Hindawi
Journal of Sensors
Volume 2023, Article ID 9754232, 1 page
https://doi.org/10.1155/2023/9754232

Retraction

Retracted: Fast Recognition Method for Multiple Apple Targets in Complex Occlusion Environment Based on Improved YOLOv5

Journal of Sensors

Received 19 December 2023; Accepted 19 December 2023; Published 20 December 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] Q. Hao, X. Guo, and F. Yang, "Fast Recognition Method for Multiple Apple Targets in Complex Occlusion Environment Based on Improved YOLOv5," *Journal of Sensors*, vol. 2023, Article ID 3609541, 13 pages, 2023.

Hindawi
Journal of Sensors
Volume 2023, Article ID 3609541, 13 pages
https://doi.org/10.1155/2023/3609541

Research Article

Fast Recognition Method for Multiple Apple Targets in Complex Occlusion Environment Based on Improved YOLOv5

Qian Hao, Xin Guo, and Feng Yang

School of Information and Communication Engineering, North University of China, Taiyuan 030051, China

Correspondence should be addressed to Xin Guo; s2005052@st.nuc.edu.cn

Received 31 August 2022; Revised 9 October 2022; Accepted 24 November 2022; Published 6 February 2023

Academic Editor: Yuan Li

The mechanization and intelligentization of the production process are the main trends in research and development of agricultural products. The realization of an unmanned and automated picking process is also one of the main research hotspots in China's agricultural product engineering technology field in recent years. The development of automated apple-picking robot is directly related to imaging research, and its key technology is to use algorithms to realize apple identification and positioning. Aiming at the problem of false detection and missed detection of densely occluded targets and small targets by apple picking robots under different lighting conditions, two different apple recognition algorithms are selected based on the apple shape features to study the traditional machine learning algorithm: histogram of oriented gradients + support vector machine (HOG + SVM) and a fast recognition method for multiple apple targets in a complex occlusion environment based on improved You-Only-Look-Once-v5 (YOLOv5). The first is the improvement of the CSP structure in the network. Using parameter reconstruction, the convolutional layer (Conv) and the batch normalization (BN) layer in the CBL (Conv + BN + Leaky_relu activation function) module are fused into a batch-normalized convolutional layer Conv_B. Subsequently, the CA (coordinate attention) mechanism module is embedded into different network layers in the improved designed backbone network to enhance the expressive ability of the features in the backbone network to better extract the features of different apple targets. Finally, for some targets with overlapping occlusions, the loss function is fine-tuned to improve the model's ability to recognize occluded targets. By comparing the recognition effects of HOG + SVM, Faster RCNN, YOLOv6, and baseline YOLOv5 on the test set under complex occlusion scenarios, the $F1$ value of this method was increased by 13.47%, 6.01%, 1.26%, and 3.63%, respectively, and the $F1$ value of this method was increased by 19.36%, 13.07%, 1.61%, and 4.27%, respectively, under different illumination angles. The average image recognition time was 0.27 s faster than that of HOG + SVM, 0.229 s faster than that of Faster RCNN, and 0.006 s faster than that of YOLOv6. The method is expected to provide a theoretical basis for apple-picking robots to choose a pertinent image recognition algorithm during operation.

1. Introduction

As a large apple producing country, China's apple production accounts for half of the global production. With population mobility and changes in land policies, the mechanization and intelligence of agricultural production are the main trends in future development [1]. Until now, the domestic apple-picking method was based on manual picking. However, mature apples are large in quantity and require much human labor and financial resources. Realize the intelligent and unmanned apple picking, thus freeing workers from repetitive manual labor [2]. Accurate and fast identification of apple targets is the premise of automatic picking. However, in the orchard environment, there are a lot of sticky, bagged, dense, occluded, and overlapping apples in long-distance pictures taken under different lighting conditions. Achieving rapid detection of multiple apple targets in a complex occlusion environment is crucial to the realization of intelligent apple picking. In the process of mechanization and intelligentization of production equipment, scholars in China and abroad have done much research on automated picking robots. Williams et al. developed a picking robot with four manipulators and a neural network-based method regarding fruit detection and picking of kiwi fruit cultivated

in a pergola [3]. In the United States, an apple-picking robot was developed in the form of a "vacuum cleaner," which sucks ripe fruits from fruit trees [4]. The fruit picking is achieved by the vacuum suction system, and the picking speed can reach up to 1 s/piece. There are also related research results in image recognition algorithms. Huang et al. proposed a multiscale feature fusion convolutional neural network for indoor small target detection. Using image enhancement technology to set up and amplify a date set, the Faster RCNN, YOLOv5, SSD, and SSD target detection models based on multiscale feature fusion were trained on an indoor scene data set based on transfer learning [5]. Lin and Li proposed an integrated circuit board (ICB) object detection and image augmentation fusion model based on YOLO. First, collect and use different types of ICBs as model training data sets and establish a preliminary image recognition model that can classify and predict different types of ICBs based on different feature points. Finally, there is a discussion of the applicability of the model to detect and recognize the ICB directionality in <1 s with a 98% accuracy rate to meet the real-time requirements of smart manufacturing [6]. Gan et al. use thermal imaging to address color similarity between immature citrus and leaves. Using the temperature difference between the fruit and the leaf surface, a tracking fruit counting algorithm is established to count the fruit in the thermal video [7]. Lei et al. used an RGB camera to collect bayberry images and established an adaptive image equalization model. The watershed transformation algorithm and convex hull theory were used to precisely localize the bayberry area against a complex background [8]. In deep learning, Yang et al. proposed an improved You-Only-Look-Once-v5 (YOLOv5) algorithm to detect the growth status of flowers at different flowering stages. The experimental results show that the accuracy of the improved algorithm is 5.4% higher than that of YOLOv5, which proves the effectiveness of the algorithm [9]. Wang et al. proposed a long-close distance-coordinated control strategy for a litchi picking robot. A long distance uses the YOLOv5 target detection network and the DBSCAN point cloud clustering method and uses the Mask RCNN instance segmentation method to segment the more distinctive bifurcated stems in the field of view. Through the processing of segmentation masks, a "Point + Line" dual reference model is proposed to guide robotic pick-up. Through the experiment, the success rate in the positioning of fruit branches reaches 88.46% [10]. Wu et al. proposed rachis detection and three-dimensional localization of cut-off point for vision-based banana robot. By building a new YOLOv5-B model and improving the loss function to improve the accuracy and speed, the contour of the rachis is segmented using an edge detection algorithm, and the optimal cut-off point is obtained as a scalar. Experiments show that the multitarget recognition rate of the YOLOv5-B model for bananas is 93.2%, and the average image processing time is only 0.009 s/piece, which can meet the robot's requirements for rachis segmentation [11]. Most of the existing apple detection algorithms cannot distinguish between the apples that are occluded by tree branches and occluded by other apples. Yan et al. proposed a lightweight apple target detection method based on improved YOLOv5s. BottleneckCSP module was improved designed to BottleneckCSP-2 module. SE module was inserted to the proposed improved backbone network. Finally, the initial anchor box size of the original network was improved [12]. Ji et al. proposed an apple object detection method based on Shufflenetv2-YOLOX to quickly and accurately detect and locate apples in the orchard natural environment. Using the lightweight network Shufflenetv2 added with the convolutional block attention module (CBAM) as the backbone, an adaptive spatial feature fusion (ASFF) module is added to the PANet to improve the detection accuracy. The trained network AP value of Shufflenetv2YOLOX is increased by 6.24%, and the detection speed is increased by 18% [13]. Yang et al. proposed a fast recognition method for multiple apple targets in dense scenes. The method drew on the idea of "point is the target" and realized the rapid identification of apple targets by predicting the center point of apple and the width and height of apple. By improving the CenterNet network, the Tiny Hourglass-24 lightweight backbone network was designed, and the residual module was optimized to improve the target recognition speed [14].

In summary, scholars in China and abroad have conducted in-depth research on the problem of fruit identification with numerous research results. Apples grow in complex agricultural environments where there are many uncertainty factors. How to select an algorithm model that is not affected by the complex environment of the orchard and train an algorithm model with high recognition accuracy and strong robustness is the research focus and difficulty in the field of fruit detection. Here, an apple detection algorithm is studied from two aspects. First, using shape features, apple recognition and localization are examined based on the HOG + SVM algorithm. Second, aiming at the problems of false detection and missed detection of sticking, bagging, dense, occluded, and overlapping apples under different lighting conditions, an improved deep learning algorithm YOLOv5 was proposed. In the fusion of convolutional layer (Conv) and batch normalization (BN) layer in CSP1_X structure, the CA mechanism module is embedded into the backbone network of the improved design to enhance the expressive ability of the features in the mobile network. Finally, for some occluded and overlapping targets, the IOU in nms is changed to DIOU_nms, which greatly improves the model's ability to recognize occluded targets. In comparison with the target detection methods using a sliding window, the proposed structure is more efficient, causes less false detection of the background, and is easier to achieve in real time. The experiments indicate that the recognition rate and operating efficiency of the developed algorithm are improved. The process steps of this study are shown in Figures 1(a) and 1(b).

2. Methods

The color features of images are easily affected by light, which causes color feature-based image recognition algorithms to make errors when detecting apples. In this section, we consider the apple shape as the key feature.

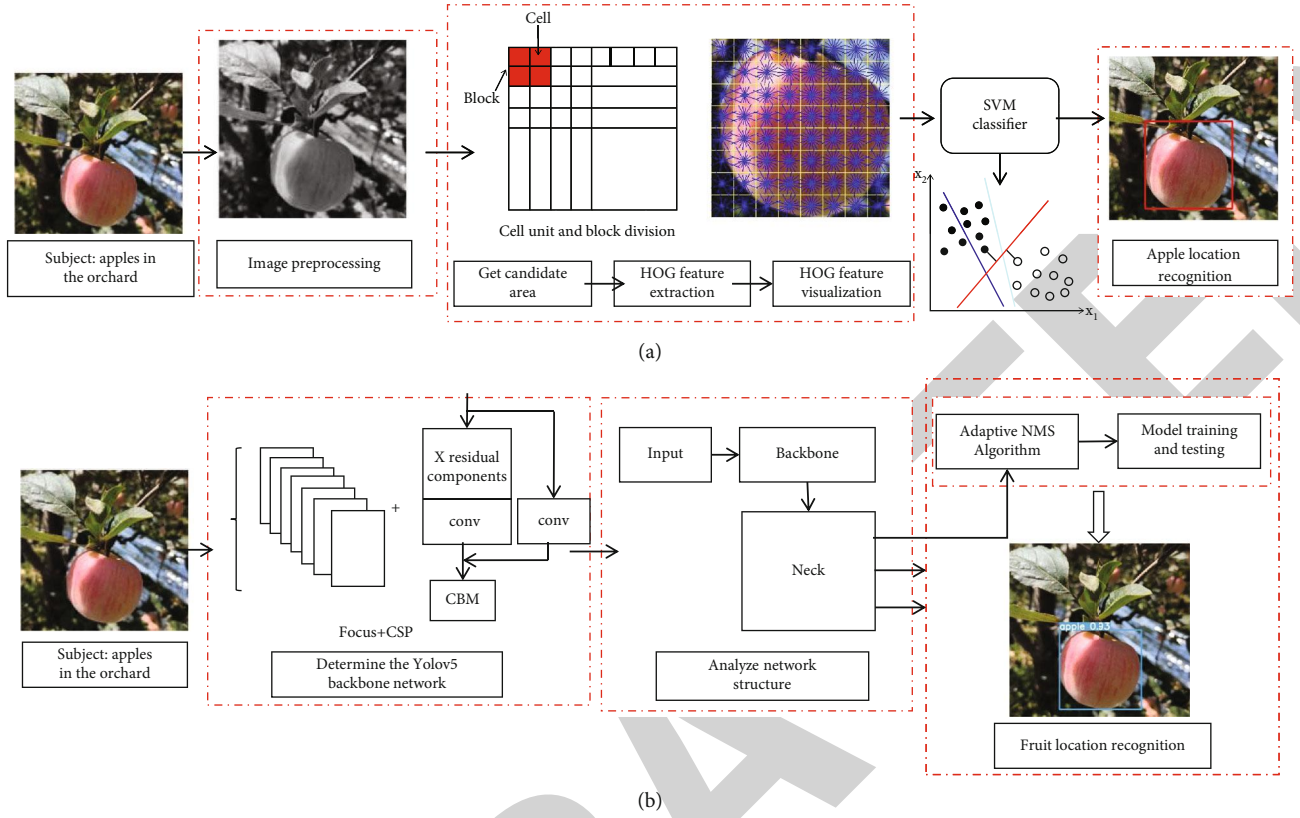

Figure 1: Diagram of two algorithms and key steps. (a) Flow chart of fruit identification and positioning based on HOG + SVM algorithm. (b) Flow chart of fruit identification and positioning based on YOLOv5 algorithm.

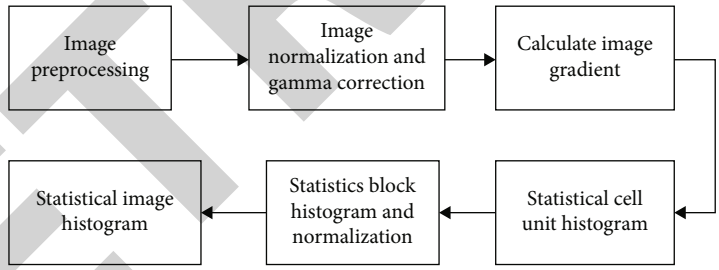

Figure 2: Schematic diagram of process for feature vector generation.

2.1. Image Normalization. During the testing process of the apples, there are obvious differences in exposure because of the influence of weather. This results in occluded views and increased background interference, making fruit detection more difficult. To solve this problem, the industry has adopted gamma correction [15]. This method can reduce the influence of shadows in some areas of the picture as much as possible, to meet the requirements. In general, there are two types of processing methods, i.e., square root and logarithmic. The square root method is used in this study, and the expression is provided in

$$I(x, y) = I(x, y)^{\gamma}, \tag{1}$$

where I represents the image, (x, y) represents the pixel point of the image, and γ generally takes on the value of 1/2.

2.2. Apple Shape Feature Extraction Based on HOG. HOG is an algorithm for summarizing features using histograms of oriented gradients. The principle is to express the gradient information of pixels or blocks of the picture and then calculate the feature information of the picture [16]. HOG is a shape feature operator, which is used to describe shape features such as image contours or edges. The basic idea of this operator is that the density of the gradient direction can describe the contour or edge of the object. The process of describing the shape of the object and generating the feature vector by the HOG operator is shown in Figure 2.

The first step is to normalize the image and perform the gamma correction operation to improve the robustness of the image to illumination changes while reducing the interference of noise. The next step is to solve the gradient information in the image, including the gradient of the image in

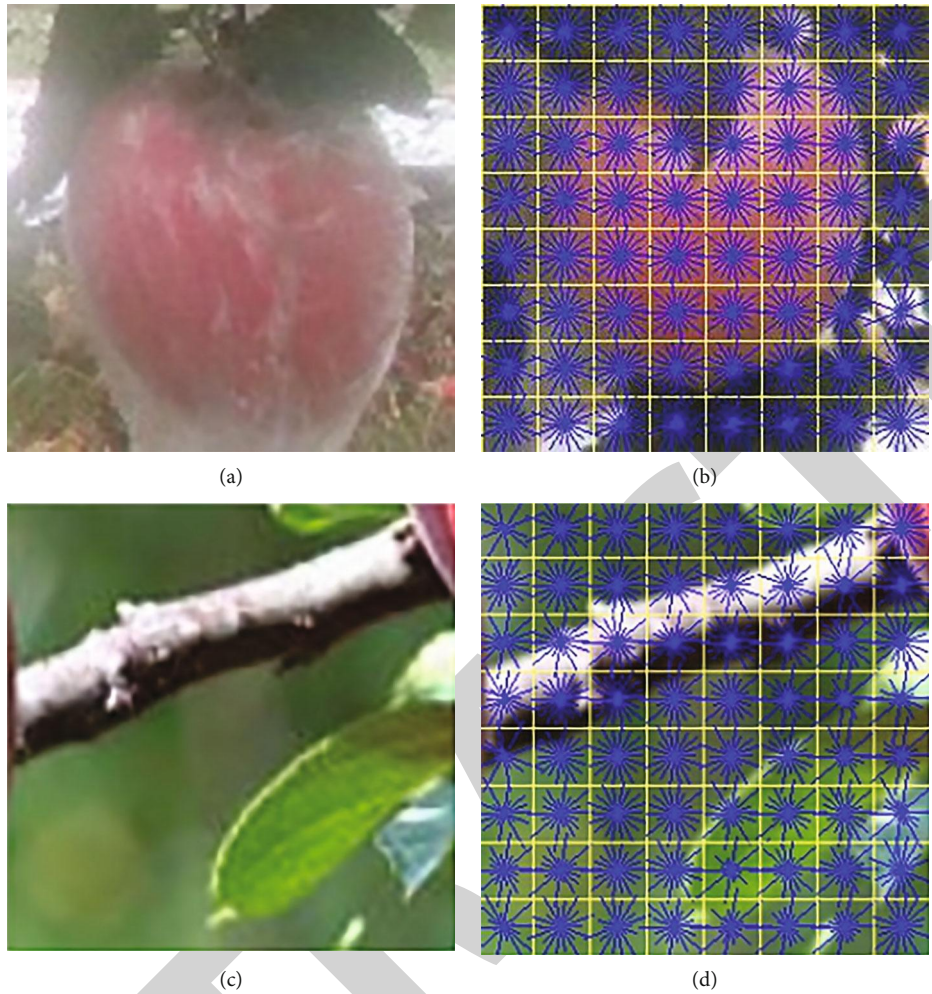

Figure 3: HOG feature visualization of target image.

the x and y directions. This is expressed in Equation (2) as follows:

$$G_x(x,y) = H(x+1,y) - H(x-1,y), \tag{2}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1). \tag{3}$$

In Equations (2) and (3), $G_x(x,y)$ is the gradient in the x direction, and $G_y(x,y)$ is the gradient in the y direction at pixel coordinates (x,y). $H(x,y)$ is the pixel value. The gradient direction and magnitude at this pixel point are obtained by calculating the gradient in each direction:

$$\begin{aligned} G(x,y) &= \sqrt{G_x(x,y)^2 + G_y(x,y)^2}, \\ G(x,y) &= \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right). \end{aligned} \tag{4}$$

The next step divides the image into a series of small cell units. Histograms are obtained for the direction and magnitude of gradients in cell units. Then, adjacent cell units are combined into blocks, i.e., statistical histograms and normalization. Next, the gradient information of each block is distributed into the direction histogram according to the gradient direction. The value range of the direction of the histogram can be between 0 and 180 degrees or 0 and 360 degrees. The specific range depends on whether the calculated gradient is positive or negative. Finally, by normalizing the block histogram, the local contrast adjustment is obtained to reduce the influence of illumination. Then, the histograms of all blocks in the image are concatenated to obtain the image histogram, i.e., the HOG feature vector of the image. According to the

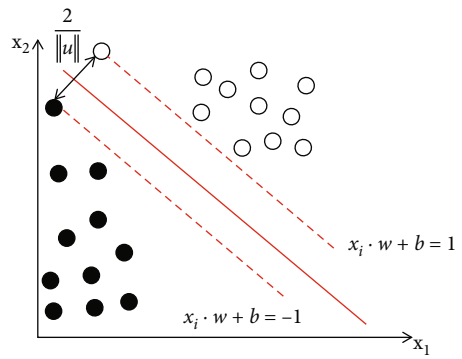

Figure 4: SVM segmentation diagram.

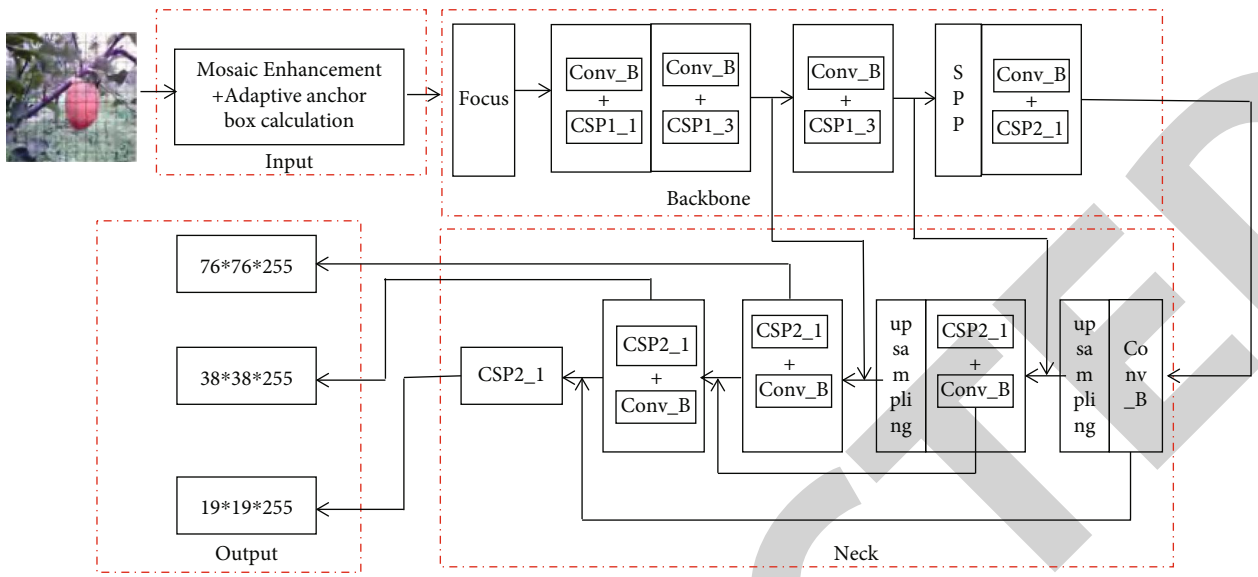

Figure 5: Framework structure of improved YOLOv5.

flowchart and aforementioned steps, after feature extraction is performed for apple and nonapple objects, the features are visualized, as shown in Figure 3.

2.3. Target Image HOG Feature Visualization. SVM is responsible for training and classifying the feature vectors extracted by HOG. The specific process is that after inputting the image to be detected, the SVM classifier performs a correlation operation on each feature in the image to obtain the category information of the features. The loss function and corresponding label are then used to calculate the parameters, and the descent method is iteratively applied to update the parameters until an optimal value is reached. The key to fruit identification and localization in the candidate area lies in the change range of the sliding window and step sizes during the traversal process. If the variation range is too small, the amount of computation increases; if the variation range is too large, the recognition and positioning accuracy may be reduced.

The principle of the SVM classifier is to find a hyperplane that separates the 2-class samples at the largest interval. We consider the set $D = \{x_i, y_i\}$, where x_i is the i^{th} eigenvector, and where y_i represents the class label corresponding to vector x_i [17]. Positive samples in the dataset are marked as +1, and negative samples are marked as -1. The set D contains two different types of data. As shown in Figure 4, a solid line is drawn to separate all points belonging to one class from those belonging to another class. Then, a hyperplane in which two-dashed lines are parallel to the solid line is drawn with a distribution symmetrical about the solid line as the center, passing through the sample points closest to the optimal hyperplane in the two types of data.

The hyperplane expression is shown in

$$wx + b = 0, \tag{5}$$

where x is a point in the set $\{x_i, y_i\}$ and w is a normal vector of the linear function. The two-dashed lines are parallel to the solid line and symmetrically distributed around it. The expressions of the two dotted lines are given by

$$x_i \bullet w + b = 1, \tag{6}$$

$$x_i \bullet w + b = -1. \tag{7}$$

The maximum distance between the two types of data is the distance between the two-dashed lines, which is $2/\|w\|$. When u is smaller, the spacing is larger.

3. Apple Recognition Algorithm Based on Improved YOLOv5

First, this section introduces and analyzes the overall improved framework structure of YOLOv5. Second, a loss function is introduced to ensure that the loss value of the network model during the training process displays a gradient descent trend.

3.1. Overall YOLOv5 Structure. The You-Only-Look-Once (YOLO) algorithm [18] was proposed at the 2016 Computer Vision Summit (CVPR); this neural network only needs to look once to identify the category and location of the target in the image. This algorithm takes an image as the input and provides the category and location of the target as output. After continuous advancements of the YOLO algorithm, the version v1 was updated to the version v7. Some improvements were added based on the original YOLOv5. The first is the improvement of the CSP structure in the network. Using parameter reconstruction, the convolutional layer (Conv) and the batch normalization (BN) layer in the CBL module were fused into a batch-normalized convolutional layer Conv_B. It is equivalent to fusing the parameters of convolution and batch normalization, so that the Conv_B convolution layer has the characteristics of batch normalization. The improved framework structure of YOLOv5 is shown in Figure 5.

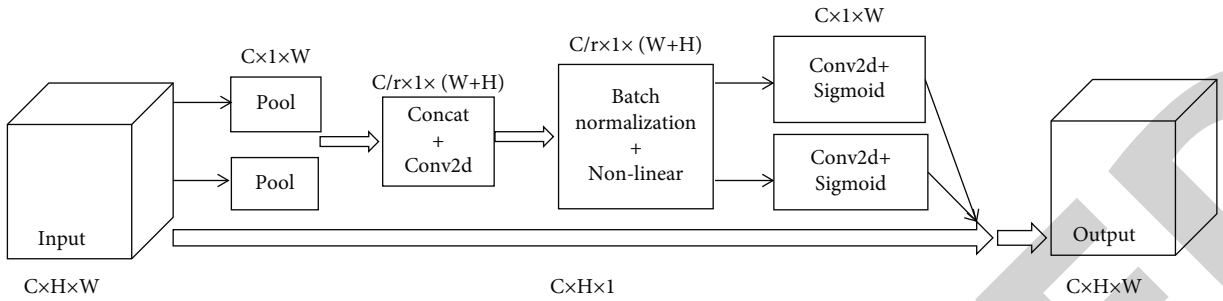

Figure 6: Structure diagram of CA module.

Table 1: Laboratory environment.

Lab environment	Configure
Operating system	64 bit windows 10
Graphics card	GTX1070
RAM	8GB
Video memory	4GB
Processor	Intel Core i7-6700 3.40GHz

The YOLOv5 network structure is mainly composed of input, backbone, neck, and output. First, when an apple image is input, the input terminal uses mosaic to expand the dataset. The pictures are randomly cropped and zoomed, and multiple photos are stitched to complete the adaptive anchor frame setting of the dataset [19]. Then, in the backbone network, the Focus module cuts the input image through the slicing operation and slices and splices the original image to a size of $608 \times 608 \times 3$. The convolutional layer outputs a feature map of $304 \times 304 \times 32$. In the CSP1_X module, X represents X residual components. The initial input is sent to the two branches, respectively, and convolution operations are performed on these two branches to reduce the number of channels of the feature map by half. Among them, one branch performs a batch-normalized convolutional layer, and after passing through the Conv2d layer, another branch is used to connect the output feature map using a concat operation. Finally, the output feature map is obtained after passing through the BN layer and convolution layer, in turn. The Neck network adopts the multiscale fusion of the FPN feature pyramid and PAN structure to better solve the scale problem of target detection. PAN is a bottom-up structure, which can strengthen the ability of network feature fusion and complete the positioning function. Finally, the output typically includes CIOU_loss and DIOU_nms. CIOU_loss generally solves the problem of nonoverlapping bounding boxes and bounding box scale information. Considering the location information of the center point of the bounding box, DIOU_nms is more suitable for target detection and provides better results for detection in complex scenes.

3.2. The Addition of the Attention Mechanism Module. Due to the complexity of the apple background, the branches and leaves in the environment interfere with the image recognition. Therefore, to improve the recognition accuracy of

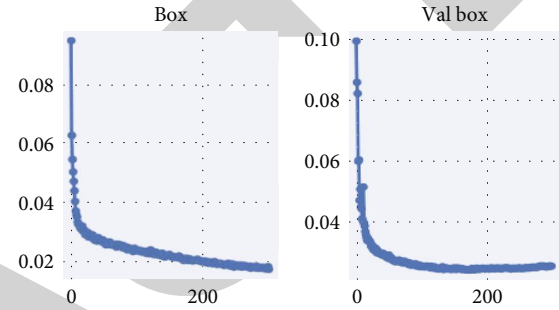

Figure 7: Model training loss curve.

the fruit picking method, the attention mechanism CA module is introduced into the YOLOv5 network to better extract different fruits. The features of the image can improve the generalization ability of the model.

The CA module is a new mobile network that decomposes the acquired two-dimensional features into two one-dimensional vector encoding processes and aggregates the features along two different directions in space. The purpose of this is to encode the generated feature maps as a pair of orientation-aware and position-sensitive feature maps, which can be complementarily applied to the input feature maps to enhance the features of objects of interest. The structure diagram of the CA module is shown in Figure 6.

The next step is to insert the CA module into the YOLOv5 network. The location where the attention mechanism is placed is not unique; thus, the location of the insertion must be first determined. We observe the network frame diagram of YOLOv5; the output has a detection layer of 76*76*255, which is concat spliced by the fifth layer of the backbone network and the seventh layer after twice upsampling in the Neck network and finally output by CSP2_1. Therefore, the CA module is added after the fifth layer of the backbone network, so that the high-dimensional feature map in the Neck network is fused with the features of the CA module, and the detection effect of large-scale objects is improved.

Subsequently, the CA module is added to the seventh layer of the backbone network, and it is spliced with the high-dimensional feature map concat obtained after the first upsampling of the Neck network, so that the output is a 38*38*255 detection layer. The detection effect of medium-sized objects is improved. Finally, the CA module is added to the eleventh layer of the backbone network to improve

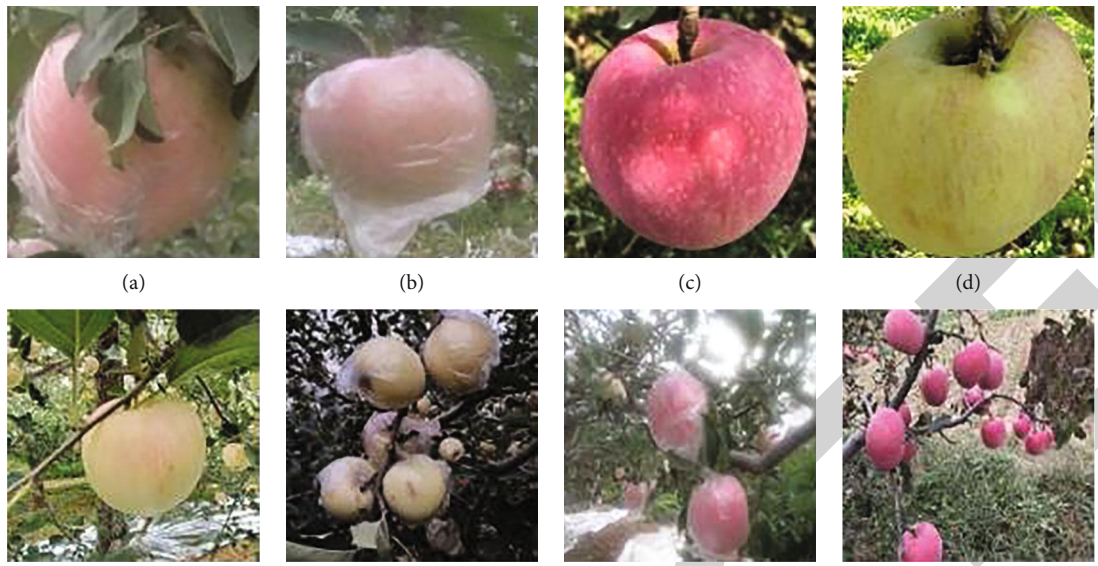

Figure 8: Different data samples for the two algorithms.

the detection ability of the 19*19*255 detection layer for small targets. The network is optimized from three scales to improve the overall recognition effect of the YOLOv5 detection model.

3.3. YOLO Loss Function. The YOLO algorithm has an important concept, i.e., the loss function, the structure of which is displayed in Figure 6. When analyzed by the YOLO algorithm, there are three types of errors, i.e., position, C value, and classification; thus, a loss function is introduced to balance these deviations. The loss of the entire algorithm consists of three parts, i.e., the coordinate error generated by the predicted and real frames, confidence error of the object, and error of predicted category.

$$\begin{aligned}
&\lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[\left(x_i - \widehat{x}_i \right)^2 + \left(y_i - \widehat{y}_i \right)^2 \right] \\
&+ \lambda_{\text{coord}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left[\left(\sqrt{\omega_i} - \sqrt{\widehat{\omega}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\widehat{h}_i} \right)^2 \right] \\
&+ \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{obj}} \left(C_i - \widehat{C}_i \right)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{s^2} \sum_{j=0}^{B} 1_{ij}^{\text{noobj}} \left(C_i - \widehat{C}_i \right)^2 \\
&+ \sum_{i=0}^{s^2} 1_i^{\text{obj}} \sum_{c \in \text{classes}} \left(p_i(c) - \widehat{p}_i(c) \right)^2.
\end{aligned} \tag{8}$$

Among them, the five parameters (x, y, w, h, and c) represent the x-axis coordinate, y-axis coordinate, target width, target height, and target category, respectively; $p_i(c)$ is the confidence level of the predicted target; λ_{coord} and λ_{noobj} are weight variables with the role of balancing the proportions. S is every grid cell; B is every bounding box; obj is the detection frame containing the target; and noobj indicates that the detection frame does not contain a target.

3.4. Experimental Evaluation. In the final result, if the apple fruit is identified as apple fruit, it means that the algorithm judges the correct result as correct; otherwise, it is defined as TP (true positives); if the background is identified as an apple fruit, it means that the algorithm judges the wrong result as correct, and it is defined as FP (false positives); If the apple fruit is judged as the background, it means that the algorithm judges the correct result as an error, and it is defined as FN (false negatives).

In model evaluation, the accuracy rate (P), recall rate (R), and $F1$ value are mainly used.

The calculation formula of P is shown in

$$P = \frac{TP}{TP + FP}. \tag{9}$$

The calculation formula of R is shown in

$$R = \frac{TP}{TP + FN}. \tag{10}$$

The formula for calculating the $F1$ value is shown in

$$F_1 = \frac{2PR}{P + R}. \tag{11}$$

After the training of the dataset is completed, the weight file obtained in this training will be generated. The obtained weight file can be selected by setting the weight path, and the apple fruit can be tested by using the trained weight file.

Figure 9: Results of identification and positioning of apples in different states by four algorithms. (a) Recognition effect of HOG + SVM algorithm in different states. (b) Recognition effect of Faster RCNN algorithm in different states. (c) Recognition effect of YOLOv6 algorithm in different states. (d) Recognition effect of baseline YOLOv5 algorithm in different states. (e) Recognition effect of the improved YOLOv5 algorithm in different states.

4. Model Training and Experimental Analysis

4.1. Training Environment. The algorithm needs to perform many operations for feature extraction and classification; thus, it requires high performance machines. The details of the implementation are presented in Table 1.

During training, pretraining weights on the official website are downloaded, and the network model is fine-tuned based on a certain pretraining weight through transfer learning. The advantage of this is that it can speed up the convergence of the model loss function and improve the detection accuracy. Setting reasonable hyperparameters also affect the quality of model training. During model training, the batch size is set to 8. The momentum is set to 0.937, the decay is set to 0.0005, the learning rate is set to 0.01, and the epoch is set to 300.

4.2. Training Results. The loss curve of the model training is shown in Figure 7. As can be seen from Figure 7, the left side is the training set iteration curve, and the right side is the validation set iteration curve. It can be seen from the figure that the loss value on the training set and the validation set drops rapidly during the first dozens of rounds of training, and after 250 rounds of training, the loss value basically tends to be stable. Therefore, in this study, the model output after 300 rounds of training is determined as the model for apple detection.

4.3. Experimental Settings. The dataset consists of real apple data captured in apple orchards; whereby, a total of 3000 apple images were collected. Mosaic was introduced to expand the dataset; particularly, random scaling adds many small targets, which improves the generalization ability of the apple detection model [20]. Then, all the images were divided into three sets: the training set, test set (70% of images), and validation set (30% of randomly selected images). When studying the apple recognition algorithm based on shape features, because the HOG shape feature operator does not need to consider the color, a variety of fruit images of different colors were introduced as positive samples. When researching the improved YOLOv5 apple recognition algorithm, pictures with different occlusion degrees, growth stages, and states as training data were found. These different data samples are shown in Figure 8.

4.4. Results and Analysis

4.4.1. Comparison of the Results of Different Target Detection Algorithms. The experiments randomly tested the recognition effect of the algorithm for various real-world images. To verify the accuracy of the algorithm for apple recognition and localization in different distribution states, in this study, the test was divided into five categories according to the above criteria, and statistics were carried out. Figures 9(a)–9(e) show the recognition effects of the HOG + SVM, Faster RCNN, baseline YOLOv5, YOLOv6, and improved YOLOv5 algorithms for images in the unobstructed, fruit sticking, and bagging states.

In the figure, it is observed that the image collected in the unobstructed state (first picture of each group) is brighter in both the fruit and background. The outline shape of the apple is very clear. In the image collected in the state of fruit adhesion (second picture of each group), the nearby fruit is easy to identify. However, after the fruit branches and leaves in the distance are blocked, the light is dark, and identification becomes difficult. The images collected in the bagging state (third picture in each group) have a relatively bright background, but the outline of the fruit is partially deformed by the bag. In the above comparison, many deep learning models are selected, and the traditional machine learning model is only HOG + SVM, because HOG + SVM is widely used in traditional target detection algorithms. HOG can better describe shape features and maintain good invariance to image geometric and optical deformation. Classify the HOG feature vector through SVM and use the trained SVM classifier to complete the classification of the localized fruit. In the detection results, the HOG + SVM algorithm displays better recognition ability for apples with clear shape features and little difference in size. However, in the state of adhesion and bagging, due to the lack of light and outline, some features of the apple are missing. The other two algorithms show stronger feature extraction ability, because the deep neural network displays better recognition effect compared to HOG + SVM.

Table 2: Evaluation results of 5 algorithms under different lighting angles.

Method	Precision/%	Recall/%	*F*1/%	Average image recognition time/s
Hog + SVM	82.23	79.85	81.5	0.296
Faster RCNN	91.76	86.34	88.96	0.255
YOLOv6	94.81	92.65	93.71	0.024
Baseline YOLOv5	90.54	92.16	91.34	0.032
Improved YOLOv5	96.35	94.6	94.21	0.026

4.4.2. Recognition of Different Lighting Angles by Different Detection Algorithms. In order to verify the superiority of the method proposed in this paper, we added the more advanced version YOLOv6 to the comparative experiment. In this test, we used the light angle of the camera as a control variable, which are three shooting lights: backlight, side light, and front light. Fifty images were randomly selected with different lighting angles in the complex occlusion environment including backlight, side light, and front light, as the test set data. The detection effects of the HOG + SVM, Faster RCNN, YOLOv6, baseline YOLOv5, and improved YOLOv5 algorithms are compared under the above conditions. The evaluation indicators are the accuracy rate precision, *F*1 comprehensive evaluation score, recall rate, and detection speed to evaluate the algorithm performance. The calculated evaluation values are weighted, averaged, and recorded in Table 2, and the overall results are shown in Figure 10.

As shown in Figure 10, apples have clear veins and average surface light intensity in the case of side light, making identification easy. In the case of backlight, the light

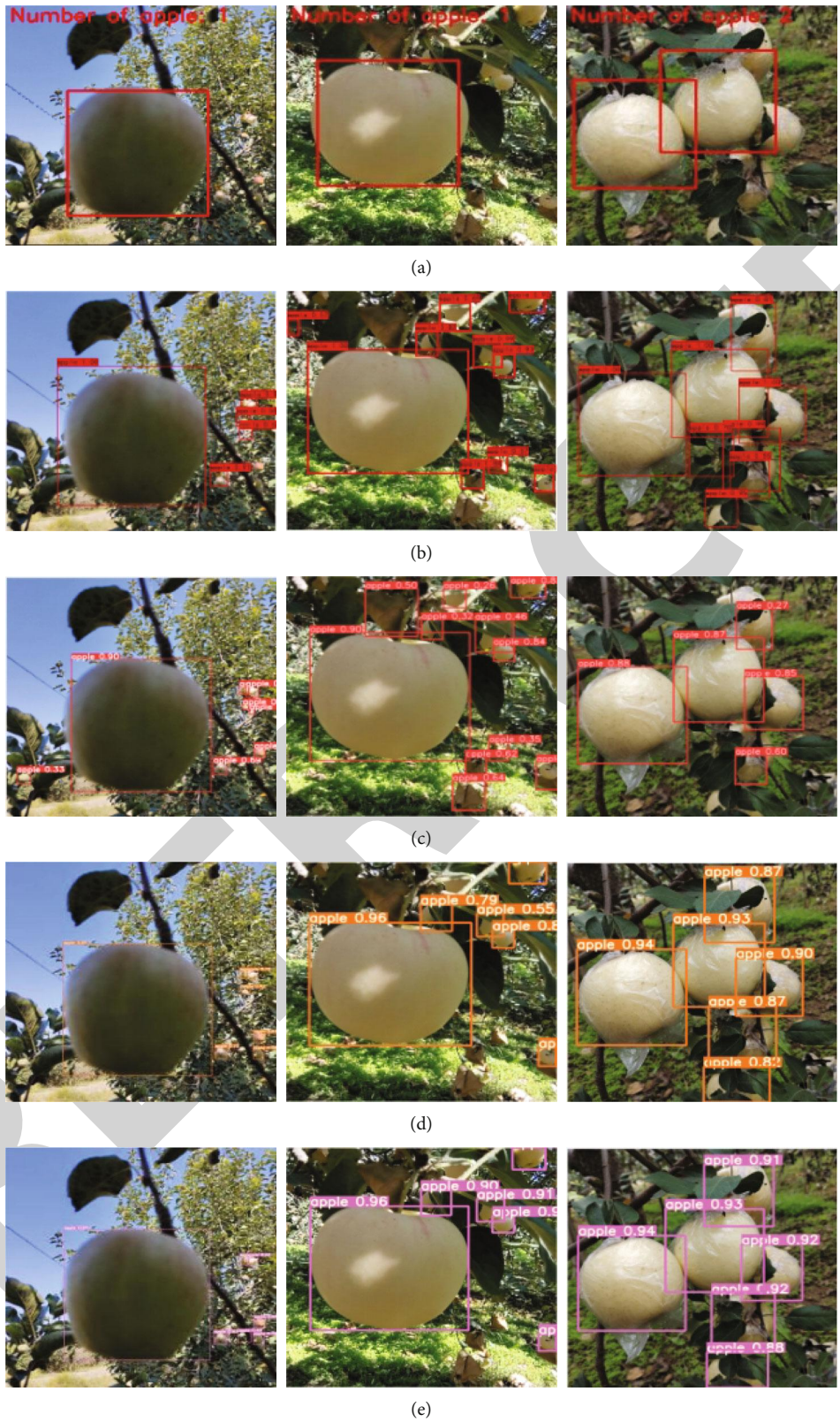

Figure 10: Results of identification and positioning of apples under different lighting angles by four algorithms. (a) Recognition effect of HOG + SVM algorithm under different illumination angles. (b) Recognition effect of Faster RCNN algorithm under different illumination angles. (c) Recognition effect of YOLOv6 algorithm under different illumination angles. (d) Recognition effect of the baseline YOLOv5 algorithm under different illumination angles. (e) Recognition effect of the improved YOLOv5 algorithm under different illumination angles.

Figure 11: Results of identification and positioning of apples under different occlusions by four algorithms. (a) Recognition effect of HOG + SVM algorithm under different occlusions. (b) Recognition effect of Faster RCNN algorithm under different occlusions. (c) Recognition effect of YOLOv6 algorithm under different occlusions. (d) Recognition effect of the baseline YOLOv5 algorithm under different occlusions. (e) Recognition effect of the improved YOLOv5 algorithm under different occlusions.

Table 3: Evaluation results of 5 algorithms under different occlusions.

Method	Precision/%	Recall/%	$F1$/%	Average image recognition time/s
Hog + SVM	73.50	75.34	74.4	0.296
Faster RCNN	81.06	80.32	80.69	0.255
YOLOv6	93.76	90.6	92.15	0.024
Baseline YOLOv5	90.23	88.76	89.49	0.036
Improved YOLOv5	94.64	92.89	93.76	0.026

intensity of the appearance of apples and branches is significantly reduced, and it is difficult to distinguish between the two. Additionally, the detection of smaller objects behind is difficult. In the case of smooth light, the light intensity of part of the surface of the apple becomes higher, and the outer surface shows a daytime color without the characteristics of veins [21].

As presented in Table 2, HOG + SVM is a fruit recognition method based on shape features, and the light has little influence on it. The disadvantage it faces is that it can only accurately identify targets with obvious shape features. The $F1$ of the improved YOLOv5 was 1.26% and 3.63% higher than that of YOLOv6 and the baseline YOLOv5, respectively, and the precision was 0.54% and 4.81% higher than that of YOLOv6 and the baseline YOLOv5, respectively. The $F1$ of Faster RCNN was 6.01% lower than that of the improved YOLOv5, and the precision was 3.59% lower than that of the improved YOLOv5. The $F1$ of the improved YOLOv5 was 13.47% higher than that of HOG + SVM, and the precision was 12.12% higher than that of HOG + SVM. It can be concluded that the evaluation index of the deep learning algorithm is generally higher than that of the traditional algorithm, because the model trained by the neural network has a strong generalization ability and still has a good effect when the color features are not obvious. The detection speed of the improved YOLOv5 and YOLOv6 is similar, leading the other two algorithms as a single-stage detection algorithm.

4.4.3. Recognition Results of Different Detection Algorithms for Different Occlusions. Next, the comparison test of the occlusion degree of the apple surface is performed. In this test, the occlusion degree of the apple surface is used as a control variable. Similarly, fifty images of fruit adhesion, bagging, dense, and overlapping situations in complex occlusion environments were randomly selected as test set data and tested through four algorithms. The overall results are shown in Figure 11, and the calculated evaluation values are weighted and averaged and recorded in Table 3.

As shown in Figure 11, the unobstructed apples have distinct veins and clear outlines and are clearly distinguished from the background. The slightly shaded apples are lighter in color, and some of their shape features are covered by branches and leaves. The heavily occluded surface of the apple interfered with the color and vein characteristics of the apple surface, and the shape of the fruit was partially deformed by the bag. In addition, the redundant part of the plastic bag made the appearance of the apple no longer round; thus, the appearance characteristics were no longer clear.

As presented in Table 3, although the HOG + SVM algorithm can identify apples, when it is blocked by branches and leaves, the detection effect is not ideal. The precision of the improved YOLOv5 was 0.88% and 4.41% higher than that of YOLOv6 and the baseline YOLOv5, respectively, and the $F1$ was 1.61% and 4.27% higher than that of YOLOv6 and the baseline YOLOv5 because the improved backbone network extracts the features of different apple targets better and improves the recognition accuracy. For bagged apples, the algorithm recognizes the plastic film as a feature of the apple during training and detects it together during detection, increasing the recognition rate instead. The precision of the improved YOLOv5 is 13.58% higher than that of Faster RCNN, and the $F1$ is 13.07% higher than that of YOLOv6 because the Faster RCNN region suggests that some small targets are ignored in network prediction, resulting in poor model detection effect.

To sum up, in the comparison between the traditional HOG + SVM algorithm and the deep learning algorithm, although the HOG + SVM algorithm can identify apples under various conditions, the precision and $F1$ value are much lower than the existing deep learning algorithms. The reason why the traditional machine learning model HOG + SVM is chosen is to verify that the traditional method HOG + SVM algorithm has the shortcomings of long detection time and low accuracy. As can be seen from Tables 2 and 3 of the revised manuscript, the improved YOLOv5 algorithm model trained in the text is better than Faster RCNN in any sample, and the gap with YOLOv6 is not obvious. However, in the comparative test, the $F1$ score is obviously better than that of YOLOv6. The improved YOLOv5 algorithm model trained in the article can be competent for the target recognition of light changes, occlusion, and bagging apples, so that the apple picking robot using this visual algorithm can realize the recognition of multiple apples in a complex occlusion environment. The superiority of the proposed YOLOv5 method is demonstrated.

5. Conclusion

In order to solve the problem of false detection and missed detection of apple picking robots in complex occlusion environments, this paper studies the traditional machine learning algorithm HOG + SVM and proposes a fast recognition method for multiple apple targets in complex occlusion environments based on improved YOLOv5. First, through the improvement of the CSP structure in the network, the convolutional layer (Conv) and the batch normalization (BN) layer in the CBL (Conv+BN + Leaky_relu activation function) module are fused into a batch-normalized convolutional layer. Subsequently, the CA (coordinate attention) mechanism module is embedded into different network layers in the improved designed backbone network to enhance the expressive ability of the features in the backbone network to better extract the features of different apple

targets. For some targets with overlapping occlusions, the IOU in nms is changed to DIOU_nms, which greatly improves the model's ability to recognize occluded targets. Finally, the experimental comparison shows that under different illumination angles, the $F1$ value of the improved YOLOv5 algorithm is 13.47%, 6.01%, 1.26%, and 3.36% higher than that of HOG + SVM, Faster RCNN, YOLOv6, and the baseline YOLOv5, respectively. The $F1$ value of the improved YOLOv5 algorithm under different occlusions is 19.36%, 13.07%, 1.68%, and 4.27% higher than that of the other algorithms, respectively. It has the ability to detect sticking and severely occluded apples, resulting in a high detection accuracy. On averaged image recognition time, the improved YOLOv5 was 0.27 s faster than HOG + SVM and 0.229 s faster than Faster RCNN and close to YOLOv6. The complexity of the occlusion of fruits and branches hinders the detector from learning an infinite variety of overlapping and occlusion situations when the training data are limited. Therefore, a direction worth exploring is the use of existing data to address this problem through unsupervised or semisupervised methods. This research can provide a theoretical basis for robots to quickly and efficiently identify apples in complex environments.

Data Availability

The data that supported the study are all in the article.

Conflicts of Interest

Qian Hao, Xin Guo, and Feng Yang declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors' Contributions

Qian Hao and Xin Guo designed the research. Qian Hao and Xin Guo processed the corresponding data. Qian Hao, Xin Guo, and Feng Yang wrote the first draft of the manuscript. Feng Yang helped to organize the manuscript. Qian Hao and Xin Guo revised and edited the final version.

Acknowledgments

This work is supported by the Shanxi Provincial Key Research and Development Project (201903D221018).

References

[1] X. Yu, Z. Fan, X. Wang et al., "A lab-customized autonomous humanoid apple harvesting robot," *Computers and Electrical Engineering*, vol. 96, article 107459, 2021.

[2] G. Zheng, "Characteristics, problems and upgrading strategies of apple export trade development in China," *China Fruits*, vol. 2021, no. 6, pp. 89–103, 2021.

[3] H. A. M. Williams, M. H. Jones, M. Nejati et al., "Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms," *Biosystems Engineering*, vol. 181, pp. 140–156, 2019.

[4] J. Tanaka, A. Ogawa, H. Nakamoto, T. Sonoura, and H. Eto, "Suction pad unit using a bellows pneumatic actuator as a support mechanism for an end effector of depalletizing robots," *ROBOMECH Journal*, vol. 7, no. 1, 2020.

[5] L. Huang, C. Chen, J. Yun et al., "Multi-scale feature fusion convolutional neural network for indoor small target detection," *Frontiers in Neurorobotics*, vol. 16, article 881021, 2022.

[6] S.-Y. Lin and H.-Y. Li, "Integrated circuit board object detection and image augmentation fusion model based on YOLO," *Frontiers in Neurorobotics*, vol. 15, article 762702, 2021.

[7] H. Gan, W. S. Lee, V. Alchanatis, and A. Abd-Elrahman, "Active thermal imaging for immature citrus fruit detection," *Biosystems Engineering*, vol. 198, no. 1, pp. 291–303, 2020.

[8] H. Lei, L. Wu, Z. Jiao, Z. Chen, J. Ma, and Z. Zhong, "Study on automatic detection method of mature bayberry in orchard environment," *Automation and Information*, vol. 42, no. 3, pp. 9–26, 2021.

[9] Q. Yang, W. Li, X. Yang, L. Yue, and H. Li, "Improved YOLOv5 method for detecting growth status of apple flowers," *Computer Engineering and Applications*, vol. 58, p. 10, 2022.

[10] H. Wang, Y. Lin, X. Xu, Z. Chen, Z. Wu, and Y. Tang, "A study on long-close distance coordination control strategy for litchi picking," *Agronomy*, vol. 12, no. 7, p. 1520, 2022.

[11] F. Wu, J. Duan, P. Ai, Z. Chen, Z. Yang, and X. Zou, "Rachis detection and three-dimensional localization of cut off point for vision-based banana robot," *Computers and Electronics in Agriculture*, vol. 198, article 107079, 2022.

[12] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sensing*, vol. 13, no. 9, p. 1619, 2021.

[13] W. Ji, Y. Pan, B. Xu, and J. Wang, "A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX," *Agriculture*, vol. 12, no. 6, p. 856, 2022.

[14] F. Yang, X. Lei, Z. Liu, P. Fan, and B. Yan, "Fast recognition method for multiple apple targets in dense scenes based on centerNet," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 53, no. 2, pp. 265–273, 2022.

[15] H. Kang and C. Chen, "Fast implementation of real-time fruit detection in apple orchards using deep learning," *Computers and Electronics in Agriculture*, vol. 168, p. 105108, 2020.

[16] P. Xu, L. Huang, and Y. Song, "An optimal method based on HOG-SVM for fault detection," *Multimedia Tools and Applications*, vol. 81, no. 5, pp. 6995–7010, 2022.

[17] X. Liu, *Research on Image Recognition Algorithm of Multifunctional Fruit and Vegetable Picking Robot, [Ph.D. thesis]*, Jiangsu University, 2020.

[18] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for kiwifruit defects based on YOLOv5," *Electronics*, vol. 10, no. 14, p. 1711, 2021.

[19] B. Tian, *Research on Apple Detection Classification and Positioning Technology in Complex Environment Based on Deep Learning, [Ph.D. thesis]*, Tianjin University of Technology, 2020.

[20] D. Zhao, R. Wu, X. Liu, and Y. Zhao, "Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background," *Transaction of The Chinese Society of Agricultural Engineering*, vol. 35, pp. 164–173, 2019.

[21] D. Wang, D. He, H. Song, C. Liu, and H. Xiong, "Combining SUN-based visual attention model and saliency contour detection algorithm for apple image segmentation," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 17391–17411, 2019.