Study on Potato Identification Algorithm Based on Improved YOLOv8s

Qing ZHAO, Caojun HUANG *

Heilongjiang Bayi Agricultural University, Daqing 163000, China

Abstract [**Objectives**] This study was conducted to propose an improved YOLOv8s-based potato identification algorithm, in order to address issues such as low detection precision and high miss detection rate in potato identification tasks. [**Methods**] The backbone network was replaced with InceptionNeXt, and the CBAM dual attention mechanism was introduced, to enhance the model's multi-scale feature extraction capability. [**Results**] The improved YOLOv8s algorithm achieved an identification precision of 94.55%, a recall of 85.34%, and an F1-score of 87.37% in potato identification. Compared with the original algorithm, it improved precision by 7.40%, recall by 2.71%, and F1-score by 2.56%. The average processing time per image was reduced by 0.12 s compared with the unimproved algorithm. The results of simulation tests showed a success rate of 98.20% in 2 000 simulated identifcation tests. [**Conclusions**] This study provides a high-precision and robust solution for potato identification tasks.

Key words Improved YOLOv8s; Potato; Object detection; Simulated identification

DOI: 10.19759/j.cnki.2164-4993.2025.04.015

In recent years, the integration of deep learning-based object detection algorithms and robotic control technology has provided new technical pathways for agricultural automation. The YOLO series of models, known for their efficient detection speed and accuracy, have become a preferred framework in the field of industrial vision. Existing research includes the study by Sun *et al.* [1], who improved the YOLOv5s model by embedding SE/CA/CBAM attention mechanisms and enhancing the SGBM algorithm. The improved model achieved a target recognition precision mAP of 99.30% and a processing speed of 63 FPS, meeting real-time requirements. However, the dataset used in their experiments was insufficient, resulting in limited model generalization and weak performance in real-world scenarios, making it unable to handle more complex situations. Wang *et al.* [2] improved YOLOv5/YOLOv8s by introducing the EMA efficient multi-scale attention mechanism, and adopting the OTA loss function and an enhanced IoU. It increased the object detection mAP by 2.50%. They also optimized the RRT algorithm, achieving a target recognition success rate of over 90.00% and a system positioning error of less than 4.00 mm. Although improvements have been made through multi-strategy fusion methods, the traditional RTT algorithm suffers from low path planning efficiency, slow search speed, and poor path quality. Wang *et al.* [3] enhanced YOLOv3 by adding 4 × sampling to the object detection network, increasing the mAP by 0.098%. Fang *et al.* [4] integrated the CA attention mechanism into YOLOv8s, improving the model's mAP by 2.30% compared with the original version. However, the generalization capability of their algorithm model remained insufficient. Li *et al.* [5] developed a YOLO-Light object detection model based on the YOLOv8 architecture. They incorporated a LightHead detection head, designed a C2f-MBConv module, and integrated the ECA attention mechanism, achieving an object detection accuracy of 93.00%. Mao *et al.* [6] proposed an ECASC_YOLO object detection model based on the YOLO framework by replacing the original C2f module with C2f_mk and incorporating the ECA attention mechanism, resulting in an object detection accuracy of 80.20%. However, their experiments did not account for external disturbance factors, resulting in weak anti-interference capability and reduced recognition accuracy. Li *et al.* [7] improved the DeepLab V3+ algorithm by incorporating a multi-scale channel attention mechanism and introducing the Focal loss function, achieving a recognition precision of 95.40%.

To address the aforementioned issues of weak anti-interference capability and insufficient recognition accuracy in existing algorithms, in this study, an improved YOLOv8s-based potato identification algorithm was proposed, aiming to achieve efficient and precise potato identification. An enhanced YOLOv8s object detection model was constructed to improve the algorithm's anti-interference capability and identification success rate and ensure that the accuracy and efficiency of the algorithm meet operating requirements in non-standard environments.

Potato Identification Model Based on Improved YOLOv8s

Dataset construction

The dataset in this study was collected using a Nikon D90 DSLR camera, yielding a total of 1 210 original potato images. To enhance the generalization capability and robustness of the constructed model, data augmentation was applied to the collected potato images. The YOLOv8 integrated a complete data enhancement process by default, which can be finely regulated through the

Received: April 7, 2025 Accepted: June 9, 2025

Qing ZHAO (2001 –), male, P. R. China, master, devoted to research about intelligent agricultural equipment.

* Corresponding author.

data. yaml file. After processing the original data with the above methods, the final dataset consisted of 8 160 images.

The collected images were annotated using the LabelImg software in YOLO format to label target categories and pixel coordinate information. The dataset was then divided into training, validation, and test sets, with 70.00% of the total samples randomly allocated as the training set, 20.00% as the validation set, and 10.00% as the test set. There was no overlap between the divided datasets. The labels were categorized into four classes: machine_damage, natural_damage, potato, and soil_on_potato.

Improved YOLOv8 and model training process

The model was trained using an Nvidia GTX 1650 graphics card, and the operating system was Windows 11. Python 3.8.2 was adopted as the programming language, and the PyTorch deep learning framework was employed. In this study, the model parameters for YOLOv8 were set as follows: from 0 to 10 epochs, the learning rate ranged from 0.001 to 0.000 5, and it was the rapid convergence phase. From 10 to 50 epochs, the learning rate ranged from 0.000 5 to 0.000 1, representing the fine-regulating phase. Beyond 50 epochs, the learning rate ranged from 0.000 1 to 0.000 01, representing the micro-adjustment phase for preventing overfitting. The batch_size was set to 1.

Replacing backbone network of model To optimize the backbone network, this study selected the InceptionNeXt backbone network. InceptionNeXt is a backbone network that combines the multi-scale feature fusion capability of the Inception structure with the efficient computation characteristics of the NeXt network. It draws on the multi-path design of Inception, and extracts features in parallel through convolutional kernels of different sizes to capture multi-scale target information. An adaptive weight allocation module was introduced to dynamically adjust the contribution of each branch based on the input features, suppressing irrelevant background noise. Redundant computations were reduced through cross-layer connections, improving gradient transmission efficiency. The original YOLOv8s backbone network's limitation lies in its single-scale dominance by CSPDarknet53, where deep layers tend to focus on large target features, causing small target information to be easily lost. The FPN-PAN structure lacks dynamic weight adjustment, leading to insufficient adaptability in complex scenes. The integration method first replaced CSPDarknet53, and adopted InceptionNeXt as the new backbone while retaining YOLOv8s's Neck and Head structures. The feature pyramid was enhanced by introducing dynamic gating to the FPN to optimize multi-scale feature fusion weights.

Introducing attention mechanism The CBAM dual attention mechanism was incorporated into the backbone network to enhance channel and spatial attention, improving the model's attention on occluded potato features. The CBAM structure consists of two parts: channel attention, which calculates inter-channel correlations through global average pooling and fully connected layers to generate channel weight matrices that amplify important channel information, and spatial attention, which computes inter-pixel correlations through convolutional layers to produce spatial weight matrices that enhance target region information. Integrating CBAM into YOLOv8s enables the model to focus more on key features, reduces noise, and improves detection accuracy for small targets. Compared with other attention mechanisms, CBAM has a simpler structure with lower computational quantity, and thus, it can maintain models' lightweight nature and is suitable for resource-constrained devices. In complex scenarios, the attention mechanism of CBAM can enable models to more accurately identify target areas.

Results and Analysis

Experimental results

The loss function of the improved YOLOv8s after 200 epochs of training is shown in Fig. 1.

In the initial phase, the training loss (train/loos) shows a rapid decline as the model quickly learns basic features, with localization and classification errors decreasing sharply. During the intermediate phase, the loss exhibits fluctuating convergence due to the cosine annealing learning rate adjustment strategy causing oscillations. In the final phase, the loss stabilizes with smooth convergence as it approaches the lower bound, indicating the model tends to converge. box_loss≈0.50, cls_loss≈0.40, indicating that the positioning accuracy is better than classification. The validation loss (val/loss) initially shows a rapid decline, indicating synchronized reduction in validation error with the training set. During the intermediate phase, an inflection point appears. The lowest point of val/box_loss is around 0.30, meaning the model achieves optimal generalization in localization, while the lowest point of val/cls_loss is around 1.30, indicating slightly weaker generalization performance in classification compared with localization.

To evaluate the accuracy of potato object detection under different experimental conditions, the captured images were divided into three scenarios: potatoes with mud on the surface, overlapping potatoes, and low-light conditions. As shown in the figure below, the detection accuracy for potatoes with mud on the surface was 97.10%. In the case of overlapping potatoes, the detection accuracy was 93.30%. Under low-light conditions, the detection accuracy was 89.70%.

To avoid code instability and conflicts caused by extensive modifications to the main branch, a progressive improvement strategy implementing only one enhancement at a time and verifying the results was adopted. The mean average precision (mAP) of the model was recorded after each improvement. The results are shown in Table 1.

Table 1 Progressive improvement test

Improvement direction	Oringinal YOLOv8s	Improved YOLOv8s
Backbone network optimization	86.90%	95.20%
Attention mechanism addition	–	A 2.10% overall improvement

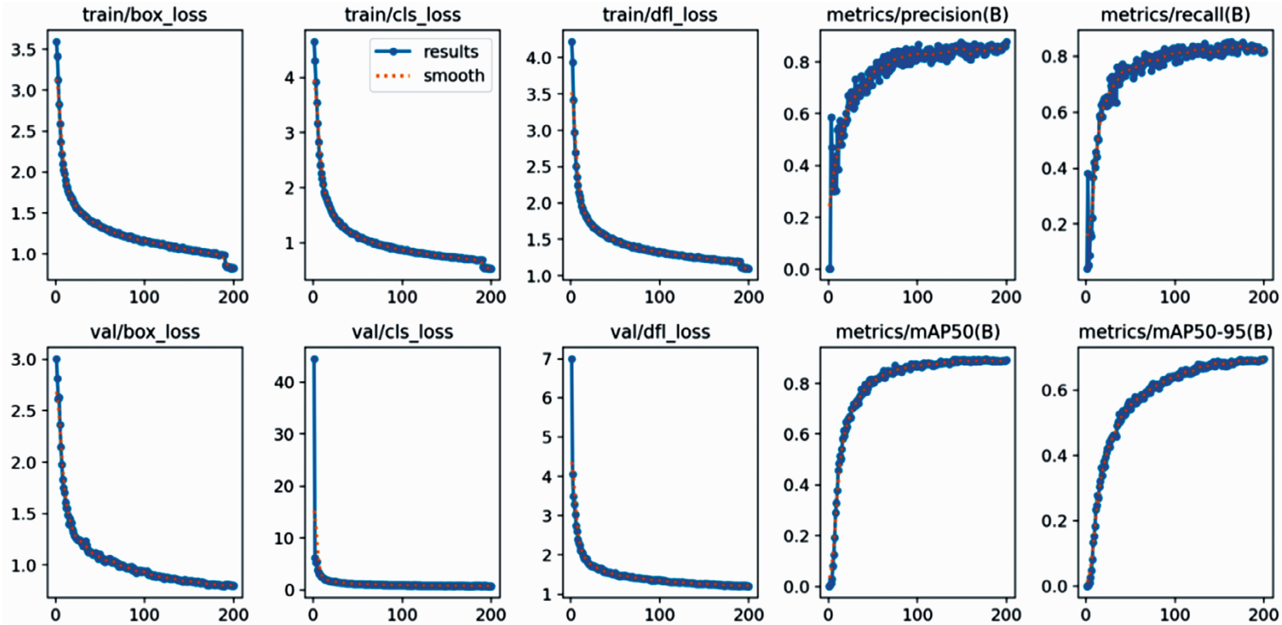

Fig. 1 Loss function of model training

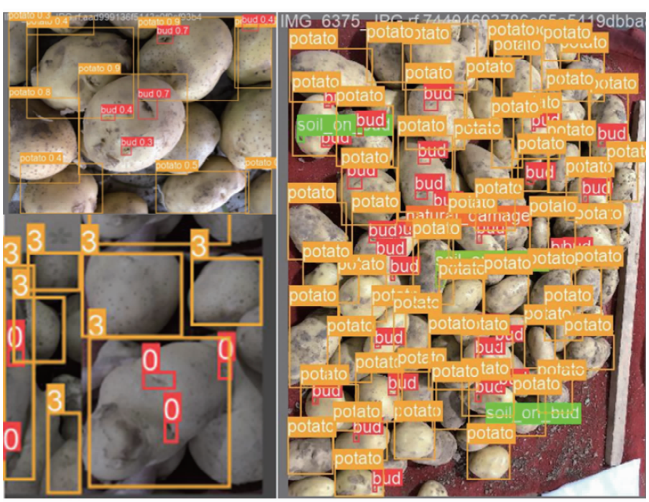

Fig. 2 Potatoes under different conditions

As shown in Table 1, the optimization of the backbone network increased the detection mean average precision (mAP) of the improved YOLOv8s by 8.30% compared with the original YOLOv8s. After introducing the CBAM dual attention mechanism, a 2.10% overall improvement was further achieved.

Comparative analysis

To validate the superiority of the improved YOLOv8s algorithm over other object detection algorithms, comparative experiments were conducted with the original YOLOv8s, SSD, Faster R-CNN, and YOLOv5 under the same dataset and runtime environment. Performance indices were evaluated and compared with the improved YOLOv8s. Precision (P), recall (R), F1-Score, mean average precision (mAP), single-image processing time and training convergence speed were used as evaluation criteria.

As shown in Table 2, the improved YOLOv8s algorithm achieved a precision of 94.55%, a recall of 85.34%, and an F1-score of 87.37% in potato identification. Compared with the original algorithm, the improved version increased precision by 7.40%, recall by 2.71%, and F1-score by 2.56%. The average processing time per image improved by 0.12 s, while the required training epochs for achieving the same precision decreased from 100 to 70 epochs.

Table 2 Performance comparison of different algorithms

Method	P//%	R//%	F1//%	mAP//%	Single-image processing time//s	Number of training convergence epochs//epochs
YOLOv8s	87.15	82.63	84.81	86.92	0.99	100
Improved YOLOv8s	94.55	85.34	87.37	95.29	0.87	70
SSD	74.78	70.84	72.67	83.55	1.44	130
Faster R-CNN	78.49	75.22	76.74	85.24	1.35	120
YOLOv5	85.44	80.34	83.12	85.62	1.05	110

Conclusions and Discussion

In this study, the problems of insufficient detection precision and high miss detection rate for small targets in potato identification

(Continued on page 80)

National Undergraduate Life Sciences Competition (2025, Scientific Inquiry Category). According to data statistics, compared with graduates of agronomy majors in previous years, students who have undergone the ideological and political education in the *Crop Breeding* course have developed a renewed understanding of the importance and necessity of agronomy majors. An increasing number of them are willing to pursue careers in agriculture, engage in agricultural research, and continue their further studies after graduation. It demonstrates that the ideological and political education in the *Crop Breeding* course has transformed students' perspectives and strengthened their sense of urgency and mission to contribute to rural revitalization.

Conclusions

The ideological and political construction in the *Crop Breeding* course should continuously refine its teaching content and methods through both teaching and learning. We must fully leverage the ideological and political elements of the course and deepen these aspects in alignment with societal needs. Through the study of the *Crop Breeding* course, students should achieve a "three-in-one" outcome integrating course knowledge, working ability, and social emotional values. This course aims to cultivate high-quality talents who study, understand, and cherish agriculture, thereby contributing to China's rural revitalization efforts.

References

[1] LIU HY, GUO XT, YAN DC, *et al.* Practice and thinking on ideological and political construction for courses of agronomy majors[J]. Shandong Jiaoyu (Gaojiao), 2020, 10: 31–32. (in Chinese).
[2] SHU YH, WANG JW, ZHANG JE. Exploration on curriculum ideological and political teaching mode and method of specialized courses in agronomy: A case study of "Agricultural Ecology"[J]. China University Teachin, 2022: 1–2. (in Chinese).
[3] WANG JC, YAO LR, SI EJ, *et al.* Thinking and exploration of student-centered curriculum ideological and political teaching under the background of rural revitalization strategy: Taking gansu agricultural university as an example[J]. Journal of Chifeng University: Philosophy and Social Science Chinese Edition, 2021,12(42): 75–78. (in Chinese).
[4] CHI HB. Rooting in the western part of Jilin and running a normal applied university that people are satisfied with[N]. Jilin Daily, 2021–12–15. (in Chinese).
[5] TANG R, ZHOU P, ZHAO ZX, *et al.* The application of literature introduction method in practice of "curriculum ideology and politics" in the course of *Crop Breeding*[J]. Journal of Yunnan Agricultural University: Social Science, 2021, 15(3): 156–161. (in Chinese).
[6] DENG JH. A wild rice plant rewrites the history of rice breeding in the world[N]. Hainan Daily, 2009–11–4. (in Chinese).
[7] FU TD, YANG GS, YANG XN, *et al.* Discovery, research and utilization of Polima cytoplasmic male sterility in Brassica napus[J]. Progress in Natural Science, 1995, 5(3): 287–293. (in Chinese).
[8] JIA SR. Progress in cotton genetic engineering in China[J]. Journal of Agricultural Science and Technology, 2000, 2(2): 18–20. (in Chinese).
[9] LI ZS. A new wheat hybrid variety—Xiaoyan 6[J]. Journal of Shanxi Agricultural Sciences, 2025, 53(2): 1–2. (in Chinese).
[10] DAI QH, QIAN Y, GU JB, *et al.* Ideological and political education in engineering courses in agricultural universities: Analysis and path[J]. China Agricultural Education, 2021, 22(3): 28–35. (in Chinese).

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU

(Continued from page 75)

tasks were addressed by proposing an improved YOLOv8s-based potato detection algorithm, and its efficiency and reliability were validated through simulation experiments. First, at the visual perception level, the introduction of the InceptionNeXt backbone network and CBAM dual attention mechanism constructed a multi-scale feature fusion framework, significantly enhancing target detection capability in complex scenarios. The improved YOLOv8s achieved an mAP of 95.29% on the dataset, representing an 8.37% increase over the original model, and the detection accuracy for small targets (AP_S) reached 98.70%. This study provides a high-precision and robust solution for agricultural automation equipment, extendable to applications such as fruit and vegetable sorting and industrial part assembly, advancing smart manufacturing technology toward practical implementation.

References

[1] SUN XD. Design and implementation of binocular vision system for sanitation robot based on YOLOv5s[D]. Xi'an: Xi'an University of Technology, 2024. (in Chinese).
[2] WANG K. Research on grasping technology of vision-guided manipulator based on ROS platform[D]. Harbin: Harbin University of Commerce, 2024. (in Chinese).
[3] WANG XQ, WANG X, YANG ZY, *et al.* Research on robot positioning and grasping based on improved YOLOv3[J]. Combined machine tool and automatic machining technology, 2024(7): 26–30. (in Chinese).
[4] FANG B, LIU F. Object recognition and grabbing detection based on improved YOLOv8s[J]. Manufacturing Technology & Machine Tool, 2024(12): 164–171. (in Chinese).
[5] LI AQ, MIN QQ, ZHOU QH. Grasping method of lightweight robot based on deep learning[J]. Equipment Manufacturing Technology, 2024(12): 1–5. (in Chinese).
[6] MAO K, FU DX, SAI QY, *et al.* Automatic capture of single-axis UAV based on target detection algorithm[J]. Control Engineering of China, 2024. (in Chinese).
[7] LI CX. Research on vision-based object detection and robot arm grasping[D]. Nanjing: Nanjing University of Information Science and Technology, 2024. (in Chinese).

Editor: Yingzhi GUANG Proofreader: Xinxiu ZHU