Recognition and localization method of maize weeding robot based on improved YOLOv5

Lijun Zhao, Yunfan Jia, Wenke Yin, Zihuan Li, Chuandong Liu, Hang Luo, Xin Hu, Hua Huang, Qiang Li, Cheng Lyu, Bin Li*

(*College of Intelligent and Manufacturing Engineering, Chongqing University of Arts and Sciences, Chongqing 402160, China*)

Abstract: In response to the challenge posed by low recognition accuracy in rugged terrains with diverse topography as well as feature recognition agricultural settings, this paper presents an optimized version of the YOLOv5 algorithm alongside the development of a specialized laser weeding experimental platform designed for precise identification of corn seedlings and weeds. The enhanced YOLOv5 algorithm integrates the effective channel attention (CBAM) mechanism while incorporating the DeepSort tracking algorithm to reduce parameter count for seamless mobile deployment. Ablation tests validated this model's achievement of 96.2% accuracy along with superior mAP values compared to standard YOLOv5 by margins of 3.1% and 0.7%, respectively. Additionally, three distinct datasets captured different scenarios, and their amalgamation resulted in an impressive recognition rate reaching up to 96.13%. Through comparative assessments against YOLOv8, the model demonstrated lightweight performance improvements, including a notable enhancement of 2.1% in recognition rate coupled with a marginal increase of 0.2% in mAP value, thus ensuring heightened precision and robustness during dynamic object detection within intricate backgrounds.

Keywords: precision agriculture, YOLOv5, weeding robot, maize, laser weeding

DOI: 10.25165/j.ijabe.20251802.9463

Citation: Zhao L J, Jia Y F, Yin W K, Li Z H, Liu C D, Luo H, et al. Recognition and localization method of maize weeding robot based on improved YOLOv5. Int J Agric & Biol Eng, 2025; 18(2): 248–258.

1 Introduction

Weed management plays a pivotal role in facility-based agriculture. The introduction of agricultural automation has enabled precise and efficient autonomous weeding, offering substantial benefits. Leading up to this study, precision agriculture technology has emerged as a transformative force in crop management and weed control within traditional farming practices. Notably, the integration of computer vision technology into automated and intelligent agricultural machinery, particularly in the advancement of crop identification and weeding robotics, holds immense potential[1].

Liu et al.[2] proposed a corn weed detection model that integrates an attention mechanism and spatial pyramid pooling structure based on YOLOv4-tiny, demonstrating real-time high efficiency and strong robustness. Meanwhile, Chen et al.[3] presented an attention mechanism along with an adaptive spatial feature fusion structure based on YOLOv4 surpassing other mainstream models in weed detection within sesame fields. However, practical application poses challenges for this model due to unique geographic and environmental conditions encountered in mountain agriculture. Fatima et al.[4] developed a computational detection system by adapting YOLOv5 onto a stand-alone device, achieving an FPS rate of 27 while maintaining compatibility with laser weeding robots. Zhu et al.[5] devised a blue-light laser weeding robot for maize seedling fields based on YOLOX technology and validated its potential as a non-contact weeding tool through triangulation-based coordinate calculation for targeted weed eradication using monocular ranging. In order to address limitations related to accuracy and speed inherent in existing methods, many scholars have opted for YOLOv5 as their primary model. Wang et al.[6] improved the performance of the modified YOLOv5 architecture in detecting small objects by incorporating spatial pyramid pooling and utilizing an attention module within YOLOv5' s framework. Jin et al.[7] introduced a bidirectional feature pyramid along with GSConv module to enhance recognition classification technique through integration of attention mechanism. Meanwhile, Ju et al.[8] proposed a real-time rice seedling recognition method based on the refined YOLOv5 algorithm, which demonstrates robust performance across diverse backgrounds and growth stages for seedlings. These methods show potential in integrating the YOLO algorithm with standalone equipment. However, most experiments were conducted under static conditions for object detection. Dynamic scenarios involving complex lighting conditions or occlusion phenomena may lead to reduced accuracy during detection as well as increased instances of false positives or missed

Received date: 2024-10-26 **Accepted date:** 2024-12-30

Biographies: **Lijun Zhao**, PhD, Professor, research interest: intelligent agricultural mechanization engineering, Email: 20190005@cqwu.edu.cn; **Yunfan Jia**, MS candidate, research interest: agricultural machinery equipment engineering, Email: 13996626836@163.com; **Wenke Yin**, MS candidate, research interest: agricultural machinery equipment engineering, Email: 17361150524@163.com; **Zihuan Li**, MS candidate, research interest: agricultural machinery equipment engineering, Email: 18285274349@163.com; **Chuandong Liu**, MS candidate, research interest: agricultural machinery equipment engineering, Email: Chuandongliu1228@163.com; **Hang Luo**, MS candidate, research interest: agricultural machinery equipment engineering, Email: 15023962569@163.com; **Xin Hu**, BS, research interest: robot engineering, Email: m15723292474@163.com; **Hua Huang**, PhD, Associate Professor, research interest: neural network, Email: hhuang@cqwu.edu.cn; **Qiang Li**, PhD, Associate Professor, research interest: intelligent agricultural equipment and robots, Email: 20200004@cqwu.edu.cn; **Cheng Lyu**, PhD, Associate Professor, research interest: assembly accuracy analysis and optimization of complex mechanical products, Email: lvcheng0424@126.com.

***Corresponding author: Bin Li**, PhD, Associate Professor, research interest: agricultural machinery equipment engineering. Chongqing University of Arts and Sciences, Chongqing 402160, China Tel: +86-15086706001, Email: cncqlibin@163.com.

detection.

Numerous scholars have leveraged target tracking technology in modern agriculture to enable the recognition and tracking detection of targets during movement, providing technical support for unmanned agricultural technology. For example, Zhang et al.[9] inserted SCAE into DeepSort architecture to measure the spatial continuous trajectory of the target, and verified its validity through experiments. Du et al.[10] combined YOLOv5 with an optimized DeepSort algorithm, resulting in an improved target detector with enhanced technical accuracy and robustness that addresses issues such as color similarity and target overlapping. Kumar et al.[11] initially detected using YOLOv5, followed by prediction and tracking using Kalman Filter and Hungarian algorithms within DeepSort. Meanwhile, Cao et al.[12] proposed a fusion concept involving an enhanced YOLOv5 model based on ECA with the DeepSort algorithm, achieving high testing accuracy and minimal error in dynamic recognition while offering a new theoretical approach for dynamic tracking recognition. Consequently, integrating YOLOv5 and DeepSort algorithm holds significant potential for laser weeders to recognize crops and weeds on the go. Furthermore, incorporating attention mechanisms into the model can aid in improving data processing and model accuracy. Zhang et al.[13] integrated the ECA Attention Mechanism Module along with ASFF Adaptive Feature Fusion into YOLOv5 to effectively address challenges related to small-sized recognition targets and limited features while enhancing average recognition accuracy. Zhang et al.[14] proposed a YOLOX algorithm that integrates the ASFF and CBAM attention mechanisms, achieving an average recognition accuracy of 99.4%, thus providing technical support for precise unmanned agricultural recognition. Meanwhile, Xu et al.[15] addressed the challenges of small sample size and category imbalance by incorporating the SE attention mechanism into ResNet as a generalized feature extractor, demonstrating superior performance compared to commonly used methods. These studies reveal diverse approaches to implementing attention mechanisms, encompassing spatial attention models, channel attention models, and hybrid spatial and channel attention models. By extracting key information from images and suppressing irrelevant details, these models enhance computational efficiency while improving model performance and accuracy in computer vision systems.

Despite extensive research on the integration of target tracking and vision algorithms, there has been limited investigation into corn seedlings in field conditions and insufficient consideration of the impact of complex environments during the recognition process. Therefore, this study aims to develop an efficient small laser weeder for mountainous areas and proposes an enhanced YOLOv5 algorithm combined with DeepSort to improve the continuity and stability of the recognition system. This paper enhances the model's generalization ability through Mosaic data augmentation and incorporates an effective channel attention (CBAM) mechanism into the YOLOv5-DeepSort model, thereby improving recognition accuracy in dynamic scenarios and enhancing performance in complex backgrounds. Finally, ablation tests, comparison tests across different datasets, and comparisons with classical learning algorithms validate the performance of the YOLOv5-DeepSort model, providing technical support for automated weeding recognition.

2 Materials and methods

2.1 Laser weeding experimental platform

A study was conducted to investigate the application of deep learning in corn seedling and weed recognition using a laser weeding experimental platform. The equipment primarily consists of a three-axis servo control system, a closed-loop stepping motor, a laser transmitter, a depth camera, and a PLC control system, as illustrated in Figure 1.

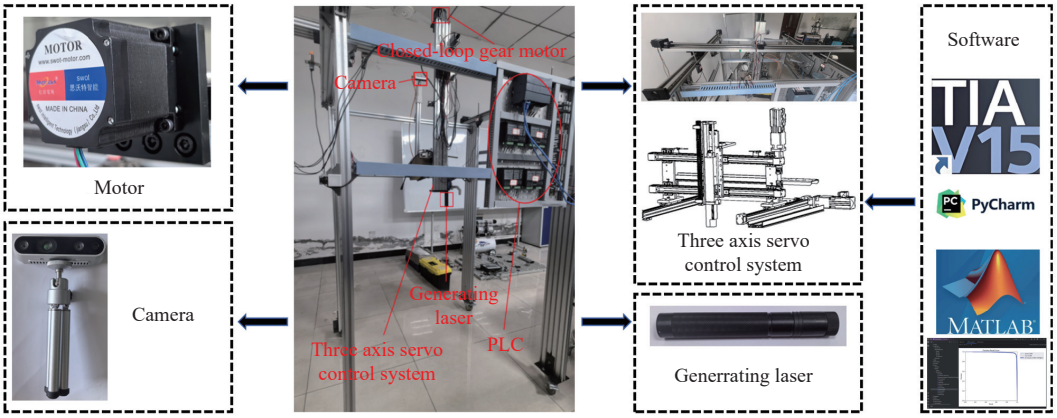


Figure 1 Laser weeding experimental platform and structural schematic

The platform is a three-axis servo control system based on PLC (Programmable Logic Controller), integrated with a depth camera and utilizing the enhanced YOLOv5 machine vision algorithm to accurately identify and localize corn seedlings. The design process involves selecting and matching servo motor parameters, creating motion control flowcharts, developing and implementing PLC programs, as well as visualizing deep learning model features. A Siemens S7-1200 PLC controller is used for single-axis and multi-axis control of the three-axis mechanism, in conjunction with a depth camera to establish TCP/IP communication protocol with Python, enabling end-effector control through communication between Python and Matlab. The three-axis coordinated control mode includes absolute motion mode and relative motion mode; the camera identifies and locates corn seedlings and weeds using data models, transmitting identified coordinate data to the PLC for automatic operation towards specified points. Software writing, execution, and testing are carried out within the PyCharm integrated development environment. The system's development environment operates on Windows 11 OS with an Intel(R) Core(TM) i9 13900HX processor running at 2.70 GHz and 16 GB DDR4 memory. Table 1 lists the PyCharm partial environment.

2.1.1 Serial communication

This research has configured the fundamental parameters of serial communication, including baud rate, data bits, stop bits, and

parity bits, to align with the requirements of the system. As a result, control commands can be transmitted from the central control unit to individual actuator modules such as laser emitters and mobile mechanisms via the serial port. Concurrently, real-time transmission of sensor and camera data back to the central processing unit facilitates analysis and decision-making processes. Additionally, the communication system incorporates error detection and recovery functions to ensure reliable data transmission while mitigating potential losses or command execution errors. To enhance communication stability and efficiency, this research has implemented buffering mechanisms and data encryption measures that safeguard against interference during transmission while ensuring accurate command execution. The integration of these technologies significantly enhances overall system performance and reliability by reducing reliance on human intervention and thereby improving agricultural automation efficiency.

Table 1 PyCharm partial environment

Condition	Version	Purpose
NumPy	1.18.5	Functional arithmetic
Opencv-python	3.8.0	Image video analysis
Torch	1.5.1	Building and training neural networks
PyYAML	5.3	Common data serialization format
Torch-vision	0.6	Processing image data

2.1.2 Camera calibration and spatial coordinate conversion

In the corn seedling and weed recognition experiment in this study, camera calibration and spatial coordinate conversion are crucial procedures for ensuring precise alignment between image data and actual physical dimensions, thereby enhancing the accuracy and reliability of the recognition system. This process involves the following key steps:

Camera calibration was performed to correct perspective distortion and determine the intrinsic. By using a standard calibration plate, multiple sets of images were captured from various angles to calculate the camera's focal length, optical center, aberration coefficient, and other intrinsic parameter information.

The experiment utilized a Homography Matrix to execute this spatial coordinate transformation. The Homography Matrix is derived by selecting several points in the image and their corresponding actual physical position points, then optimizing the calculation using the least squares method. The establishment of the coordinate system in the camera imaging process is illustrated in Figure 2.

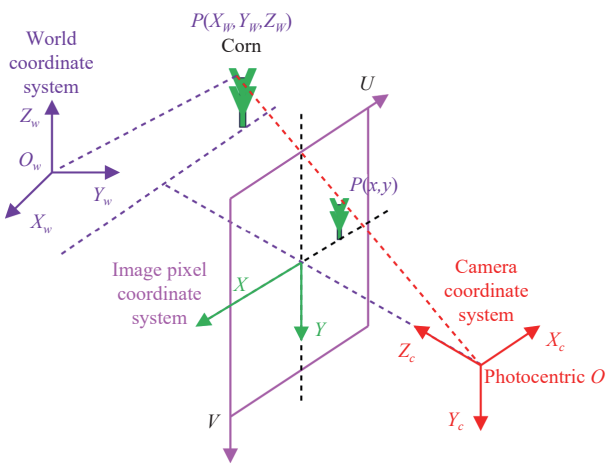


Figure 2 Coordinate system establishment during camera imaging

The $X_WY_WZ_W$ framework delineates an accurate three-dimensional spatial reference frame for precisely defining spatial element positions and their interconnections. The camera's coordination follows a defined $X_CY_CZ_C$ scheme based on an XY plane with its perpendicular Z-axis orientation. Furthermore, ensuring orthogonality with respect to image planes, this setup guarantees that all camera coordinates remain orthogonal. Both global and local coordination systems manifest as three-dimensional structures within 3D space; they can undergo transformation via translation or rotational operations, and the calculation is shown in Equation (1).

$$\begin{bmatrix} x_c \\ y_c \\ z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ \vec{0} & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \quad (1)$$

In the provided equation, R denotes the rotation matrix and t represents the translation matrix, collectively constituting the external parameters of the camera to be determined. These parameters define the transformation process from the camera's coordinate system to the image coordinate system. The Realsense depth camera at a resolution of 640x480 was used for data acquisition and subsequently calibrated using the Camera Calibration Toolbox in MATLAB. The right focal length was set, and the camera's optical axis coordinates were set at (u_0, v_0) in the image coordinate system. A suitable photograph was selected for extracting the external parameter matrices and obtaining details of the rotation matrix.

As the aforementioned data utilized, it is possible to convert pixel coordinates into world coordinates using a MATLAB program. By computing the Euclidean distance between two world coordinate points, precise measurements can be obtained.

2.2 Laser weeding principle

As illustrated in Figure 3, the laser weeding process begins with capturing images of corn seedlings and weeds using the Intel RealSense depth camera. Subsequently, these images are input into the model to extract the coordinates of the corn seedlings and weeds. The PLC then automatically moves to the corresponding points based on this coordinate data. Following this, precise control is exerted over the laser transmitter to target and eliminate the weeds.

2.3 Dataset construction

2.3.1 Dataset acquisition

This study entailed the collection of maize seedling data from diverse environmental settings, resulting in a total of four thousand photos. The image collection site was located within Yongchuan District, Chongqing Municipality (105°38′-106°05′E longitude; 28°56′-29°34′N latitude), with images captured between June 15th and August 10th, 2024. The dataset encompasses Pearl Glutinous 8 maize grown under controlled laboratory conditions as well as those subjected to August mulching or natural outdoor environments during June. These varied conditions facilitated the construction of distinct maize identification datasets focusing on three-to-five leaf stage seedlings. To enable comparative analysis, three separate datasets were established, one for each environmental condition—constant temperature lab setting, August mulching scenario, and integrated experimental setup, respectively—leading to the development of individual recognition models whose performance was compared for selection purposes. A subset of these seedling maize datasets is illustrated in Figure 4, which served as input for real-time detection algorithms.

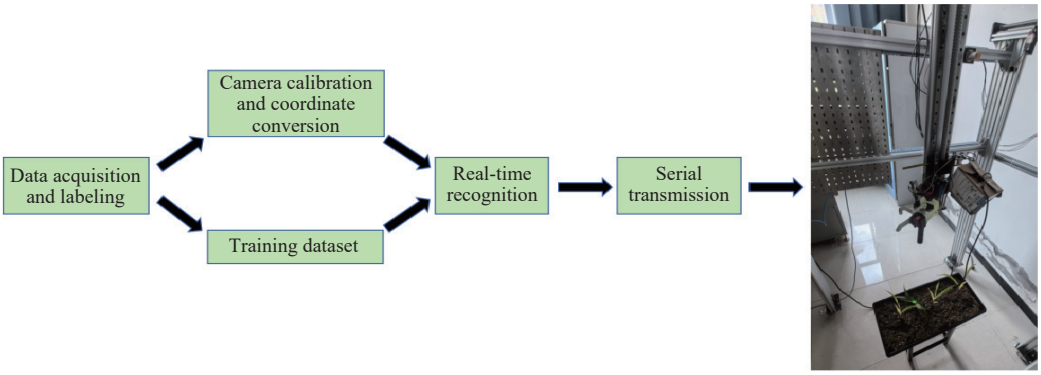


Figure 3 Laser weeding principle

a. Laboratory glutinous corn 8 b. August mulching field environment c. June natural environment

Figure 4 Seedling maize dataset

2.3.2 Dataset labeling

In this investigation, maize labeling was conducted manually using the Labelimg tool to establish two distinct cohorts: a training cohort (train) and a validation cohort (val) for data annotation and label preservation, respectively. Specifically, 80% of the images from the current dataset were assigned to the training cohort folder, while the remaining 20% were designated to the validation cohort folder. This allocation ensures comprehensive data coverage and enhances model training effectiveness.

2.3.3 Data enhancement

In the process of identifying seedling maize, conventional single image inputs may result in insufficient model learning due to the morphological similarities between weeds and maize seedlings in their early growth stages. By utilizing Mosaic data augmentation, the model is compelled to focus on features from various regions of the image during training, thereby enhancing its sensitivity to local features. For example, even when corn seedlings are partially obscured or entangled with weeds, the model can effectively distinguish and identify targets[16]. Moreover, Mosaic data augmentation also promotes the model's adaptation to diverse lighting and environmental conditions, which is crucial for practical field applications. Traditional data augmentation methods such as random rotation, scaling, or color transformation may not be comprehensive enough to address all scenarios. The Mosaic method enhances the robustness of the model in rapidly changing environments by incorporating multiple background and lighting conditions within a single training sample[17]. The Mosaic data enhancement picture is shown in Figure 5.

2.4 Lightweight YOLOv5 target detection model

2.4.1 YOLOv5 network model

In this study, YOLOv5s was chosen as the primary recognition framework due to its optimal balance between speed and accuracy, as well as its suitability for deployment in resource-constrained embedded systems. Among the YOLOv5 family, YOLOv5s stands out as the lightest model and offers comparable performance to YOLOv5m, YOLOv5l, and YOLOv5x. While it may demonstrate slightly lower recognition accuracy, its rapid processing speed and reduced computational requirements are particularly crucial for real-time image data processing in complex and dynamic external environments such as small laser weeders in mountainous areas [18]. Considering the practical considerations of mountain operations, device portability and energy consumption are key design factors. The utilization of YOLOv5s ensures high recognition rates while maintaining system efficiency and enabling prolonged operation. Furthermore, through custom optimization and fine-tuning of the YOLOv5s model, we have enhanced its performance for specific tasks to better align with the needs of seedling maize and weed recognition[19].

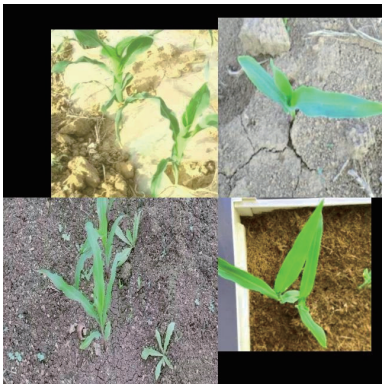


Figure 5 Mosaic data enhancement picture

2.4.2 Recognition model fusing CBAM-YOLOv5

The CBAM (Convolutional Block Attention Module) attention mechanism integrates both channel attention and spatial attention. As illustrated in Figure 6, the CBAM implementation process initially adjusts the input feature layer using channel attention and subsequently applies spatial attention processing. This combination

ensures that the model not only emphasizes critical channel features but also optimizes their essential regions in the spatial dimension, thereby significantly enhancing feature representation[20].

In the channel attention mechanism depicted in Figure 7, initial operations entail global average pooling and global maximum pooling on the input feature layers. Subsequently, the results of average pooling and maximum pooling are integrated with the output of the shared connectivity layer, and weights for the feature layers are computed using a sigmoid function. These weights are then element-wise multiplied with the original input feature layer to modulate its significance.

As depicted in Figure 8, this strategy involves applying an attention mechanism to two pivotal feature layers derived from a backbone network and integrating it during up-sampling procedures with the objective of enhancing both feature processing and recognition efficiency. This arrangement aims to fortify model capabilities by concentrating on extracting crucial features while preserving the structural integrity of pre-trained models.

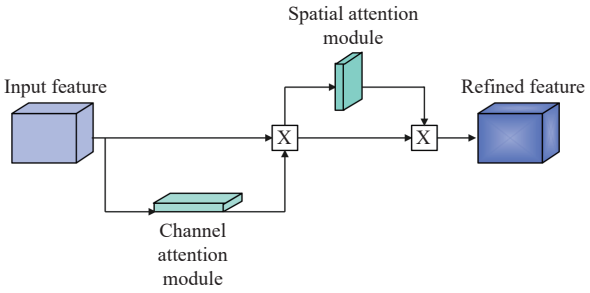


Figure 6 Convolutional attention module implementation process

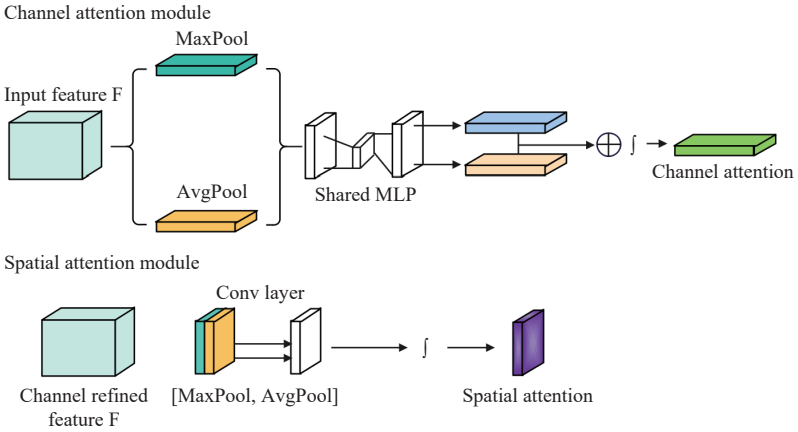


Figure 7 Diagram of composition of channel attention mechanism and spatial attention mechanism

2.4.3 Integration of DeepSort_YOLOv5 for corn and weed recognition

This study integrates the DeepSort algorithm into the YOLOv5-based maize and weed recognition model to enhance target tracking accuracy and continuity (as depicted in Figure 9). The DeepSort (Deep Learning Object Sorting) algorithm expands upon the SORT algorithm by incorporating a deep learning network's feature extraction sub-network, thereby significantly improving targeted object discrimination—particularly in complex scenarios characterized by frequent occlusions and interactions. Leveraging a convolutional neural network within the feature extraction process generates high-dimensional vectors describing individual targets' appearance information. During tracking procedures, after initial identification via YOLOv5, DeepSort employs predictive updating utilizing Kalman filtering for position-velocity estimation while leveraging these high-dimensional vectors for precise data correlation. Furthermore, effective management of target lifecycles through track creation policies reduces identity switching issues during lost track terminations—resulting in more stable tracking outcomes. This amalgamation of advanced feature acquisition alongside precise state estimation facilitates enhanced accuracy as well as reliability when dynamically monitoring corn seedlings and weeds[21].

In a specific application context, this algorithm is well-suited for effectively managing the movement of corn and weeds in intricate farmland environments, where the appearance of the target may be influenced by changing weather or lighting conditions. DeepSort achieves efficient target tracking through a fusion of the Kalman filter and the Hungarian algorithm. The Kalman filter predicts the position of the target between video frames, while the Hungarian algorithm matches the expected position with new detections. This combination not only enhances tracking accuracy but also improves the system's capability to handle occluded and interacting dynamic targets. Figure 10 illustrates a flowchart depicting multi-target tracking using YOLOv5.

The specific process is as follows:

Step 1: YOLOv5 performs target detection on the first frame of the video sequence;

Step 2: Initialize the Kalman filter, set the initial state vector and covariance matrix;

Step 3: Detect the target in the current frame;

Step 4: Using the prediction equation of the filter, combined with the previous information, predict the new position of the target in the current frame;

Step 5: Implement the Hungarian algorithm for data association, match the detection results in the current frame with the predicted results, and ensure the correct tracking of the target;

Step 6: According to the number of matches, make the next decision;

Step 7: If no match is found, or if the match does not meet the conditions, the system will re-initialize the Kalman filter;

Step 8: Check whether the number of matching times exceeds 30. If the number of matching times exceeds the threshold, the tracking is successful. If the number of mismatch times exceeds the threshold, proceed to the next step.

Step 9: If the match is successful, the process will point to the Track Success node. If more than 30 matches are detected, delete the match, indicating that the target is no longer being tracked.

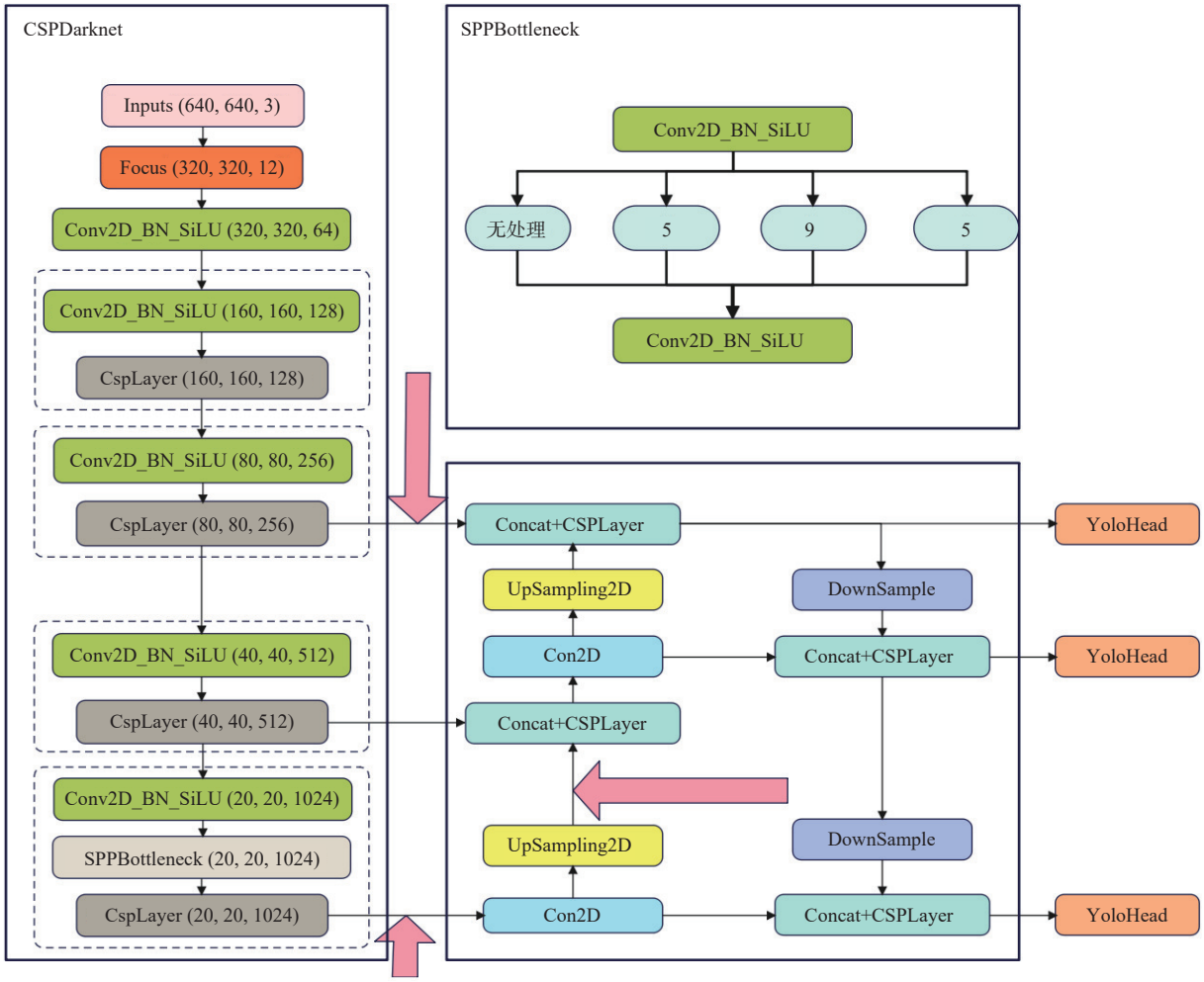


Figure 8 Structural diagram of recognition network incorporating channel attention mechanism

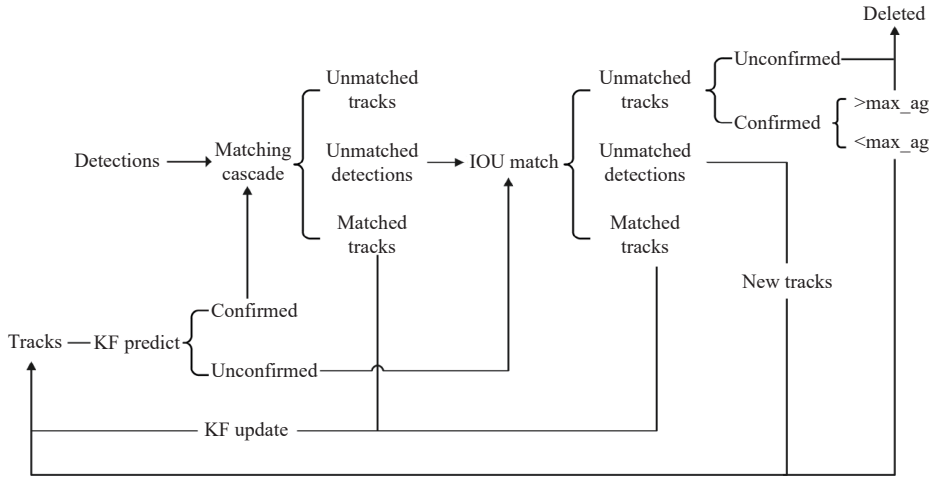


Figure 9 Basic flowchart of DeepSort algorithm

3 Results

3.1 Evaluation indicators

In order to comprehensively and objectively evaluate the performance of both the original model and the improved model in weed detection, this research conducted a comparative analysis of their detection results under identical conditions. This research employed metrics such as Recall (R), Precision (P), Average Precision (AP), mean Average Precision (mAP), and F_1 score[22] to offer a quantitative overview of weed target detection. The calculation is shown in Equation (2):

$$\text{Recall}=\frac{TP}{TP+FN} \tag{2}$$

Precision quantified the performance of the model in terms of false positives, and the calculation is shown in Equation (3):

$$\text{Recall}=\frac{TP}{TP+FP} \tag{3}$$

As a balance between recall and accuracy achieved, the F_1 score is employed as the harmonic mean of precision and recall, where TP denotes the number of true positives and FP denotes the number of false positives, and the calculation is shown in Equation (4):

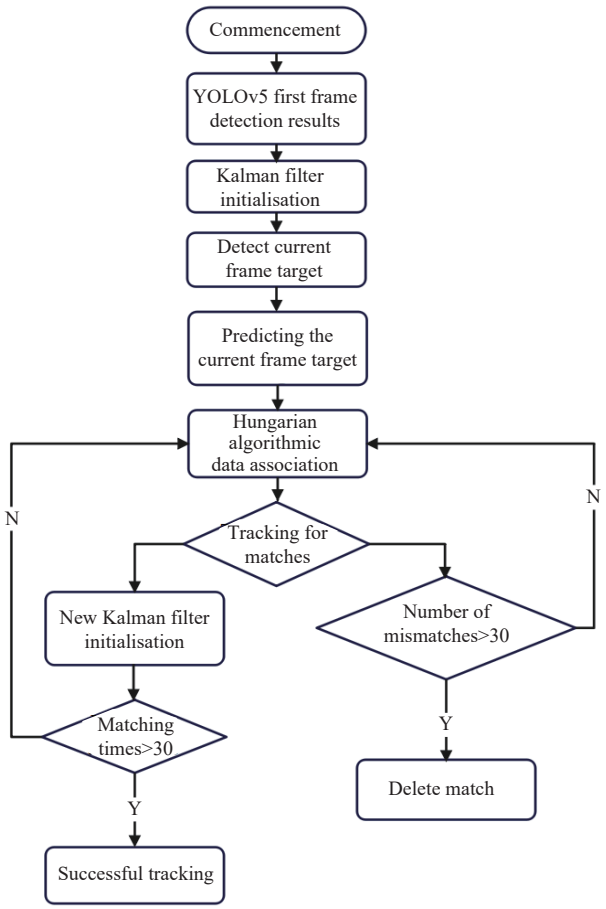


Figure 10 Multi-target tracking YOLOv5 flowchart

$$F_1=2\times\frac{Precision\times Recall}{Precision+Recall} \quad (4)$$

The average of the mean accuracies of all classes *mAP* can be calculated by Equation (5):

$$mAP=\frac{1}{N}\sum_{i=1}^{N}AP_i \quad (5)$$

where, AP_i is the average precision of the i^{th} class and N is the total number of classes.

3.2 Ablation test

To evaluate the performance of the different modules of the proposed DeepSort_YOLOv5 network, ablation experiments were conducted and the results are listed in Table 2. The components compared include the introduced DeepSort algorithm, CBAM, and the WIoU loss function. YOLOv5 integrates the DeepSort algorithm and achieves an accuracy, recall, and mAP value of 93%, 96.8%, and 97.8%, respectively, with a mAP value of 0.1% less than the baseline YOLOv5. Afterward, the CBAM attention network is added to the backbone network. The corresponding YOLOv5 achieved 96.2%, 95.8%, and 98.6% in accuracy, recall, and mAP values, respectively. Compared to Method 2, the accuracy and mAP values increased by 3.2% and 0.8%, respectively, and decreased by 1% in recall. The results show that fusing the CBAM hybrid attention module into the DeepSort_YOLOv5 network significantly improves the accuracy and robustness of model detection. In addition, the WIoU loss function was replaced with Method 3 to form Method 4. Method 4 achieved 95%, 96.1%, and 98.2% on the accuracy, recall, and mAP values, respectively. Compared with Method 3, the accuracy and mAP values of Method 4 decreased by 1.2% and 0.4%, respectively, and only the recall increased by 0.3%. The results show that fusing the CBAM module into the network is the key to improve the model precision, while replacing the loss function with WIoU only improves the prediction rate of positive samples, and performs poorly in terms of recognition precision and stability. The ablation experiment verified the effectiveness of the improved components in improving the performance of maize seedlings and weed detection.

Table 2 Results of ablation experiments

Method	DeepSort	CBAM	WIoU	Precision/%	Recall/%	mAP@0.5/%
1				93.1	96.3	97.9
2	√			93.0	96.8	97.8
3	√	√		96.2	95.8	98.6
4	√	√	√	95.0	96.1	98.2

Figure 11 shows the mAP of the improved YOLOv5 model in detecting maize seedlings and weeds. Overall, these models show good training performance in distal convergence and high detection accuracy, achieving over 94% mAP in 20 training cycles. The training curves show that the accuracies of all models are stable above 100 calendar elements, which confirms that 200 calendar elements of training is sufficient in this study. In Figure 12, the improved YOLOv5 model outperforms YOLOv5, Method 2, and Method 4 in terms of performance in addition to recall, validating the effectiveness of the enhanced DeepSort_YOLOv5 network for corn seedling and weed identification.

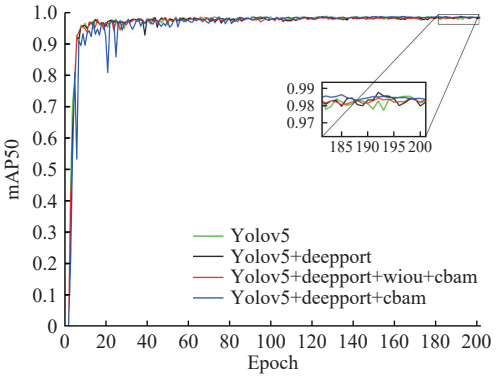


Figure 11 Training curves of mAP for improved YOLOv5 models

3.3 Comparison experiments

To comprehensively evaluate the effectiveness of the DeepSort_YOLOv5 model, this study conducts a comparative analysis with other classical deep learning models, including Faster R-CNN, Cascade R-CNN, YOLOv3, YOLOv4, YOLOv6, YOLOv7, and YOLOv8. While Faster R-CNN represents a two-stage deep learning model category, the rest belong to the one-stage models in the YOLO series. All models are trained and tested on identical datasets under consistent environmental and parameter settings. The comparative results are presented in Table 3. The two-stage model category represented by Faster Cascade R-CNN achieves accuracy, recall, and mAP values of 92.3%, 91.5%, and 96.5%, respectively; however, single-stage models outperform it significantly. For instance,YOLOv8 attains accuracy, recall, and mAP values of 94.1%, 96.7%, and 98.4%, marking an improvement of 1.8%, 5.2%, and 1.9% over Faster Cascade R-CNN's performance metrics. In addition, the DeepSort_YOLOv5 model surpasses other one-stage models, as indicated in Table3, where it demonstrates improvements in accuracy by 2.1% and 0.2% compared to YOLOv8, and by 2.2% and 0.4% compared to YOLOv7. Based on the experimental results presented in Table 3, it is evident that the R-CNN network model exhibits a significantly

a. FasterR-CNN

b. Cascade R-CNN

c. YOLOv5

d. YOLOv8

e. Enhanced YOLOv5

Note: Yellow circles represent false detections, duplicate detections, and missed detections.

Figure 12 Detection results of models

larger number of parameters. This observation leads to the analysis that the two-stage target detection model may not be well-suited for this dataset. In contrast, the YOLOv5 network model demonstrates a lower parameter count compared to Faster R-CNN and Cascade R-CNN, resulting in an increased mAP of 97.9%. This finding suggests that for this dataset, the structure of the one-stage target detection model outperforms that of the two-stage model. Furthermore, the higher number of parameters contributes to improved accuracy and mAP for YOLOv7. On the other hand, both YOLOv3 and YOLOv4 network models possess a significant number of parameters, requiring high computational effort while delivering poor accuracy, thus rendering them unsuitable for maize seedling and weed identification. The more recent algorithms proposed by the YOLO family—namely YOLOv7 and YOLOv8—demonstrate promising performance with fewer parameters and reduced computational effort requirements. It is worth noting that compared with the optimized version that meets the lightweight requirements for mobile detection, the DeepSort_YOLOv5 model proposed by us has 7.01M parameter and 15.8 GFlops, which are 8.6 and 7.6 higher than YOLOv5 and YOLOv8, respectively, and has faster detection speed. Although the detection speed is lower than other algorithms such as YOLOv7, the network structure is more lightweight. Additionally, the proposed DeepSort_YOLOv5 model achieves an F_1 value identical to those of YOLOv7, YOLOv8, and YOLOv5 at 95%, thereby validating its exceptional detection performance.

Due to the relative regularity of the laboratory environment and crop cultivation in the mulched corn field, four images were selected from the natural environment in order to demonstrate the superiority of the proposed model in detecting weeds in complex

natural environments and to confirm the effectiveness of the improved DeepSort_YOLOv5 network model. Comparisons were made with four classical models—Faster R-CNN, Cascade R-CNN, YOLOv5, and YOLOv8—and the detection results are shown in Figure 12. For both Faster R-CNN and Cascade R-CNN, a large number of miss-detections were observed, which suggests that both target detection models are not applicable to this study. YOLOv8 and YOLOv5 also had the problem of slight missed detections and performed poorly in terms of accuracy compared to YOLOv5 and YOLOv8. YOLOv8 also showed duplicate detections, which may affect accuracy and increase detection time, making it impractical for real-world applications.

Table 3 Performance comparison of DeepSort_YOLOv5 and other deep learning models

Models	Precision/ %	Recall/ %	F_1-Score/%	Param (106)	GFlops (109)	mAP@0.5/ %
Faster R-CNN	90.2	89.4	91.0	42.5	156.2	93.8
Cascade R-CNN	92.3	91.5	92.0	64.2	237.5	96.5
YOLOv3	91.8	90.6	91.0	103.6	283.0	95.7
YOLOv4	93.6	92.5	93.0	54.5	119.4	96.7
YOLOv5	93.1	96.3	95.0	2.5	7.2	97.9
YOLOv6	93.9	94.0	94.0	4.2	11.9	97.7
YOLOv7	94.0	95.8	95.0	37.2	105.2	98.2
YOLOv8	94.1	96.7	95.0	10.0	8.2	98.4
Ours	96.2	95.8	95.0	7.01	15.8	98.6

Overall, DeepSort_YOLOv5 was beneficial for reducing common problems in agricultural production applications, such as corn seedlings and weed shading, and significantly improved identification and localization accuracy.

3.4 Experimental results on different datasets

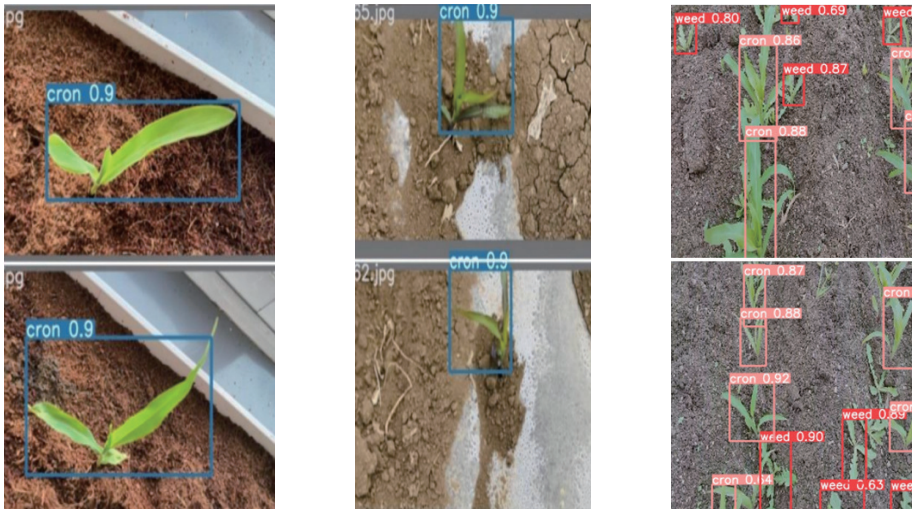
During phase three, this research employed an advanced attention mechanism integrated with DeepSort_YOLOv5 for object recognition purposes. The model underwent rigorous training using four diverse datasets, including controlled laboratory conditions featuring glutinous maize 8 cultivation; outdoor environments with mulched fields observed during August; natural outdoor settings documented in June; as well as a comprehensive combined dataset analysis which is detailed in Table 4 alongside specific training outcomes. The laboratory-based dataset for glutinous corn 8 demonstrated an impressive accuracy rate of 93.64%, showcasing the model's proficiency in identifying targets within controlled settings characterized by consistent lighting and background conditions conducive to feature recognition, thereby yielding high precision levels. Notably, the model achieved higher accuracy (95.27%) when tested under challenging conditions within outdoor environments in August, where factors like light reflections and shadow effects from mulching could potentially disrupt performance, indicating its robustness against environmental complexities. Conversely, performance dipped slightly to an accuracy of 87.90% when operating within natural outdoor settings in June. This could be attributed to the multitude of variables present in this setting, including fluctuating light conditions, changes in weather, and potential obstructions, all of which contribute to heightened recognition challenges. This finding suggests that further refinement of the model or adjustment of the training approach may be necessary to improve recognition accuracy in natural settings. Moreover, the self-constructed fusion datasets exhibit a peak recognition rate of 96.13%, indicating that training the model on integrated datasets containing diverse environmental factors can enhance its capacity for generalization and adaptation across varied scenarios, thereby bolstering overall resilience and adaptability.

Table 4 Experimental results of different datasets

Dataset	Precision/%	Recall/%	mAP@0.5/%
Laboratory glutinous corn 8	93.64	93.01	96.06
August mulching field conditions	95.27	94.79	96.38
June natural environment	87.90	86.52	89.45
Self-built fusion datasets	96.13	95.11	98.41

Additionally, these findings demonstrate that the DeepSort_YOLOv5 model, when integrated with the enhanced attention mechanism, exhibits robust adaptability and recognition capabilities across diverse environments. However, the diminished performance in natural settings observed in June suggests potential challenges for real-world applications, necessitating further adaptation and optimization.

The combined datasets experiment demonstrated an impressive recognition rate of 96.13%. These integrated datasets were employed as the primary data source for input into the DeepSort_YOLOv5 network recognition model, which incorporates a refined attention mechanism for real-time corn seedling detection. The findings of the detection are depicted in Figure13, highlighting the effectiveness and precision of this approach in practical scenarios.



a. Laboratory glutinous Corn 8 b. August mulching field environment c. June natural environment

Figure 13 Schematic representation of recognition of different datasets under the improved network

4 Discussion

In this study, an improved maize seedling and weed detection method based on DeepSort_YOLOv5 was proposed. PLC and depth camera were integrated to establish an experimental platform of three-axis servo control system. At present, many researchers are also conducting research in the field of deep learning. Li et al.[23] proposed an improved YOLOv5 algorithm based on shallow feature layer to solve the problem of gradient disappearance in training process by introducing CBAM attention mechanism. The improved algorithm has a mAP value of 94.3% and a *p*-value of 88.5%. Lin et al.[24] proposed a citrus fruit number counting method based on the combination of improved YOLOv5 algorithm and DeepSort tracking algorithm. CBAM and Contextual Transformer attention mechanisms were incorporated, and SIoU loss function was used to replace GIoU, with a tracking accuracy of 90.83%. However, the real-time performance of the improved algorithm is reduced. Garcia-Navarrete et al.[25] implemented an artificial vision system based on the YOLOv5 model to distinguish corn from four kinds of weeds, which played a certain role in promoting the construction of an accurate weeding system. The recognition *p*-value of corn was 97%, and the mAP value was 97.5%. After introducing the CBAM attention mechanism and integrating the DeepSort algorithm, the new method proposed in this study can identify datasets under three different scenarios, with a *p*-value of 96.13% and a mAP value of 98.41%, which can provide technical support for the development of precision agriculture technology to adapt to the diversified and challenging agricultural environment in the future. However, the enhanced YOLOv5 algorithm still faces challenges in accurately identifying small targets against complex backgrounds, as well as being affected by light, rainfall, and overlapping of recognized objects. Li et al.[26] proposed a fusion design of MCD-YOLOv5, and also established an unmanned aerial vehicle (UAV) for crop pest detection, providing ideas for identification stability and accuracy. Therefore, in the subsequent stage, more images of corn seedlings and weeds will be collected, the algorithm structure will be further improved, and low-altitude drones will be combined with airborne cameras to improve the recognition accuracy and stability under complex terrain.

5 Conclusions

This study faced challenges in achieving precise results using existing detection methods due to the complexities associated with variations in weed density within diverse environmental backgrounds. In response, this research proposed an improved approach for maize seedling and weed detection based on DeepSort_YOLOv5 to enhance both precision and resilience within an improved algorithm framework. The refined methodology resulted in the following:

1) Four distinct datasets were curated, while model resilience was enhanced through data augmentation techniques capable of accommodating varied image scales under intricate environmental scenarios. Notably, our enhanced DeepSort_YOLOv5 achieved an impressive accuracy rate of 96.13% on internally constructed datasets—outperforming all other sets—validating its effectiveness across practical applications amidst challenging environmental conditions.

2) The CBAM attention mechanism was incorporated alongside integrating DeepSort into the YOLOv5 architecture. This research observed a significant improvement, with a 3.1% increase in both accuracy rates and mean Average Precision (mAP) values during ablation testing when compared against the standard YOLOv5 setup, an essential capability for effectively distinguishing between corn seedlings and weeds. In comparative assessments against traditional target-detection models, the DeepSort_YOLOv5 demonstrated outstanding performance metrics, an impressive 96.2% accuracy rate coupled with a high mAP value at 98%. Furthermore, its compact parameter count of just 7.01 M makes it suitable for mobile deployment.

3) A PLC-based three-axis servo control system was developed and integrated with a depth camera, and an enhanced YOLOv5 algorithm was utilized to achieve precise positioning of corn seedlings. The effectiveness of the experimental platform was validated through practical experiments.

Acknowledgements

This work was financially supported by Chongqing Science and Technology Bureau Key R&D Projects in Agriculture and Rural Areas (Grant No. cstc2021jscx-gksbX0003), Chongqing Municipal Education Commission Science and Technology Research Project (Grant No. KJZD-M202201302), Chongqing Municipal Science and Technology Bureau Excellence Programme Project (Grant No. 20231102), Chongqing Municipal Science and Technology Bureau Innovation and Development Joint Fund Project (Grant No. CSTB2022NSCQ-LZX0024), and the 2024 Chongqing Natural Science Foundation Joint Fund for Innovation and Development (Municipal Education Commission) Project (Grant No. CSTB2024NSCQ-LZX0091).

[References]

[1] Jiang H, Murengami B G, Jiang L, Chen C, Johnson C, Cheein F A, et al. Automated segmentation of individual leafy potato stems after canopy consolidation using YOLOv8x with spatial and spectral features for UAV-based dense crop identification. Computers and Electronics in Agriculture, 2024; 219: 108795.

[2] Liu S Q, Jin Y S, Ruan Z W, Ma Z, Gao R, Su Z B. Real-time detection of seedling maize weeds in sustainable agriculture. Sustainability, 2022; 14(22): 15088.

[3] Chen J Q, Wang H B, Zhang H D, Luo T, Wei D P, Long T, et al. Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. Computers and Electronics in Agriculture, 2022; 202: 107412.

[4] Fatima H S, ul Hassan I, Hasan S, Khurram M, Stricker D, Afzal M Z. Formation of a lightweight, deep learning-based weed detection system for a commercial autonomous laser weeding robot. Applied Sciences, 2023; 13(6): 3997.

[5] Zhu H B, Zhang Y Y, Mu D L, Bai L Z, Zhuang H, Li H. YOLOX-based blue laser weeding robot in corn field. Frontiers in Plant Science, 2022; 13: 1017803.

[6] Wang M J, Li Y, Meng H W, Chen Z W, Gui Z Y, Li Y P, et al. Small target tea bud detection based on improved YOLOv5 in complex background. Frontiers in Plant Science, 2024; 15: 1393138.

[7] Jin X, Jiao H W, Zhang C, Li M Y, Zhao B, Liu G W, et al. Hydroponic lettuce defective leaves identification based on improved YOLOv5s. Frontiers in Plant Science, 2023; 14: 1242337.

[8] Ju J Y, Chen G Q, Lv Z Y, Zhao M Y, Sun L, Wang Z T, et al. Design and experiment of an adaptive cruise weeding robot for paddy fields based on improved YOLOv5. Computers and Electronics in Agriculture, 2024; 219: 108824.

[9] Zhang T, Zhao D F, Chen Y S, Zhang H L, Liu S L. DeepSORT with siamese convolution autoencoder embedded for honey peach young fruit multiple object tracking. Computers and Electronics in Agriculture, 2024; 217: 108583.

[10] Du P C, Chen S, Li X, Hu W W, Lan N, Lei X M, et al. Green pepper fruits counting based on improved DeepSort and optimized Yolov5s. Frontiers in Plant Science, 2024; 15: 1417682.

[11] Kumar S, Singh S K, Varshney S, Singh S, Kumar P, Kim B G, et al.

Fusion of deep sort and Yolov5 for effective vehicle detection and tracking scheme in real-time traffic management sustainable system. Sustainability, 2023; 15(24): 16869.

[12] Cao Y Y, Chen J, Zhang Z C. A sheep dynamic counting scheme based on the fusion between an improved-sparrow-search YOLOv5x-ECA model and few-shot deepsort algorithm. Computers and Electronics in Agriculture, 2023; 206: 107696.

[13] Zhang D Y, Zhang W H, Cheng T, Zhou X G, Yan Z H, Wu Y H, et al. Detection of wheat scab fungus spores utilizing the Yolov5-ECA-ASFF network structure. Computers and Electronics in Agriculture, 2023; 210: 107953.

[14] Zhang P, Li D L. CBAM+ASFF-YOLOXs: An improved YOLOXs for guiding agronomic operation based on the identification of key growth stages of lettuce. Computers and Electronics in Agriculture, 2022; 203: 107491.

[15] Xu X L, Li W S, Duan Q L. Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification. Computers and Electronics in Agriculture, 2021; 180: 105878.

[16] Asadi B, Shamsoddini A. Crop mapping through a hybrid machine learning and deep learning method. Remote Sensing Applications: Society and Environment, 2024; 33: 101090.

[17] Ahmed Z, Nalley L, Brye K, Green V S, Popp M, Shew A M, et al. Winter-time cover crop identification: A remote sensing-based methodological framework for new and rapid data generation. International Journal of Applied Earth Observation and Geoinformation, 2023; 125: 103564.

[18] Chen R Q, Sun L, Chen Z X, Wuyun D J, Sun Z. Early identification of corn and soybean using crop growth curve matching method. Agronomy, 2024; 14(1): 146.

[19] Tian Y H, Zhang K, Hu X B, Lu Y. Crop type recognition of VGI road-side images via hierarchy structure based on semantic segmentation model Deeplabv3+. Displays, 2024; 81: 102574.

[20] Lv M, Su W H. YOLOV5-CBAM-C3TR: an optimized model based on transformer module and attention mechanism for apple leaf disease detection. Frontiers in Plant Science, 2024; 14: 1323301.

[21] Xu H Y, Song J, Zhu Y Q. Evaluation and Comparison of Semantic Segmentation Networks for Rice Identification Based on Sentinel-2 Imagery. Remote Sensing, 2023; 15(6): 1499.

[22] Wang Z, Liu D, Wang Z, Liao X, Zhang Q. A new remote sensing change detection data augmentation method based on mosaic simulation and haze image simulation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023; 16: 4579–4590.

[23] Li R, Wu Y P. Improved YOLO v5 wheat ear detection algorithm based on attention mechanism. Electronics, 2022; 11(11): 1673.

[24] Lin Y H, Hu W X, Zheng Z H, Xiong J T. Citrus identification and counting algorithm based on improved YOLOv5s and DeepSort. Agronomy, 2023; 13: 1674.

[25] García-Navarrete O L, Santamaria O, Martín-Ramos P, Valenzuela-Mahecha M Á, Navas-Gracia L M. Development of a detection system for types of weeds in maize (*Zea mays* L.) under greenhouse conditions using the YOLOv5 v7. 0 model. Agriculture, 2024; 14(2): 286.

[26] Li L P, Zhao H, Liu N. MCD-Yolov5: accurate, real-time crop disease and Pest identification approach using UAVs. Electronics, 2023; 12(20): 4365.