ITC 3/55
Information Technology and Control
Vol. 55 / No. 3/ 2026
pp. 1467-1489
DOI 10.5755/j01.itc.55.3.44394

Agricultural Mechanized Picking and Sorting Technology Based on Object Detection and Improved MobileNet V1

Received 2026/02/26 | Accepted after revision 2026/06/03

HOW TO CITE: Zhu, Y., Tang, J., Jing, R., Zhao, X., Yang, D. (2026) Agricultural Mechanized Picking and Sorting Technology Based on Object Detection and Improved MobileNet V1. *Information Technology and Control*, 55(3), 1466-1488. https://doi.org/10.5755/j01.itc.55.3.44394

Agricultural Mechanized Picking and Sorting Technology Based on Object Detection and Improved MobileNet V1

Yupeng Zhu*
School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, China; School of Electronic Science and Technology, Hainan University, Haikou, 570208, China; e-mail: yupengj@outlook.com

Jun Tang
School of Mechanical & Electrical Engineering, Hainan Vocational University of Science and Technology, Haikou 571126, China; e-mail: 18756988312@163.com

Ru Jing
School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, China; e-mail: rufeng25@163.com

Xiaowei Zhao
School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, China; e-mail: zhaoxw_garfield@126.com

Darun Yang
School of Information Engineering, Hainan Vocational University of Science and Technology, Haikou, 571126, China; e-mail: 15519315102@163.com

Corresponding author: yupengj@outlook.com

With the rapid development of computer vision technology, intelligent and mechanized agricultural production will inevitably introduce visual image technology to replace manual picking and sorting work. A two-stage detection and sorting integrated framework based on feature reuse is proposed to meet the dual requirements

of high-precision target positioning and lightweight real-time classification for agricultural mechanized picking and sorting tasks. This framework uses the improved YOLOv8 as a spatial position perception module and enhances robustness to low-quality samples by introducing deep convolution and Wise IoU V3 loss function. At the same time, the adapted and improved MobileNet V1 is used as a lightweight mass sorting module, and Adam parameter optimization and Maxout activation function are adopted to improve classification efficiency. The results showed that in crop object detection in the experimental field, the mean detection accuracy of the YOLOV8+MobileNet V1 model robot was 0.954, and the average sorting error rate was 4.779%. The improved YOLOV8+MobileNet V1 model had an average sorting recognition accuracy of 0.983 and a picking rate of 0.933 in specific applications. The average picking time for each crop was 4.47 seconds, which was better than other models. The research findings denote that the intelligent crop picking and sorting method constructed by the research has high object detection and sorting accuracy, and the picking efficiency of the model is higher, which has good application prospects in mechanized agricultural production.

KEYWORDS: Crops, Object detection, MobileNet V1, Sorting technology, YOLOv8

1. Background

In the last few years, with the continuous growth of agricultural production sales and planting scale, the development of agricultural automation and modernization is the key to future intelligent construction. In the process of agricultural production, the picking and sorting of crops require high precision in visual object detection. Currently, intelligent detection equipment has low efficiency and high cost [15]. The shelf life of most agricultural products after harvesting is only 5-7 days, and manual sorting can easily cause damage and shorten the shelf life. Traditional sorting methods will be eliminated, and existing machine vision-based automatic sorting methods operate under ideal conditions, which have errors compared to practical applications. Additionally, increasing the amount of network layers increases computational costs, which is not conducive to agricultural modernization [6]. A core contradiction in the current field of agricultural visual computing is that high-precision object detection and sorting models often require expensive GPUs and large amounts of memory, making it difficult to deploy in embedded controllers of agricultural machinery. Although lightweight models meet resource constraints, their detection and sorting accuracy in complex agricultural environments is far from reaching practical levels. How to design a visual perception framework that achieves the optimal balance between accuracy, computational efficiency, and adaptability to agricultural scenes in object detection and sorting under embedded resource constraints? The key to answering this scientific question is to break through the linear trade off curve of traditional single model "precision efficiency" and explore the mechanism of collaborative decomposition and feature reuse of detection and sorting tasks.

Vijh et al. [26] introduced an improved feature extraction algorithm for plant leaf image classification. The algorithm determined the optimal visual word through a random feature extraction algorithm, represented feature occurrence or repetition using discrete bi complex wavelet transform, and finally used a multi-layer perceptron for classification. The experiment showed that the average accuracy (AA) of this method was 0.7552, which was superior to several other algorithms [26]. Ferrag et al. [11] reviewed the intrusion detection systems for agricultural 4.0 network security and classified them based on machine learning techniques. Meanwhile, public datasets and implementation frameworks were brought in, and the challenges and future research directions of intrusion detection in network security in Agriculture 4.0 were discussed [11]. Yeh et al. [31] employed the AlexNet-spatial pyramid pooling network to classify mango grades. The system contains user interface, object detection, image preprocessing, and fruit grading classification. The experimental results showed that this method was superior to the traditional AlexNet method [31]. Cheng [5] designed a smart production line for cherry tomatoes that uses artificial intelligence, the IoT and data analysis to detect and classify visual defects in 360°. With an

AA of 97.4% in practical applications, this method promoted the growth of intelligent agriculture by enabling defects to be detected quickly, ensuring the quality of sorting and improving crop management [5]. Donapati et al. [10] proposed a hybrid model that combines U-Net and CNN for seed detection and germination analysis, achieving effective results. In the experiment, the pixel accuracy of this method was 0.91, the intersection was 0.84, and the accuracy was 0.90 [10]. Singh et al. [25] focused on the integration of IoT sensor data collection, cloud software data processing, and machine learning algorithm data analysis, aiming to improve crop yield and sustainability. Simultaneously, it comprehensively explored the technological landscape and research level, compared and analyzed different machine learning algorithms, and proposed an integrated sensor and machine learning method for predicting crop types [25].

Yang et al. [29] proposed a parallel processing CNN accelerator grounded in an optimized version of the MobileNet architecture, which is used in machine vision and pattern recognition applications. The experimental findings denoted that the classification performance of this method on ImageNet reached 580.6 fps, with a power consumption of only 6.51 W [29]. Yang et al. [30] designed an automatic scene recognition method based on channel state information, suitable for vehicle to vehicle systems. This method used a lightweight CNN inspired by MobileNet to reduce computational complexity (CC) and model size, while compensating for performance loss through knowledge distillation. The findings denoted that the CC and model size of this method were only 30% of the original method, and the performance loss was limited [30]. Gava et al. [12] designed a multi-threaded implementation optimization for MobileNet to address the resilience issue of CNNs in industrial IoT. The findings denoted that multi-threaded execution increased the critical failure rate by 5 times, and some input images were more robust to silent data damage events [12]. Ryu et al. [22] proposed the Mobileware architecture of MobileNet, which uses channel fixed data streams and places distributed static random access memory near each processing element to achieve high bandwidth communication. This architecture had a throughput 1.4 × -29.5 × higher than traditional weight fixed architectures, and has been validated on the Xilinx ZCU102 FPGA evaluation board [22]. Wang et al. [27] proposed a cone-shaped microfiber Bragg grating spectrum demodulation method based on MobileNet CNN. The results showed that compared with five classic CNN models, MobileNet achieved 100% testing accuracy and reduced testing time by 20% -40% [27]. Mai et al. [18] designed a malware detection method based on Markov images and MobileNet, providing an economically efficient and high-performance solution, and innovatively analyzing the robustness of opcode sequences.

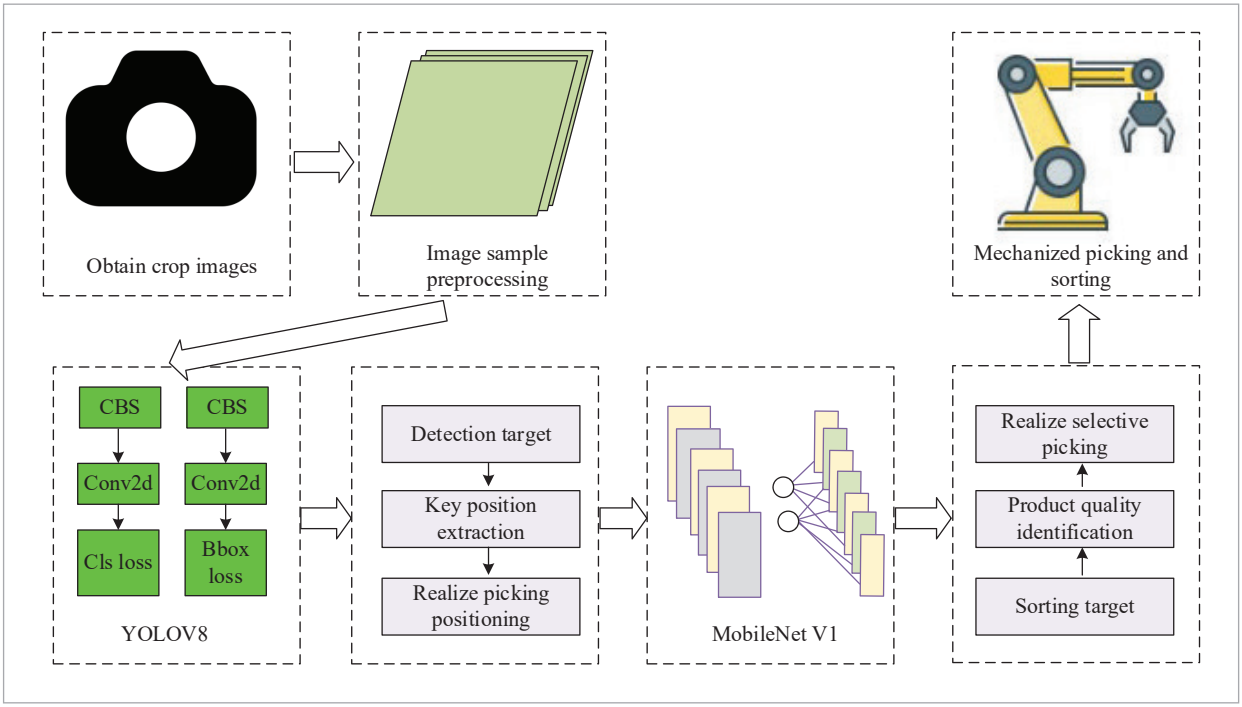
In summary, there are still two key gaps in the current research of computer vision in agricultural picking and sorting: firstly, most studies consider object detection and quality sorting as two independent stages, calling different models to execute sequentially, resulting in wasted computing resources and decreased real-time performance. Secondly, existing lightweight models (such as the MobileNet series) have insufficient positioning accuracy in complex agricultural environments, while high-precision models (such as the YOLO series) are difficult to meet the computational constraints of edge devices. To fill the gap mentioned above, the main contribution of the research is to first propose a detection sorting collaborative framework, which improves YOLOv8 for spatial position perception and MobileNet V1 for semantic quality discrimination. Both are used to complete target monitoring and sorting work on the experimental platform, significantly reducing overall computational latency. At the same time, the study verified the collaborative paradigm of "large model positioning+small model sorting" in the agricultural picking scene, and proved that the paradigm is superior to single model or non collaborative combination in accuracy and efficiency.

2. Research Design

2.1. Object Detection Technology for Mechanized Harvesting of Crops Based on YOLOv8

In agricultural production and marketing, the quality of crop products is closely related to the efficiency of picking and sorting. Mechanized production methods can greatly increase the efficiency of

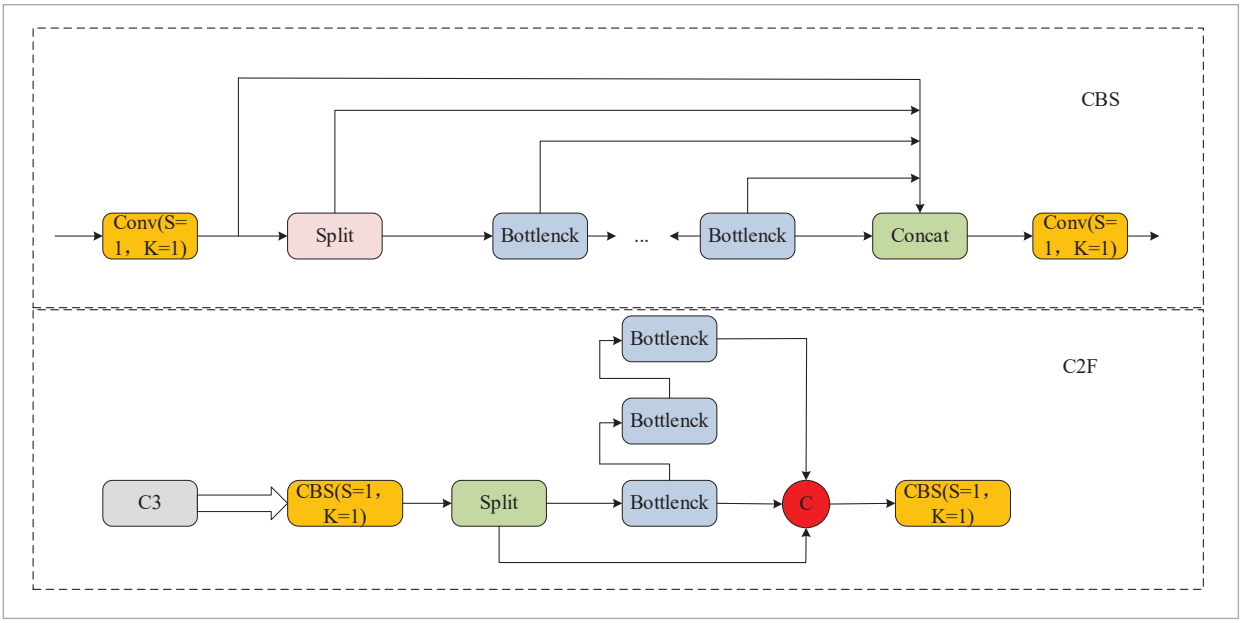
Figure 1
Framework of agricultural mechanized picking and sorting technology.



crop production, thereby extending the shelf life in subsequent marketing processes [4]. This study explores the optimization of crop target detection and sorting technology in agricultural mechanization picking processes to enhance their efficiency and accuracy of intelligent picking and sorting [20]. During the mechanized picking process, synchronously completing the sorting of agricultural products can further enhance the speed of agricultural intelligent production and reduce the cost of agricultural product damage. The YOLOv8 module not only completes the target bounding box regression, but also passes the intermediate layer feature map in real-time to the MobileNet V1 module, which only needs to perform lightweight quality classification on the located area without repeatedly extracting global features. This mechanism of task decoupling and feature reuse not only retains the spatial perception ability of large models, but also leverages the efficient classification advantages of small models. The framework of agricultural mechanized picking and sorting technology developed through research is shown in Figure 1.

As shown in Figure 1, the study first constructs a crop object detection model based on YOLOv8, preprocesses the collected crop images, and then inputs them into the YOLOv8 model to complete the object detection task. YOLOv8 is a newly developed object detection algorithm in the You Only Look Once (YOLO) series, which modernizes the object detection workflow by introducing multiple architectural and functional improvements. The YOLOv8 model uses a bottom-up spatial refinement path to downsample shallow features and fuse them with deep features, building multi-scale collaborative perception capabilities [14]. By using independent regression branches to predict the coordinates and aspect ratio of the target bounding box, the classification branch outputs the probability of the target category. The advantage of YOLOv8 is that it can effectively alleviate the parameter coupling problem between classification and localization tasks. The YOLOv8 model includes a backbone network structure, a Neck network structure, and a header network structure. The convolution module is the main operational module of each network structure, and the structures of the

Figure 2
Structure of CBS Convolution Module and C2F Connection Module.



CBS convolution module and C2F connection module are shown in Figure 2.

As shown in Figure 2, both the CBS convolution module and the C2F connection module rely on convolution operations to complete feature extraction. They are the core of the YOLOV8 model. CBS stands for Conv BN SiLU, which sequentially executes Convolution, Batch Normalization, and SiLU activation function. The input feature map is segmented along the channel dimension via C2F, followed by convolutional operations on each segmented feature map. In the output stage, the output feature map is formed by concatenating and integrating the results. This enables the model to capture multi-scale feature information, thereby enhancing its ability to detect small targets. Therefore, in the Neck module of YOLOv8, the up sampling or down sampling path for feature map fusion within the FPN/PAN structure of Neck and the deformable convolution introduced in the 19th layer of the model are used to enhance the detection accuracy of the model for small target crops. The convolution sampling position set is defined as the following Equation (1).

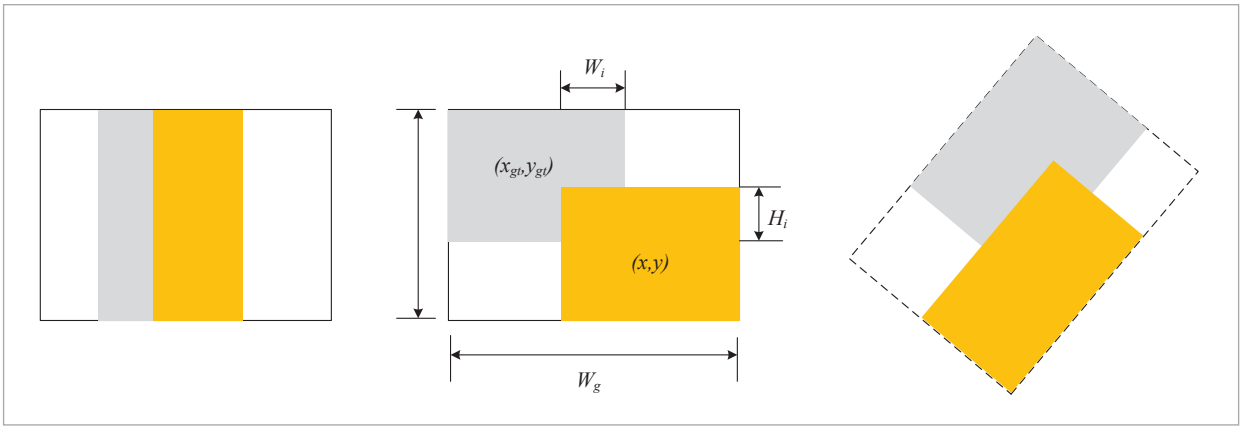
$$\theta = \{(-1,-1),(-1,0),...,(0,1),(1,1)\} \quad (1)$$

In Equation (1), θ represents sampling location point set, which contains a series of coordinate points representing the key feature positions in the crop feature map. Based on the convolution kernel θ, the output expression of the convolution module is denoted in Equation (2).

$$y(p_0) = \sum_{p_n \in \theta} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) . \quad (2)$$

In Equation (2), $y(p_0)$ represents the output at position p_0, w represents the weight, p_n represents all points in the convolution kernel operation, x represents the input, and Δp_n represents the offset. Through the operation of the above convolution module, the YOLOv8 model can adapt to the irregular structure of products in the recognition of crops with different shapes, and make small shape adjustments based on local information, thereby enhancing the ability to capture key position features of crops. In the YOLOv8 model, the LF of Bounding Box Regression (BBR) is crucial for object detection. Training YOLOv8 may lead to a decrease in the model's generalizxation ability due to the impact of various factors, such as weather and environment, on the quality of the sample during crop image shooting

Figure 3
The boundary box principle of WIoU's monotonic focusing mechanism.



[7]. However, the traditional YOLOv8 model cannot solve the problem of the influence of low-quality samples on model performance by using a loss function. Therefore, the study adopted Wise-IoU V3 (WIoU V3). WIoU is a bounding box loss function that uses a dynamic, non-monotonic focusing mechanism to address the negative impact of noisy samples on the model during object detection training. The boundary box principle of WIoU's monotonic focusing mechanism is shown in Figure 3.

As shown in Figure 3, WIoU evaluates anchor box quality by introducing outlier degree instead of IoU, while WIoUV3 allocates gradient gains grounded on anchor box quality. This method minimizes the gradient of low-quality samples while retaining the gradient of high-quality samples, avoiding overfitting of the model to low-quality data [1]. First, a distance attention mechanism needs to be constructed, and then this mechanism is obtained through a two-layer attention mechanism.

$$\begin{cases} L_{WIoUv1} = R_{WIoU} L_{IoU} \\ R_{WIoU} = \exp((x-x_{gt})^2 + (y-y_{gt})^2 / (W_g^2 + H_g^2)^*) \end{cases}. \quad (3)$$

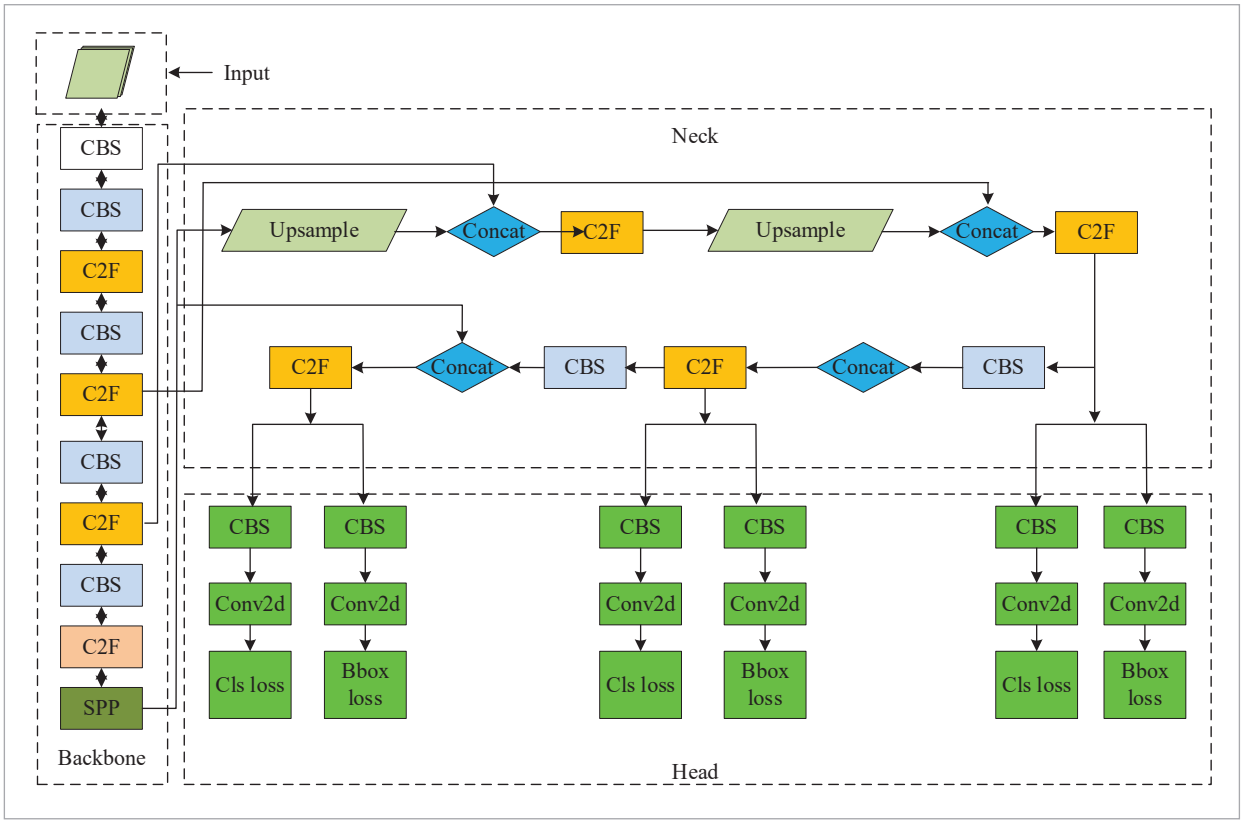
In Equation (3), L_{WIoUv1} represents the WIoU V1 LF, R_{WIoU} represents the significantly enlarged L_{IoU} of ordinary quality anchor boxes, and L_{IoU} represents the significantly reduced R_{WIoU} of high-quality anchor boxes. W_g and H_g denote the length and width of the smallest closed box. To prevent R_{WIoU} from generating gradients that hinder convergence, W_g and H_g must be separated from the computational graph, with $(W_g^2 + H_g^2)^*$ representing the separation operation. (x, y) represents the target box; (x_{gt}, y_{gt}) represents the anchor box boundary. After obtaining the WIoU V1 loss function, to strengthen the object detection classification effectiveness of YOLOv8, the monotonic focus coefficient constructed by WIoU V2 is used to minimize the contribution of abnormal samples to the loss value during model training. The monotonic focus coefficient is shown in Equation (4).

$$L_{WIoUv2} = \left(\frac{L_{IoU}^*}{\bar{L}_{IoU}}\right)^{\gamma} L_{WIoUv1}, \gamma > 0. \quad (4)$$

In Equation (4), L_{WIoUv2} represents the increased WIoU V2 LF, $\bar{L}_{IoU}$ represents the mean of L_{IoU}, and γ represents the parameter that controls the strength of the monotonic focusing coefficient. On the basis of the WIoU V2 LF mentioned above, a new mechanism of dynamic non-monotonic focusing coefficient is designed to dynamically adjust the model's attention to different quality samples, thereby constructing the WIoU V3 LF, as denoted in Equation (5).

$$\begin{cases} \beta = \dfrac{L_{IoU}^*}{\bar{L}_{IoU}} \in [0, +\infty) \\ L_{WIoUv3} = \gamma L_{WIoUv1} \\ \gamma = \dfrac{\beta}{\delta \alpha^{\beta-\delta}} \end{cases}. \quad (5)$$

Figure 4
The structure of the YOLOv8 model.



In Equation (5), β represents the outlier degree, which is employed to show the relationship between the model's predicted outlier degree and the true outlier degree. The bigger the number, the smaller the outlier predicted by the model, and the easier it is to sort the sample. δ and α represent parameters that control the magnitude of gradient gain. To make the YOLOv8 object detection model focus on samples with high outlier levels, the WIoU V3 LF assigns smaller gradient gains to these samples to prevent them from having a significant negative impact on model training. Based on the above methods, the structure of the YOLOv8 model is denoted in Figure 4.

As shown in Figure 4, the backbone network of YOLOv8 generates feature maps of different scales through layer by layer convolution and downsampling. The neck network adopts the Path Aggregation Network Feature Pyramid Network (PAN-FPN) structure, which first transfers deep semantic features to shallow layers through a top-down path, and then fuses shallow positional information to deep layers through a bottom-up path, thus achieving multi-scale feature enhancement. The head network adopts a Decoupled Head design and consists of two independent branches: the classification branch outputs the probability of the target category, and the regression branch predicts the coordinates and aspect ratio of the bounding box. This decoupling design effectively alleviates the parameter coupling problem between classification and localization tasks. The pseudocode for the improved YOLOV8 object detection model is shown in Table 1.

Therefore, to address challenges such as geometric deformation of targets, leaf coverage, fruit overlap, and dense small targets in agricultural detection scenarios, the traditional convolution operation of YOLOv8 model is replaced by deep convolution, and the WIoU loss is improved by recalibrating outliers.

Table 1
Pseudocode for the improved YOLOV8 object detection model.

Input: Crop image I, anchor box set A, real bounding box B_gt
Output: Predict bounding box B_pred, category probability P
1. Feature extraction
   For each convolutional layer:
      Perform deep convolution on the input feature map
      Perform point by point convolution
   Output feature map F
2. Multi scale feature fusion
   Downsampling shallow features and concatenating them with deep features → F_fused
3. Boundary box regression and classification:
   Predicting offset Δ and category score S from F_fused
4. Calculate WIoU V3 loss
   For each anchor box a ∈ A:
   Calculate the outlier degree β=(IoU_Loss (a, B_gt) - μ)/σ
   Calculate the dynamic gradient gain r=β ^ δ/(β ^ δ+ε)
   WIoU V1 = 1 - IoU(a, B_gt) + R(a, B_gt)
   L_box = r * WIoU V1
   Total loss L_total=L_box+L_CLS (cross entropy)+L_dfl
6. Reverse propagation updates network parameters
7. Return B_pred, P

2.2. Crop Sorting Method Based on Improved MobileNet V1

After optimizing the YOLOv8 object detection model by improving the LF, the study applies the model to mechanized harvesting activities of crops. However, in the mechanized production process of crops, there are some normal agricultural products, immature agricultural products, and spoiled agricultural products [17]. If all harvested crops are placed together, it may increase the overall spoilage rate of agricultural products and reduce the preservation time. Therefore, the research considers integrating agricultural product sorting into the process of mechanized harvesting of crops to achieve efficient mechanized agricultural production [32]. The study selects the MobileNet V1 model for crop sorting. MobileNet V1 is a lightweight CNN whose core idea is to reduce computation time through depthwise separable convolutions (DSCs). The deep separable convolution architecture of MobileNet V1 is chosen as the basis for the study, but to adapt to the feature reuse mechanism, the network is structurally adjusted and named MobileNet V1. This module no longer receives cropped blocks from the original image, but instead receives feature blocks from the YOLOv8 neck feature map that have been cropped by RoIAlign and channel adapted, thus skipping the underlying feature extraction layer of the original network and directly performing DSC operations [19]. In deep convolution operations, each input channel is independently filtered in space to ensure that the amount of output channels matches the amount of input. In point-by-point convolution, 1 × 1 convolution is used to combine information from different channels and adjust the output dimensions [13]. Firstly, the CC of standard convolution can be expressed as Equation (6).

$$N_s = D_K \cdot D_K \cdot M \cdot N \cdot D_W \cdot D_H . \quad (6)$$

In Equation (6), N_s represents the standard convolution computation, D_K means convolution kernel size, D_W and D_H indicate the length and width of the input and output feature maps, and M and N denote the amount of channels in the input and output feature maps, respectively. The calculation method for the CC of DSC is shown in Equation (9) below, based on the calculation method of Equation (7).

$$N_d = D_K \cdot D_K \cdot M \cdot D_W \cdot D_H + M \cdot N \cdot D_W \cdot D_H . \quad (7)$$

In Equation (7), N_d represents the computational cost of DSC. The MobileNet V1 consists of 28 layers, with the first layer being a 3x3 standard convolution, followed by stacking 13 sets of DSC modules. Comparing Equations (6)-(7), when using a 3 * 3 convolution kernel, the computational cost of DSC is reduced to 1/9 of that of standard convolution. In the improvement of the MobileNet V1 model, the optimization of model parameters is the key factor affecting model performance. The conventional parameter optimization method for MobileNet V1 model is the exponential weighted moving average method to calculate gradients, which can stabilize the learning process while avoiding the problem of gradient explosion or disappearance [28]. However, traditional models are prone to the problem of kernel holes during deep convolutional layer training. Therefore, the research adopts the Adaptive

Moment Estimation (Adam) algorithm to optimize the parameters of the MobileNet V1 network model. Adam can automatically adjust the learning rate based on the gradient to handle the sparse part of the gradient, avoid gradient sparsity and oscillation problems, and strengthen the training efficiency and performance of the model [2]. In the Adam algorithm, the gradient calculation formula is first given as Equation (8).

$$g \leftarrow + \frac{\nabla_{\varphi} \sum_{i-1}^{m} L(f(x_i;\varphi) y_i)}{m}. \tag{8}$$

In Equation (8), g means the gradient cumulative variable used to store the first-order moment estimate (FOME) of the gradient. m means the amount of samples, x_i means the sample with the number i, and y_i represents the corresponding target for that sample. φ represents the model parameters that need to be updated through the Adam algorithm, represents the LF, and f represents the predicted output of the model. ∇_{φ} represents the gradient of the parameter, which is the derivative of the LF with respect to the parameter. Based on Equation (8), the FOME and second-order moment estimate (SOME) are updated, as shown in Equation (9).

$$\begin{cases} s \leftarrow \tau_1 s + (1-\tau_1) g \\ r \leftarrow \tau_2 r + (1-\tau_2) g \odot g \end{cases}. \tag{9}$$

In Equation (9), s and r denote the FOME and SOME, respectively, while τ_1 and τ_2 represent the decay rates of the FOME and SOME. $g \odot g$ represents the square of the elements of the gradient cumulative variable. Afterwards, the deviation between the first-order moment estimation and the SOME is corrected according to Equation (9), as shown in Equation (10).

$$\begin{cases} \hat{s} \leftarrow s / (1-\tau_1^t) \\ \hat{r} \leftarrow r / (1-\tau_2^t) \end{cases}. \tag{10}$$

In Equation (10), $\hat{s}$ and $\hat{r}$ represent the modified FOME and SOME, respectively. Afterwards, the parameter update amount is calculated based on the corrected deviation 93, as shown in Equation (11).

$$\Delta\varphi = -\eta \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}. \tag{11}$$

In Equation (11), $\Delta\varphi$ represents the amount of model parameter updates, η represents the learning rate, and δ represents a very small constant used to stabilize the denominator value. Based on this, the improved MobileNet V1 network adopts the Adam algorithm to effectively adjust the learning rate by combining the advantages of momentum and adaptive learning rate, thereby accelerating convergence speed and improving model performance.

Afterwards, research is conducted to strengthen the activation function of the MobileNet V1 network model. The conventional MobileNet V1 network model uses ReLU activation function, but ReLU may introduce unstable factors during the training process, resulting in gradient vanishing and the parameters of related neurons may no longer be updated. Therefore, the study uses Maxout activation function instead of ReLU activation function. Maxout activation function avoids the problem of neuron deactivation that may be caused by ReLU by selecting the maximum value in the input, thereby strengthening the training efficiency and performance of the model. The neural computation method for the Maxout hidden layer is shown in Equation (12).

$$h_i(x) = \max_{j \in [1,k]} z_{ij}. \tag{12}$$

In Equation (12), k represents the Maxout hidden layer parameters, x is the hidden layer output, and z_{ij} represents the Maxout hidden layer input. i represents neuron number, j represents hidden layer number. The input expression is shown in Equation (13).

$$z_{ij} = x^T W_{ij} + b_{ij}. \tag{13}$$

In Equation (13), W represents weight, b represents bias term, and T represents transpose operator. After optimizing the activation function, the MobileNet V1 network structure is studied to minimize the CC and complexity of the model while ensuring information capacity and accuracy [8]. The optimized MobileNet V1 network structure is denoted in Fig-

Figure 5
Optimized MobileNet V1 network structure.

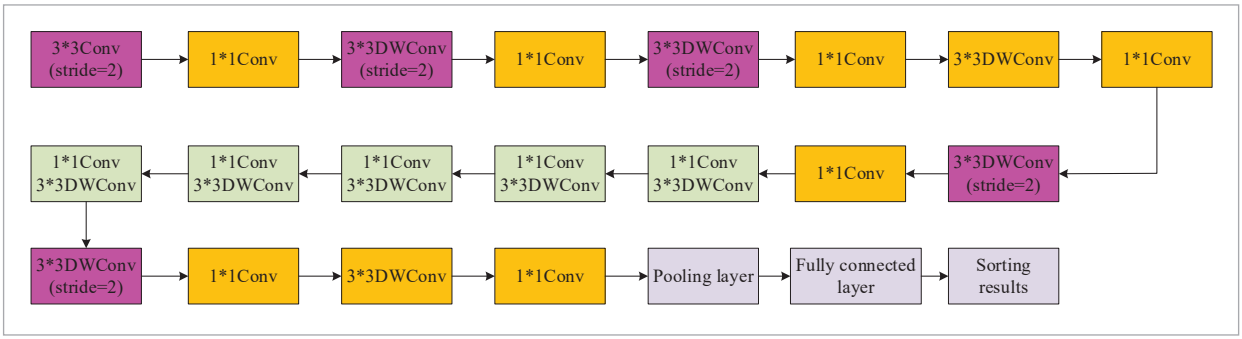
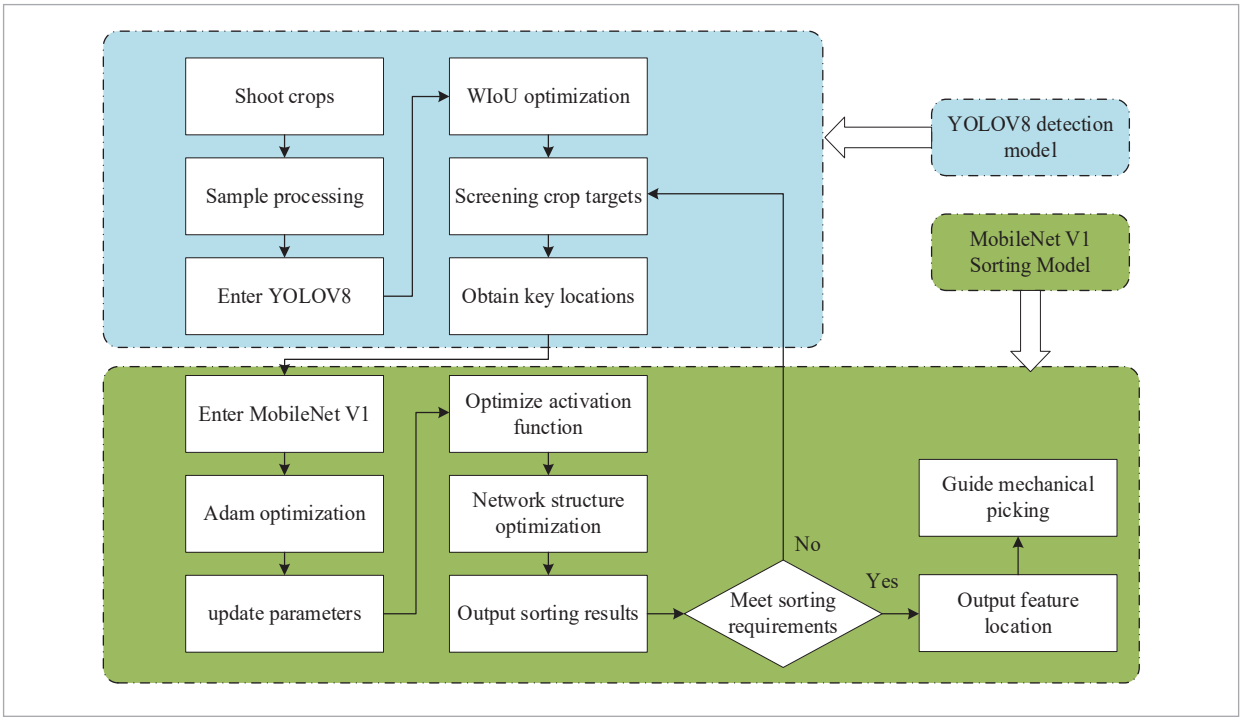


Figure 6
Intelligent mechanized picking and sorting model integrating improved YOLOV8 model and improved MobileNet V1 network model.



ure 5. From Figure 5, the study specifically adjusts the CC of the first few layers by using deep convolution and pointwise convolution operations to improve efficiency and performance [3]. After structural optimization, the size and CC of the MobileNet V1 network model are reduced, which means that the optimized model is more suitable for efficiently completing product sorting work in agricultural production environments [16]. Finally, the construction process of the intelligent mechanized picking and sorting model that integrates the improved YOLOV8 model and the improved MobileNet V1 network model is denoted in Figure 6.

From Figure 6, after deep convolution operation and WIoU LF optimization, the improved YOLOV8 model completes the target detection of crops and the extraction of picking area positions. After completing object detection, the efficiency and accuracy of the

MobileNet V1 network model in sorting work are improved through Adam optimization, activation function optimization, and network structure optimization operations. Through the improved MobileNet V1 network model, crop targets can be sorted. Unqualified sorting results will be removed from the crop samples, while qualified agricultural product feature maps will serve as guidance information for subsequent mechanized harvesting. In the traditional concatenated model, MobileNet V1 needs to re extract features from the original image blocks, resulting in a large amount of redundant computation. A feature reuse bridging module is designed to directly transfer the deep semantic feature maps generated in the YOLOv8 backbone network to MobileNet V1, replacing its previous convolutional operations. Specifically, for each candidate region detected by YOLOv8, feature blocks corresponding to the region are cropped from the output feature map of YOLOv8's Neck network using RoIAlign operation. These feature blocks are then fed into the feature adapter of MobileNet V1, where their channel numbers are mapped to the required dimensions of MobileNet V1. The adapted features are directly used as input for MobileNet V1, replacing the forward propagation of the original image blocks. This mechanism eliminates the need for MobileNet V1 to re extract underlying visual features.

3. Results and Analyses

3.1. Performance Analysis of Improved YOLOV8 Model in Crop Object Detection

To assess the performance of the enhanced YOLOV8 model and MobileNet V1 network model in agricultural product target detection and sorting, image samples were taken from five crops in an agricultural experimental base of a university. The five crops were grapes, cherry tomatoes, mushrooms, cucumbers, and mangoes. During the growth process of crops, 5000 images of harvestable samples and 500 images of non-harvestable samples were taken for each crop. A total of 27500 images were captured to form a sample dataset, which was broken into a training dataset and a testing dataset in a 7:3 ratio. The test dataset contains 19250 samples of mixed agricultural product types, while the test dataset contains 8250 samples of mixed agricultural product types. The study completed performance testing by constructing an improved YOLOV8 model, with specific experimental parameters and environmental conditions denoted in Table 2.

Based on the experimental environment and model parameters in Table 1, the study first compared the performance differences of the YOLOV8 model before and after improvement. The study analyzed the target detection accuracy and loss of different models in the training dataset, and the specific research findings are denoted in Figure 7.

Figure 7(a) showcases that the accuracy of the pre-optimized YOLOv8 model fluctuated between 0.7 and 0.83, while the optimized YOLOv8 model achieved accuracy ranging from 0.85 to 0.95. In Figure 7(b), the loss of the optimized YOLOV8 model approached zero after approximately 100 iterations. In contrast, the loss of the pre-improved YOLOV8 model gradually converged to zero only after exceeding 100 iterations. Experimental results demonstrated that the optimized YOLOV8 network model exhibited higher efficiency than the pre-optimized model. Subsequently, the study compared the mAP of the traditional YOLOv8 model and the improved model during training. The specific test findings are denoted in Figure 8.

From Figure 8(a), as the number of iterations increased, the mAP_{50} of the optimized YOLOV8 model gradually approached 1, and the mAP_{50} curve of the optimized YOLOV8 was relatively smooth. The mAP_{50} curve of the pre-optimized YOLOv8 model fluctuated significantly before 200 iterations, and ultimately the mAP_{50} of the traditional YOLOV8 model was lower than that of the optimized model. In Figure 8(b), as the number of iterations increased, the $mAP_{50:90}$ curve of the optimized YOLOV8 model became relatively smooth and gradually approached 0.9, while the $mAP_{50:90}$ curve of the pre-optimized YOLOv8 model fluctuated greatly. At the same time, the $mAP_{50:90}$ of the pre-optimized YOLOv8 model was ultimately lower than that of the optimized model. The findings denote that the optimized YOLOV8 network model has higher object detection accuracy than pre-optimized model. Afterwards, the study conducted ablation experiments on the test dataset to analyze the object detection effectiveness of the YOLOV8 model under different optimization measures. The specific outcomes are denoted in Table 3.

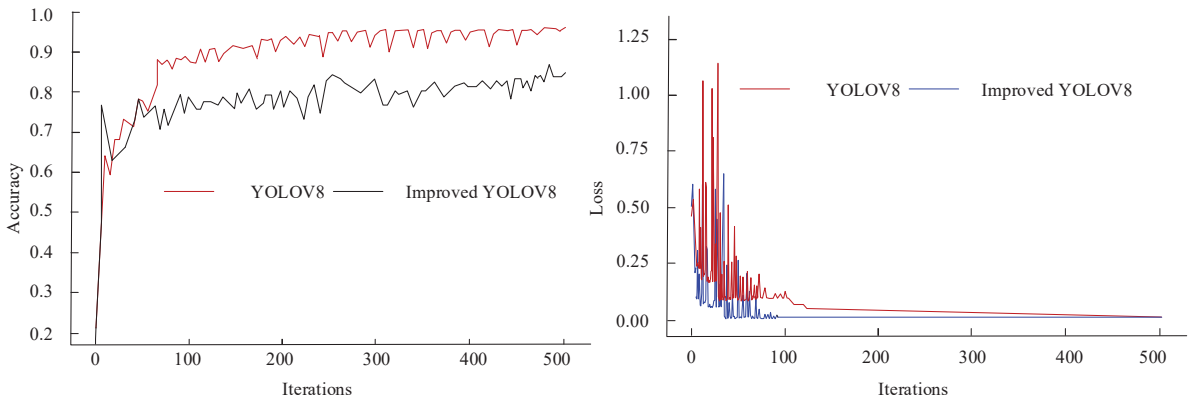
Table 2

Experimental parameters and environmental conditions.

Environmental parameters		Experimental data parameters	
Operating system	Windows 10	Learning rate	0.001
Central processing unit	Intel(R) Core(TM) i9-12900KF	First-order moment estimation attenuation rate	0.9
Memory	64G	Second order moment estimation attenuation rate	0.999
Graphics processing unit	NVDIA GeForce RTX3090	constant	10-8
Convolution operation programming	CUDA v11.6	image size	640*640
Model building platform	Pytorch 1.13.1	Input batch	32
Model training platform	Python 3.8	training iterations	500

Figure 7

The accuracy and loss of object detection during the training process of different models.

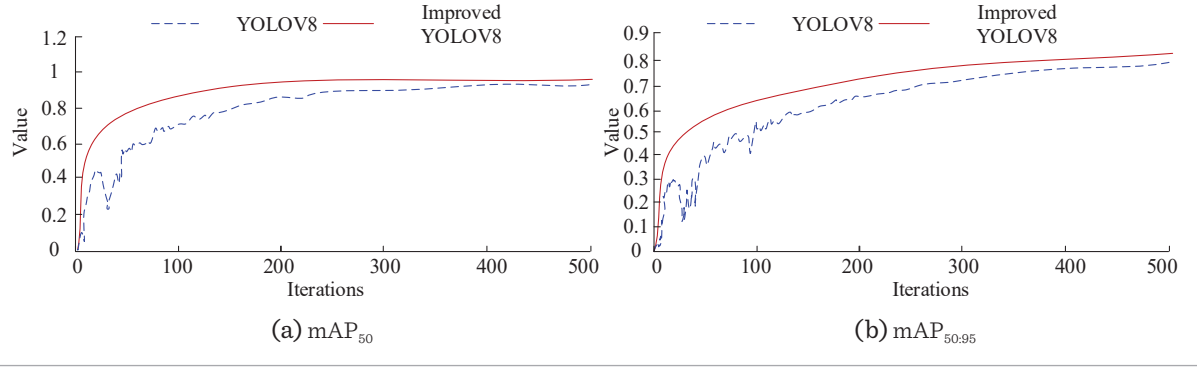


(a) The difference in accuracy between traditional and improved models

(b) The difference in losses between traditional and improved models

Figure 8

Comparison of mAP results between traditional and improved YOLOV8 during training process.



(a) mAP_{50}

(b) $mAP_{50:95}$

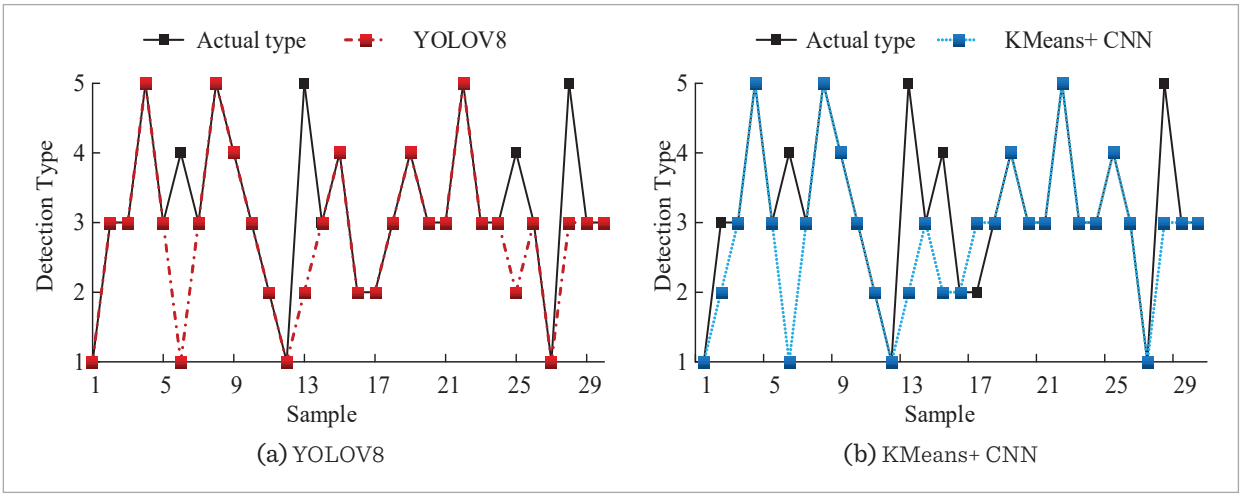
Table 3
Target detection performance of YOLOV8 model under different optimization measures.

YOLOV8		Accuracy			mAP_{50}		
Deformable Convolution	WIOU	Crop testing	Key position detection	Average	Crop testing	Key position detection	Average
NO	NO	0.956	0.925	0.9405	0.952	0.883	0.9175
YES	NO	0.954	0.963	0.9585	0.982	0.961	0.9715
NO	YES	0.965	0.944	0.9545	0.971	0.939	0.955
YES	YES	0.976	0.967	0.9715	0.984	0.964	0.974

In Table 3, WIOU represents LF optimization. From the data in Table 3, the AA of traditional YOLOV8 was 0.9405, and the average mAP_{50} was 0.9175 After optimizing the deformable convolutional operation, the AA of the YOLOV8 model was 0.9585, and the average mAP_{50} was 0.9715. After optimizing the LF, the AA of the YOLOV8 model was 0.9545, and the average mAP_{50} was 0.955. Finally, under the dual optimization of deformable convolutional operation and LF, the AA of YOLOV8 model was 0.9715, and the average mAP_{50} was 0.974. The experimental results show that the deformable convolutional operation and LF optimization used in the study can greatly enhance the crop target detection effectiveness of the YOLOV8 model. The sampling points of deep convolution can cover the main texture areas of the fruit. The geometric deformation problem of target objects in agricultural scenes was solved by calculating the average Euclidean distance between sampling points and target contours. WIoU V3 significantly suppressed the gradient contribution of high-quality samples and extreme noise samples, while enhancing the gradient of moderately occluded samples, thereby improving object detection performance. Finally, the study randomly selected 30 crop images during the growth process from the test sample dataset and compared the crop species recognition performance of different detection methods. The study compared the improved YOLOV8 model with the K-means algorithm optimized CNN (KMeans+CNN). The specific test findings are denoted in Figure 9.

The vertical axes 1-5 in Figure 9 represent five types of crops: grapes, cherry tomatoes, mushrooms, cucumbers, and mangoes. From Figure 9(a), among the 30 sample tests, the improved YOLOV8 model only

Figure 9
Performance of different algorithms in object detection and recognition.



(a) YOLOV8 (b) KMeans+ CNN

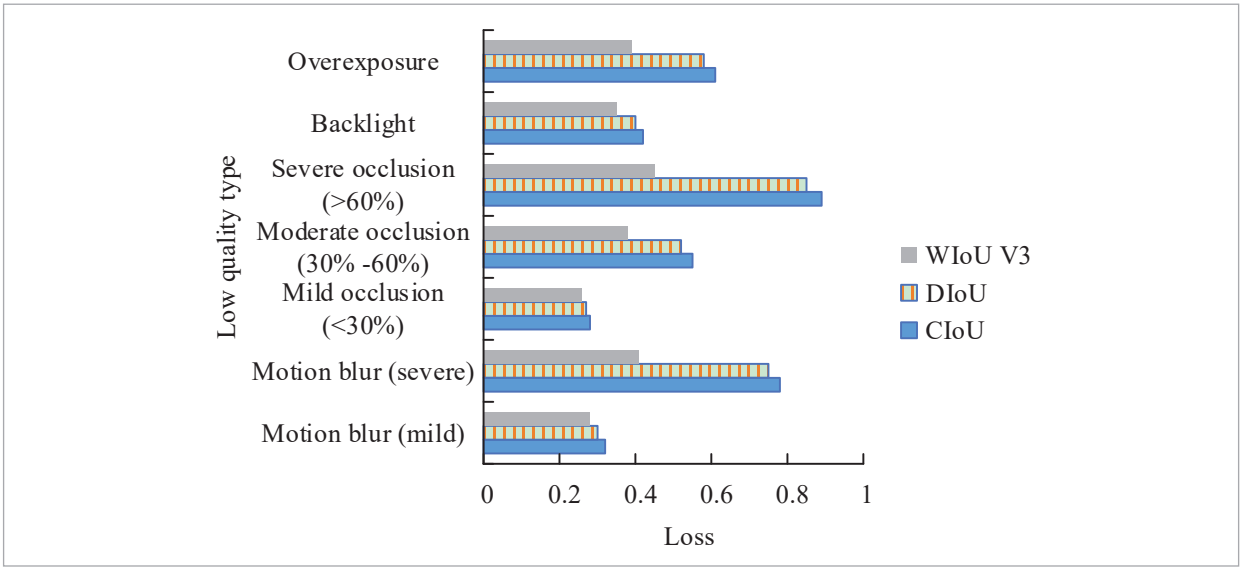
Table 4

mAP_{50} performance of various models in agricultural detection tasks.

Model	mAP_{50}	Parameter count (M)	FPS
YOLOv12m	93.1%	/	/
RT-DETRv2-X	93.6%	/	/
SE-YOLO	96.1%	2.05	/
YOLOv7-LEES	95.6%	35.5	116.3
CTDA	95.3%	6.7	154.1
Improved YOLOv8	98.8%	3.41	64.7

Figure 10

Statistics of loss values of different loss functions on low-quality samples.



had 3 recognition errors, with a recognition accuracy of 90%. From Figure 9(b), KMeans+CNN made a total of 6 recognition errors, with a recognition accuracy of 80%. The experiment showcases that the improved YOLOV8 model can maintain the classification performance in object detection and recognition in crop mechanized harvesting, and is more reliable than other models.

To further verify the competitiveness of the improved YOLOV8+MobileNet V1 model proposed in this study in practical agricultural application scenarios, a comparative analysis was conducted with state-of-the-art methods published in recent years in terms of object detection accuracy and sorting recognition performance. Due to differences in datasets, hardware environments, and evaluation metrics used in different studies, direct horizontal comparisons pose certain challenges. However, by synthesizing multiple publicly available experimental results, the performance level of this method can be evaluated within a relatively consistent framework. In the object detection task, SE-YOLO, YOLOv7-LEES, and the CTDA model improved based on YOLOv8 were selected as the latest SOTA comparison methods. At the same time, the public evaluation results of the latest detectors such as YOLOv12m and RT-DETRv2-X were included as references. Table 4 summarizes the mAP50 performance indicators of each model in agricultural testing tasks.

From Table 4, the method achieved an AA of 98.8% on the mixed dataset of five types of crops, which is higher than other comparative methods. At the same time, the research method performed competitively in the framework of lightweight design, with a parameter count of 3.41M lower than the 6.7M of the CTDA model. In terms of FPS indicators, the YOLOv8 improved model proposed in the study had only 64.7, which was lower than other models, but could meet the needs of real-time detection.

To verify the superiority of WIoU V3 on low-quality agricultural samples, this study trained YOLOv8 models using CIoU, DIoU, and WIoU V3 in the test set, and compared the loss distribution and detection performance of the three models on three types of low-quality samples (motion blur, leaf occlusion, and abnormal lighting). The specific results are shown in Figure 10.

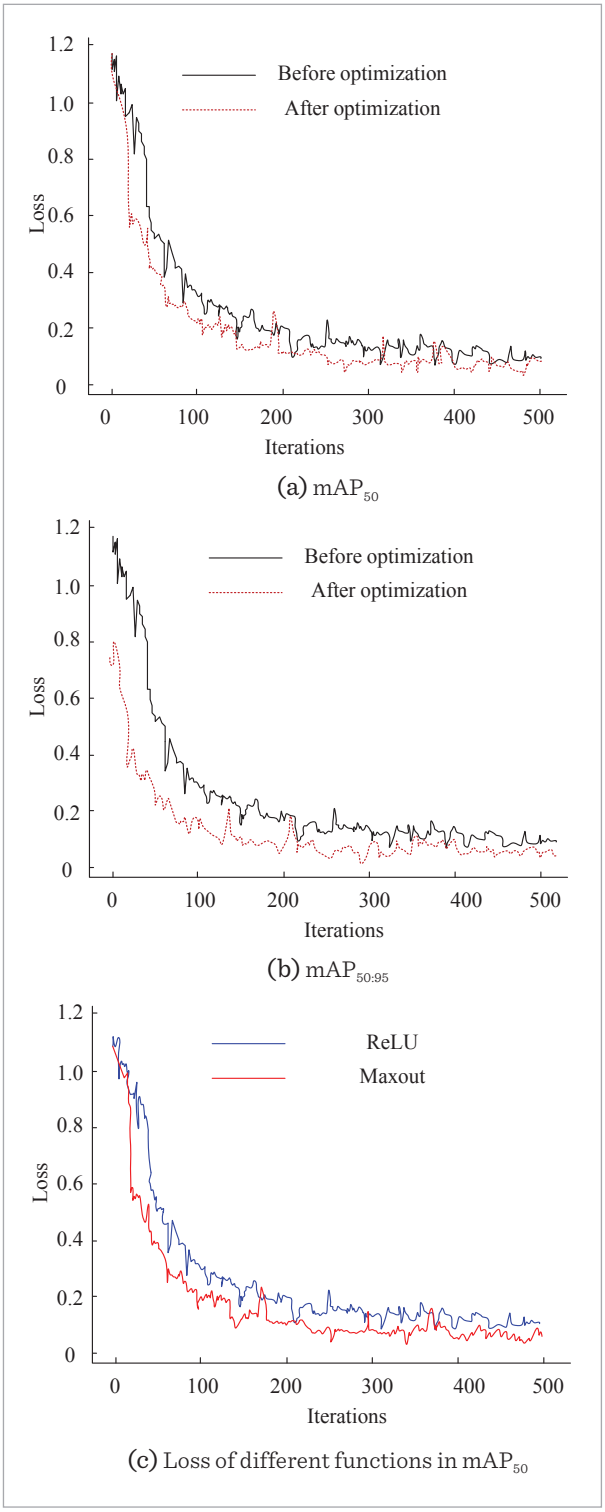
From Figure 10, the loss values of CIoU and DIoU on heavily occluded, and heavily motion blurred samples were significantly higher than 0.75, while WIoU V3 reduced the loss values of these extremely low-quality samples to below 0.45 through outlier recalibration. This is because the dynamic focusing mechanism of WIoU V3 identified that the outliers of these samples were too high, automatically reducing their gradient gain. For moderately occluded samples (30% -60% occlusion), the loss value of WIoU V3 was 0.38, which was lower than the loss values of CIoU (0.55) and DIoU (0.52).

3.2. Performance Analysis of Improved MobileNet V1 Model in Crop Sorting

To assess the sorting effect of the improved MobileNet V1 model in agricultural mechanized harvesting, MobileNet V1 was trained and tested on the same crop dataset samples. Firstly, to demonstrate the impact of studying the optimization behavior of replacing the LF on MobileNet V1, the study analyzed the loss of MobileNet V1 during the training process before and after improving the LF. The specific results are shown in Figure 11.

From Figure 11(a), before optimizing the LF, the mAP_{50} loss of the MobileNet V1 model mostly fluctuated within the range of 0.15-0.2. After optimizing the LF, the mAP_{50} loss of the MobileNet V1 model mostly fluctuated within the range of 0.1-0.15. From Figure 11(b), after optimizing the LF,

Figure 11
The loss of MobileNet V1 during training before and after optimizing the LF.



the $mAP_{50:90}$ loss of the MobileNet V1 model was basically below the range of 0.2. Before optimizing the LF, the $mAP_{50:90}$ loss of the MobileNet V1 model fluctuated around 0.2. From Figure 11(c), overall, the Maxout activation function has a lower overall loss value than the ReLU function. The findings denote that after improving the LF, the MobileNet V1 model has smaller losses and is more stable. Afterwards, the study compared the sorting accuracy and iteration time of different models during the training process. The study compared three different CNN structures, VGG-16, ShuffleNet, and MobileNet V1. The specific findings are denoted in Table 5.

From the data in Table 5, in the first iteration, the VGG-16 model had an initial accuracy of 0.51 and took 243 seconds. The ShuffleNet model had an initial accuracy of 0.63 and took 66 seconds. The initial accuracy of improved MobileNet V1 was 0.71 and took 52 seconds. In the final iteration, the VGG-16 model achieved a final accuracy of 0.95 and took 230 seconds. The final accuracy of the ShuffleNet model was 0.97, taking 57 seconds. The initial accuracy of MobileNet V1 was 0.98, taking 48 seconds. The findings denote that the MobileNet V1 model has higher accuracy performance in crop image sorting compared to other models, while the iterative training time is the shortest and the efficiency is the highest. Afterwards, the study randomly extracted some samples from the test dataset to further test the sorting performance of ShuffleNet model and improved MobileNet V1 for different crops. The specific results are shown in Figure 13.

To verify the independent contributions of the three improvements (Adam optimizer, Maxout activation function, and network structure lightweighting) to MobileNet V1, this study trained MobileNet V1 sorting models with different configurations under the premise of fixed YOLOv8 detection results. The results are shown in Table 6

Table 5

Accuracy and time consumption of different network structures in crop sorting.

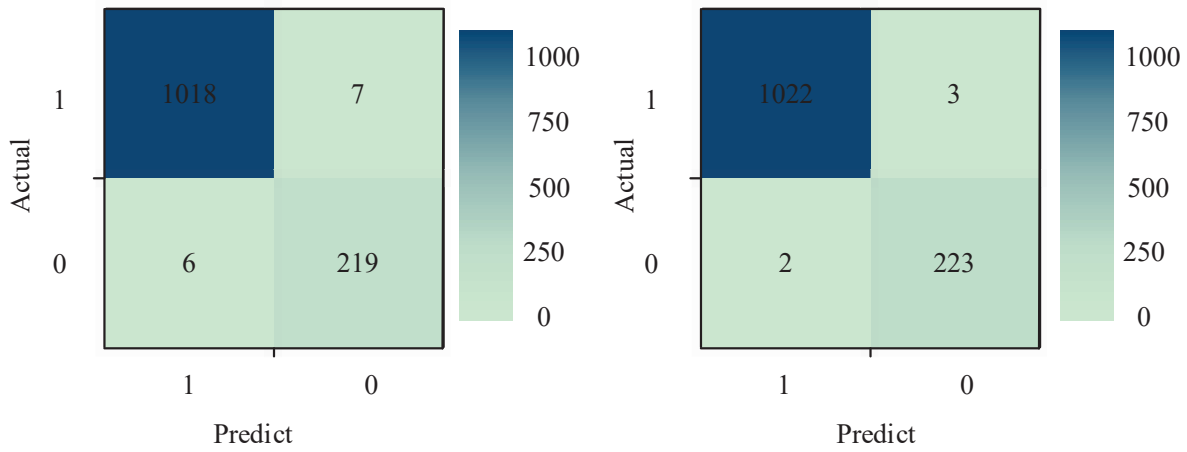
Algorithm	VGG-16		ShuffleNet		MobileNet V1	
Iteration rounds	Accuracy	Time(s)	Accuracy	Time(s)	Accuracy	Time(s)
1	0.51	243	0.63	66	0.71	52
100	0.75	251	0.73	56	0.85	47
200	0.87	240	0.80	58	0.90	49
300	0.95	246	0.89	58	0.95	47
400	0.95	239	0.95	57	0.98	49
500	0.95	230	0.97	57	0.98	48

Table 6

Accuracy and time consumption of different network structures in crop sorting.

Model	Sorting accuracy (%)	Parameter count (M)	Single sample inference time (ms)
MobileNet V1	93.12 ± 0.51	4.23	12.4 ± 0.8
Adam	95.08 ± 0.42	4.23	12.5 ± 0.7
Maxout	96.23 ± 0.38	5.12	15.1 ± 0.9
Structural optimization	94.56 ± 0.47	2.41	8.3 ± 0.5
Adam + Maxout	97.45 ± 0.33	5.12	15.3 ± 0.8
Adam + Maxout + Structural optimization	98.31 ± 0.28	2.89	9.1 ± 0.6

Figure 12
ShuffleNet model and MobileNet V1 for crop sorting performance.



(a) ShuffleNet (b) Improved MobileNet V1

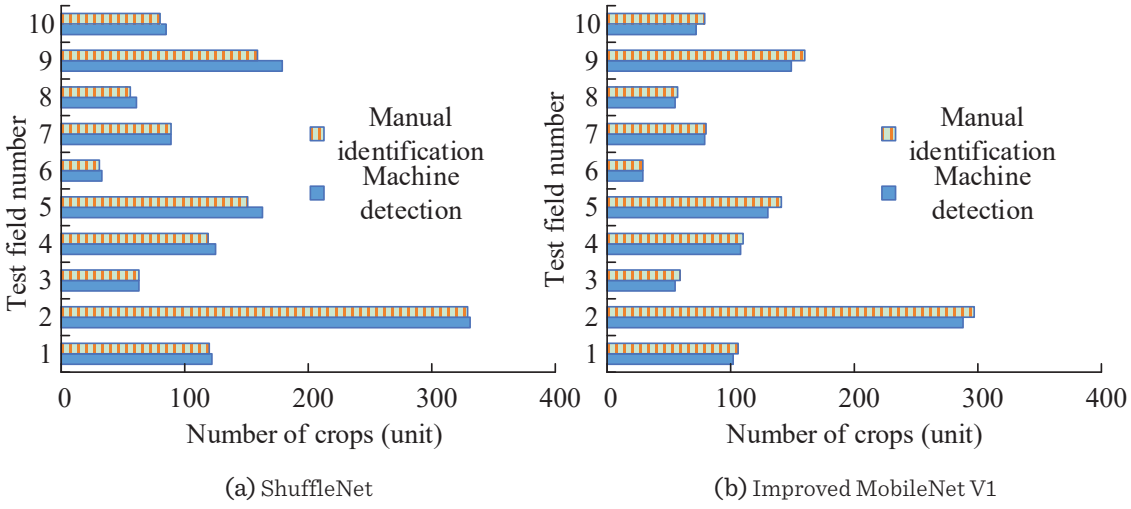
From Table 6, the Adam optimizer improved the sorting accuracy from 93.12% to 95.08%, mainly due to the enhancement of gradient stability. The Maxout activation function brought a greater improvement (+3.11%), verifying its advantages in avoiding neuronal death and enhancing feature selectivity, but increasing the number of parameters and inference time. Although structural optimization only improved accuracy by 1.44%, it significantly reduced the number of parameters and inference time, making it the key to implementing embedded deployment. When the three improvements were combined, the accuracy reached 98.31%, which was higher than any single or double combination, and the inference time was 9.1ms, far lower than the baseline, proving the synergistic effectiveness of the three.

The 1 and 0 on the horizontal and vertical axes in Figure 12 represent the sorting results for harvestable and non-harvestable items, respectively. From Figure 12(a), in the sorting recognition of the ShuffleNet model, the number of correct sorting for harvestable samples was 1018, and the number of incorrect sorting was 7. The ShuffleNet model had a correct sorting quantity of 219 and an incorrect sorting quantity of 6 for non-harvestable samples. From Figure 12(b), in the improved MobileNet V1 model for sorting and recognition, the amount of correctly sorted samples was 1022, and the amount of incorrectly sorted samples was 3. The improved MobileNet V1 model had a correct sorting quantity of 223 and an incorrect sorting quantity of 2 for non-harvestable samples. Among the erroneous samples, the false positive samples of the improved MobileNet V1 model include one case of immature cherry, one case of immature grape, and one case of locally damaged mushroom. In fruit judgment, immature fruits appear light green and differ greatly from mature red fruits, but the model still judges them as harvestable with a confidence level of 0.62. Perhaps due to fully learning the discriminative features of the green stage or color confusion with the background leaves, the model may misjudge. The findings indicate that the enhanced MobileNet V1 model has higher accuracy performance in crop sorting, and can accurately identify whether growing crops should be picked.

Finally, the study analyzed the overall performance of improved YOLOV8+MobileNet V1. Ten experimental fields were used to conduct practical application analysis in the mango planting area of a university's agricultural experimental base. The study utilized an Intel RealSense D455 depth camera and the computer listed in Table 1 to develop

Figure 13

Comparison of crop quantity collection in experimental fields under different methods.



(a) ShuffleNet (b) Improved MobileNet V1

an image acquisition system and control analysis system for an agricultural intelligent harvesting robot. At the same time, a Selective Compliance Assembly Robot Arm (SCARA) with four degrees of freedom was used to form the collection system of the agricultural robot. Claws and cutting knives were installed on the SCARA robot arm, and the clamping force range was set to 2.12N-2.67N. In field applications, experiments were conducted on weather conditions with sufficient light. In the experimental field, the fruits were obstructed by branches, leaves, other fruits, or supports by more than 30%. At the same time, there were leaves with colors and textures similar to the fruits in different fields, as well as interference from other crops. This was used to test the model's ability to distinguish between small and overlapping targets. Afterwards, the study obtained the quantity of mango products and the number of harvestable mango products in 10 experimental fields through manual recognition and robot recognition, which were used to analyze the improved object detection and harvestable sorting ability of YOLOV8+MobileNet V1 under robot recognition mode. The specific results are denoted in Figure 12.

From Figure 13(a), under manual recognition, the number of crops in fields 1-10 was 122, 331, 63, 125, 163, 33, 89, 61, 179, and 85, respectively. The crop recognition numbers of robots equipped with YOLOV8+MobileNet V1 model were 120, 329, 63, 119, 151, 31, 89, 56, 159, and 80, respectively. The average detection accuracy of the YOLOV8+MobileNet V1 model robot for crop target detection in all farmland was 0.954. From Figure 13(b), under manual recognition, the number of harvestable crops in fields 1-10 was 102, 288, 55, 108, 130, 29, 79, 55, 149, and 72, respectively. The robots equipped with YOLOV8+MobileNet V1 model could recognize 106, 297, 59, 110, 141, 29, 80, 57, 160, and 79 harvestable crops, respectively. Therefore, in the detection of crop targets in all farmland, the average sorting error rate of the YOLOV8+MobileNet V1 model robot was 4.779%. To compare the mango recognition performance, the recognition performance of KMeans+ShuffleNet model and YOLOV8+MobileNet V1 model under different weather conditions was studied and compared, as shown in Figure 14.

From Figure 14, for partially occluded fruits, YOLOV8+MobileNet V1 had stronger recognition ability, and under clear weather conditions, the model could recognize fruits that are farther away and have smaller targets. Afterwards, the performance and time consumption of robots equipped with different models in picking were compared, and the specific findings are denoted in Table 7.

Figure 14
Visualization images of different model detection results.



Table 7
Performance and time consumption of robots equipped with different models in harvesting.

YOLOV8+CNN			YOLOV8+MobileNet V1			KMeans+ShuffleNet		
Accuracy of detecting the quantity of harvestable crops	Picking rate	Picking time (s)	Accuracy of detecting the quantity of harvestable crops	Picking rate	Picking time (s)	Accuracy of detecting the quantity of harvestable crops	Picking rate	Picking time (s)
89%	86%	13.3	96%	93%	4.2	92%	88%	10.6
92%	89%	15.6	99%	93%	4.6	93%	91%	11.4
98%	91%	14.5	100%	94%	4.6	94%	93%	11.1

In experimental fields numbered 1-10, due to the significantly larger number of harvestable crops in experimental field 2 compared to other fields, this experimental sample was removed. Fields numbered 1, 3, and 7 were sorted and picked using robots equipped with YOLOV8+CNN. Fields numbered 4, 8, and 10 were sorted and picked using robots equipped with YOLOV8 and MobileNet V1. The farmlands numbered 5, 6, and 9 were sorted and picked using robots equipped with YOLOV8+-MobileNet V1. The YOLOV8+CNN model had an average sorting recognition accuracy of 0.93 and a picking rate of 0.887 in specific applications, with an average picking time of 14.47 seconds. The improved YOLOV8+MobileNet V1 model had an average sorting recognition accuracy of 0.983 and a

picking rate of 0.933 in specific applications, with an average picking time of 4.47 seconds per crop. The KMeans+ShuffleNet model had an average sorting recognition accuracy of 0.93 and a picking rate of 0.907 in specific applications, with an average picking time of 11.03 seconds per crop. The experimental results show that in practical applications, robots equipped with YOLOV8+MobileNet V1 model also have the highest accuracy and picking ability in agricultural picking selection, while taking the shortest time.

4. Discussion

Agriculture is accelerating its transformation towards digitization and intelligence, and deep learning, especially convolutional neural networks (CNN), has become the core driving technology for agricultural visual tasks. From seed quality inspection, soil analysis, disease detection to yield estimation, CNN and its derived models have demonstrated excellent performance in multiple agricultural processes [9]. In this context, the research focuses on the two key links of picking and sorting, proposing a collaborative paradigm of "big model localization+small model classification", further expanding the application boundaries of CNN in agricultural mechanization. In the field of plant disease recognition, existing research such as MASK-RCNN combined with contrast enhancement and feature screening has achieved high-precision recognition of apple leaf diseases [21]. Similarly, the improved YOLOv8 module in this study also undertakes the tasks of target localization and region perception, and enhances its robustness to low-quality samples such as occlusion and deformation through deep convolution and WIoU V3 loss function. Compared with MASK-RCNN, YOLOv8 has stronger real-time and end-to-end deployment capabilities, making it more suitable for agricultural production scenarios. Random forest and CNN have both achieved excellent performance in yield prediction tasks, but deep learning models have more advantages in multivariate and nonlinear feature modeling [24]. This study further demonstrates that in agricultural picking and sorting tasks, lightweight models can significantly reduce inference latency while maintaining high accuracy after Adam optimization, Maxout activation function, and structural compression. Although this study achieved good detection and sorting results in the experimental field, it still faces challenges in practical promotion. The current model training has not fully considered complex agricultural environmental factors such as changes in lighting, leaf occlusion, and fruit overlap. In the future, data augmentation and domain adaptation techniques such as Generative Adversarial Networks (GANs) or Visual Transformers (ViTs) should be introduced to improve the generalization ability of models [23].

5. Conclusion

A picking and sorting model based on an improved YOLOV8+MobileNet V1 network was developed to deal with the issues of low precision and insufficient efficiency in agricultural mechanization. This model was used to achieve efficient and high-precision intelligent agricultural production. The performance of the model was tested on a crop dataset, and the findings denoted that the accuracy of the optimized YOLOV8 model fluctuated within the range of 0.85-0.95 during the training, while the traditional YOLOV8 model varied within the accuracy value range of 0.7-0.83. In the ablation experiment, the AA of traditional YOLOV8 was 0.9405, and the average mAP_{50} was 0.9175. Under the dual optimization of deformable convolutional operation and LF, the AA of YOLOV8 model was 0.9715, and the mAP_{50} was 0.974. In object detection of sample types, the improved YOLOV8 model only had 3 recognition errors, while KMeans+CNN had a total of 6 recognition errors. Afterwards, the performance of the MobileNet V1 network was analyzed, and the outcomes denoted that in the improved MobileNet V1 model for sorting and recognition, the correct sorting quantity for harvestable samples was 1022, and the incorrect sorting quantity was 3. The improved MobileNet V1 model had a correct sorting quantity of 223 and an incorrect sorting quantity of 2 for non-harvestable samples. Finally, in field experiments, the improved YOLOV8+MobileNet V1 model achieved an average sorting recognition accuracy of 0.983 and a picking rate of 0.933 in specific applications, with an average picking time of 4.47 seconds per crop. The outcomes denote that

the enhanced YOLOV8+MobileNet V1 network has high efficiency and recognition accuracy in crop picking and sorting work. In the model training, the study did not account for the influence of lighting and environmental factors on the model's picking and sorting accuracy. Therefore, future research will incorporate samples under conditions such as leaf occlusion and low light levels to strengthen the model's generalization capability.

Fundings

This research is supported by: The Ministry of Education's Industry-University Cooperation and Collaborative Talent Cultivation Project-Huawei's New Engineering Project Construction Initiative, Exploring the Hainan Free Trade Port's HarmonyOS Emerging Engineering Initiative through the "Smart Terminal Application Software Development" Course, (NO. 250600007194058); The Hainan Provincial Natural Science Foundation of China under Grant, Research on Surface Defect Detection of Electronic Components in Complex Backgrounds Based on Machine Vision, (NO.625QN350); The Hainan Philosophy and Social Science Planning Project, Sustainable Development Study of Hainan Marine Protected Areas, (NO. HNSK(ZX)24-255); The 2023 Scientific Research Project of Hainan Vocational University of Science and Technology under Grant, Research on Defect Identification, Detection, and Application of Electronic Components Based on Intelligent Algorithms, (NO.HKKY2023-21).

References

1. Ali, N., Ullah, M., Mohammed, A., Bais, A., Berraies, S., Ruan, Y., Sangha, J. S. Advancing Fusarium Head Blight Detection in Wheat Crop: A Review and Future Directions to Sustainable Agriculture. IEEE Transactions on Consumer Electronics, 2025, 71(3), 7473-7491. https://doi.org/10.1109/TCE.2025.3549057
2. Anthoniraj, S., Karthikeyan, P., Vivek, V. Weed Detection Model Using the Generative Adversarial Network and Deep Convolutional Neural Network. Journal of Mobile Multimedia, 2022, 18(2), 275-291. https://doi.org/10.13052/jmm1550-4646.1826
3. Bello, R. W., Oladipo, M. A. Mask YOLOv7-Based Drone Vision System for Automated Cattle Detection and Counting. Artificial Intelligence Applications, 2024, 2(2), 129-139. https://doi.org/10.47852/bonviewAIA42021603
4. Chakraborty, D., Deka, B. Deep Learning-Based Selective Feature Fusion for Litchi Fruit Detection Using Multimodal UAV Sensor Measurements. IEEE Transactions on Artificial Intelligence, 2025, 6(7), 1932-1942. https://doi.org/10.1109/TAI.2025.3532205
5. Cheng, Y. H., Kuo, C. N., Lin, Y. D. An Artificial IoT-Enabled Smart Production Line for 360° Visual Defect Detection and Classification of Cherry Tomatoes. IEEE Internet of Things Journal, 2024, 11(10), 18082-18098. https://doi.org/10.1109/JIOT.2024.3360715
6. Cui, B., Liang, L., Ji, B., Zhang, L., Zhao, L., Zhang, K., Shi, F., Créput, J. C. Exploring the YOLO-FT Deep Learning Algorithm for UAV-Based Smart Agriculture Detection in Communication Networks. IEEE Transactions on Network and Service Management, 2024, 21(5), 5347-5360. https://doi.org/10.1109/TNSM.2024.3424232
7. de Jong, S., Baja, H., Tamminga, K., Valente, J. APPLE MOTS: Detection, Segmentation and Tracking of Homogeneous Objects Using MOTS. IEEE Robotics and Automation Letters, 2022, 7(4), 11418-11425. https://doi.org/10.1109/LRA.2022.3199026
8. Deka, B., Chakraborty, D. UAV Sensing-Based Litchi Segmentation Using Modified Mask-RCNN for Precision Agriculture. IEEE Transactions on Agri-Food Electronics, 2024, 2(2), 509-517. https://doi.org/10.1109/TAFE.2024.3420028
9. Dhanya, V. G., Subeesh, A., Kushwaha, N. L., Vishwakarma, D. K., Kumar, T. N., Ritika, G., Singh, A. N. Deep Learning-Based Computer Vision Approaches for Smart Agricultural Applications. Artificial Intelligence in Agriculture, 2022, 6, 211-229. https://doi.org/10.1016/j.aiia.2022.09.007
10. Donapati, R. R., Cheruku, R., Kodali, P. Real-Time Seed Detection and Germination Analysis in Precision Agriculture: A Fusion Model with U-Net and CNN on

Jetson Nano. IEEE Transactions on AgriFood Electronics, 2023, 1(2), 145-155. https://doi.org/10.1109/TAFE.2023.3332495

11. Ferrag, M. A., Shu, L., Friha, O., Yang, X. Cyber Security Intrusion Detection for Agriculture 4.0: Machine Learning-Based Solutions, Datasets, and Future Directions. IEEE/CAA Journal of Automatica Sinica, 2022, 9(3), 407-436. https://doi.org/10.1109/JAS.2021.1004344
12. Gava, J., Sherjil, A., Laurini, L., Atukpor, E., Bastos, R., Moraes, F., Reis, R., Ost, L. Radiation-Induced Soft Error Assessment of a Multithreaded MobileNet CNN Model Running in a Multicore Edge Processor. IEEE Transactions on Nuclear Science, 2025, 72(8), 2821-2829. https://doi.org/10.1109/TNS.2025.3585859
13. Hong, H., Choi, D., Kim, N., Kim, H. Mobile-X: Dedicated FPGA Implementation of the MobileNet Accelerator Optimizing Depthwise Separable Convolution. IEEE Transactions on Circuits and Systems II: Express Briefs, 2024, 71(11), 4668-4672. https://doi.org/10.1109/TCSII.2024.3440884
14. Kong, W., Li, Y., Hu, J., Luo, S., Rui, J. A Shadow-Robust Pavement Damage Detection Framework Based on RA-Cycle-GAN and DDE-YOLOv8. IEEE Transactions on Intelligent Transportation Systems, 2025, 26(8), 11342-11355. https://doi.org/10.1109/TITS.2025.3556941
15. Lin, I. A., Cheng, Y. W., Lee, T. Y. Enhancing Smart Agriculture with Lightweight Object Detection: MobileNetV3-YOLOv4 and Adaptive Width Multipliers. IEEE Sensors Journal, 2024, 24(23), 40017-40028. https://doi.org/10.1109/JSEN.2024.3478810
16. Lin, N., Zhang, L., Wu, T., Hawbani, A., Zhou, H., Zhao, L. Surface Multiple Object Tracking: An Accurate HAT-YOLOv8-ADT Tracking Model. IEEE Internet of Things Journal, 2025, 12(11), 18266-18278. https://doi.org/10.1109/JIOT.2025.3539852
17. Liu, Y., Zhao, Y., Yu, B., Zhu, C., Huo, G., Li, Q. An Improved YOLOv8-Based Shallow Sea Creatures Object Detection Method. IEEE Journal of Oceanic Engineering, 2025, 50(2), 817-834. https://doi.org/10.1109/JOE.2025.3538954
18. Mai, C., Liao, R., Ren, J., Gong, Y., Zhang, K., Zhang, C. MobileNet-Based IoT Malware Detection with Opcode Features. Journal of Computer and Information Networks, 2023, 8(3), 221-230. https://doi.org/10.23919/JCIN.2023.10272350
19. Mishra, D., Chemweno, S., Hadar, O., Ben-Tovim, O., Lazarovitch, N., Ephrath, J. E. Unsupervised Image Super-Resolution for Root Hair Enhancement and Improved Root Traits Measurements. IEEE Transactions on AgriFood Electronics, 2024, 2(1), 81-90. https://doi.org/10.1109/TAFE.2024.3359660
20. Ranjith, C. P., Hardas, B. M., Mohideen, M. S. K., Raj, N. N., Robert, N. R., Mohan, P. Robust Deep Learning Empowered Real-Time Object Detection for Unmanned Aerial Vehicles Based Surveillance Applications. Journal of Mobile Multimedia, 2023, 19(2), 451-476. https://doi.org/10.13052/jmm1550-4646.1925
21. Rehman, Z. U., Khan, M. A., Ahmed, F., Damaševičius, R., Naqvi, S. R. W., Nisar, K., Javed, K. Recognizing Apple Leaf Diseases Using a Novel Parallel Real-Time Processing Framework Based on Mask R-CNN and Transfer Learning: An Application for Smart Agriculture. IET Image Processing, 2021, 15(10), 2157-2168. https://doi.org/10.1049/ipr2.12183
22. Ryu, S., Jang, J., Oh, Y., Kim, J.-J. Mobileware: Distributed Architecture with Channel Stationary Dataflow for MobileNet Acceleration. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2024, 43(9), 2661-2673. https://doi.org/10.1109/TCAD.2024.3380555
23. Shah, A., Khan, M. A., Alzahrani, A. I., Alalwan, N., Hamza, A., Manic, S., Damaševičius, R. FuzzyShallow: A Framework of Deep Shallow Neural Networks and Modified Tree Growth Optimization for Agriculture Land Cover and Fruit Disease Recognition from Remote Sensing and Digital Imaging. Measurement, 2024, 237, 115224. https://doi.org/10.1016/j.measurement.2024.115224
24. Sharma, P., Dadheech, P., Aneja, N., Aneja, S. Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. IEEE Access, 2023, 11, 111255-111264. https://doi.org/10.1109/ACCESS.2023.3321861
25. Singh, A. K., Fru Junior, F. N., Mainsah, N. L., Abdoul-Rahmane, B. Enabling Data Collection and Analysis for Precision Agriculture in Smart Farms. IEEE Transactions on AgriFood Electronics, 2025, 3(1), 69-85. https://doi.org/10.1109/TAFE.2024.3454644
26. Vijh, S., Gaurav, P., Kumar, S., Bansal, P., Singh, M., Khan, M. A., Palade, V. USMA-BOF: A Novel Bag-of-Features Algorithm for Classification of Infected

Plant Leaf Images in Precision Agriculture. IEEE Robotics and Automation Magazine, 2023, 30(4), 30-40. https://doi.org/10.1109/MRA.2023.3315929

27. Wang, H., Cao, S., Yue, X., Huang, Y., Yang, Q., Zhang, J., Ran, Y. Spectral Demodulation of Tapered Microfiber Grating Using MobileNet. IEEE Sensors Journal, 2025, 25(2), 2791-2797. https://doi.org/10.1109/JSEN.2024.3507099
28. Xiong, B., Fan, S., He, X., Xu, T., Chang, Y. Small Logarithmic Floating-Point Multiplier Based on FPGA and Its Application on MobileNet. IEEE Transactions on Circuits and Systems II: Express Briefs, 2022, 69(12), 5119-5123. https://doi.org/10.1109/TCSII.2022.3205945
29. Yang, D., Luo, Z. A Parallel Processing CNN Accelerator on Embedded Devices Based on Optimized MobileNet. IEEE Internet of Things Journal, 2023, 10(21), 18844-18852. https://doi.org/10.1109/JIOT.2023.3277869
30. Yang, J., Wang, Y., Zhao, H., Gui, G. MobileNet and Knowledge Distillation-Based Automatic Scenario Recognition Method in Vehicle-to-Vehicle Systems. IEEE Transactions on Vehicular Technology, 2022, 71(10), 11006-11016. https://doi.org/10.1109/TVT.2022.3184994
31. Yeh, J. F., Lin, K. M., Lin, C. Y., Kang, J. C. Intelligent Mango Fruit Grade Classification Using AlexNet-SPP with Mask R-CNN-Based Segmentation Algorithm. IEEE Transactions on AgriFood Electronics, 2023, 1(1), 41-49. https://doi.org/10.1109/TAFE.2023.3267617
32. Yu, J., Zhang, L., Cheng, D., Bu, C., Wu, H., Song, A. RepMobile: A MobileNet-Like Network with Structural Reparameterization for Sensor-Based Human Activity Recognition. IEEE Sensors Journal, 2024, 24(15), 24224-24237. https://doi.org/10.1109/JSEN.2024.3412736

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/).