IMPROVED YOLO11-BASED ALGORITHM FOR SOYBEAN SEEDLING RECOGNITION IN MECHANICAL WEEDING ROBOTS

/

基于改进 YOLO11 的机械除草机器人识别大豆苗算法

Shuai ZANG[1)], Lin WAN[*1,2)], Gang CHE[1,2)], Nai-chen ZHAO[1)], Chun-sheng WU[3)] Jia-yu WANG [1)]

[1)] College of Engineering, Heilongjiang Bayi Agricultural University, Daqing 163319, China

[2)] Key Laboratory of Intelligent Agricultural Machinery Equipment in Heilongjiang Province, Daqing 163319, China

[3)] Jiamusi Branch of Heilongjiang Agricultural Machinery Research Institute, China

Tel:+86-459-13555523188; E-mail: 381995603@qq.com

Corresponding author: Lin Wan

DOI: https://doi.org/10.35633/inmateh-77-84

Keywords: *Mechanical weeding; YOLO11; soybean seedling detection; deep learning.*

ABSTRACT

Addressing issues such as high soybean seedling detection omission rates and inaccurate target recognition during mechanical weeding operations in soybean fields, which lead to low weeding efficiency, this paper proposes a lightweight convolutional model based on an improved YOLO11 model. Deployed on an intelligent mechanical soybean weeding robot, it utilizes precisely identified soybean seedling coordinates to perform mechanical weeding operations, thereby enhancing weeding efficiency. Building upon the original YOLO11 architecture, this model replaces standard convolutional blocks with Deep Separable Convolution (DWconv) modules. It performs channel pruning on the C3K2 lightweight convolutional module and employs Point-Shuffle operations for channel mixing to enhance feature map information flow, thereby improving edge feature recognition for small targets. The introduction of an Efficient Channel Attention (ECA) mechanism increases channel selectivity for large target features, enhancing sensitivity to critical semantic information. The original loss function is optimized by incorporating an improved bounding box loss function (SIOU), accelerating model convergence and strengthening generalization capabilities. The improved YOLO11 model achieved a 2.0 percentage point increase in mAP50% on the self-built soybean dataset compared to the original YOLO11, reaching 94%. Model parameters and floating-point operations were reduced from 2.59MB and 6.4×10⁶ to 1.97MB and 5.0×10⁶ respectively, representing decreases of 23.9% and 21.9%. This achieves synergistic optimization of model lightweighting and computational efficiency while maintaining detection accuracy.

摘要

针对大豆田间机械除草作业时识别大豆苗漏检率高,识别目标不准确等导致除草效率低等问题,本文提出了一种基于改进 YOLO11 模型的轻量化卷积模型,部署在智能机械式大豆除草机器人上,利用识精准别到的大豆苗坐标来进行机械除草作业以提高除草效率.该模型在原 YOLO11 网络架构基础上,使用深度可分离卷积模块 DWconv 替代普通卷积块,对 C3K2 轻量级卷积模块进行通道裁剪,使用 Point-Shuffle 操作进行通道混洗提高特征图间的信息流动,提高对小目标的边缘特征识别效果.引入高效通道注意力机制(ECA),增大对大目标特征的通道选择性,提高对关键语义信息的敏感度.对原损失函数进行优化,引入改进的边界框损失函数（SIOU）,提高模型收敛速度,增强模型泛化性.改进后的 YOLO11 模型,相较于原 YOLO11 在自建大豆数据集上 mAP50%提高了 2.0 个百分点,达到了 94%.模型参数量、浮点计算量由 2.59MB、6.4×106 降低至 1.97MB、5.0×106 同比减少了 23.9%和 21.9%,在保证检测精度的同时,实现了模型轻量化与计算效率的协同优化.

INTRODUCTION

China's annual soybean consumption reaches 120 million tons. In soybean fields, weeds compete with soybean seedlings for sunlight and nutrients, significantly reducing yield. With the rapid development of smart agriculture, accurate soybean seedling detection—an essential component of intelligent weeding operations—faces dual challenges arising from the complexity of field environments and the limited computing capacity of edge devices.

During field operations, image acquisition is easily affected by mechanical vibrations, illumination changes, and soil moisture variations, causing key features of small weed targets—such as leaf edges and stems—to be obscured by noise. Moreover, the severe overlap between soybean leaves and weeds, along with their similar textures, further complicates discrimination, making it difficult for traditional vision algorithms to accurately distinguish between them.

Edge-deployment platforms such as unmanned agricultural machinery and handheld devices impose stringent requirements on the lightweight design and low-latency performance of target detection models. It is essential to reduce model parameters and computational complexity—while maintaining detection accuracy—in order to accommodate the limited resources of embedded processors. However, existing object detection algorithms exhibit notable limitations in soybean-field scenarios.

Two-stage detectors offer high localization accuracy, but their region-proposal mechanisms introduce substantial computational overhead, making it difficult to meet real-time operational demands. Single-stage detectors, such as SSD, provide faster inference but suffer from insufficient capability in small-object feature extraction, and their multi-scale feature fusion strategies lack dynamic channel allocation mechanisms, resulting in suboptimal balance between fine-grained detail capture and large-target recognition

For example, recent studies have explored crop–weed detection using deep learning–based object detection frameworks. An improved Faster R-CNN–based weed detection algorithm was proposed to enhance detection accuracy, achieving a mean average precision (mAP) of 81.3% with a processing time of 0.132 s per image *(Huang et al., 2024)*. In addition, a deep convolutional neural network incorporating color-based features was employed for segmentation, and the improved ResNet model achieved an accuracy of 97.2% on the test set with a detection speed of 78.34 frames per second *(Jin et al., 2024)*. Furthermore, attention mechanisms have been introduced to single-stage detectors, where an enhanced YOLOv8 model integrating an improved convolutional block attention module (CBAM) achieved a mean detection accuracy of 98.2% on the test dataset *(Gao et al., 2024)*.

Although these approaches demonstrate notable improvements in detection accuracy, their applicability to real-world soybean field mechanization remains limited. Two-stage detectors such as Faster R-CNN exhibit advantages in localization precision; however, the computational overhead associated with region proposal generation and refinement makes them unsuitable for deployment on resource-constrained edge devices commonly used in agricultural machinery. In addition, the standard feature pyramid network (FPN) employed in such frameworks tends to suppress fine-grained features of small objects during multi-scale feature fusion, thereby constraining its effectiveness in detecting small soybean seedlings and weeds under field conditions. Single-stage detectors, such as SSD, offer faster inference speeds but suffer from insufficient receptive field design for small targets and the absence of dynamic channel allocation mechanisms. These limitations hinder the model's ability to simultaneously preserve detailed features of small seedlings and maintain robust recognition performance for larger plant structures, leading to degraded performance in complex agricultural environments characterized by uneven illumination, soil background interference, and plant overlap.

From the perspective of soybean production, mechanized field operations—particularly during the early growth stages—face unique challenges, including narrow row spacing, high plant density, and strong sensitivity of seedlings to mechanical disturbance. These characteristics impose strict requirements on perception accuracy, real-time responsiveness, and computational efficiency for onboard vision systems. Therefore, existing detection models, which are primarily optimized for generic scenarios or laboratory conditions, fail to fully meet the practical demands of soybean field mechanization. This highlights the necessity of developing a lightweight, high-precision, and edge-deployable detection framework specifically tailored to soybean field environments, capable of supporting intelligent operations such as precision weeding and autonomous field management.

YOLO-series algorithms, with their end-to-end detection pipelines and strong multi-scale feature learning capabilities, have demonstrated outstanding performance in real-time detection tasks. However, YOLO11 still faces challenges such as a relatively large model size and insufficient accuracy in small-object detection. Thus, further optimization is required to meet the specific demands of agricultural field scenarios.

To address the above issues, this study proposes a lightweight convolutional model based on an improved YOLO11 architecture. The model employs depthwise separable convolutions (DWConv) and channel pruning to reduce its overall size, enabling deployment on a self-propelled mechanical soybean-weeding robot for precise identification of soybean seedlings during field weeding operations, thereby improving the overall weeding rate. A Point-Shuffle channel-mixing operation is introduced to enhance feature flow and improve the recognition of edge features in small targets. Furthermore, an Efficient Channel Attention (ECA) mechanism is incorporated to increase sensitivity to key semantic information of large targets, and an improved bounding-box regression loss (SIOU) is adopted to enhance localization accuracy.

This study aims to achieve collaborative optimization between model lightweighting and detection performance, providing an efficient solution for real-time and accurate identification of soybean seedlings and

weeds in the field, and promoting the large-scale application of intelligent weeding technologies in agricultural production.

IMAGE ACQUISITION

The images used in this study were collected from the experimental fields of Heilongjiang Bayi Agricultural University and the Nenjiang Farm. The imaging system consisted of a camera-equipped acquisition vehicle. Data collection was conducted in June 2023 and July 2024, covering both the cotyledon and true-leaf stages of soybean seedlings. After screening and filtering, a total of 30,000 valid images were obtained, encompassing diverse conditions such as cloudy, rainy, and sunny weather, as well as various occlusions. Representative collected images are shown in Figure 1, where soybean seedlings are highlighted with green bounding boxes. The image acquisition height was 86 cm, and the acquisition platform operated at a speed of 0.5 m/s. The imaging equipment is presented in Figure 2.

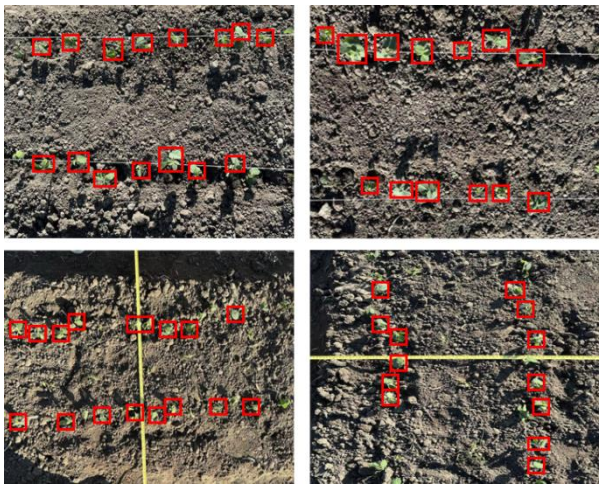

Fig. 1 - Dataset Schematic Diagram

Fig. 2 - Front View of the Acquisition Vehicle

MATERIALS AND METHODS

YOLO11 Object Detection Model

YOLO11 is an optimized extension of YOLOv8, designed to operate efficiently on edge devices and in complex environments. It adopts an anchor-free detection head, eliminating the anchor-box mechanism and directly predicting the coordinates and dimensions of bounding boxes, thereby significantly reducing image processing time. Compared with the C2f module used in the original YOLO models, YOLO11 employs the C2PSA module—an enhanced version of C2f that integrates the PSA module to strengthen feature extraction and attention mechanisms. By incorporating the PSA module into the standard C2f structure, YOLO11 achieves a more powerful attention mechanism, improving its ability to capture critical features. Additionally, owing to the optimized CSPNet backbone, the overall model size is reduced by approximately 23–24% compared with YOLOv8.

Building upon the YOLO11 framework, this study introduces several targeted improvements. First, depthwise separable convolutions and channel-pruning techniques are incorporated to substantially reduce the number of parameters and computational cost while preserving feature-extraction capability, enabling efficient deployment on various edge-computing hardware platforms and lowering power consumption and inference latency. Second, a Point-Shuffle channel-mixing mechanism is integrated to enhance the interaction and flow of information across channels, thereby improving the model's adaptability to multi-view and multi-resolution images and strengthening the robustness of feature representation. Third, the ECA attention mechanism is introduced to dynamically adjust channel weights and emphasize multi-scale target features, enabling the model to better focus on small seedlings and weeds as well as key features embedded in complex backgrounds, effectively mitigating the effects of occlusion and noise. Finally, SIoU loss is adopted to optimize the bounding-box regression process, enhancing localization accuracy in scenarios involving plant occlusion and target confusion and ensuring precise discrimination between seedlings and weeds. Through these improvements, the enhanced YOLO11 model achieves efficient and accurate detection of seedlings and weeds under constrained hardware conditions and complex field environments. The improved model architecture is illustrated in Fig. 3.

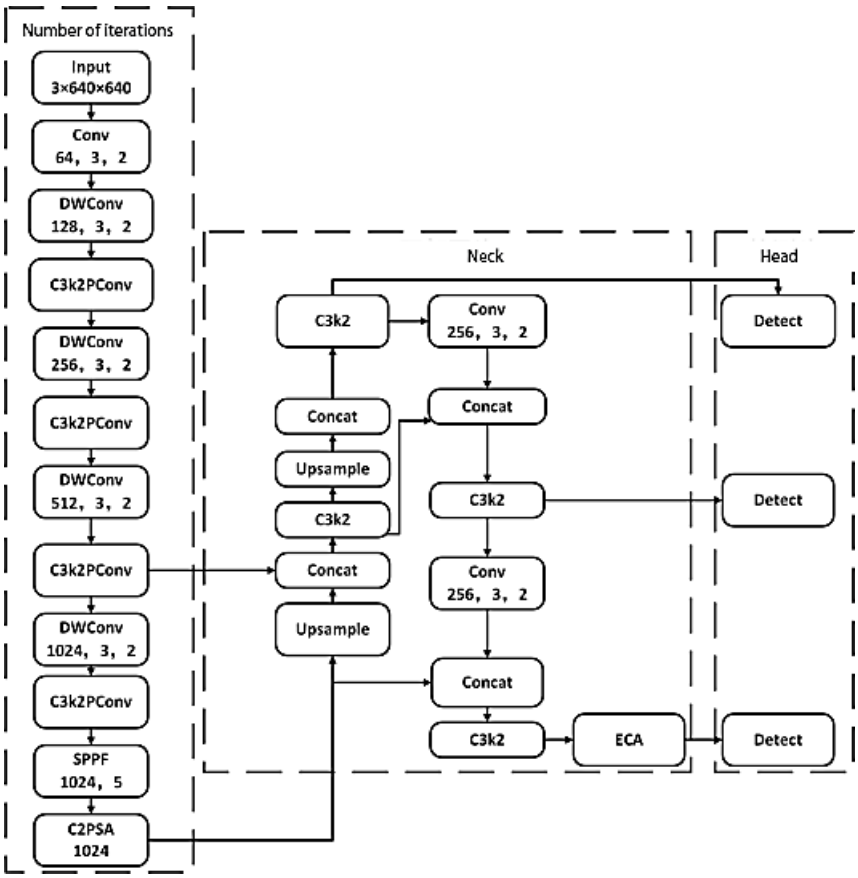

Fig. 3 - YOLO11 Model Architecture Diagram

Note: Conv denotes the convolution module; Concat represents the feature-fusion module; Upsample refers to the upsampling module; Detect indicates the detection head; DWConv denotes the depthwise separable convolution module; C3K2PConv represents the C3K2 module based on partial convolution; and ECA refers to the efficient channel attention mechanism.

C3K2-PConv Module

In the operational scenario of mechanical weeding in soybean fields, the working platform requires high detection accuracy to ensure efficient field operations. At the same time, due to the limited space available on mobile devices, the detection model must remain lightweight to fit within the constrained storage capacity of embedded processors, ensuring accurate soybean–weed detection without increasing hardware costs or computational power consumption. Moreover, mechanical weeding demands strong real-time performance, enabling rapid differentiation and localization of weeds and soybeans so that the operating speed of the weeding mechanism can be effectively synchronized with the movement speed of the machine. To address these requirements, this study applies a lightweight pruning strategy to the C3K2 module and employs the improved C3K2-PConv module, in which channel pruning reduces the number of model parameters and enhances the detection frame rate.

The C3K2-PConv module is an optimized variant based on the C2P architecture. It divides a single input channel into multiple channel groups and applies depthwise convolution to a subset of channels. By adopting a hybrid architecture that integrates standard convolution, partial convolution, and residual connections, the module achieves a balance between feature representation capability and computational efficiency. Parallel processing of multiple channel groups enables the capture of feature information at different scales. As illustrated in Fig. 4, when using an input with a batch size of 4, 32 channels, and a feature-map resolution of 160×160, the feature maps are first normalized by a batch normalization layer. After the SIoU activation function introduces nonlinearity, the data dimensions remain (4×16×160×160). The two branches obtained from channel splitting are processed by two sequential PConv partial-convolution modules. In each PConv module, the feature-map channels are segmented so that only part of the channels undergo 3×3 convolution while the remaining channels are preserved, enabling progressive feature extraction through multiple partial-convolution operations. This approach reduces computational cost while maintaining expressive capability. After passing through the partial-convolution operations, the two branches are fused via channel concatenation, restoring the channel count to 32.

Compared with the original C3K2 module used in YOLO11, the application of the C3K2-PConv module reduces the number of model parameters from 2.59×10^6 to 2.41×10^6, representing a decrease of approximately 6.95%.

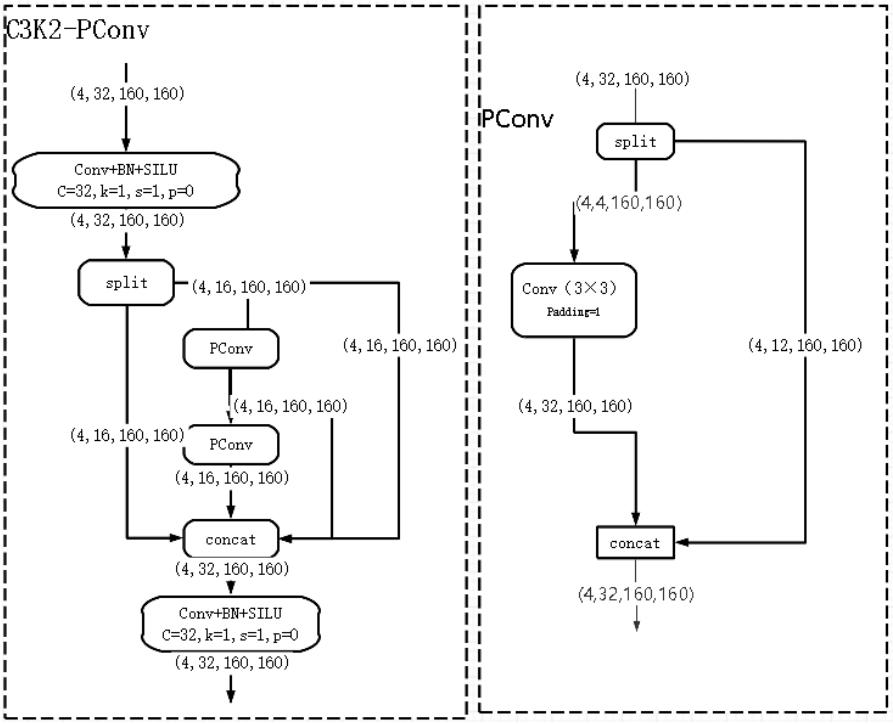

Fig. 4 - C3K2-PConv Architecture Diagram

Note: Conv denotes the convolution module; Concat represents the feature-fusion module; split refers to the channel-splitting operation; PConv denotes the partial convolution module; BN indicates the batch normalization layer; SiLU is the activation function; and Padding refers to the feature-padding operation.

ECA Efficient Channel Attention Mechanism

The ECA module assigns weights to different feature channels through a channel-wise attention mechanism, suppressing irrelevant channels and enhancing the representation of key features such as soybean leaf textures and contours. This addresses challenges in soybean-field weeding scenarios, where severe occlusion between soybean plants and weeds, as well as overlapping visual characteristics—including shape, leaf structure, texture, and color—between soybean seedlings and weeds, make accurate differentiation difficult. The ECA channel-attention mechanism first applies adaptive average pooling to all channels to obtain corresponding descriptors, which represent each channel's average response over the entire spatial feature map and serve as a form of global information aggregation. A 3×3 convolution is then applied along the channel dimension to capture local inter-channel dependencies, reducing the number of parameters relative to fully connected operations. The convolution output is mapped to the range [0, 1] via a Sigmoid activation function to generate attention scores for each channel. These attention weights are subsequently fed back to the spatial dimension of the original feature maps, producing weighted feature representations that emphasize important channels while suppressing irrelevant ones. The attention architecture is illustrated in Fig. 5.

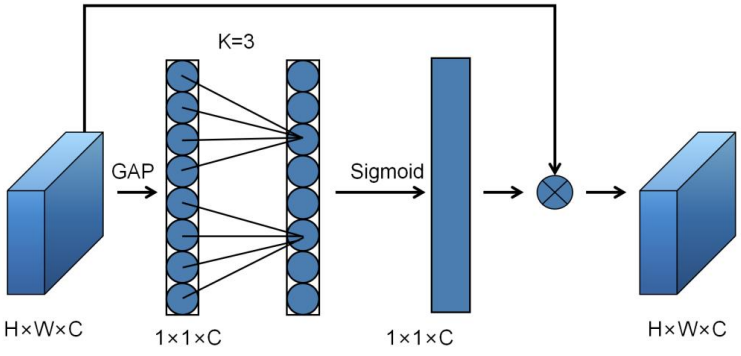

Fig. 5 - ECA Attention Architecture Diagram

SIoU Loss Function

The SIoU loss function is an improved version of the traditional IoU loss, incorporating geometric constraints and a scale-invariant deformation design. In the context of this study, the detection model must output the center point of soybean seedlings to prevent potential crop damage. Traditional IoU loss focuses solely on the overlap ratio between the predicted and ground-truth bounding boxes and is therefore unable to reflect information such as center-point distance or differences in aspect ratio. SIoU enhances the original IoU formulation by introducing three additional components—distance loss, angle loss, and shape loss—providing a more comprehensive measure of the discrepancy between predicted and ground-truth boxes and offering more reasonable gradient guidance. This helps ensure that the predicted bounding box aligns more closely with the center of the soybean plant. The angle component penalizes cases where the angle between the center-to-center line and the coordinate axes is excessively large, guiding the predicted box toward a more optimal direction of movement. The shape component suppresses predicted boxes with large aspect-ratio deviations, ensuring that the bounding-box shape remains close to the ground-truth geometry.

（1）The definition of the angle loss is shown in Fig. 6, and its corresponding calculation formula is given in Equation (1).

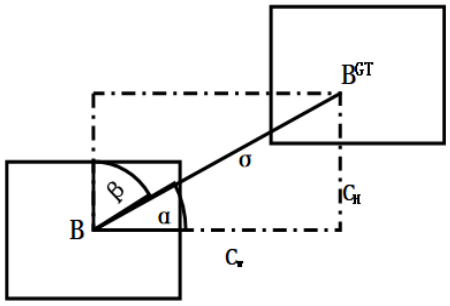

Fig. 6 - Definition of Angle Loss

Note: B denotes the predicted bounding box; BGT represents the ground-truth box; σ is the Euclidean distance between their center points; CW is the horizontal distance difference; and CH is the vertical distance difference.

$$L_{ALoss} = \cos\left(2 \times \left(\arcsin\left(\frac{C_H}{\sigma}\right) - \frac{\Pi}{4}\right)\right) \tag{1}$$

where:

L_{ALoss} denotes the localization-aware loss value; C_H is the vertical distance difference.; σ denotes the classification confidence score output by the detector.

（2）The definition of the distance loss is illustrated in Fig. 7, and its corresponding calculation formula is provided in Equation (2).

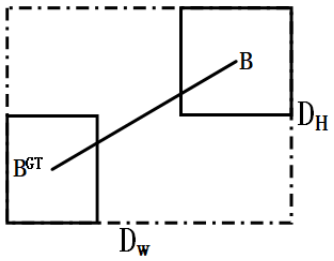

Fig. 7 - Distance Loss

Note: DW and DH denote the width and height of the minimum enclosing rectangle of the ground-truth and predicted bounding boxes.

$$L_{DLoss} = \sum_{i=D_W D_H,} (1-\exp(-\gamma \frac{pi}{ci})), \gamma = 2 - L_{ALoss} \tag{2}$$

where:

L_{DLoss} denotes the localization-driven loss for the predicted bounding box; γ is the scaling factor for the distance loss; i is the index of the bounding box or element in the calculation; pi is the predicted probability of the target class; ci is the ground-truth label of the target class.

（3）The calculation formula for the shape loss is shown in Equation (3).

$$L_{SLoss} = \sum_{i=W,H} (1-\exp(-\theta \frac{\delta i}{mi})) \tag{3}$$

where:

L_{SLoss} denotes the localization shape loss for the predicted bounding box; θ is the coefficient controlling the contribution of the shape loss; δi is the deviation of the predicted box from the ground-truth box in width or height; *mi* is the scalar to amplify the penalty for shape error.

（4）The overall SIoU loss is given by the calculation formula shown in Equation (4).

$$L_{SIOU}=(1-IOU)+L_{DLoss}+L_{SLoss} \quad (4)$$

where: IOU denotes the localization-driven loss.

During soybean recognition in open-field environments, seedlings are often affected by various disturbances such as weed occlusion, unstable lighting conditions, and changes in camera viewpoint, all of which can easily cause fluctuations in detection performance. The SIoU loss function addresses these challenges through a dual-core mechanism. On one hand, its distance penalty term precisely quantifies the center-point deviation between the predicted box and the ground-truth box, ensuring that the model's localization accuracy remains robust against environmental interference. On the other hand, by incorporating the IoU-based penalty term, SIoU effectively constrains the overlap ratio between the predicted and ground-truth boxes, reducing the risk of missed or incorrect detections caused by occlusion or similar disturbances.

EXPERIMENTS AND ANALYSIS

The hardware environment used in this study consisted of a Windows 11 (64-bit) operating system, Anaconda version 24.9.2, CUDA version 12.8, an NVIDIA GTX 1650 GPU, 4 GB of system memory, and Python 3.9, with PyTorch serving as the deep learning framework. The YOLO11n architecture with YOLO11n pretrained weights was employed, and the model was trained for 100 epochs using a self-constructed soybean field dataset. The batch size was set to 16, and the input image resolution was 640×640. The optimization strategy adopted the SCD optimizer with an initial learning rate of 0.01 kept constant, along with a momentum of 0.937 and a weight decay of 0.0005. A warm-up phase was applied during the first three epochs to stabilize early training. Data augmentation included Mosaic processing, HSV color transformation, translation–scaling operations, and other techniques, while a bounding-box loss weight of 7.5 was used to enhance localization accuracy. During training, validation was performed at each epoch using an IoU threshold of 0.7, and automatic mixed precision was enabled to accelerate computation.

Baseline Model Comparison Experiments

A baseline model comparison experiment was conducted to evaluate the performance improvements achieved through model iteration by comparing multiple models from the YOLO family. Compared with other models in the same series, YOLO11 demonstrates significant enhancements in detection accuracy, parameter count, model size, and floating-point computational cost. In terms of detection accuracy, YOLO11-n achieves an mAP@50 of 92, outperforming YOLOv5-n (80.0), YOLOv7-n (85.2), and YOLOv8-n (90.4). Its mAP@50–95 reaches 64.6, exceeding that of YOLOv5-n (57.5), YOLOv7-n (60.7), and YOLOv8-n (64.4). Regarding model size, YOLO11-n is only 4.29 MB, smaller than YOLOv5-n (64.39 MB), YOLOv7-n (5.94 MB), and YOLOv8-n (5.61 MB). In terms of floating-point operations, YOLO11-n requires 2.59 GFLOPs, which is lower than YOLOv5-n (7.22 GFLOPs), YOLOv7-n (36.9 GFLOPs), and YOLOv8-n (3.16 GFLOPs). For the parameter count, YOLO11-n contains 6.4 million parameters, fewer than YOLOv5-n (16.4 million), YOLOv7-n (104.5 million), and YOLOv8-n (8.9 million).

SSD and Faster R-CNN were selected as comparison models for soybean seedling–weed detection experiments. In terms of detection accuracy, YOLO11-n achieves an mAP@50 of 92, which is significantly higher than SSD (70.1) and Faster R-CNN (73.3). Its mAP@50–95 reaches 64.6, also surpassing SSD (53.6) and Faster R-CNN (55.1), indicating a clear advantage in identifying targets in soybean field environments. From the perspective of model size, SSD (90.6 MB) and Faster R-CNN (108 MB) are both much larger than YOLO11-n (4.29 MB), making them unsuitable for deployment on resource-constrained platforms such as mechanical weeding systems. Regarding computational complexity, SSD requires 26.3 GFLOPs and Faster R-CNN requires 137.1 GFLOPs, both considerably higher than YOLO11-n's 2.59 GFLOPs. In terms of parameter count, SSD contains 62.7 million parameters and Faster R-CNN contains 370.2 million, far exceeding YOLO11-n's 6.4 million. These results indicate that SSD and Faster R-CNN impose much higher computational demands and struggle to meet the real-time processing requirements of weeding machinery. The baseline model comparison results are summarized in Table 1, and the model accuracy curves are shown in Fig. 8.

In summary, YOLO11-n demonstrates significant advantages over SSD and Faster R-CNN in terms of detection accuracy, model size, and computational resource consumption, making it well suited to the practical operational requirements of soybean-field weeding. Therefore, YOLO11-n was selected as the baseline model for optimization in this study.

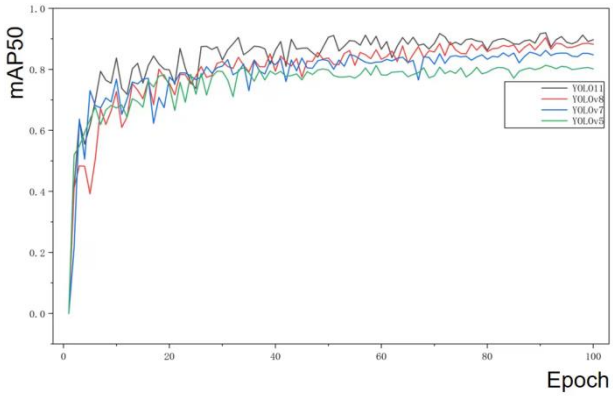

Fig. 8 - Baseline Model Comparison Experiment

Table 1

Baseline Model Comparison Experiments

Model	Detection Accuracy		Model Size/MB	GFLOPS	Number of Parameters/million
	mAP@50/%	mAP@50～95/%			
YOLOV5-n	80.0	57.5	64.39	7.22	16.4
YOLOV7-n	85.2	60.7	5.94	36.9	104.5
YOLOV8-n	90.4	64.4	5.61	3.16	8.9
YOLO11-n	92	64.6	4.29	2.59	6.4
SSD	70.1	53.6	90.6	26.3	62.7
Farster-RCNN	73.3	55.1	108	137.1	370.2

Loss Function Comparison Experiments

The design of the loss function has a direct impact on both the model's localization accuracy and training efficiency. Specifically, mAP@50 reflects localization performance under a low IoU threshold, whereas mAP@50–95 evaluates comprehensive accuracy across multiple thresholds (0.5 to 0.95), imposing substantially stricter requirements on bounding-box regression. Under a unified experimental setting, five representative loss functions—IoU, GIoU, DIoU, CIoU, and SIoU—were compared to assess their influence on model performance.

The results show that SIoU achieves an mAP@50 of 93.4, outperforming IoU (92.0), GIoU (91.8), DIoU (92.1), and CIoU (92.4). This indicates that SIoU already provides superior detection capability under relatively relaxed localization conditions. More critically, SIoU attains an mAP@50–95 of 67.3, nearly 3 percentage points higher than the second-best CIoU (64.4), demonstrating its markedly improved bounding-box prediction accuracy under stricter localization requirements.

In terms of training dynamics, SIoU converges within 62 epochs, fewer than IoU (80), GIoU (75), DIoU (70), and CIoU (68). This suggests that the SIoU formulation better aligns with the model's optimization process, enabling more efficient parameter updates and faster convergence to the optimal solution. Moreover, SIoU achieves the lowest average regression error (8.2), substantially outperforming the other loss functions (IoU: 12.4; GIoU: 11.1; DIoU: 10.3; CIoU: 9.8), indicating more accurate spatial localization and reduced deviation in bounding-box coordinate prediction.

The comparative results of the loss-function experiments are summarized in Table 2.

Table 2

Loss Function Comparison Experiments

Loss Function	mAP@50/%	mAP@50～95/%	Number of Convergence Epochs	Average Regression Error/pixels
IOU	92.0	62.7	80	12.4
GIOU	91.8	63.5	75	11.1
DIOU	92.1	64	70	10.3
CIOU	92.4	64.4	68	9.8
SIOU	93.4	67.3	62	8.2

Ablation Experiments

Ablation experiments were conducted to evaluate the effectiveness of the proposed improvements. Individual components—including DWConv, C3K2-PSConv, ECA attention, and the SIoU loss function—as well as the fully optimized model were systematically analyzed. The evaluation metrics included mAP@50, mAP@50–95, model parameter count, and floating-point operations, with training hyperparameters kept consistent across all experiments.

Comparative analysis of different YOLO11 variants shows that the YOLO11-C3K2PSConv-DWConv-ECA-SIoU model, which integrates DWConv, C3K2-PSConv, ECA, and SIoU, significantly outperforms the baseline YOLO11 model. In single-component experiments, DWConv reduces parameters by 17.4% and FLOPs by 18.8% while improving detection accuracy; C3K2-PSConv enhances feature fusion, increasing mAP@50 by 1.0; ECA attention improves mAP@50–95 by 1.5 through channel-wise attention; and SIoU substantially boosts mAP@50–95 by 2.7 without increasing model complexity. When multiple components are fused, the collaborative effect enables the model to achieve an mAP@50 of 94.0 and mAP@50–95 of 68.1, while reducing the parameter count to 1.97 million and GFLOPs to 5.0. These results validate the effectiveness of the multi-component complementary fusion strategy in enhancing both detection accuracy and computational efficiency.

The results of the ablation experiments are summarized in Table 3.

Table 3

Ablation Experiments

Model	DW conv	C3K2 PSCONV	ECA	SIoU	mAP50/ %	mAP50 ～95/%	Number of Parameters / million	GFLOPs
YOLO11					92	64.6	2.59	6.4
YOLO11-DWconv	√				92.3	65.8	2.14	5.2
YOLO11-C3K2PSCONV		√			93	65.2	2.45	6.4
YOLO11-eca			√		93.5	66.1	2.62	6.6
YOLO11-siou				√	93.4	67.3	2.59	6.4
YOLO11-C3K2PSCONV-DWconv-eca	√	√	√	√	94	68.1	1.97	5.0

Model Visualization Analysis

To evaluate the model's recognition performance, Grad-CAM was employed for heatmap visualization, complemented by field experiments. First, input images were resized and padded to 640×640 pixels using the letterbox algorithm to satisfy the stride constraints of the YOLO series. During forward propagation, intermediate feature maps were captured, and gradients corresponding to class classification and bounding-box regression were accumulated over multiple backward passes. Grad-CAM was then applied to compute channel-wise weights. After ReLU activation and normalization, a single-channel color map was generated. In the resulting heatmaps, regions of high to low response were mapped from red to blue and overlaid semi-transparently onto the original image. The heatmap visualization is shown in Fig. 9. The optimized model exhibits strong responses in both weed and soybean regions while effectively suppressing irrelevant background areas, thereby validating the effectiveness of the proposed model improvements.

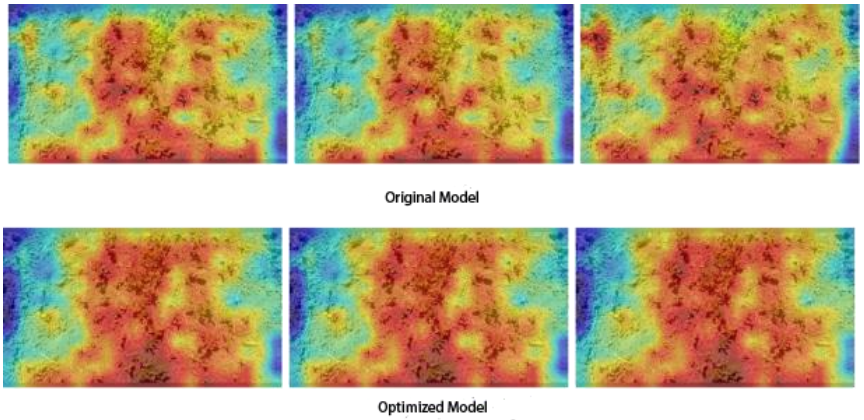

Fig. 9 - Heatmap Analysis

Field Experiments

The optimized object detection model was deployed on an intelligent soybean field weeding machine and validated in the experimental soybean field at Bayi Agricultural University, Heilongjiang. The soybean variety used was Suinong 26, and the growth stage was the first trifoliate leaf (V1) stage. The average plant height ranged from 50 to 80 mm, with ridge height of 200 mm, ridge width of 1100 mm, intra-row spacing of 46 mm, and inter-row spacing of 450 mm. The vehicle speed was controlled at 1 m/s. Predictions were performed on the experimental field using the optimized model, and the detection results are shown in Figure 10. By obtaining the precise coordinates of each detected soybean seedling and using the fixed camera height along with a scaling factor, the spacing between individual seedlings was calculated, as illustrated in Figure 11. This spacing information was transmitted via serial communication to a microcontroller to control the motion of the weeding mechanism, enabling precise inter-row weeding. The expected weeding rate and seedling injury rate were 89.54% and 2.51%, respectively, satisfying the agronomic requirements for soybean weeding operations. The overall weeding performance is demonstrated in Figure 12.

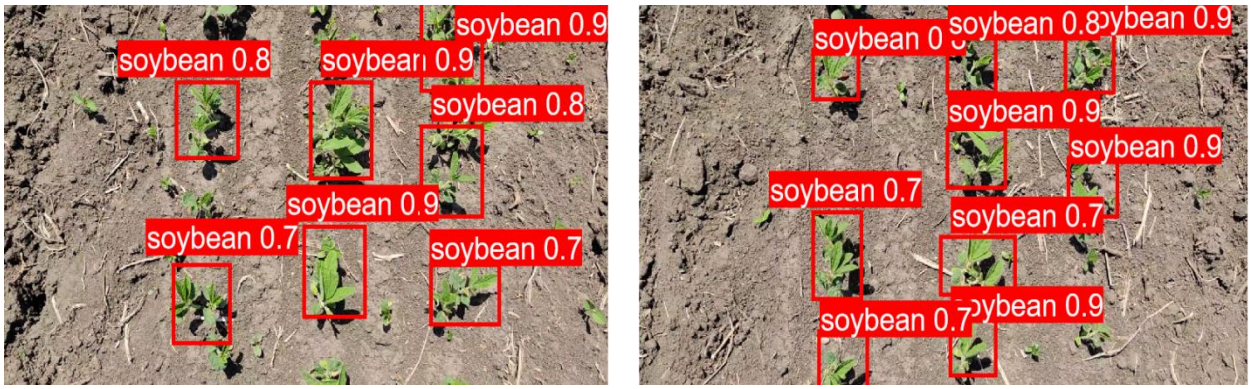

Fig. 10 - Field Detection Results

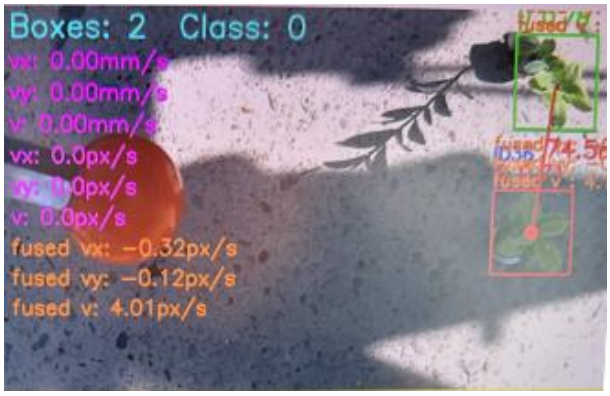

Fig. 11 - Recognition Interface

Fig. 12 -Weeding Effect Diagram

CONCLUSIONS

This study proposes a lightweight convolutional model based on an improved YOLO11 framework for weed detection in soybean fields. Depthwise separable convolutions (DWConv) were employed to replace standard convolutional blocks, and channel pruning was applied to the C3K2 lightweight convolution modules. In addition, Point-Shuffle operations were used for channel shuffling to enhance edge feature recognition of small targets. An efficient channel attention mechanism (ECA) was introduced to improve channel selectivity for large target features, and the loss function was optimized by incorporating SIOU to accelerate model convergence and improve generalization. Experimental results on a self-constructed soybean dataset show that the improved model achieves a mAP@50% of 94%, representing a 2.0% improvement over the original YOLO11 model. The model parameters and floating-point operations were reduced by 23.9% and 21.9%, respectively, while achieving a weeding rate of 89.54% and a seedling injury rate of 2.51%, resulting in enhanced average weeding efficiency. This work achieves a synergistic optimization of detection accuracy, model lightweight design, and computational efficiency, providing an effective solution for real-time soybean seedling and weed recognition and enabling the large-scale application of intelligent weeding technology in soybean fields.

ACKNOWLEDGEMENT

This work was supported by Heilongjiang Provincial Natural Science Foundation Joint Guided Project (LH2023E105)

REFERENCES

[1] Cao, Y.C., Deng, S.P., Zhang, X.L. (2025). Research on Pixel-Level Grasp Detection Method Integrating Attention Mechanism（融合注意力机制的像素级抓取检测方法研究）[J]. *Robot Technique and Application*, No.02, pp. 23-26.

[2] Chen, X., Tan, F. (2025). Research on Lightweight Soybean Field Weed Recognition Method Based on Improved YOLOv5（基于改进 YOLOv5 的轻量级大豆田间杂草识别方法研究）[J]. *Farm Machinery Using & Maintenance*, No.02, pp. 1-7. DOI: 10.14031/j.cnki.njwx.2025.02.001

[3] Cui, J. (2023). *Research on Weed Recognition Method at Soybean Seedling Stage Based on Deep Learning（基于深度学习的大豆幼苗期杂草识别方法研究）*[D]. Jilin Agricultural University. Changchun/China. DOI: 10.27163/d.cnki.gjlnu.2023.001002

[4] Cui, L.Q., Li, W.X. (2024). Contraband Detection in Complex Scenes Integrating Efficient Channel Attention（融合高效通道注意力的复杂场景违禁品检测）[J]. *Journal of Liaoning Technical University (Natural Science),* Vol.43, No.04, pp. 494-505.

[5] Gao, F.R., Gu, H.N., Zhang, Q.L.(2024). Paddy Field Weed Identification Based on Agricultural Big Data and Deep Learning（基于农业大数据和深度学习的稻田杂草识别）[J]. *Jiangsu Agricultural Sciences*, Vol.52, No.18, pp. 215-221. DOI: 10.15889/j.issn.1002-1302.2024.18.028

[6] Han, K.L., Wang, Z.K., Yu, Y.F.(2025). Cotton Field Complex Environment Obstacle Detection Method Based on Improved YOLO11n Model（基于改进 YOLO11n 模型的棉花田间复杂环境障碍物检测方法）[J]. *Transactions of the Chinese Society for Agricultural Machinery*, Vol.56, No.05, pp. 111-120. Beijing/China.

[7] Hou, Y. (2020). *Research on Soybean Weed Recognition Method Based on Image Processing（基于图像处理的大豆杂草识别方法研究）*[D]. Jilin Agricultural University. Changchun/China. DOI: 10.27163/d.cnki.gjlnu.2020.000346

[8] Hu, Y.R., Tian, S.H.M., Wang, X.Y. (2025). Design of Mature Grape Cluster Recognition and Positioning System Based on YOLO Model（基于 YOLO 模型的成熟葡萄簇识别定位系统设计）[J]. *Communication & Information Technology*, No.04, pp. 7-12.

[9] Huang, M.J., Cai, W.Q., Zhang, Z.J.(2025). Real-Time Accurate Recognition Algorithm for Lychee Fruit Varieties Based on Improved YOLO11（基于改进 YOLO11 的荔枝果实品种实时精准识别算法）[J]. *Transactions of the Chinese Society of Agricultural Engineering*, Vol.41, No.11, pp. 156-164. Beijing/China.

[10] Huang, S.Q., Huang, F.L., Luo, L.M.(2024). Research on Sugarcane Field Weed Detection Algorithm Based on Faster R-CNN（基于 Faster R-CNN 的蔗田杂草检测算法研究）[J]. *Journal of Chinese Agricultural Mechanization*, Vol.45, No.06, pp. 208-215. DOI: 10.13733/j.jcam.issn.2095-5553.2024.06.031

[11] Jin, H.P., Mou, H.W., Liu, T.,(2024). Recognition of Vegetables and Weeds Based on Deep Convolutional Neural Network（基于深度卷积神经网络的青菜和杂草识别）[J]. *Journal of Agricultural Science and Technology*, Vol.26, No.08, pp. 122-130. DOI: 10.13304/j.nykjdb.2023.0873

[12] Liu, Y.J., Zhang, K., Wang, L., Chen, X. (2024). Weed Detection Method in Field Crops Based on Improved YOLO Network (基于改进 YOLO 网络的田间作物杂草检测方法) [J]. Transactions of the Chinese Society of Agricultural Engineering, 40(12), pp. 156–164.

[13] Lin, Z.M., Ma, C., Hu, D. (2024). Paddy Field Seedling Stage Weed Detection Method Based on Improved YOLOv8 Convolutional Neural Network（基于改进 YOLOv8 卷积神经网络的稻田苗期杂草检测方法）[J]. *Hubei Agricultural Sciences*, Vol.63, No.08, pp. 17-22. DOI: 10.14088/j.cnki.issn0439-8114.2024.08.004

[14] Peng, Z.X., Ying, Z.F., Ge, H.(2025). ADE-YOLO: Small Object Detection Algorithm in Degraded Environment Based on Improved YOLO11 (ADE-YOLO: 基于改进 YOLO11 的退化环境下小目标检测算法）[J]. *Computer Engineering and Applications*, pp. 1-13. [2025-07-31].

[15] Shu, A.J., Zhang, Y.C. (2024). Lightweight Weed Detection Model Based on Improved YOLOv8（基于改进 YOLOv8 的轻量化杂草检测模型）[J]. *Software Engineering*, Vol.27, No.10, pp.18-22. DOI: 10.19644/j.cnki.issn2096-1472.2024.010.004

[16] Wu, S.C., Mao, Y.M., Hu, H.Z. (2025). Grape Fruit and Leaf Disease Detection Method Based on Improved YOLO11n (基于改进 YOLO11n 的葡萄果叶病害检测方法) [J]. *Transactions of the Chinese Society of Agricultural Engineering*, pp.1-8. https://link.cnki.net/urlid/11.2047.s.20250721.1535.010

[17] Wu, Z., Chen, Y., Zhao, B., Kang, X., Ding, Y. (2021). Review of Weed Detection Methods Based on Computer Vision [J]. *Sensors*, 21(11), 3647. DOI: 10.3390/s21113647.

[18] Wu, Z.K., Zhang, W., Qi, L.Q. (2025). Research on Weed Distribution in Soybean Field Based on Improved YOLOv5（基于改进 YOLOv5 的豆田杂草分布研究）[J]. *Journal of Agricultural Mechanization Research*, Vol.47, No.04, pp. 77-82+91. DOI: 10.13427/j.issn.1003-188X.2025.04.011

[19] Yan, S.Y., Lu, Y.L. (2025). Weed and Soybean Seedling Detection at Early Soybean Growth Stage Based on Lightweight YOLOv5n（轻量化 YOLOv5n 的大豆幼苗期杂草与豆苗检测）[J]. *Sino-Global Food Industry*, No.01, pp. 57-59.

[20] Ye, S.H. (2024). *Design and Experimental Study of Inter-Row Weeding Device for Soybean*（*大豆株间除草装置的设计与试验研究*）[D]. *Chinese Academy of Agricultural Sciences*. Beijing/China. DOI: 10.27630/d.cnki.gznky.2024.000791.