POTATO APPEARANCE DETECTION ALGORITHM BASED ON IMPROVED YOLOv8
/
基于改进 YOLOV8 的马铃薯外观品相检测算法

Huan ZHANG[1)], Zhen LIU[1)], Ranbing YANG[1,2)], Zhiguo PAN[*1)], Zhaoming SU[1)], Xinlin LI [1)], Zeyang LIU [1)], Chuanmiao SHI [1)], Shuai WANG[1)], Hongzhu WU[3)]

[1)] College of Electrical and Mechanical Engineering, Qingdao Agricultural University, Qingdao/ China
[2)] College of Mechanical and Electrical Engineering, Hainan University, Haikou/ China
[3)] Qingdao Hongzhu Agricultural Machinery Co., Ltd., Qingdao/ China
Tel: +8615318715305; E-mail: peter_panzg@163.com
Corresponding author: Zhiguo Pan
DOI: https://doi.org/10.35633/inmateh-74-76

Keywords: *Classification of potato species, Automatic sorting, Target detection, YOLOv8, MobileNetV4*

ABSTRACT
To meet the demands for rapid and accurate appearance inspection in potato sorting, this study proposes a potato appearance detection algorithm based on an improved version of YOLOv8. MobileNetV4 is employed to replace the YOLOv8 backbone network, and a triple attention mechanism is introduced to the neck network along with the Inner-CIoU loss function to accelerate convergence and enhance the accuracy of potato appearance detection. Experimental results demonstrate that the proposed YOLOv8 model achieves precision, recall, and mean average precision of 91.4%, 87.7%, and 93.7% respectively on the test set. Compared to YOLOv5s, YOLOv7tiny, and the original base network, it exhibits minimal memory usage while improving the mAP@0.5 by 1.1, 0.9, and 0.3 percentage points respectively, providing a reference for potato quality inspection.

摘要
为满足马铃薯分拣过程中对外观品相检测快速、准确的需求，本研究提出了一种基于改进 YOLOv8 的马铃薯外观品相检测算法。使用 MobileNetV4 替换 YOLOv8 主干网络，颈部网络引入三重注意力机制，Inner-CIoU 损失函数，加速收敛，提升马铃薯品相检测准确率。实验结果表明，提出的 YOLOv8 模型在测试集上的精确率、召回率和平均精度分别为 91.4%、87.7%、93.7%，相比较 YOLOv5s、YOLOv7tiny 和原基础网络，模型内存占用最少的同时的 mAP@0.5 分别提升了 1.1、0.9、0.3 个百分点，为马铃薯品质检测提供参考。

INTRODUCTION
Potatoes are the fourth largest staple crop in China. In 2023, the potato planting area in China was approximately 4,621 thousand hectares, with an annual production of fresh potatoes reaching 18.909 million tons. Due to the consumption habits of the Chinese population, fresh consumption accounts for more than 70% of the total potato consumption. Implementing scientific grading and classification strategies is crucial for the commercialization of fresh potatoes, meeting personalized needs, and enhancing economic value. Potato sprouting and surface damage are key factors affecting their market value and are important standards for determining potato grades. Therefore, timely and accurate detection of potato sprouting and surface damage is of significant importance in the grading, classification, and commercialization processes.

In most potato-producing regions of China, the sorting of fresh potatoes is primarily performed manually, resulting in high labor intensity and low efficiency. Therefore, the mechanization of rapid grading and sorting at production sites has become an inevitable trend. With the rapid advancement of computer technology, machine vision and deep learning have been widely applied in the agricultural sector. These methodologies are primarily founded on Convolutional Neural Networks (CNNs) and can be broadly categorized into two-stage detectors and one-stage detectors based on their processing workflows. They have significantly ameliorated the issues inherent in traditional object detection approaches.

In the realm of two-stage detection, the R-CNN series algorithms stand out as the most representative (*Girshick R. et al., 2014; Ren S. et al., 2016; He K. et al., 2017*). In the first stage, regions of interest (ROIs) are extracted from the input image, while in the second stage, each ROI undergoes object classification and bounding box regression. *Arshaghi et al., (2023)*, proposed a deep learning model based on CNN for potato defect detection and classification.

Through large-scale image data training, it can accurately identify disease spots, cracks, insect eyes, etc. Data augmentation technology is introduced to improve the generalization ability of the model. Experiments show that the performance of the model is significantly better than the traditional method, and the recognition accuracy, speed and stability are excellent in complex environments.

Md et al (2022) applied K-means segmentation technology combined with deep learning networks to predict and classify potato leaf diseases. The experimental results indicate that by utilizing the VGG16 model, the accuracy of their model reached 97%.

Geng et al., (2024) conducted research on an accurate and non-destructive detection method for potato sprouts that focuses on deformable attention. By embedding DAS (Deformable Attention Sampling), the approach enhances focus on relevant pixel image areas, thereby providing theoretical support for the non-destructive detection of sprouts in automated seed potato slicing.

Hatice et al., (*2022*) proposed a novel deep learning model called MDSCIRNet, which is based on the transformer architecture and deep separable convolutions for classifying potato leaf diseases. The experimental results show that this model improves the accuracy to 99.11% compared to the original model. *Zhao Yue et al., (2022),* employed the Faster-RCNN model to identify potato leaf diseases, achieving early and accurate diagnosis, thereby enhancing diagnostic efficiency and ensuring the yield and quality of potatoes. Although the R-CNN series models boast high accuracy, their large size and slow detection speed render them unsuitable for real-time detection. In contrast, single-stage detection models, known for their rapid detection and strong scalability, are more suitable for practical applications. Among these, the YOLO series of models stands out as the most renowned and widely applied, with numerous researchers employing the YOLO series within the agricultural sector.

Zhang W. et al., (2022), utilized the YOLOv3-tiny network model to detect potato seed tuber eyes, achieving high precision in detection. On the NVIDIA Jetson Nano platform, the model achieves a real-time detection speed of 40 FPS, meeting the demands of embedded systems and providing technical support for subsequent automated segmentation detection.

Dai et al., (2022), introduced an optimized YOLO v5 model, named DA-ActNN-YOLOV5, aimed at researching potato diseases across multiple regions and periods. This model achieved a remarkable recognition accuracy of 99.81% for early and late blight on the test set, representing a significant improvement of 9.22% in average accuracy compared to the original model.

ElMasry et al., (2012), developed a fast and accurate system based on machine vision, which constructs an image database and extracts geometric features and Fourier shape parameters. The study identified roundness, range, and four Fourier descriptors as key classification features. Experiments demonstrated that the system achieved an average correct classification rate of 96.5% on the training set and 96.2% on the test set. Additionally, it achieved a 100% accuracy rate in classifying the size of potatoes, showcasing its significant potential in the automatic detection and sorting of deformed potatoes.

Yue et al., (2024), employed the YOLOv8n network for citrus detection, enhancing the feature extraction network to achieve a detection accuracy of 96.9%. These methods have garnered notable success in agricultural target detection, providing robust support for the realization of intelligent agricultural machinery.

The aforementioned scholars conducted a feasibility analysis of the application of object detection technology in the agricultural sector, highlighting certain challenges when identifying surface defects on potatoes. They noted that while object detection technology is effective, it is limited by the types and accuracy of defect identification. The subtle differences between various potato varieties make it difficult to extract more effective features, and the system's stability needs improvement. To address these issues, this paper proposes an improved model based on YOLOv8, named MTI-YOLOv8.

MATERIALS AND METHODS

Image acquisition system

To capture images of potato appearance features, an online detection device for potato appearance grading was designed, as shown in Figure 1. This device primarily comprised a potato conveying system and an image processing system. The potato conveying system consisted of a motor, roller conveyor belt, and belt conveyor. The image processing system included a depth camera, dedicated camera lighting, a Jetson TX2 board, data cables, power supply, and detection algorithms. The detection algorithm was deployed on the Jetson TX2 board, with images captured by the camera being transmitted to the TX2 board via data cables.

Fig.1 – Schematic diagram of the depth camera

Dataset producing

On the experimental platform at Qingdao Agricultural University in Shandong Province, a depth camera was used to collect image data of potatoes. The experiment selected the early-maturing and high-yielding Holland 15 potato variety from Heilongjiang, Inner Mongolia, and the Central Plains region as the test subjects. A total of 560 potato samples were randomly selected and placed in different environments for a period of time to observe potential surface damage. To obtain representative samples of surface damage, potatoes with varying degrees and areas of damage were sampled periodically. Given the significant impact of image quality on the model's detection performance, the collected images were screened to remove blurry ones, resulting in a selection of 2120 images that met the experimental standards. These images were saved in jpg format with a resolution of 2592×2048 pixels. For each potato, two clear and unblurred photos—one of the front and one of the back—were taken to ensure the completeness and accuracy of the data.

In the process of annotating surface damage on potatoes, any mislabeling of defects could have resulted in non-defective areas being incorrectly marked as defective. This could have led the model to learn inaccurate information, thereby reducing its recognition accuracy and affecting its performance in practical applications. In this study, when annotating potato images, sprouting or rotting potatoes were labeled as "rot," green discoloration on the surface was labeled as "green," surface cracks were marked as "break," and whole, undamaged potatoes were labeled as "intact" to differentiate between categories, as illustrated in Figure 2.

a) ***b)*** ***c)*** ***d)***

Fig. 2 – Potatoes with different degrees of sprouting and damage

a) Rot potato; b) Green potato; c) Break potato; d) Intact potato

Image preprocessing

Using the open-source software Labelimg, potato surface damage information was annotated, and the annotations were uniformly saved in the PASCAL VOC dataset standard format. To further enhance the model's generalization performance and robustness, as well as to reduce the risk of overfitting, data augmentation techniques were applied. These techniques included horizontal and vertical flipping, brightness enhancement and reduction, motion blur, and contrast enhancement, among others. After augmentation, a total of 4,071 sample images were obtained. These images were randomly divided into training, testing, and validation sets in a 7:2:1 ratio, respectively resulting in 2,849 images for the training set, 815 images for the testing set, and 407 images for the validation set. The training set was used to train the model, the validation set was employed to adjust the model's hyperparameters and conduct preliminary evaluations of the model's capabilities, and the testing set was utilized to assess the model's detection accuracy and evaluate its generalization ability.

THE IMPROVEMENT METHOD OF POTATO QUALITY DETECTION

My question is: When using the traditional YOLOv8 model for classifying potatoes, several challenges are encountered, such as the insignificant features affecting recognition accuracy, slow detection speed, poor system stability, and large weight file sizes impacting efficiency. To enhance detection performance and reduce computational costs, it is necessary to make deep optimizations and improvements across multiple areas. Firstly, to address the issue of excessive model computational load, the MobileNetV4 network structure is introduced to reduce the size of the weight files and improve the model's lightness. In addition to introducing the MobileNetV4 network structure, the integration of a Triplet Attention module is proposed to emphasize attention computation across dimensional interactions. This enhances feature representation and improves efficiency, aiding in the precise localization and identification of detection objects. Furthermore, by incorporating the Inner-CIoU loss function, classification information is embedded into the IoU (intersection over union) calculation to enhance the model's bounding box regression performance. These optimizations and improvements are based on a comprehensive analysis of current challenges, aiming to tackle each issue methodically, thereby bringing substantial improvements to the detection of potato appearance quality. The improved model is named the MTI-YOLOv8 network, with its model structure depicted in Figure 3.

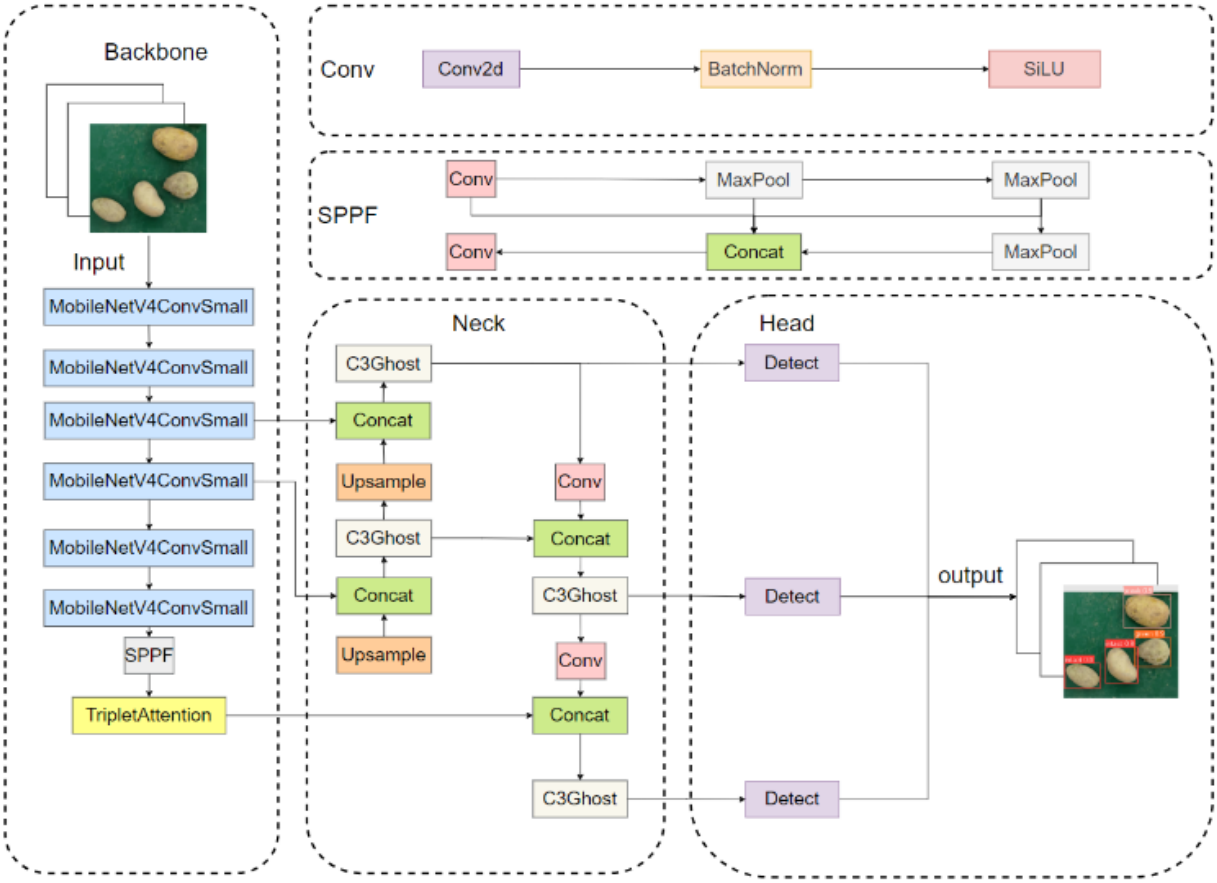

Fig. 3 – The structure of MTI-YOLOv8 model

MobileNetV4 backbone network

The original YOLOv8 model, while comprehensive in its capabilities, encounters challenges with a large number of parameters and accuracy that could be enhanced in the process of potato recognition. Experiments have revealed that the original YOLOv8 model exhibits high computational resource consumption and limited inference speed when handling complex and variable potato images. In response, an improvement strategy was implemented by replacing the backbone network of YOLOv8 with MobileNetV4, aiming to mitigate the issues of high model complexity and computational demand.

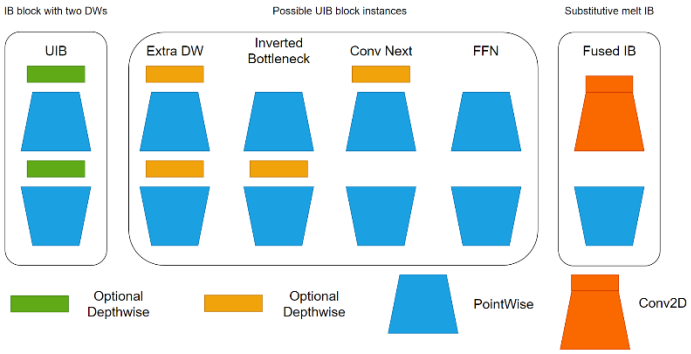

Fig. 4 – Universal Inverted Bottleneck (UIB) blocks

The improved model, by incorporating the MobileNetV4 backbone network, significantly reduces parameter count and computational load, enhancing inference speed and making it more amenable for deployment on edge computing and mobile devices. MobileNetV4 achieves lightweight performance while maintaining high efficiency through the innovative introduction of Universal Inverted Bottleneck (UIB) search blocks, Mobile MQA attention modules, and an optimized Neural Architecture Search (NAS) strategy. This enhances feature extraction capabilities and improves the accuracy and robustness of potato recognition. Additionally, the model demonstrates greater flexibility and reduced power consumption, eliminating the reliance on high-performance hardware and showcasing exceptional performance across diverse hardware platforms.

Triplet Attention module

In addressing the potato recognition task, the original YOLOv8 model, despite its comprehensive functionality, has shown significant shortcomings when faced with high environmental complexity, large weight files that impose storage burdens, and slow recognition speeds, as revealed by experiments. To mitigate these issues, this paper introduces the Triplet Attention mechanism as an improvement strategy.

In addressing the potato recognition task, the original YOLOv8 model, despite its comprehensive functionality, has shown significant shortcomings when faced with high environmental complexity, large weight files that impose storage burdens, and slow recognition speeds, as revealed by experiments. To mitigate these issues, this paper introduces the Triplet Attention mechanism as an improvement strategy. The Triplet Attention mechanism, with its unique tri-branch design, effectively captures cross-dimensional feature interactions between the spatial dimensions (H, W) and the channel dimension (C) of the input tensor. This endows the model with a more comprehensive image understanding capability, enabling it to precisely capture subtle and critical features in potato images. Compared to traditional attention mechanisms, Triplet Attention demonstrates significant advantages in computational efficiency by focusing on the interaction analysis among three elements rather than performing global computation on the entire input data. This greatly reduces the computational burden and time cost, making it especially suitable for the rapid processing of high-resolution potato images.

In complex scenarios where potatoes may exhibit different growth stages, morphological changes, and varying lighting conditions, the Triplet Attention mechanism enhances the model's ability to analyze cross-dimensional features. This significantly improves recognition accuracy and robustness, providing strong technical support for potato recognition tasks.

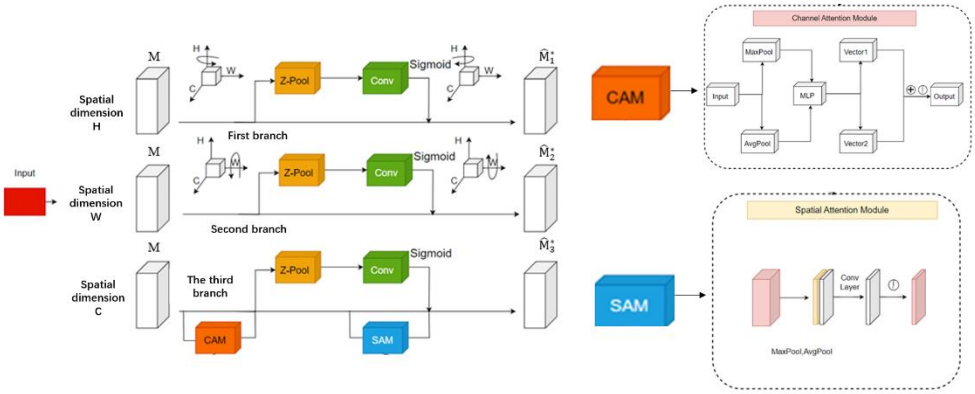

Fig. 5 – Triplet Attention structure

Triplet Attention consists of three parallel branches, and the input tensor $M \in \mathbf{R}^{C\times H\times W}$ is divided into three branches. In the first branch, the tensor M rotates 90°counterclockwise around the dimension H to obtain the rotation tensor $\hat{M}_1$. After pooling, the tensor shape is 2×H×C, and then the convolution operation is performed. The attention weight is generated by the Sigmoid activation function. Finally, the rotation is 90° clockwise around the dimension H, and the output tensor $\hat{M}_1^*$ is completed. The interaction between channel C and dimension H is completed. In the second branch, the tensor M rotates 90°counterclockwise around the dimension W, and the rotation tensor $\hat{M}_2$ is obtained.

After the pooling layer, convolution, and Sigmoid activation function, it rotates 90° clockwise around the dimension W, and the output tensor $\hat{M}_2^*$ is completed. The channel C interacts with the dimension W. In the third branch, after the tensor M passes through the pooling layer, convolution, and Sigmoid activation function, the output tensor $\hat{M}_3^*$ is obtained. Finally, the three tensors are averaged and aggregated to produce an output tensor y, as shown in equation (1).

$$y = \frac{1}{3}\left(\hat{M}_1\sigma(\psi_1(\hat{M}_1^*)) + \hat{M}_2\sigma(\psi_2(\hat{M}_2^*)) + M\sigma(\psi_3(\hat{M}_3^*))\right) \quad (1)$$

In the formula: $\sigma(\cdot)$ is the Sigmoid activation function; $\psi_1(\cdot), \psi_2(\cdot), \psi_3(\cdot)$ are standard convolutions.

Inner-CIoU loss function

In analyzing the process and patterns of bounding box regression, the unique nature of the bounding box regression problem was observed. During the model training phase, using smaller auxiliary boxes to calculate loss can positively impact the regression of high IoU samples, while low IoU samples exhibit the opposite trend. This addresses the increased difficulty in potato recognition caused by occlusion of loss features due to complex environments during potato harvesting.

To enhance recognition accuracy and accelerate the speed of bounding box regression, the Inner-IoU Loss method is innovatively proposed. This method introduces a scale factor, ratio, to regulate the generation of auxiliary boxes of different sizes, which are then used for loss calculation. By integrating Inner-IoU Loss into existing IoU-based loss functions, faster and more accurate regression results can be achieved, as illustrated in Figure 6.

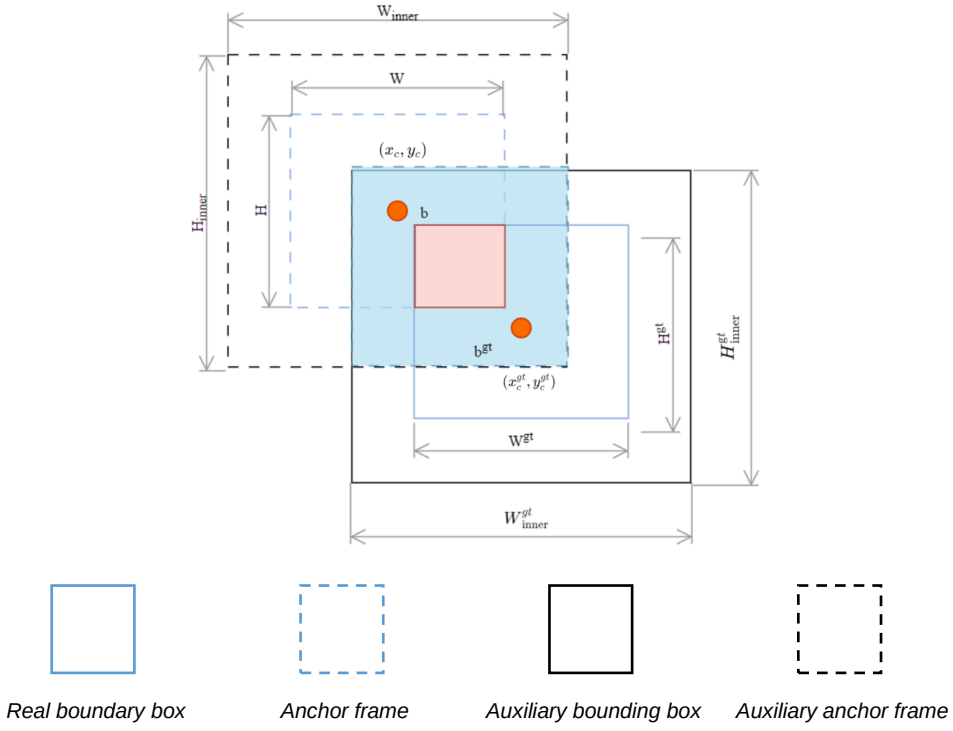

Fig. 6 – Inner-IoU loss function

Note : The real bounding box and the anchor box are represented by b^{gt} and b, respectively. The center point coordinates of the real bounding box is (x_c^{gt}, y_c^{gt}), the center point coordinates of the anchor box is (x_c, y_c), the width of the real bounding box is W^{gt}, the height of the real bounding box is H^{gt}, the width of the anchor box is W, the height of the anchor box is H, the width of the auxiliary bounding box is W_{inner}^{gt}, the height of the auxiliary bounding box is H_{inner}^{gt}, the width of the auxiliary anchor box is W_{inner}, and the height of the auxiliary anchor box is H_{inner}.

To enhance the generalization and convergence speed of the CIoU loss function, Inner-IoU is introduced, utilizing auxiliary bounding boxes to calculate loss and thereby accelerating bounding box regression. Inner-IoU adjusts the auxiliary boxes through a scaling factor, allowing it to flexibly adapt to different datasets and detectors, and enhancing generalization capabilities in complex scenes.

RESULTS AND ANALYSIS

Test environment and parameter configuration

All experiments conducted in this study were trained on the same server, with the experimental platform operating on Windows 11. The CPU used was an Intel(R) Core(TM) i5-13500HX, supported by 16GB of memory. The graphics card was an NVIDIA GeForce RTX 4060 with 6GB of VRAM. The development language employed was Python 3.10, while the software environment comprised Pytorch 2.0.1, CUDA version 12.1, and CUDNN version 8.9. All other parameters adhered to the official default settings of YOLOv8.

Evaluating indicator

In order to comprehensively evaluate the detection performance of the proposed model, the following evaluation indicators were used: Precision, Recall, mAP @ 0.5, mAP @ 0.5 ~ 0.95, number of model parameters, and number of model floating-point calculations.

The calculation formula is as follows:

$$P = \frac{T_p}{T_p + F_p} \tag{2}$$

$$R = \frac{T_p}{T_p + F_n} \tag{3}$$

$$AP = \int_1^0 P(R)dR \tag{4}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{5}$$

In the formula, T_P represents the number of positive samples correctly predicted by the model; F_P represents the number of negative samples that are predicted to be positive by the model error. F_N represents the number of positive samples that are incorrectly predicted as negative by the model. AP_i denotes the AP value of category i; n represents the number of data set categories.

Analysis of backbone network effectiveness

To evaluate the enhancement in performance of the MobileNetV4 backbone network within the MTI-YOLOv8 detection model, a comparative experiment was conducted. In this study, the conventional C3 and C2F detection backbones used in YOLO series models V5 and V8 were directly compared with the MTI-YOLOv8 model integrated with MobileNetV4 under identical testing conditions. The experimental data presented in Table 1 demonstrate that the MTI-YOLOv8 model, utilizing MobileNetV4 as its backbone network, exhibits significant advantages across several key performance metrics.

Specifically, the model achieved an accuracy of 91.4%, representing an improvement of 2.3% and 2.0% over the C3 and C2F detection backbones, respectively. In terms of recall, MobileNetV4 excelled with an impressive score of 87.7%, surpassing the C2F detection backbone by a notable 5.5%. Furthermore, regarding the mean Average Precision (mAP)—a critical metric for assessing overall detection accuracy—the MobileNetV4 model attained a score of 93.7%. This reflects a modest yet consistent increase compared to the 92.6% and 93.4% achieved by the C3 and C2F models, respectively.

In addition, regarding detection speed—an extremely important metric in practical applications—the model employing MobileNetV4 demonstrated faster processing speed, achieving 56 frames per second (FPS). This represents an increase of 6 FPS and 9 FPS compared to the C3 and C2F models, respectively, further validating its efficiency and superiority in the task of potato seedling detection. In summary, MobileNetV4 not only enhances detection accuracy and recall rate while maintaining a high mean average precision but also significantly accelerates detection speed, delivering a comprehensive performance improvement for the potato appearance detection model.

Table 1

Comparison test results of different detection backbones

backbone	P/%	R/%	mAP@0.5%	FPS
MobileNetV4	91.4	87.7	93.7	56
C3	89.1	92.6	92.6	50
C2F	89.4	82.2	93.4	47

Effectiveness analysis of attention mechanism

In our experiments on optimizing attention mechanisms for potato appearance detection tasks, it was observed that although CBAM, as an attention module integrating spatial and channel dimensions, has demonstrated excellent performance in other object detection studies, it did not significantly enhance accuracy in this study. Meanwhile, both the Global Attention Mechanism (GAM) and Dynamic Attention Transformer (DAT) contributed to considerable accuracy improvements; however, GAM resulted in increased model parameters and computational load, whereas DAT effectively alleviated these burdens. The EMA attention mechanism achieved a notable enhancement in detection accuracy, leveraging its ability to maintain the correlation between spatial and channel information and offering more stable feature representations. However, this improvement came with an approximate 9.09% increase in parameters and about a 20.5% rise in GFLOPs. In contrast, the Triplet Attention mechanism effectively enhanced feature representation by directly modeling inter-channel dependencies while maintaining lower model parameters and GFLOPs. This resulted in a 0.3% improvement in mAP@0.5, optimizing the model's detection accuracy. Therefore, for the task of potato appearance detection, the Triplet Attention mechanism demonstrates greater practicality and efficiency, making it a more suitable choice.

Table 2

Comparison of the results of different attention mechanisms

Attention Mechanism	mAP@0.5	mAP@0.5~0.95/%	Parameters /M	Floating Point Operation Quantity /G
——	93.4	73.1	3.01	8.1
Triplet Attention	93.7	72.8	2.61	7.0
CBAM	93.2	71.7	3.01	8.1
GAM Attention	93.3	71.7	4.62	9.4
EMA	94.2	73.3	12.10	28.6
DAT	93.3	71.4	2.46	6.8

Comparative test analysis of different algorithms

In order to verify the superiority of the research algorithm, the research algorithm MTI-YOLOv8 was compared with Faster R-CNN, Tood, YOLOv5s, YOLOv7-tiny, YOLOv8n, and another algorithm under the same conditions. The test results were shown in Table 3.

Table 3

Different model experiment results

Models	Recall/%	mAP@0.5	mAP@0.5~0.95/%	FLOPs/G	Parameters/MB	Model Size/MB
Faster R-CNN	65.1	71.5	38.4	137.20	41.36	315.46
Tood	67.1	87.8	54.7	126.96	32.03	244.13
YOLOv5s	88.8	92.6	70.9	15.8	7.02	14.5
YOLOv7-tiny	85.2	92.8	71.5	13.3	5.8	11.7
YOLOv8n	82.2	93.4	73.1	8.1	3.01	6.2
MTI-YOLOv8	87.7	93.7	72.8	7.0	2.61	5.54

Based on the experimental results presented in Table 3, it can be concluded that the two-stage object detection algorithm Faster R-CNN involves a relatively high number of floating-point operations and parameters, resulting in a larger model weight file. Therefore, it is considered that the two-stage object detection algorithm is not well-suited for the lightweight real-time detection requirements of this dataset.

YOLOv8 outperforms YOLOv7-tiny in terms of recall rate and mAP, while having fewer parameters and a smaller model size than other networks. The improved algorithm MTI-YOLOv8 enhances key metrics compared to the original YOLOv8n, with a 5.5% increase in recall rate and a 0.3% increase in average precision, while reducing the number of parameters and model size by 0.4 and 0.66, respectively. The experimental results clearly indicate that the MTI-YOLOv8 algorithm demonstrates excellent performance in potato object detection.

Fig. 7 – Comparison of model detection effects

Ablation test

In order to verify the effectiveness of each improved module of the algorithm in this study, the original model YOLOv8n was used as the baseline model, and the accuracy, recall rate, mAP @ 0.5, floating point operation amount and parameter number were used as evaluation indexes. The ablation test was carried out in different combinations of multiple improved modules, and the results were shown in table 4.

Table 4

Ablation experiment

MobileNetV4	Triplet Attention	Inner-CIoU	P/%	R/%	mAP@0.5%	Parameters	FLOPs(G)
×	×	×	89.4	82.2	93.4	3264396	12.1
√	×	×	89.2	81.2	89.3	1740491	9.6
√	√	×	91.0	89.5	90.4	1632198	8.1
√	√	√	91.4	87.7	93.7	1632198	5.9

According to the analysis of the experimental results in Table 4, it can be seen that by improving the backbone network of the original YOLOv8n model using the MobileNetV4 model, the lightweight structure significantly reduces the number of floating-point operations and parameters of the model. However, this reduction comes at the cost of a decrease in detection accuracy, with mAP@0.5% dropping by 4.1%. The ablation experiments also show that on the basis of the improved backbone network model, adding the Triplet Attention mechanism significantly enhances recognition accuracy, while slightly reducing the model's parameters and floating-point operations. This indicates that the Triplet Attention mechanism performs better in balancing model performance and accuracy improvement. When the Inner-CIoU loss function is introduced, it aids in bounding box regression, improving the model's localization ability and accelerating model convergence, resulting in a 3.3% increase in mAP@0.5%.

In conclusion, compared to the original YOLOv8n baseline network model, the enhanced MTI-YOLOv8 model, while showing a modest improvement in mAP@0.5, achieved significant enhancements of 2 percentage points in precision and 6.5 percentage points in recall. Moreover, the model's floating-point operations and parameter counts were reduced by 1,632,198 and 6.2 GFLOPs, respectively, strongly demonstrating the effectiveness and efficiency of the improved algorithm proposed in this study.

CONCLUSIONS

This study developed a potato appearance detection model based on the improved YOLOv8n network. Detection results on the same dataset indicated that the enhanced model achieved an accuracy of 91.4%, a recall rate of 87.7%, and an average precision of 93.7% in potato detection, surpassing the original YOLOv8n network and other detection models. Field experiments conducted on a laboratory-constructed testing platform demonstrated that the improved YOLOv8n model can effectively detect potatoes in motion within a conveyor speed range of 3 to 5 m/min, providing a novel solution for rapid potato detection.

ACKNOWLEDGEMENT

The author has been supported by the “National Key R&D Program of China” (Project number: 2023YFD2000904), the “Industry Technology System of Modern Agriculture” (Project number: CARS-09-P32), and the “Qingdao Science and Technology Huimin Demonstration Project” (Project number: 23-2-8-xdny-2-nsh).

REFERENCES

[1] Xia, Z., Pan, X., Song, S., Li, L. E., & Huang, G. (2022). Vision transformer with deformable attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 4794-4803.

[2] Arshaghi, A., Ashourian, M., & Ghabeli, L. (2023). Potato diseases detection and classification using deep learning methods. *Multimedia Tools and Applications*, 82(4), 5725-5742.

[3] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arxiv preprint arxiv:2004.10934*

[4] Dai, G., Hu, L., & Fan, J. (2022). DA-ActNN-YOLOV5: Hybrid YOLO v5 Model with Data Augmentation and Activation of Compression Mechanism for Potato Disease Identification. *Computational Intelligence and Neuroscience*, 6114061.

[5] ElMasry, G., Cubero, S., Moltó, E., & Blasco, J. (2012). In-line sorting of irregular potatoes by using automated computer-based machine vision system. *Journal of Food Engineering*, 112(1–2), 60-68. https://doi.org/10.1016/j.jfoodeng.2012.03.027

[6] Geng B, Dai G, Zhang H, Qi S, Christine DEW. (2024). Accurate non-destructive testing method for potato sprouts focusing on deformable attention. *INMATEH - Agricultural Engineering*: Vol.72, pp. 402-413. DOI: 10.35633/inmateh-72-36.

[7] Girshick, & Ross. (2015). [IEEE 2015 IEEE international conference on computer vision (ICCV) - Santiago, Chile (2015.12.7-2015.12.13)]. *IEEE international conference on computer vision (ICCV) - fast R-CNN*. 1440-1448.

[8] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* pp. 580-587.

[9] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In Proceedings of the IEEE international conference on computer vision. 2961-2969.

[10] Jocher, G., Stoken, A., Borovec, J., Changyu, L., Hogan, A., Diaconu, L., ... & Rai, P. (2020). YOLOv5: V3. 1-bug fixes and performance improvements. *Zenodo*. https://zenodo.org/records/4154370

[11] Li, N., Wang, M., Yang, G., Li, B., Yuan, B., & Xu, S. (2024). DENS-YOLOv6: A small object detection model for garbage detection on water surface. Multimedia Tools and Applications, 83(18), 55751-55771.

[12] Liu, Y., Shao, Z., & Hoffmann, N. (2021). Global attention mechanism: Retain information to enhance channel-spatial interactions. arxiv preprint arxiv:2112.05561.

[13] Nishad, M. A. R., Mitu, M. A., & Jahan, N. (2022). Predicting and classifying potato leaf disease using k-means segmentation techniques and deep learning networks. Procedia Computer Science, 212, 220-229.

[14] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., & Huang, Z. (2023). Efficient multi-scale attention module with cross-spatial learning. In ICASSP 2023-2023 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp.1-5. *IEEE*.

[15] Qin, D., Leichner, C., Delakis, M., Fornoni, M., Luo, S., Yang, F., & Howard, A. (2025). MobileNetV4: Universal Models for the Mobile Ecosystem. In *European Conference on Computer Vision*. pp. 78-96. Springer, Cham.

[16] Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

[17] Redmon, J. (2018). Yolov3: An incremental improvement. *arxiv preprint arxiv:1804.02767*.

[18] Reis, H. C., & Turk, V. (2024). Potato leaf disease detection with a novel deep learning model based on depthwise separable convolution and transformer networks. *Engineering Applications of Artificial Intelligence*, 133, 108307.

[19] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.

[20] Wang, C.Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 7464-7475.

[21] Yue, K., Zhang, P., Wang, L., Guo, Z., & Zhang, J. (2024). Recognizing citrus in complex environment using improved YOLOv8n (基于改进 YOLOv8n 的复杂环境下柑橘识别). *Transactions of the Chinese Society of Agricultural Engineering*, 40(8), 152-158.

[22] Zhang, W., Han Y., Huang, C., & Chen Z. (2022). Recognition method for seed potato buds based on improved YOLOv3-tiny. *INMATEH-Agricultural Engineering*, 67(2), pp.364-373

[23] Zhao Yue, Zhao Hui, Jiang Yongcheng, Ren Dongyue, Li Yang, Wei Yong. (2022). Potato Leaf Disease Detection Method Based on Deep Learning(基于深度学习的马铃薯叶片病害检测方法). *Journal of Chinese Agricultural Mechanization*, (10), 183-189. doi:10.13733/j.jcam.issn.2095-5553.2022.10.026.