International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 8256-8267

https://doi.org/10.70517/ijhsa464706

Research on citrus target recognition based on the improved YOLOv8 algorithm

Danyi Zhang[1], Jun Li[2,*] and Zhengshun Fei[1]
[1] School of Automation and Electrical Engineering, Zhejiang University of Science & Technology, Hangzhou, Zhejiang, 310000, China
[2] School of Aeronautical Engineering, Taizhou University, Taizhou, Zhejiang, 318000, China
Corresponding authors: (e-mail: 19818518892@163.com).

Abstract This paper leverages the lightweight characteristics of the YOLOv5 algorithm to enhance the performance of citrus fruit picking point detection by optimizing the enhanced feature representation of the YOLOv5 algorithm. In the original YOLOv5 network model, to improve the prior boxes obtained from the K-means clustering algorithm, the binary K-means+IoU algorithm is used to update the prior boxes for citrus fruit target detection. The ECANet attention mechanism is added to enhance the algorithm's ability to focus on important features and eliminate interference from irrelevant features. Combining WIoU-Loss as the loss function for the candidate boxes in the citrus fruit recognition network model achieves more precise citrus fruit target recognition. We analyze the optimization effects of the three strategies—the ECANet module, the K-means+IoU algorithm, and the WIoU loss function—on the YOLOv5 algorithm. Using the citrus fruit image data constructed in this paper under natural environmental conditions, we analyze the improved YOLOv5 algorithm's performance in detecting targets when citrus fruits overlap or are occluded. The experimental results of citrus fruit recognition show that the mAP value, precision P, recall R, and F1 value of the proposed recognition and detection method are 94.86%, 93.49%, 89.26%, and 0.88%, respectively. Moreover, the positioning error of citrus fruit targets does not exceed 2 mm. The proposed algorithm is proven to be effective and can provide reference for the motion target points of the end-effector of citrus picking robots.

Index Terms YOLOv5, ECANet, WIoU-Loss, K-means, fruit target recognition

I. Introduction

Citrus fruits are the world's largest category of fruit, with 138 countries worldwide producing them. Among these, countries such as China, Brazil, India, Mexico, and the United States lead in production volume [1], [2]. China's citrus industry is structured into five major production zones, with both area and production volume ranking first globally, accounting for nearly one-third of the world's citrus production [3], [4]. However, citrus harvesting currently relies primarily on manual labor, which poses challenges such as high labor intensity, high production costs, and low productivity [5], [6]. Therefore, developing intelligent harvesting robots to replace manual labor and free humans from complex agricultural production holds significant importance [7], [8].

In recent years, advancements in computer vision and deep learning technologies have driven the widespread application of automated citrus classification and recognition systems based on image processing and machine learning [9], [10]. These systems can rapidly, efficiently, and accurately classify and identify citrus characteristics such as variety, grade, size, and color, thereby enhancing citrus quality assessment and market value, and providing robust technical support and assurance for agricultural production and marketing [11]-[13]. Deep learning methods demonstrate superior application performance in this field, with the most commonly used object detection algorithm being the YOLO series, which has been widely adopted [14]-[16]. YOLO is an object detection algorithm proposed by Redmon in 2016. Its core principle is to reformulate the object detection problem as a regression problem, using a single convolutional neural network structure to predict bounding boxes and class probabilities, thereby achieving faster processing speeds than other algorithms [17]-[20]. By integrating sensor and intelligent robotics technologies, efficient citrus harvesting and classification can be achieved, further enhancing agricultural production efficiency and quality [21]-[23]. However, the initial YOLO algorithm suffered from severe localization errors and low detection accuracy, necessitating improvements to achieve more precise citrus object recognition [24], [25].

Reference [26] conducted experiments using 1,200 citrus images and proposed a lightweight YOLO-MECD model based on the YOLOv11s architecture. Through comparative analysis, it was demonstrated that the YOLO-MECD model achieves significant improvements in detection performance and computational efficiency. Reference [27] highlights the importance of citrus fruits in agriculture and proposes a fast detection algorithm for citrus fruits based on global context fusion. By introducing the AG-YOLO network to integrate context information, it effectively addresses the issues of low detection accuracy and missed detections in citrus detection algorithms. Reference [28] proposes the YOLO-CIT model and integrates the R-LBP method to accurately identify citrus fruits at different maturity stages. The study indicates that the R-LBP algorithm can amplify the texture features of citrus fruits at different maturity stages, and the YOLO-CIT model combined with the R-LBP algorithm can identify the maturity of citrus fruits in complex environments. Literature [29] proposed a citrus recognition method based on the YOLOv4 neural network and trained the network model under the Darknet framework, revealing that this recognition model can meet the real-time image recognition speed and accuracy requirements of citrus harvesting robots. Literature [30] proposes a high-precision, lightweight YOLOv4 detection method that can accurately and quickly detect citrus fruits in complex growth environments, providing support for the development of citrus harvesting robots. Literature [31] proposes a citrus detection and localization method based on improved YOLOv5s stereo vision technology. Through experiments, it demonstrates the recall rate of citrus detection under three different conditions: uneven lighting, weak lighting, and good lighting, and can achieve accurate and rapid detection and localization of citrus in complex environments. Literature [32] uses oranges as the experimental subject and proposes a deep learning-based orange counting algorithm. This algorithm includes two sub-algorithms: OrangeYolo for detection and OrangeSort for tracking. It verifies that this method outperforms manual counting based on video in terms of fruit detection accuracy. Literature [33] utilized panoramic photography to collect images of citrus fruit trees and proposed an AC-YOLO-based citrus identification method in natural orchard environments, verifying that this method demonstrated good performance in identifying citrus fruits in natural orchard environments. Literature [34] addresses issues such as missed detections and false positives in citrus detection in complex orchard environments, proposing a citrus detection model based on an improved YOLOv5 algorithm. This model overcomes the issue of parameter sharing in convolutional operations, effectively improving detection accuracy, and provides important support for citrus localization and harvesting. Literature [35] proposes an improved multi-scale YOLO algorithm (improved-YOLOv3) aimed at achieving rapid and accurate identification of citrus fruits in field environments. Experimental validation demonstrates that this algorithm possesses strong robustness and higher detection accuracy, enabling citrus identification in complex environments. Literature [36] proposed the dense-truu-yolo model, which can effectively improve detection accuracy in cases of severe occlusion and overlap of citrus fruits. Literature [37] designed the citrus picking point localization workflow CPPL. Based on extensive experiments, CPPL achieved high citrus recognition accuracy, providing an efficient method for real-time citrus harvesting. The above studies emphasize the important role of citrus detection and recognition in effective citrus picking and propose improved algorithms and models for detecting and recognizing citrus, such as YOLOv5s, YOLOv4 neural networks, and YOLO-CIT, which can achieve precise recognition of citrus in complex environments.

This paper collects citrus image data from natural environments to establish a citrus image dataset. Data augmentation techniques such as flipping and scaling are applied to optimize the training dataset. Evaluation metrics for citrus fruit object detection and algorithm localization-related evaluation metrics are proposed. The structure of the YOLOv5 algorithm is analyzed. Based on the YOLOv5 network model, the WIoU loss function is modified, the ECA attention mechanism is added, and the binary K-means+IoU algorithm is used to update the prior boxes for citrus fruit target detection. An improved YOLOv5-based citrus target recognition algorithm is proposed. Analyze the effectiveness of the improved YOLOv5 network model. Combine the constructed real citrus fruit dataset to analyze the actual detection performance and spatial localization error data of the improved YOLOv5 algorithm.

II. Materials and Methods

II. A. Dataset Construction

The image data used in this experiment was obtained from on-site photography in an orchard. The photography location was the Citrus Research Institute of a certain university, and the equipment used was a Canon 60D and a smartphone. The images were saved in JPG format with a resolution of 4032×3072.

To detect and identify target citrus fruits in a real harvesting environment, 2,134 images of citrus fruits at different angles, lighting conditions, distances, and sizes were collected. The original images were annotated using the Label Img tool and randomly divided into a training set of 1,400 images, a validation set of 400 images, and a test set of 334 images.

Additionally, to enhance the model's robustness and generalization capabilities and prevent overfitting, data augmentation was applied to the dataset prior to model training. The augmentation methods included flipping, scaling, random translation, blurring, and brightness adjustment. After data augmentation, 4,570 images of citrus fruits ready for picking were obtained, including 3,000 images for the training set, 1,000 images for the validation set, and 570 images for the test set.

II. B.Test Environment and Training Strategy

The hardware environment used for training in this paper is a CPU Intel(R) Xeon(R) Gold 6242R CPU @ 3.10 GHz, GPU Tesla V100-PCIE 64GB×2, and 1TB of memory.

The software environment is Ubuntu 20.04.1. The operating system is Python 3.8, PyTorch 1.10, Torchvision 0.11, CUDA 11.2, and cuDNN 8.2.0.

The computer hardware environment for the recognition experiment is an Intel(R) Core(TM) i7-10875H CPU @ 2.30 GHz, a GeForce RTX 2060 GPU, and 32GB × 2 of memory. The software environment consists of a Windows 10 operating system, Python 3.8, PyTorch 1.10, Torchvision 0.11, CUDA 11.3, and cuDNN 8.2.0.

After multiple parameter adjustment tests, the final model parameters selected for training are shown in Table 1.

Table 1: Model parameter

Parameter	Numerical value	Parameter	Numerical value
Image size	640×640	Maximum learning rate	0.001
Optimizer	Adam	Momentum	0.6
Batch Size	64	Epoch	500

II. C.Evaluation Indicators

Recognition accuracy is an important evaluation metric for citrus fruit target detection. Therefore, this paper selects accuracy rate P, recall rate R, average precision A_P, and mean average precision m_{AP} as evaluation metrics for the target detection model. Among these, P reflects the model's precision, R reflects the model's recall, A_P reflects the average precision for a single category, and m_{AP} reflects the mean average precision across all categories. The specific calculation methods are as follows:

$$P = \frac{T_p}{\left(T_p + F_p\right)} \times 100\% \tag{1}$$

$$R = \frac{T_p}{\left(T_p + F_N\right)} \times 100\% \tag{2}$$

$$A_p = \int_0^1 P \cdot (R) dR \tag{3}$$

$$m_{AF} = \frac{1}{N} \sum_{i=1}^{N} A_{r_i} \tag{4}$$

Among these, T_p denotes the number of correctly identified target fruits where the model predicts a positive sample and the actual result is also a positive sample. F_p denotes the number of false positives where the model predicts a positive sample but the actual result is a negative sample. F_N represents the number of samples incorrectly identified as negative samples when the model predicts a negative sample and the actual result is a positive sample.

The primary function of the algorithm in this paper is to achieve more accurate identification and spatial localization of obscured citrus fruits, obtaining the spatial coordinates of the center of mass and the fruit diameter of the target fruit, thereby providing reference points for the motion targets of the end-effector of the citrus harvesting robot.

The evaluation metric for the algorithm's fitting accuracy is the overlap ratio between the fitted contour of the target fruit and the area of pixels within the manually annotated fruit contour region, calculated using the following formula:

$$C = \frac{\left|Q \cap Q_r\right|}{Q} \times 100\% \tag{5}$$

In the equation, C represents the overlap degree, Q represents the number of pixels within the manually annotated fruit contour region, and Q_r represents the number of pixels within the fitted contour.

In actual experiments, it is difficult to accurately obtain the actual spatial coordinates of the target fruit, so it is not suitable to evaluate the positioning accuracy of the algorithm based on the spatial position error of the centroid. Therefore, the spatial position of the occluded citrus fruit is compared with its spatial position when not occluded to evaluate the positioning effect of the algorithm.

The formula for calculating the positioning error of the algorithm is:

$$\Delta = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{6}$$

In the equation, Δ represents the algorithm's positioning error, and x_i, y_i, z_i represent the estimated three-dimensional spatial coordinates of the citrus fruit in an unobstructed state. x_j, y_j, z_j are the algorithm-estimated three-dimensional spatial coordinates of the same citrus fruit in an occluded state.

III. Algorithm

III. A. YOLOv5 Algorithm

YOLOv5 is a relatively mature algorithm in the YOLO algorithm series, characterized by its highly lightweight model. It not only excels in object detection but also performs exceptionally well in classification and localization tasks involved in object recognition. As a result, subsequent iterations of the YOLO series have been developed with minor modifications based on YOLOv5 [38]-[40].

Based on different network depths and widths, YOLOv5 can be categorized into YOLOv5m, YOLOv5s, YOLOv5l, and YOLOv5x. This paper focuses on research based on YOLOv5s.

The YOLOv5s model consists of four core components: the input layer, the backbone, the neck, and the detection component.

(1) The input layer preprocesses the detection images and uniformly adjusts the format of all images to a size of 640 pixels × 640 pixels, which are then input into the next component (backbone).

(2) The backbone section primarily includes multiple structures such as the Focus structure, convolutional structure, C3, and SPP. The primary task of the Focus structure is slicing, which expands the number of channels. This not only improves the model's computational speed but also ensures that image information is not lost. The convolutional structure performs convolution processing on the image to enhance the network's expressive capabilities. While extracting image features, it introduces nonlinear activation function feature information into deeper layers of the network. The C3 structure performs residual convolution on images to avoid the vanishing gradient problem that occurs as the network deepens. The SPP structure primarily performs multi-scale feature fusion operations on images.

(3) The Neck section also includes multiple main structures, such as convolution, upsampling, downsampling, and C3. The C3 structure is primarily composed of residual convolution structures. Additionally, the Neck section enables the transmission of image feature information from high-level to low-level and from low-level to high-level, thereby enabling the YOLOv5 network to simultaneously detect large and small objects.

(4) The Output section consists of three Head structures. The Head structure outputs the object's class information and corresponding confidence scores, as well as the corresponding positions of different predicted bounding boxes.

Since all convolutions in the YOLOv5 network are standard convolutions, and citrus fruits have small volumes, occupy few image pixels, and suffer from occlusion issues, as well as uneven scale distributions at different distances, the object detection model based on the YOLOv5 algorithm performs poorly. Based on the aforementioned issues, this paper further improves the model's feature extraction performance and enhances the feature fusion performance across different scale channels.

III. B. Improvements to YOLOv5

III. B. 1) Optimization of A Priori Boxes

The selection of a priori boxes is a crucial component in object detection, as appropriate a priori boxes can enhance the detection accuracy of the algorithm. This paper employs the binary K-means + IoU algorithm to recalculate prior boxes suitable for citrus object detection, thereby improving the accuracy and speed of the detection algorithm. IoU stands for Intersection-over-Union, which represents the ratio of the intersection between the predicted box and the true box to the union between the predicted box and the true box, where A denotes the predicted box and B denotes the true box. The formula for IoU is given in Equation (7):

$$IoU = \frac{A \cap B}{A \cup B} \tag{7}$$

Due to the limitations of the K-means algorithm, this paper adopts the binary K-means algorithm as a replacement for the K-means algorithm. Additionally, the Euclidean distance criterion in the algorithm is modified to the 1-IoU distance to achieve better clustering results. The citrus dataset is re-clustered to reduce errors and enhance the accuracy of the object detection algorithm.

The computational process of the binary K-means algorithm is shown in Figure 1, and the steps are as follows:

(1) Treat all sample points in the dataset as a single cluster.
(2) Calculate the total error for each cluster.
(3) Select a cluster and perform K-means clustering with k=2.
(4) Calculate the total error after dividing the cluster into two parts.
(5) Select the cluster with the smallest SSE error for the division operation.
(6) Repeat steps 2–3 until the number of clusters reaches the specified k value.

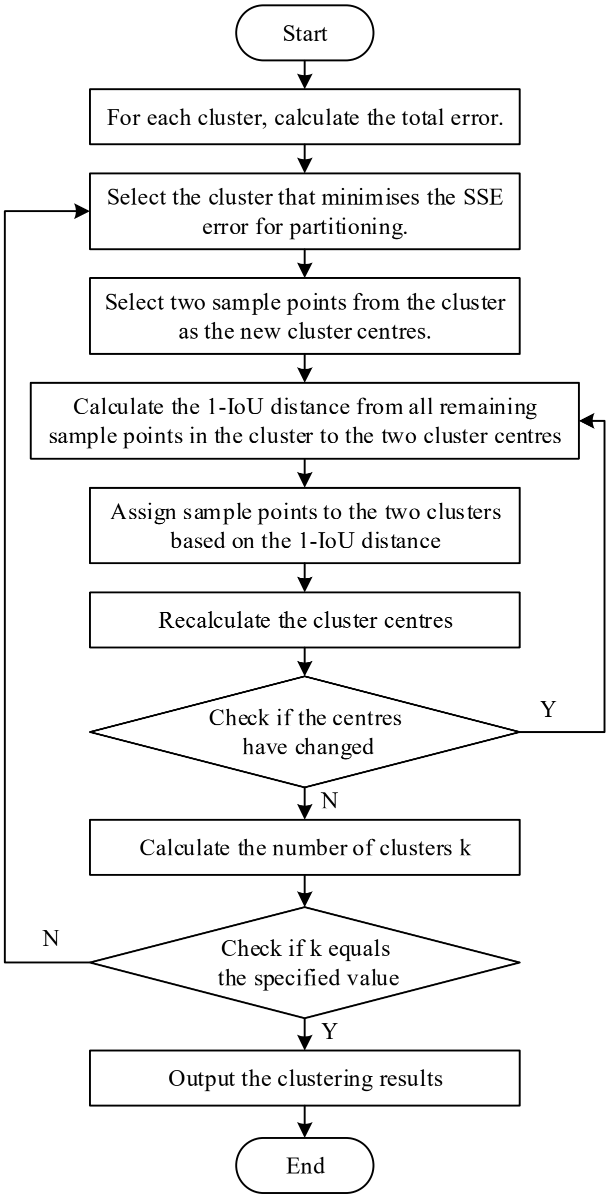

Start
For each cluster, calculate the total error.
Select the cluster that minimises the SSE error for partitioning.
Select two sample points from the cluster as the new cluster centres.
Calculate the 1-IoU distance from all remaining sample points in the cluster to the two cluster centres
Assign sample points to the two clusters based on the 1-IoU distance
Recalculate the cluster centres
Check if the centres have changed
Y
N
Calculate the number of clusters k
Check if k equals the specified value
N
Y
Output the clustering results
End

Figure 1: Binary k-means computing process

III. B. 2) Integration of Attention Module

Currently, widely used attention mechanisms include SENet, ECANet, and CBAM.

SENet is an attention mechanism that focuses on the channel dimension.

CBAM typically first applies a channel attention module to the input feature map, then applies a spatial attention module to the newly processed feature map.

ECANet is another implementation of a channel attention mechanism and can be viewed as an improved version of SENet [41]. Unlike SENet, ECANet directly uses a 1×1 convolution layer after global average pooling, eliminating the fully connected layer, thereby avoiding dimension reduction. ECANet achieves cross-channel information interaction through one-dimensional convolution, with the size of the convolution kernel adaptively varying via a function, whose expression is given by Equation (8):

$$k = \left| \frac{\log_2(c)}{\gamma} + \frac{b}{\gamma} \right| \tag{8}$$

In the equation, $\gamma = 2$, and $b = 2$. ECANet performs global average pooling on the input feature map, transforming the feature map from a matrix of $[h, w, c]$ to a vector of $[1,1,c]$. The adaptive one-dimensional convolution kernel is then applied to the one-dimensional convolution to obtain the weights for each channel of the feature map. Finally, the normalized weights are multiplied by the original input feature map channel by channel to generate the weighted feature map. The ECANet attention mechanism uses convolution layers with very few parameters to replace fully connected layers with high parameter requirements, enabling the network model to have appropriate cross-channel interaction capabilities while significantly reducing complexity.

After analyzing the attention modules of SENet, ECANet, and CBAM, this paper adopts the ECANet attention mechanism as a method to improve the YOLOv5 model, enhancing the ability to focus on important features and reducing interference from irrelevant features through the ECANet module.

III. B. 3) Improvements to the IoU-Loss Algorithm

In the non-maximum suppression algorithm, the localization loss function calculates the distance deviation between the final output prediction box and the true box. Through error backpropagation, the weight parameter values are adjusted so that the output prediction box continuously approaches the true box.

In the YOLOv5 algorithm, the localization loss function uses CIoU-Loss. In addition, there are other loss functions: GIoU-Loss, DIoU-Loss, etc.

GIoU-Loss is an improvement over IoU-Loss, with its calculation formula given by Equation (9). C represents the minimum bounding rectangle between bounding boxes A and B and effectively measures the similarity of non-overlapping regions between bounding boxes. When two bounding boxes are in an inclusive relationship, GIoU degenerates into IoU and cannot distinguish relative similarity. The formula is:

$$L_{GIoU} = 1 - \left(IoU - \frac{C - A \cup B}{C} \right) \tag{9}$$

To address the issues associated with GIoU-Loss, DIoU-Loss omits the calculation of the minimum bounding rectangle area and introduces two bounding box distance variables to assist in measuring the similarity between two bounding boxes. The calculation formula is as follows:

$$L_{DIoU} = 1 - \left(IoU - \frac{\rho^2(b, b^{gt})}{c^2} \right) \tag{10}$$

In the equation, b is the center point of the predicted box, b^{gt} is the center point of the target box, ρ is the Euclidean distance between the two center points, and c is the diagonal distance of the smallest rectangle that can simultaneously cover both the predicted box and the true box.

Similar to GIoU-Loss, DIoU-Loss has a non-zero gradient when there is no intersection with the ground truth box, enabling optimization. By introducing a size difference term, CIoU can better handle changes in object shape and size differences, thereby providing a more accurate similarity metric. Its loss formula is defined as follows:

$$L_{CIoU} = 1 - \left(IoU - \frac{\rho^2(b, b^{gt})}{c^2} - \alpha v \right) \tag{11}$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \tag{12}$$

$$\alpha = \frac{v}{(1 - IoU) + v} \tag{13}$$

A good loss function should reduce the penalty for geometric factors when the ground truth boxes and predicted boxes overlap well, and less intervention during training can help the model achieve better generalization capabilities. WIoU-Loss builds on this by incorporating distance attention, which includes a dynamic non-monotonic mechanism. It designs a reasonable gradient gain distribution strategy that reduces large or harmful gradients in extreme samples. The formula for WIoU-Loss is shown in Equation (14):

$$L_{WIoU} = \frac{\beta}{\delta \alpha^{\beta - \delta}} RW_{IoU} L_{IoU} \tag{14}$$

$$R_{WIoU} = \exp\left(\frac{(x - x^{gt})^2 + (y - y^{gt})^2}{(w_s^2 + h_s^2)} \right) \tag{15}$$

$$\beta = \frac{L_{IoU}^*}{L_{IoU}} \in [0, +\infty) \tag{16}$$

$$L_{IoU} = 1 - IoU \tag{17}$$

In the equation, δ and α are adjustable hyperparameters, L_{IoU} is the IoU-Loss, and β represents the outlier degree of the candidate box, where a smaller outlier degree indicates higher quality of the candidate box.

After analyzing and comparing the advantages and disadvantages of various IoU loss functions, this paper decides to adopt WIoU-Loss as the loss function for candidate boxes in the citrus recognition network model.

IV. Test results and analysis

IV. A. Improved network model performance

IV. A. 1) Attention Module Comparison Test

To verify the impact of different attention mechanisms on the recognition performance of the proposed model, five attention mechanisms—ECANet, SE, CA, CBAM, and ECA—were respectively inserted into the network without modification, yielding the results of the attention mechanism comparison experiment. The results of the attention mechanism comparison experiment are shown in Table 2.

As shown in the table, the average precision and accuracy of the SE and CA modules inserted into the network are significantly lower than those of the ECANet and ECA modules. However, the recall rate improvement of the ECANet module is significantly higher than that of the ECA module. Compared to the other three networks, ECANet can consider both the spatial features of citrus target images and the channel features of images. Therefore, ECANet was selected for insertion into the backbone network.

Table 2: The attention mechanism compares the results of the test

Attention model	Accuracy rate/%	Recall rate/%	Mean accuracy/%	Model memory usage/MB
SE	89.63	85.23	90.24	15.20
CA	87.04	84.75	91.69	15.20
CBAM	92.22	89.14	92.75	15.20
ECA	93.75	90.25	94.39	15.20
ECANet	96.39	93.67	97.07	15.20

IV. A. 2) Loss function comparison test

A comparison of the five loss functions—WIoU, CIoU, GIoU, SIoU, and EIoU—is shown in Table 3.

For complex citrus features, GIoU only considers distance loss, resulting in a comprehensive decline in metrics such as precision and recall. EIoU separates the aspect ratio influence factors for separate calculation, slightly improving the model's average accuracy, with an average accuracy of 92.69%. SIoU considers multiple costs simultaneously and defines the width and height of the target box as consistency loss during loss calculation, improving the model's average accuracy by 1–2 percentage points. Based on comprehensive model performance evaluation metrics, this paper selects WIoU as the network loss function.

Table 3: Loss function comparison test

Loss function	Accuracy rate/%	Recall rate/%	Mean accuracy/%	Model memory usage/MB
CIoU	89.64	91.35	90.17	15.6
GIoU	83.77	85.82	86.44	15.2
SIoU	91.35	92.27	93.25	17.3
EIoU	86.52	89.64	92.69	16.8
DIoU	87.81	89.53	91.04	19.4
WIoU	92.02	93.17	94.58	15.7

IV. A. 3) Ablation test

To further validate the optimization effects of each improvement method on the final algorithm, each was added to the model for ablation testing.

The three improvement methods are modifying the WIoU loss function, the ECA attention mechanism, and the K-means+IoU distance. The testing method is as follows: each of the three improvement methods is added to the original YOLOv5 algorithm, and the effects are observed separately. All improvement modules are added, and the innovative algorithm is compared with the original YOLOv5.

The results of the ablation experiments are compared as shown in Table 4. As can be seen from the table, compared to the YOLOv5 model before improvement, the three improvement methods each improved the overall network performance in different aspects.

The improved K-means+IoU showed the most significant improvement in recall rate and average precision, with increases of 3.6 percentage points and 2.49 percentage points, respectively.

In the improved network model after applying the three improvement strategies, the average precision in the test set increased from 88.54% to 95.81%, while precision and recall also improved by 5.87 percentage points and 7.26 percentage points, respectively, indicating that the improved model exhibits good convergence.

Table 4: The ablation test results were compared

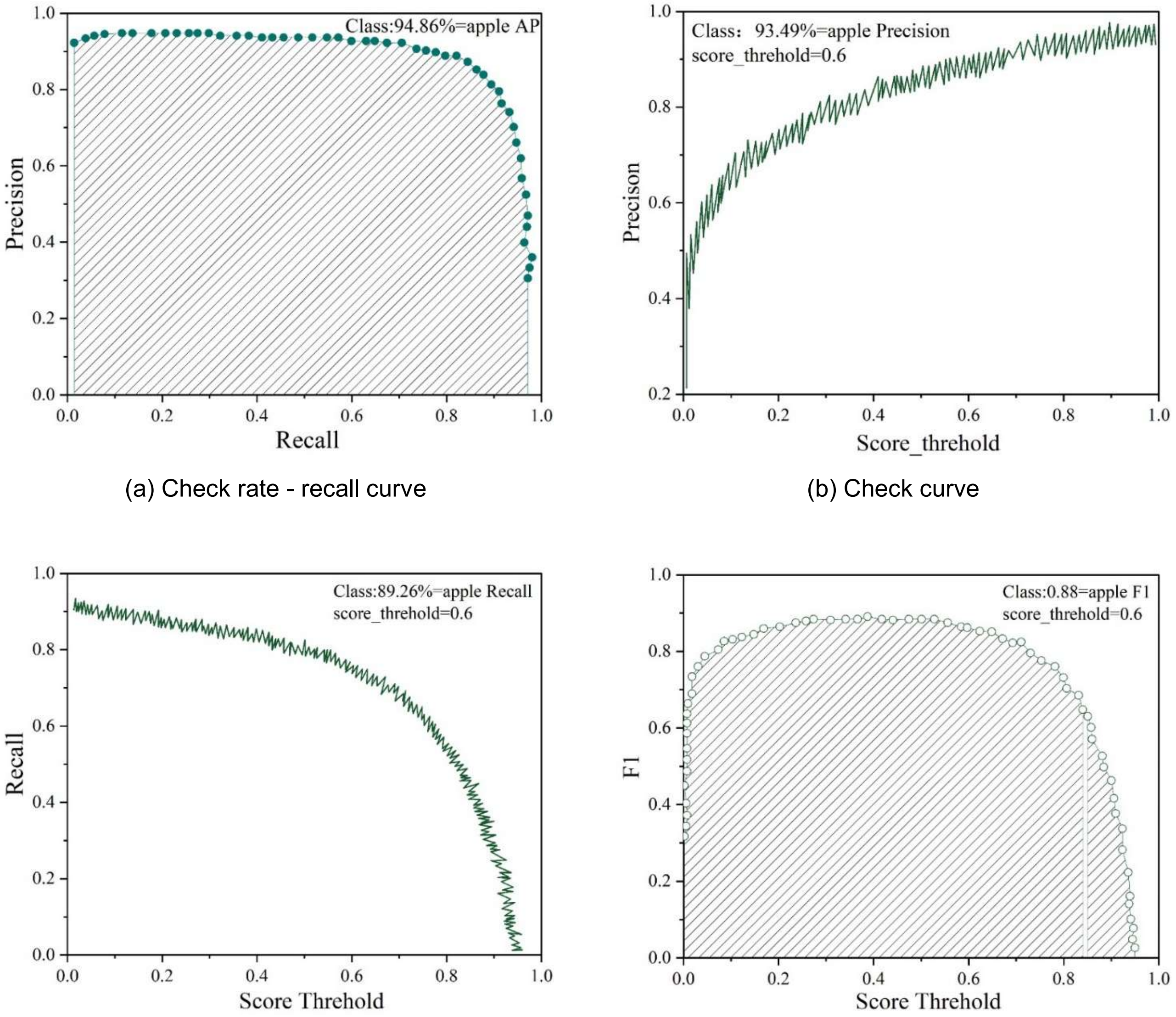
WIoU	ECA	K-means+IoU	Accuracy rate/%	Recall rate/%	Mean accuracy/%	Model memory usage/MB
×	×	×	87.65	89.01	88.54	15.4
√	×	×	88.21	86.64	89.62	15.7
×	√	×	90.36	89.53	91.87	14.6
×	×	√	91.25	90.85	91.03	14.3
√	√	×	90.03	91.63	90.85	15.2
√	×	√	89.78	91.45	90.19	15.0
×	√	√	92.36	95.32	94.23	14.8
√	√	√	93.52	96.27	95.81	14.1

IV. B. Comparison of recognition results

IV. B. 1) Analysis of experimental results

The improved network model YOLOv5 was trained using iterative autonomous deep learning on the citrus image dataset established in this paper. After training, the recognition results of the improved YOLOv5 algorithm were tested to verify whether the optimized neural network model has advantages.

The experimental results of the improved YOLOv5 network model for citrus image recognition are shown in Figure 2. The results indicate that the mAP value (mean average precision) of the proposed recognition and detection method is 94.86%, the precision rate P is 93.49%, the recall rate R is 89.26%, and the F1 value is 0.88.



(a) Check rate - recall curve

(b) Check curve

(c) Recall curve

(d) F1 curve

Figure 2: Improved yolov5 network model for the identification of citrus images

IV. B. 2) Comparison of experimental results

To further test and improve the accuracy of YOLOv5's citrus image recognition results, a comparative analysis was conducted between the recognition results of YOLOv5 and those of the known image recognition network models SSD, YOLOv4, and YOLOv5 for the same citrus image. The recognition prediction results of each network model for the same citrus image are shown in Table 5. The experimental results, including mAP values, precision P, recall R, and F1 values, were compared among the four algorithms.

The SSD model achieved an mAP value of 76.84%, a precision rate (P) of 94.21%, and a recall rate (R) of 70.63% for the citrus dataset in this study. The YOLOv5 model achieved an mAP value of 93.71%, a precision rate (P) of 92.36%, and a recall rate (R) of 90.63% for the citrus dataset in this study.

Based on the priorities set in this paper, the mAP value of the improved YOLOv5-based detection method proposed in this paper is the highest at 96.83%, which is 19.99% higher than SSD, 4.79% higher than YOLOv4, and 3.12% higher than YOLOv5.

The computational results indicate that the improved YOLOv5 network model achieves the highest accuracy in citrus recognition.

Table 5: The results of the model of the same citrus image recognition

Model	mAP/%	Accuracy ratio/%(P)	Recall rate/%(R)	F1 value
SSD	76.84	94.21	70.63	0.73
YOLOv3	89.23	92.24	79.52	0.86
YOLOv4	92.04	91.09	87.14	0.89
YOLOv5	93.71	92.36	90.63	0.91
Improved YOLOv5	96.83	94.67	95.76	0.93

IV. C. Three-dimensional coordinate acquisition experiment

The 3D coordinate acquisition experiment uses a Kinect V2 depth camera and a program written in the Python programming language. First, the camera is used to capture color images and depth images of the scene, and then these two types of images are fused. Then, the improved YOLOv5 model proposed in this paper is used to perform object detection on the fused images, yielding the bounding box information of the citrus objects in the images. The center point of this bounding box is calculated, representing the position of the citrus object's center point in the image. Subsequently, the transformation relationship between the pixel coordinate system and the camera coordinate system is calculated, and combined with the depth information stored in the fused images, the relative coordinates of the target fruit in the actual space with respect to the camera are determined.

The localization experiment results are shown in Table 6. The table displays the calculated localization coordinates, actual coordinates, and the coordinate error between the two.

The experimental results in the table indicate that the combination of the citrus recognition and spatial localization algorithms proposed in this paper achieves an average error of less than 2 mm between the calculated three-dimensional coordinates of the fruit and the actual coordinates, which is sufficient to meet practical application requirements.

Table 6: Location test results

Serial number	Location coordinates (X, Y, Z)/mm	Actual coordinates X, Y, Z)/mm	Coordinate error $(\Delta X, \Delta Y, \Delta Z)$ /mm
1	(30.2,16.8,86.9)	(29.7,15.9,87.3)	1.8
2	(18.9,60.4,72.5)	(19.4,60.6,72.1)	1.1
3	(23.7,7.9,98.3)	(24.4,8.2,97.7)	1.6
4	(-8.9,15.8,114.4)	(-8.3,15.4,113.8)	1.6
5	(36.4,31.2,62.1)	(36.7,31.5,62.5)	1.0
6	(-81.2,104.5,426.9)	(-81.3,103.8,427.1)	1.0
7	(104.7,-103.4,454.3)	(104.3,-103.2,454.1)	0.8
8	(135.1,352.1,521.6)	(134.7,351.3,521.2)	0.8
9	(-87.3,312.4,324.8)	(-87.1,312.2,324.6)	0.6
10	(75.8,-67.3,121.3)	(75.4,-67.1,120.9)	1.0

V. Conclusion

To address the issues of machine vision localization errors and poor citrus target recognition caused by overlapping or obstructed citrus fruits in natural environments, this paper proposes a citrus target recognition algorithm based on an improved YOLOv5 algorithm. By optimizing the prior boxes, the algorithm determines the more precise spatial coordinates of the target citrus fruits, thereby meeting the requirements for citrus target recognition.

After adding the ECANet module, WIoU loss function, and K-means+IoU distance to the base YOLOv5 algorithm, the average accuracy of the experiments improved from 88.54% to 95.81%, with precision and recall rates increasing by 5.87% and 7.26%, respectively. This demonstrates that the improved YOLOv5 network model exhibits good convergence. Experiments using the improved YOLOv5 algorithm for citrus image recognition achieved an mAP value and precision rate P value both exceeding 92%, with a recall rate R value reaching 89.26%. Multiple localization experiments showed that the three-dimensional coordinate errors of citrus fruit picking points were all within 2 mm, demonstrating excellent localization performance.

The YOLOv5 network model optimization designed in this paper can achieve automated recognition and spatial localization of citrus fruits in natural environments, meeting the visual system requirements for citrus picking robots to perform automated picking tasks, and can be further optimized for application.

References

[1] Lu, X., Zhao, C., Shi, H., Liao, Y., Xu, F., Du, H., ... & Zheng, J. (2023). Nutrients and bioactives in citrus fruits: Different citrus varieties, fruit parts, and growth stages. Critical Reviews in Food Science and Nutrition, 63(14), 2018-2041.

[2] Zou, Z., Xi, W., Hu, Y., Nie, C., & Zhou, Z. (2016). Antioxidant activity of Citrus fruits. Food chemistry, 196, 885-896.

[3] Huang, Z., Li, Z., Yao, L., Yuan, Y., Hong, Z., Huang, S., ... & Ding, J. (2024). Geographical distribution and potential distribution prediction of thirteen species of Citrus L. in China. Environmental Science and Pollution Research, 31(4), 6558-6571.

[4] Cong, L., Kaiwei, L., Jiquan, Z., Yueting, Y., Sicheng, W., & Chunyi, W. (2021). Refined climatic zoning for citrus cultivation in Southern China based on climate suitability. Journal of Applied Meteorological Science, 32(4), 421-431.

[5] Lv, H., Wu, S., Xie, Y., Liao, Y., & Zhang, S. (2022). Design of a Tracked Citrus Picking Robot. Open Access Library Journal, 9(3), 1-16.

[6] Yanqing, W., Yang, T., & Guangyou, Y. (2023). Design and experiment of control system for robot citrus picking. Journal of Chinese Agricultural Mechanization, 44(9), 146.

[7] Xiao, X., Wang, Y., Zhou, B., & Jiang, Y. (2024). Flexible Hand Claw Picking Method for Citrus-Picking Robot Based on Target Fruit Recognition. Agriculture, 14(8), 1227.

[8] Liu, Y. P., Yang, C. H., Ling, H., Mabu, S., & Kuremoto, T. (2018, November). A visual system of citrus picking robot using convolutional neural networks. In 2018 5th international conference on systems and informatics (ICSAI) (pp. 344-349). IEEE.

[9] Song, C., Wang, C., & Yang, Y. (2020, October). Automatic detection and image recognition of precision agriculture for citrus diseases. In 2020 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE) (pp. 187-190). IEEE.

[10] Zhuang, J. J., Luo, S. M., Hou, C. J., Tang, Y., He, Y., & Xue, X. Y. (2018). Detection of orchard citrus fruits using a monocular machine vision-based method for automatic fruit picking applications. Computers and Electronics in Agriculture, 152, 64-73.

[11] Cubero, S., Aleixos, N., Albert, F., Torregrosa, A., Ortiz, C., García-Navarrete, O., & Blasco, J. (2014). Optimised computer vision system for automatic pre-grading of citrus fruit in the field using a mobile platform. Precision agriculture, 15, 80-94.

[12] Safdar, A., Khan, M. A., Shah, J. H., Sharif, M., Saba, T., Rehman, A., ... & Khan, J. A. (2019). Intelligent microscopic approach for identification and recognition of citrus deformities. Microscopy research and technique, 82(9), 1542-1556.

[13] Tang, Y., Chen, M., Wang, C., Luo, L., Li, J., Lian, G., & Zou, X. (2020). Recognition and localization methods for vision-based fruit picking robots: A review. Frontiers in Plant Science, 11, 510.

[14] Xiao, X., Jiang, Y., & Wang, Y. (2024). A method of robot picking citrus based on 3D detection. IEEE Instrumentation & Measurement Magazine, 27(3), 50-58.

[15] Peng, K., Ma, W., Lu, J., Tian, Z., & Yang, Z. (2023). Application of machine vision technology in citrus production. Applied Sciences, 13(16), 9334.

[16] Daming, W., Yifei, H., Huaying, L., Yujiang, G., & Huibo, H. (2023). Research on image recognition algorithm of citrus picking robot. Journal of Chinese Agricultural Mechanization, 44(9), 222.

[17] Liu, C., Tao, Y., Liang, J., Li, K., & Chen, Y. (2018, December). Object detection based on YOLO network. In 2018 IEEE 4th information technology and mechatronics engineering conference (ITOEC) (pp. 799-803). IEEE.

[18] Jiang, P., Ergu, D., Liu, F., Cai, Y., & Ma, B. (2022). A Review of Yolo algorithm developments. Procedia computer science, 199, 1066-1073.

[19] Gothane, S. (2021). A practice for object detection using YOLO algorithm. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(2), 268-272.

[20] Wen, H., Dai, F., & Yuan, Y. (2021, January). A study of YOLO algorithm for target detection. In 26th international conference on artificial life and robotics (ICAROB) (Vol. 26, pp. 622-625).

[21] Zhu, Y., Zhou, J., Yang, Y., Liu, L., Liu, F., & Kong, W. (2022). Rapid target detection of fruit trees using UAV imaging and improved light YOLOv4 algorithm. Remote Sensing, 14(17), 4324.

[22] Wang, S., Li, Q., Yang, T., Li, Z., Bai, D., Tang, C., & Pu, H. (2024). LSD-YOLO: Enhanced YOLOv8n Algorithm for Efficient Detection of Lemon Surface Diseases. Plants, 13(15), 2069.

[23] Chen, W., Lu, S., Liu, B., Li, G., & Qian, T. (2020). Detecting citrus in orchard environment by using improved YOLOv4. Scientific Programming, 2020(1), 8859237.

[24] Mirhaji, H., Soleymani, M., Asakereh, A., & Mehdizadeh, S. A. (2021). Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. Computers and Electronics in Agriculture, 191, 106533.

[25] Lv, Q., Sun, F., Bian, Y., Wu, H., Li, X., Li, X., & Zhou, J. (2025). A Lightweight Citrus Object Detection Method in Complex Environments. Agriculture, 15(10), 1046.

[26] Liao, Y., Li, L., Xiao, H., Xu, F., Shan, B., & Yin, H. (2025). YOLO-MECD: citrus detection algorithm based on YOLOv11. Agronomy, 15(3), 687.

[27] Lin, Y., Huang, Z., Liang, Y., Liu, Y., & Jiang, W. (2024). Ag-yolo: A rapid citrus fruit detection algorithm with global context fusion. Agriculture, 14(1), 114.

[28] Wang, C., Han, Q., Li, C., Zou, T., & Zou, X. (2024). Fusion of fruit image processing and deep learning: a study on identification of citrus ripeness based on R-LBP algorithm and YOLO-CIT model. Frontiers in Plant Science, 15, 1397816.

[29] Wang, C., Luo, Q., Chen, X., Yi, B., & Wang, H. (2021, March). Citrus recognition based on YOLOv4 neural network. In Journal of Physics: Conference Series (Vol. 1820, No. 1, p. 012163). IOP Publishing.

[30] Xu, L., Wang, Y., Shi, X., Tang, Z., Chen, X., Wang, Y., ... & Zhao, Y. (2023). Real-time and accurate detection of citrus in complex scenes based on HPL-YOLOv4. Computers and Electronics in Agriculture, 205, 107590.

[31] Hou, C., Zhang, X., Tang, Y., Zhuang, J., Tan, Z., Huang, H., ... & Luo, S. (2022). Detection and localization of citrus fruit based on improved You Only Look Once v5s and binocular vision in the orchard. Frontiers in Plant Science, 13, 972445.

[32] Zhang, W., Wang, J., Liu, Y., Chen, K., Li, H., Duan, Y., ... & Guo, W. (2022). Deep-learning-based in-field citrus fruit detection and tracking. Horticulture Research, 9, uhac003.

[33] Xiao, X., Wang, Y., Jiang, Y., Wu, H., Zhang, Z., & Wang, R. (2024). AC-YOLO: citrus detection in the natural environment of orchards. Journal of Agricultural Engineering, 55(4).

[34] Yu, Y., Liu, Y., Li, Y., Xu, C., & Li, Y. (2024). Object Detection Algorithm for Citrus Fruits Based on Improved YOLOv5 Model. Agriculture; Basel, 14(10).

[35] Xiao, X., Huang, J., Li, M., Xu, Y., Zhang, H., Wen, C., & Dai, S. (2022). Fast recognition method for citrus under complex environments based on improved YOLOv3. The Journal of Engineering, 2022(2), 148-159.

[36] Zheng, T., Zhu, Y., Liu, S., Li, Y., & Jiang, M. (2025). Detection of citrus in the natural environment using Dense-TRU-YOLO. International Journal of Agricultural and Biological Engineering, 18(1), 260-266.

[37] Liang, Y., Jiang, W., Liu, Y., Wu, Z., & Zheng, R. (2025). Picking-Point Localization Algorithm for Citrus Fruits Based on Improved YOLOv8 Model. Agriculture, 15(3), 237.

[38] Johann S. J. C. Amorim,Accacio F. S. Neto,Rafael S. Chaves,Alessandro R. L. Zachi,Josiel A. Gouvêa,Fabio A. A. Andrade & Milena F. Pinto. (2025). Collaborative Inspection of Solar Panel Farms Using YOLOv5-Based Computer Vision and UGV-UAV Integration. Journal of Intelligent & Robotic Systems,111(2),66-66.

[39] Zhuoyuan Tang,Md Maruf Hasan & Thilo Strauss. (2025). Optimized YOLOv5 model for safety helmet and flame detection system. Signal, Image and Video Processing,19(8),595-595.

[40] Zidong Nie,Yitian Xu,Jie Zhao & Min Yuan. (2025). Fire classification and detection in imbalanced remote sensing images using a three-sphere model combined with YOLOv5. Applied Soft Computing,177,113192-113192.

[41] Wei Tian,Bazhou Li,Jingjing Cao,Feichao Di,Yang Li & Jun Liu. (2024). An Improved YOLOv5 Model for Concrete Bubble Detection Based on Area K-Means and ECANet. Mathematics,12(17),2777-2777.