. These results demonstrated the effectiveness of the proposed model in enhancing regression performance.

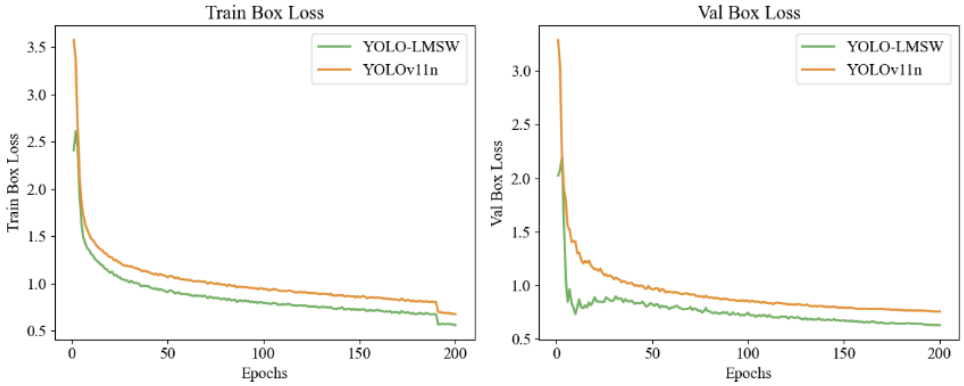

Fig. 7 - Convergence curve of loss function

Comparison experiments

To comprehensively evaluate the detection performance of the proposed YOLO-LMSW algorithm, a series of comparative experiments were conducted against several state-of-the-art one-stage object detection models. The results were summarized in Table 2. The findings indicated that YOLO-LMSW outperformed all competing methods across key evaluation metrics. Compared with YOLOv5, YOLO-LMSW achieved improvements of 0.9%, 2.2%, and 1.0% in P, R, and mAP50, respectively. When compared to YOLOv8, the respective gains were 0.7%, 1.2%, and 0.9%. Relative to YOLOv7-tiny, the performance advantages were even more pronounced, with increases of 4.2% in P, 4.9% in R, and 2.9% in mAP50. Furthermore, YOLO-LMSW demonstrated superior performance over both YOLOv11 and YOLOv12, achieving gains of 1.2%, 1.0%, and 0.6%, and 0.7%, 0.9%, and 0.2%, respectively, across the same three metrics. Compared to RT-DETR, YOLO-LMSW consistently exhibited outstanding results in all evaluated aspects, confirming its competitive advantage.

Table 2

Comparison experiments

Model	P(%)	R(%)	mAP50(%)	Params(M)	FLOPs(G)
YOLOv5n	94.1	90.1	96.0	1.8	4.1
YOLOv7-tiny	90.8	87.4	94.1	6.0	13.0
YOLOv8n	94.3	91.1	96.1	3.0	8.1
YOLOv11n	93.8	91.3	96.4	2.6	6.3
YOLOv12n	94.3	91.4	96.8	2.6	6.3
RT-DETR	87.4	81.6	89.9	27.9	89.2
YOLO-LMSW	95.0	92.3	97.0	1.7	4.0

As illustrated in Figure 8, a qualitative analysis was conducted using several high-performing detection models. In the visualizations, red bounding boxes denoted false positives, while black boxes represented false negatives. The results showed that YOLO-LMSW exhibited markedly superior localization accuracy and recognition performance, particularly in detecting small objects. It also significantly reduced both false positives and false negatives. For large object detection tasks, the model's performance was comparable to that of other benchmark algorithms. These findings clearly indicated that YOLO-LMSW not only maintained strong accuracy for large-scale targets but also significantly enhanced the detection of small-scale objects, thereby validating the robustness and adaptability of the proposed method in multi-scale field scenarios.

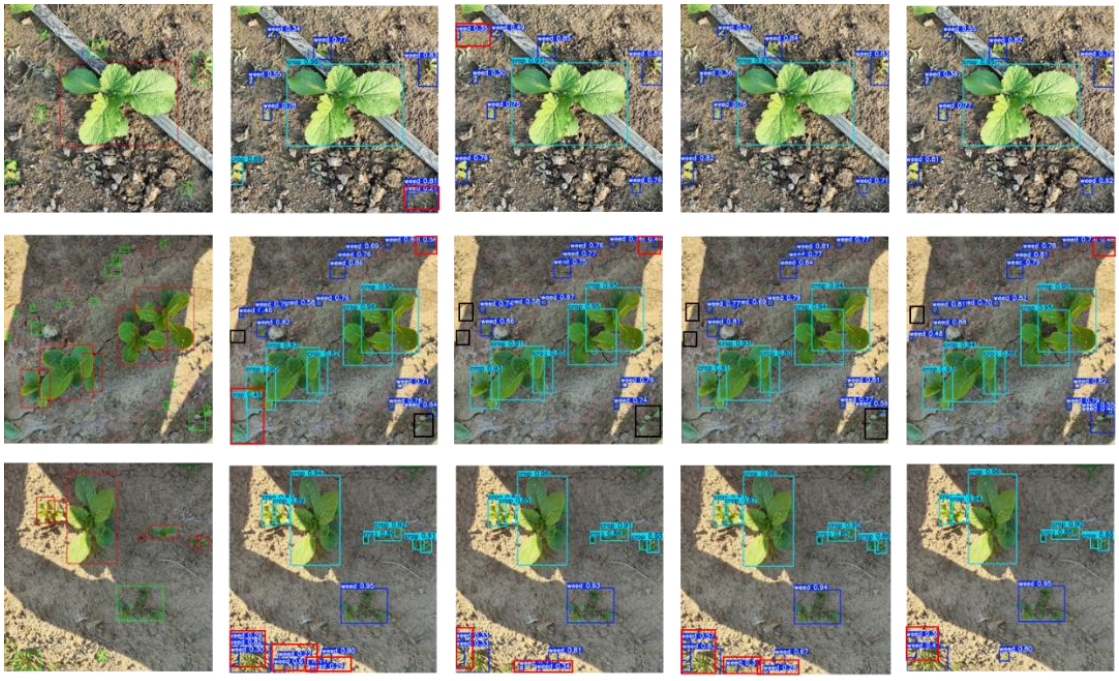

Fig. 8 - Detection effect diagram of different models

Deployment of edge devices

To further evaluate the practical deployment potential of the proposed model, it was deployed on a Jetson Orin Nano (4GB) embedded device. During the deployment test, an independent test set unrelated to the training data was used for inference validation, with the input image size set to 640×640. The inference time for each image and the number of detected instances for each category were recorded throughout the test. The experimental results showed that the model achieved an average inference time of 117 ms per image on the test set. The number of detected targets for each category closely matched the actual quantities, demonstrating the model's capability for real-time detection and its strong robustness in terms of accuracy. The inference performance is illustrated in Figure 9.

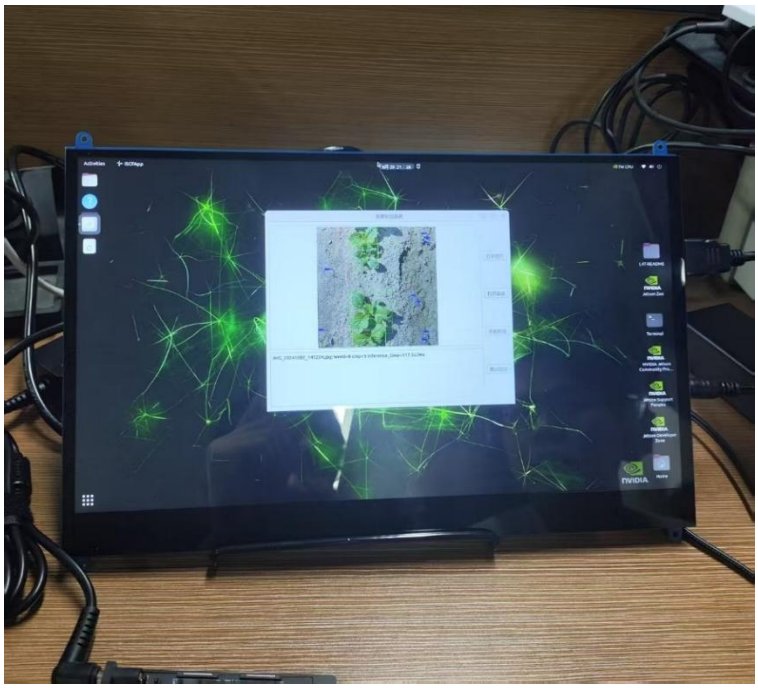

Fig. 9 - Edge device inference effect diagram

CONCLUSIONS

In this study, an improved YOLO-LMSW algorithm based on YOLOv11 was proposed to achieve a balance between speed and accuracy. Firstly, LMSConv was introduced, and the LMSC3k2 module was designed to enhance feature extraction capability and reduce computational load. Subsequently, LMSNet was constructed as the lightweight backbone network of YOLOv11. At the same time, LMSHead was developed as the detection head to make the model structure more efficient. To improve target localization accuracy without increasing computational cost, the WIoU loss function was adopted. The effectiveness of each module was verified through comprehensive ablation and comparison experiments. The experimental results showed that YOLO-LMSW outperformed the baseline models in terms of P, R, and mAP50, while maintaining significantly fewer Params and lower FLOPs. The model achieved an average inference time of 117 ms per image when deployed on a Jetson Orin Nano (4GB) embedded device, confirming its capability for real-time detection. Furthermore, the detection results closely matched the ground truth annotations, demonstrating the robustness and reliability of the model in practical scenarios. These results highlighted the strong potential of YOLO-LMSW for real-time weed detection applications, especially in resource-constrained environments. In future work, the model is planned to be integrated into a weeding robot to further evaluate its performance under real-world agricultural conditions.

ACKNOWLEDGEMENT

This research was supported by the development of a laser weed removal technology system for farmland (381941).

REFERENCES

[1] Cai, Y. L., Zeng, F. G., Xiao, J. Y., Ai, W. J., Kang, G. B., Lin, Y. D., Cai, Z. X., Shi, H. W., Zhong, S. X.,Yue, X. J., (2023). Attention-aided semantic segmentation network for weed identification in pineapple field. *Computers and Electronics in Agriculture, 210*, 107881. doi:https://doi.org/10.1016/j.compag.2023.107881

[2] Chen, J. Q., Wang, H. B., Zhang, H. D., Luo, T., Wei, D. P., Long, T.,Wang, Z. K., (2022). Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. *Computers and Electronics in Agriculture, 202*, 107412. doi:https://doi.org/10.1016/j.compag.2022.107412

[3] Chollet, F., (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *Paper presented at the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2017.195

[4] Darbyshire, M., Coutts, S., Bosilj, P., Sklar, E.,Parsons, S., (2024). Review of weed recognition: A global agriculture perspective. *Computers and Electronics in Agriculture, 227*, 109499. doi:https://doi.org/10.1016/j.compag.2024.109499

[5] Hasan, A. S. M. M., Diepeveen, D., Laga, H., Jones, M. G. K.,Sohel, F., (2023). Image patch-based deep learning approach for crop and weed recognition. *Ecological Informatics, 78*, 102361. doi:https://doi.org/10.1016/j.ecoinf.2023.102361

[6] Hu, R., Su, W. H., Li, J. L.,Peng, Y. K., (2024). Real-time lettuce-weed localization and weed severity classification based on lightweight YOLO convolutional neural networks for intelligent intra-row weed control. *Computers and Electronics in Agriculture, 226*, 109404. doi:https://doi.org/10.1016/j.compag.2024.109404

[7] Hua, C. J., Zhang, A. R., Jiang, Y., Yu, J. F.,Chen, Y., (2022). Lawn weed recognition based on improved fuzzy C-means clustering algorithm(基于改进模糊C均值聚类算法的草坪杂草识别). *Journal of South China Agricultural University, 43*(3), 107-115. doi:10.7671/j.issn.1001-411X.202109005

[8] Ji, W. L., Liu, Z.,Xing, H. H., (2024). Lightweight Method for Identifying Farmland Weeds Based on YOLO v5(基于YOLO v5的农田杂草识别轻量化方法研究). *Transactions of the Chinese Society for Agricultural Machinery, 55*(1), 212-222，293. doi:10.6041/j.issn.1000-1298.2024.01.020

[9] Khan, S., Tufail, M., Khan, M. T., Khan, Z. A., Anwar, S., (2021). Deep learning-based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer. *Precision Agriculture, 22*(6), 1711-1727. doi:https://doi.org/10.1007/s11119-021-09808-9

[10] Ma, N., Zhang, X., Zheng, H. T.,Sun, J., (2018). ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. *Paper presented at the Computer Vision – ECCV 2018*, Cham. doi:https://doi.org/10.1007/978-3-030-01264-9_8

[11] Praveenraj, R., Ramya, R., Thanu, S.,Aswinkumar, R., (2024). Computer Vision based Smart Bot for Weed Detection and Removal in Vegetable Crop Fields. *Paper presented at the 2024 International Conference on Inventive Computation Technologies (ICICT)*. doi:10.1109/ICICT60155.2024.10544594

[12] Redmon, J., Divvala, S., Girshick, R.,Farhadi, A., (2016). You Only Look Once: Unified, Real-Time Object Detection. *Paper presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2016.91

[13] Sabzi, S., Abbaspour-Gilandeh, Y.,Arribas, J. I., (2020). An automatic visible-range video weed detection, segmentation and classification prototype in potato field. *Heliyon, 6*(5), e03685. doi:https://doi.org/10.1016/j.heliyon.2020.e03685

[14] Shang, W. Q.,Qi, H. B., (2022). Identification algorithm of field weeds based on improved Faster R-CNN and transfer learning (基于改进Faster R-CNN与迁移学习的农田杂草识别算法). *Journal of Chinese Agricultural Mechanization, 43*(10), 176-182. doi:10.13733/j.jcam.issn.2095-5553.2022.10.025

[15] Shao, Y. Y., Guan, X. L., Xuan, G. T., Gao, F. R., Feng, W. J., Gao, G. L., Wang, Q. Y., Huang, X. C.,Li, J. C., (2023). GTCBS-YOLOv5s: A lightweight model for weed species identification in paddy fields. *Computers and Electronics in Agriculture, 215*, 108461. doi:https://doi.org/10.1016/j.compag.2023.108461

[16] Tong, Z. J., Chen, Y. H., Xu, Z. W.,Yu, R., (2023). Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. doi:abs/2301.10051

[17] Xiang, M., Gao, X., Wang, G., Qi, J., Qu, M., Ma, Z., Chen, X., Zhou, Z.,Song, K., (2024). An application oriented all-round intelligent weeding machine with enhanced YOLOv5. *Biosystems Engineering, 248*, 269-282. doi:https://doi.org/10.1016/j.biosystemseng.2024.11.009

[18] Yan, J. L., Zhai, Z. K., Feng, Z. X.,Yang, J., (2025). A lightweight millet downy mildew spore detection method based on improved YOLOv8s. *INMATEH-Agricultural Engineering, 75*(1).

[19] Yu, Y., Zhang, Y., Cheng, Z. Y., Song, Z.,Tang, C. K., (2023). MCA: Multidimensional collaborative attention in deep convolutional neural networks for image recognition. *Engineering Applications of Artificial Intelligence, 126*, 107079. doi:https://doi.org/10.1016/j.engappai.2023.107079

[20] Zhang, J. W., Peng, Q. J., Kang, J. M., Di, Z. F., Wang, S. W., Chen, Y. K.,Chen, Y. L., (2025). Design and experiment of a corn inter-plant weeding machine based on visual recognition. *INMATEH-Agricultural Engineering, 75*(1).

[21] Zhang, X., Wang, Q., Wang, C., Wang, X., Xu, Z.,Lu, C., (2024). Guidelines for mechanical weeding: Developing weed control lines through point extraction at maize root zones. *Biosystems Engineering, 248*, 321-336. doi:https://doi.org/10.1016/j.biosystemseng.2024.11.003

[22] Zheng, Z. H., Wang, P., Ren, D. W., Liu, W., Ye, R. G., Hu, Q. H.,Zuo, W. M., (2022). Enhancing Geometric Factors in Model Learning and Inference for Object Detection and Instance Segmentation. *IEEE Transactions on Cybernetics, 52*(8), 8574-8586. doi:10.1109/TCYB.2021.3095305