




Communication

Tree Trunk and Obstacle Detection in Apple Orchard Based on Improved YOLOv5s Model

Fei Su¹, Yanping Zhao¹, Yanxia Shi², Dong Zhao¹, Guanghui Wang¹, Yinfan Yan^{1,3} , Linlu Zu^{1,3,*} 
and Siyuan Chang⁴ 

¹ College of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an 271018, China
² Intelligent Manufacturing College, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China
³ Shandong Provincial Key Laboratory of Horticultural Machineries and Equipment, Tai'an 271018, China
⁴ School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
* Correspondence: zulinlu@sdau.edu.cn

Abstract: In this paper, we propose a tree trunk and obstacle detection method in a semistructured apple orchard environment based on improved YOLOv5s, with an aim to improve the real-time detection performance. The improvement includes using the K-means clustering algorithm to calculate anchor frame and adding the Squeeze-and-Excitation module and 10% pruning operation to ensure both detection accuracy and speed. Images of apple orchards in different seasons and under different light conditions are collected to better simulate the actual operating environment. The Gradient-weighted Class Activation Map technology is used to visualize the performance of YOLOv5s network with and without improvement to increase interpretability of improved network on detection accuracy. The detected tree trunk can then be used to calculate the traveling route of an orchard carrier platform, where the centroid coordinates of the identified trunk anchor are fitted by the least square method to obtain the endpoint of the next time traveling route. The mean average precision values of the proposed model in spring, summer, autumn, and winter were 95.61%, 98.37%, 96.53%, and 89.61%, respectively. The model size of the improved model is reduced by 13.6 MB, and the accuracy and average accuracy on the test set are increased by 5.60% and 1.30%, respectively. The average detection time is 33 ms, which meets the requirements of real-time detection of an orchard carrier platform.

Keywords: semistructured orchard environment; tree trunk and obstacle detection; improved YOLOv5s; carrier platform



Citation: Su, F.; Zhao, Y.; Shi, Y.; Zhao, D.; Wang, G.; Yan, Y.; Zu, L.; Chang, S. Tree Trunk and Obstacle Detection in Apple Orchard Based on Improved YOLOv5s Model. *Agronomy* **2022**, *12*, 2427. <https://doi.org/10.3390/agronomy12102427>

Academic Editor: Lihong Xu

Received: 28 August 2022

Accepted: 1 October 2022

Published: 6 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Currently, the orchard operation mainly depends on manual labor; however, with the high labor intensity and increasingly serious aging population, it is imperative to develop intelligent equipment suitable for orchard operation [1]. Apple orchard is a semistructured environment, an automatic mobile robot must be able to detect the traveling route as well as to avoid obstacles in time when operating [2]. To ensure the operation safety, it is necessary not only to accurately plan the path when the equipment works alone but also to accurately identify the operators for man-machine collaboration. The detection of obstacles in apple orchard (mainly refers to operators) [3], the path planning of equipment, and the deployment of model in embedded equipment are the main factors hindering the wide application of intelligent equipment.

Lidar sensor detection [4–6], ultrasonic sensor detection [7], infrared sensor detection [8] and computer vision detection [9] are commonly used obstacle detection methods [10]. Chen et al. presented a novel tree trunk detection method based on the integration of multiple cameras and ultrasonic sensors and reduced the error of mobile robot localization identification by a moving average filter in orange orchard. The recall rate and

accuracy of the trunk recognition were approximately 93.62% and 98.96% [7]. Shalal et al. presented a method for local-scale orchard mapping based on tree trunk detection and utilized modest-cost color vision and laser-scanning technologies, where color and edges detection methods were used for tree trunk detection [11]. Freitas et al. proposed a classification and clustering of registered 3D points to generate the obstacle position offline, which could detect people and bins with no false positives [12]. The accuracy of laser and ultrasonic sensors can meet the requirements of orchard operations, but the cost is high, which may not be suitable for wide application [13].

The cost of computer vision based on detection method is lower than other detection methods. With the development of deep learning theory, detection algorithm based on convolutional neural network (CNN) has been widely used in the intelligent agriculture area [14]. At present, there are two types of target detection method: one is two-stage target detection method, which is represented by R-CNN [15], Fast R-CNN [16], and Faster R-CNN [17]; the other is the one-stage target detection method, which is represented by YOLO series [18,19] and SSD [20]. Li et al. proposed a lightweight MobileNetV2 object detection method based on YOLOv3 to identify typical obstacles in orchards. For the test sets, the mean average precision is 88.64% [2]. However, the study did not combine the detection results with path planning and did not consider changes in the orchard environment in different seasons. Yang et al. proposed a method called yolov3_trunk_model (Y3TM) to detect trunks using transfer learning. The results showed that the Y3TM model improved recall rate of over 93% with the average detection time of 0.3 s [21]. Different management needs to be carried out in different seasons of the orchard, including pruning, thinning, spraying, and harvesting [22].

In view of the above problems and the large environmental differences of orchards in different seasons, in this paper, we collected images of different seasons in the apple orchards. An improved YOLOv5s target detection model was proposed, with the K-means clustering algorithm was used to adjust the size of the anchor frame and the attention mechanism module was added to improve the detection accuracy of the trunk and obstacle people. Finally, the lightweight model was obtained through model pruning, which could make sure the detection speed as well as keep the detection accuracy. The Gradient-weighted Class Activation Map (Grad-CAM) technology was used to visualize the network before and after improvement [23]. The least square method [24,25] was used to fit the centric coordinates to get the end point of the route, which could be used on future route planning for a carrier platform.

2. Materials and Methods

2.1. Data Acquisition and Dataset Preparation

The data used in this paper was collected in an Apple Base of City Tai'an, China. The resolution of the collected images was 4096×3072 pixels. In the data acquisition stage, the camera was fixed at about 50 cm from the moving bracket to the ground. Data were collected in February, June, September, and December, which could represent spring, summer, autumn, and winter seasons. A total of 1800 images were obtained, including both front-light and back-light conditions. Figure 1 shows part of the collected images.

PASCAL VOC format is used to make datasets, and the LabelImg tool is used to annotate data. The generated xml file is used to store the data label information, including the location coordinate information of tree trunks and obstacles. Note that only the first two rows of trees are marked because the back rows can easily be mistaken as branches and two rows are enough to fit a line. In this paper, the obstacles are people performing farming activities, 600 pieces of data with obstacles, accounting for a third of the total dataset. Since during the apple harvest season, farmers need to complete the picking work and have large number of workers, while at other times, the number of orchard workers is low, so images with obstacle people are mainly concentrated in summer and autumn. In the data set, 1200 images are randomly selected as the training set, 300 images as the validation set, and 300 images as the test set in the ratio of 1:1:1 among the four seasons

dataset. The dataset includes 971 images in the front-light condition and 829 images in the back-light condition. The ratio between the back-light condition and the front-light condition is roughly 1:1.17. Mosaic data augmentation is used in the network training process, which splice four images into one image to improve the detection effect. The validation set is only used to evaluate the performance of the model during training and do not participate in the training process, and the test set is used to test the generalization performance of the model.

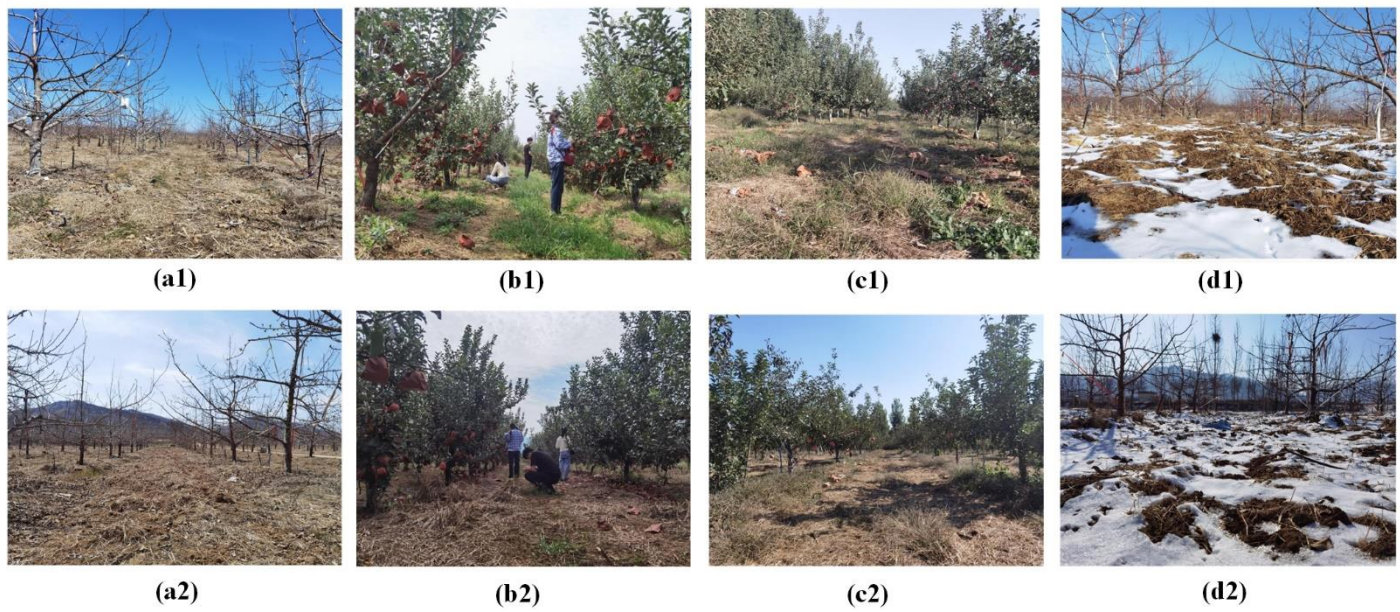


Figure 1. Images under different conditions. (a1) Spring, Front-light. (a2) Spring, Back-light. (b1) Summer, Front-light. (b2) Summer, Back-light. (c1) Autumn, Front-light. (c2) Autumn, Back-light. (d1) Winter, Front-light. (d2) Winter, Back-light.

2.2. Improvement of YOLOv5s Network Architecture

The size of the model directly affects deployment on mobile devices and real-time detection, which is the prerequisite for tree trunk and obstacle detection in orchard. According to the different width and depth, the YOLOv5 series model is divided into four kinds: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, where YOLOv5s is one of the most lightweight models, therefore is adopted in the paper [26]. The improved YOLOv5s network is shown in Figure 2. The YOLOv5s algorithm is composed of three parts. The first part is the input, and the input size of the training image is 608×608 . The second part is the backbone network, which uses CSPDarkNet53 network to extract rich information features from the input images. The third part is the detection layer, which adopts multiscale detection. The path aggregation network (PAN) is used after the feature pyramid network (FPN) to realize the fusion of feature information of different scales, and then the three generated feature maps are predicted. The Squeeze-and-Excitation (SE) module [27] is added to the last layer of the backbone network to improve feature extraction capability and detection accuracy.

As shown in Figure 3, SE is a channel attention mechanism, which is composed of Squeeze-and-Excitation module. In the Squeeze stage, the $W \times H \times C$ feature vector obtained is compressed to $1 \times 1 \times C$ through global average pooling, the calculation equation of a feature map is as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (1)$$

where u_c is the feature map of each channel, and c is the location of the channel.

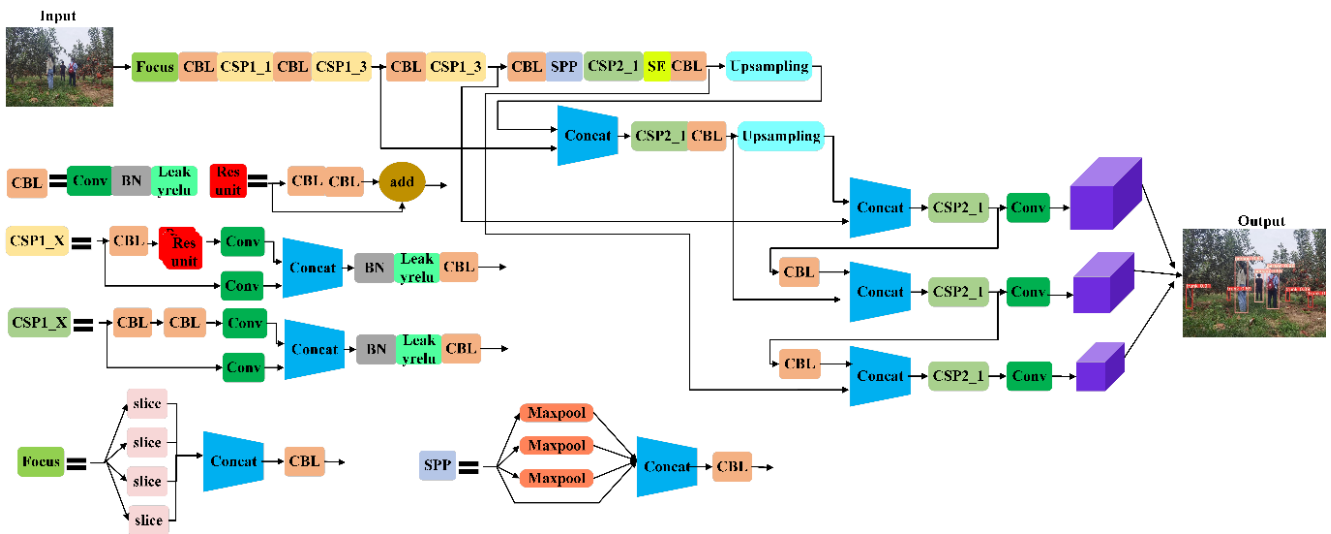


Figure 2. The improved YOLOv5s. SPP: Spatial Pyramid Pooling, SE: Squeeze-and-Excitation, CSP1_X is applied to the backbone, and CSP2_X is applied to the neck.

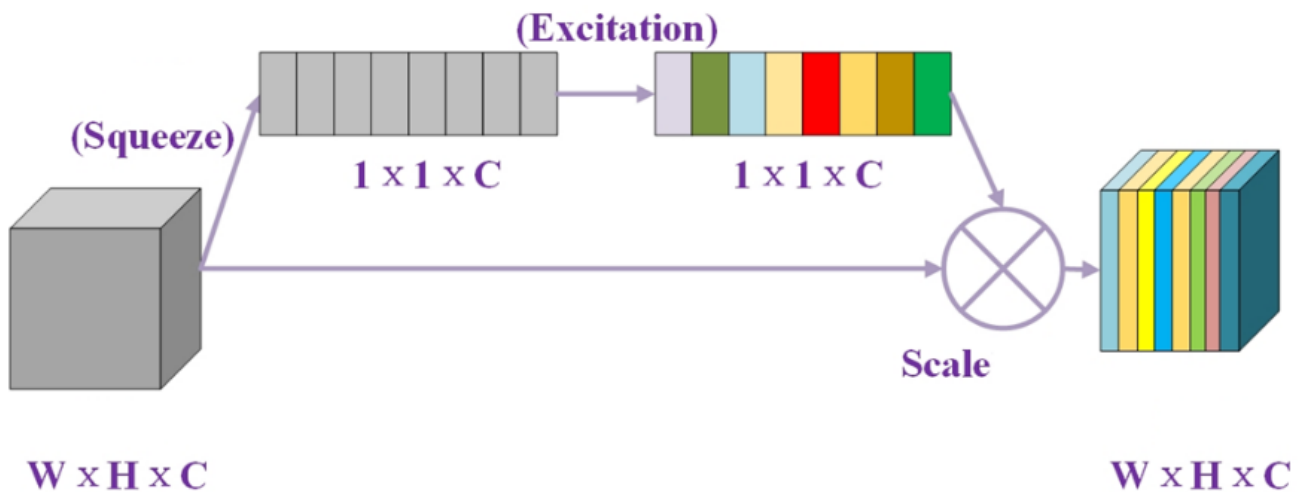


Figure 3. The Squeeze-and-Excitation Module. W, H, and C represent the width, height, and number of channels of the feature map respectively.

In the Excitation stage, the full connection layer and activation function is added, the full connection layer can fuse the input characteristic information, and the activation function can map the input between 0 and 1. The operation equation is as follows:

$$S_c = \sigma(g(z_c, W)) = \sigma(W_2\delta(W_1z_c)) \tag{2}$$

where z_c is the global description obtained by Squeeze operation, δ represents ReLU function, ensuring positive output, W_1, W_2 are two full connection layers.

After the SE operation, the weight of each channel in the input feature map is obtained, and then the original feature and weight are fused through Scale operation, which is the general multiplication operation, the equation is as follows:

$$X = S_c \times u_c \tag{3}$$

The SE module enables the network to pay different attention to different channel features, that is, to give more weight to important feature channels and less weight to irrelevant feature channels to compress useless information and improve detection accuracy.

2.3. Pruning of the YOLOv5s Model

A neural network can achieve higher accuracy with a deeper layer; however, the model parameters also increase with the layer depth. Therefore, the computation requirements will be more with deep depth model and the model size of obtained model will become larger. For model implementation in mobile devices, a small model size is important for the real-time detection of trunks and obstacles. Therefore, after the SE is added, the improved model is pruned as shown in Figure 4.

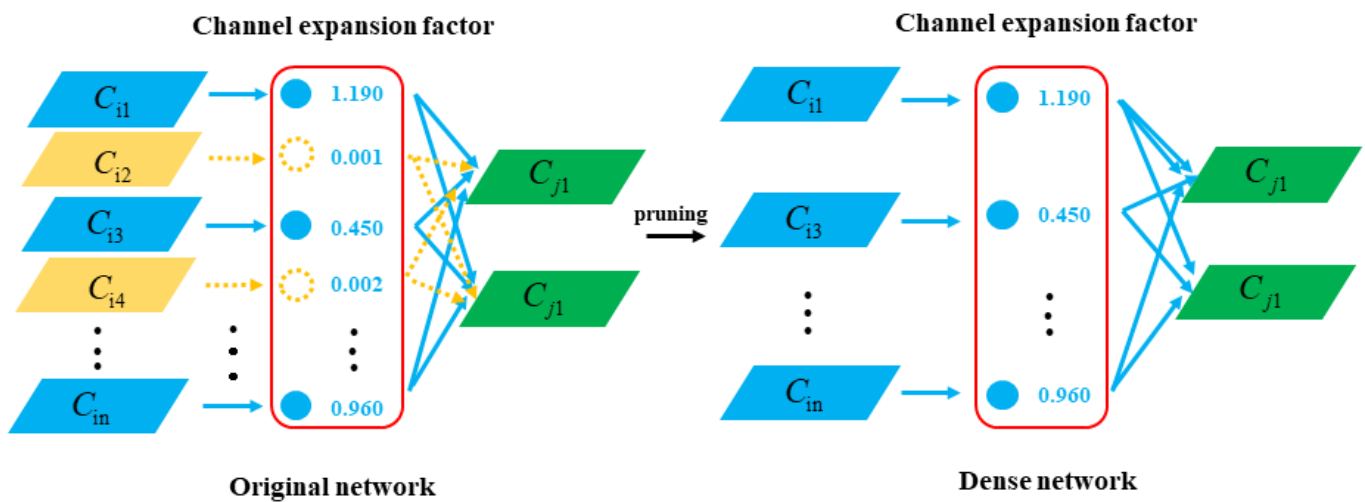


Figure 4. The pruning operation. C_i denotes i^{th} conv layer, C_j denotes $(i + 1)^{th}$ conv layer.

In this paper, the Batch Normalization (BN) layer coefficient γ pruning is used [28,29]. The pruning method restrains the BN layer coefficients by adding L1 regularization to make the corresponding coefficients sparser. After the corresponding sparse training, the layer with small sparsity is pruned and the relative activation is also very small, so the influence on the following is small. The optimal model can be obtained through repeated iteration.

2.4. The Route Direction Determination of the Carrier Platform

The centric coordinate point of the recognition frame is obtained by the coordinates of the recognition box, and the endpoint of the path information in the next stage is obtained by fitting the centroid coordinate line. To ensure the intersection of the two fitted lines, the identified centroid coordinates are projected onto the same plane. When the improved YOLOv5s model is used to identify tree trunks, the position information of the recognition frame can be recorded, as shown in Figure 5, the recognition frame is selected as the matching target to load, and the barycenter of the recognition frame is obtained by using the barycentric solution function. The solution equation is as follows:

$$x_z = \frac{\sum_{i=1}^2 x_i}{2} \tag{4}$$

$$y_z = \frac{\sum_{i=1}^2 y_i}{2} \tag{5}$$

From the above two equations, the coordinates of the centric point (x_z, y_z) , where x_i and y_i are the horizontal and vertical coordinate values of the upper left and lower right corners of each recognition frame.



Figure 5. Center of mass Coordinate fitting.

The 2D coordinates (x_z, y_z) of the barycenter of the tree trunks on both sides are fitted by the least square method, and the slopes and intercepts of the two lines are $a_1, b_1, a_2,$ and $b_2,$ respectively. The fitting results are shown in Figure 5. The equations of the two lines are:

$$y = a_1x + b_1 \quad (6)$$

$$y = a_2x + b_2 \quad (7)$$

The intersection of the two fitting lines (x_0, y_0) is solved, and the intersection is the end point of the current planned path [30]. Then, the boundary of the path and the direction of the carrier platform are determined.

2.5. Network Visualization Based on Grad-CAM

As an end-to-end network, the intermediate learning process is poorly understood, resulting in poor interpretability of the network. To improve the performance of the network, it is necessary to visualize the network, so the Grad-CAM is used in this paper for visualization [23]. The specific operation is forward propagation of the network to obtain the feature map and the predicted output value, the gradient information of the feature map is obtained by backward propagation of the output predicted value. By averaging the gradient information value on W and H , the importance of each channel of the feature map of this category and the weighted sum of the data of each channel of the feature map can be obtained, and the ReLU activation function can finally be used.

3. Results

The evaluation indexes of the model on the training set include Precision (P), Recall (R), and mean Average Precision (mAP). The model training and testing processes are carried out on the same computer, the experimental environment is summarized in Table 1. The hyperparameters of the model are set to 4 samples per batch, with a learning rate of 0.001, momentum of 0.937 and weight decay of 0.0005. The K-means clustering algorithm is used to calculate the size of the initial anchor frame on the data set of trunk and obstacle,

and the model convergence is accelerated in the process of model training. The sizes of the 9 anchors are, respectively, (8, 34), (10, 46), (13, 54), (15, 67), (18, 83), (23, 100), (33, 132), (50, 213), and (78, 314).

Table 1. Experimental environment.

Configuration	Parameter
CPU	Intel(R) Core(TM) i5-9300H CPU @ 2.40 GHz
GPU	NVIDIA GeForce GTX 1650
Development environment	Python 3.7, PyTorch 1.6, Anaconda3
Operating system	Windows10 (64-bit)

3.1. Model Pruning Operation

As shown in Figure 6a, pruning is done after model training. During the normal network training process, the coefficient γ of BN layer is approximately normal distribution as the number of training round goes on. The coefficient that does not participate in training is also selected due to the small number of values around 0, so the prune operation cannot be carried out under this situation.

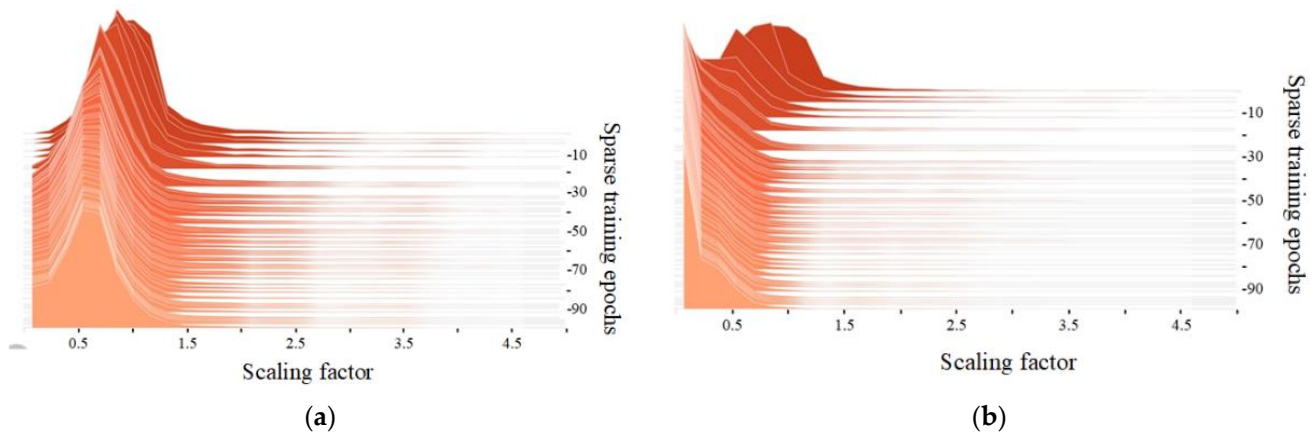


Figure 6. (a) The BN layer coefficient of normal training changes. (b) The BN layer coefficient of sparse training changes.

First, the sparse training is carried out. By adding the L1 regular constraint of γ to the loss function, the sparsity of γ can be realized; thus, the unimportant weight can be removed and the model size is reduced. The equation is as follows:

$$L1 = \lambda \sum_{\gamma \in \Gamma} g(\gamma) \quad (8)$$

Here, $g(\gamma)$ represents a sparsity-induced penalty on γ , and λ is the penalty factor.

As shown in Figure 6b, more and more γ approaches to 0 during the sparse training, indicating that the weight parameters of BN layer that do not participate in training are filtered out and pruning operation can be carried out. The model parameters are fine-tuned after the preset 10% pruning operation, such that the improved YOLOv5s model is obtained.

3.2. Model Performance Comparison before and after Improvement

Figure 7a,b show the P-R curve area of the improved YOLOv5s recognition model in the test set. The P-R curve area of people (obstacle) is the largest, which indicates that the model has better screening ability when features are obvious. The comparison of the model memory, P value, R value, mAP and average detection time indexes between the original YOLOv5s model and the improved YOLOv5s model is given in Table 2. Due to the addition of the SE attention mechanism the detection accuracy could be improved, where the improved model has a mAP% value 1.3% higher than the original YOLOv5s

model. However, the added SE module would increase the parameters of the network. Thus, another improvement is the pruning operation to obtain a more lightweight model. In order to better keep the detection accuracy parameters of the pruned model is further fine-tuned. Therefore, the size of the improved YOLOv5s model is 13.9 MB, with a 13.6 MB reduction, and the average detection time is 33 ms with 2 ms reduction.

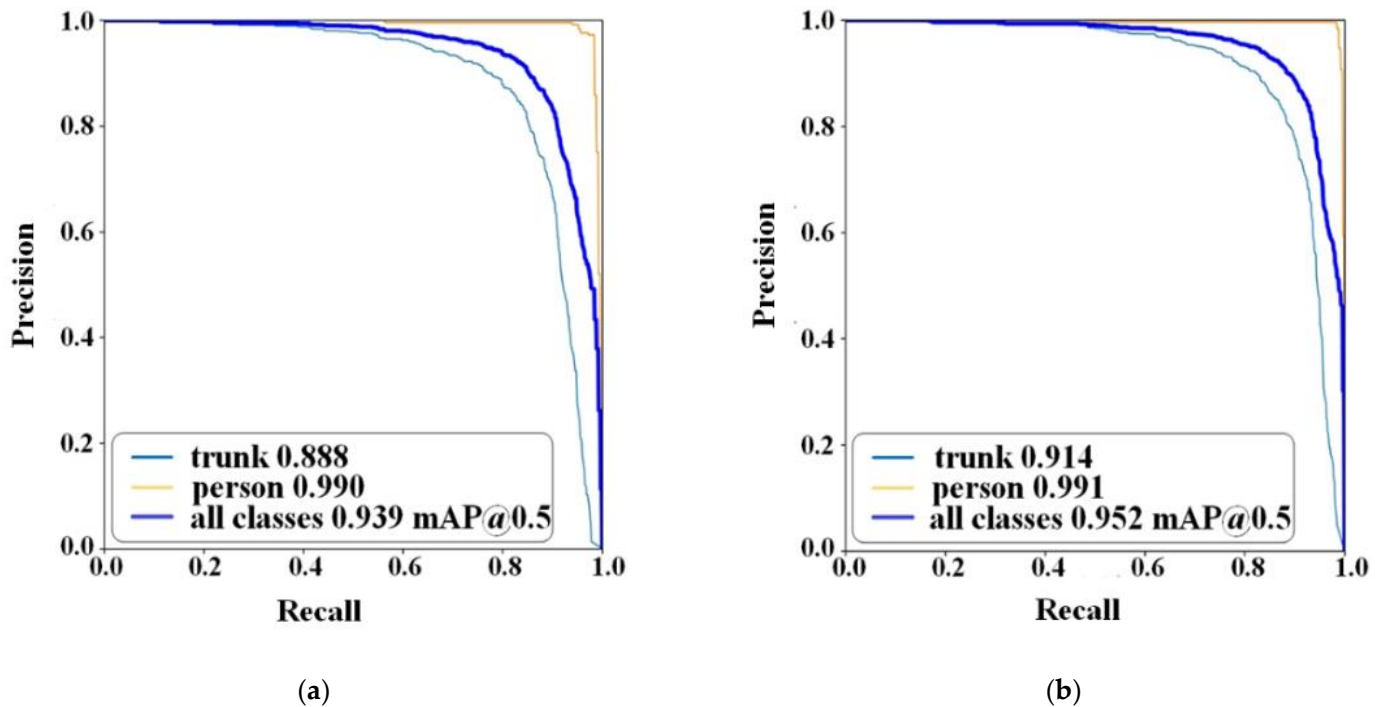


Figure 7. P-R curve area of test set. (a) The original YOLOv5 model. (b) The improved YOLOv5 model.

Table 2. Indexes before and after model improvement.

Model	Model Size/MB	Precision/%	Recall/%	mAP/%	Detection Time/ms
YOLOv5s	27.50	86.20	99.00	93.90	35.00
Improved YOLOv5s	13.90	91.80	99.00	95.20	33.00

The improved YOLOv5s model can pay more attention to useful features in feature extraction due to the addition of SE attention mechanism. Figure 8 shows the detection results of the model under different conditions before and after improvement. Although both models can correctly detect tree trunks and obstacles, the confidence score of the improved model to the tree trunk and the human is 93% and 97%, respectively, with 8% and 3% higher than that of the original model. To analyze the reasons for this result, the detection results of these two networks are visualized in Figure 9. The improved YOLOv5s model focuses more on the target and does not pay attention to the tree trunk with a longer distance (the tree trunk with a longer distance is the background by default in this paper).

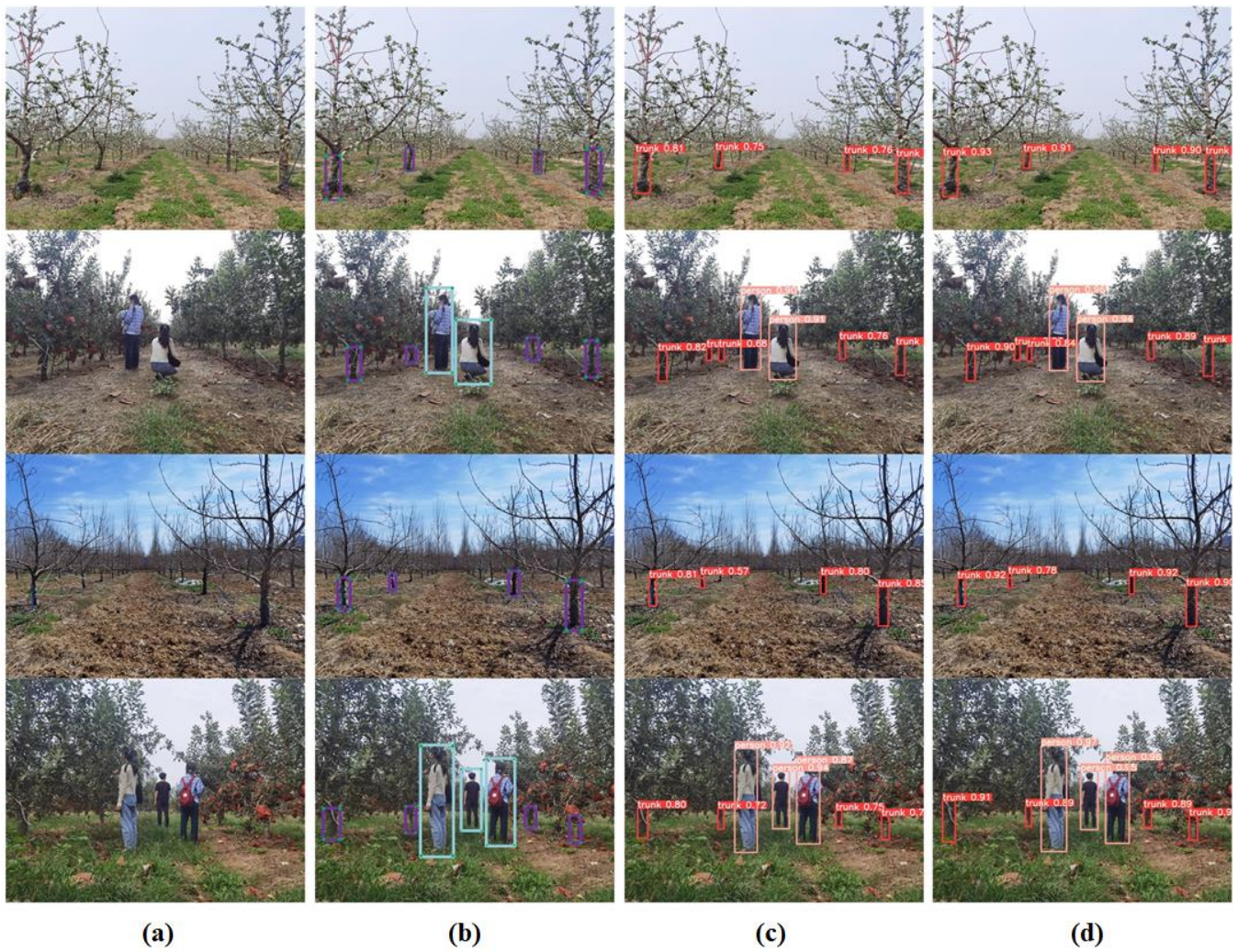


Figure 8. Comparison of model detection results before and after improvement. (a) Original image. (b) Labeling images. (c) YOLOv5s model. (d) Improved YOLOv5s model.

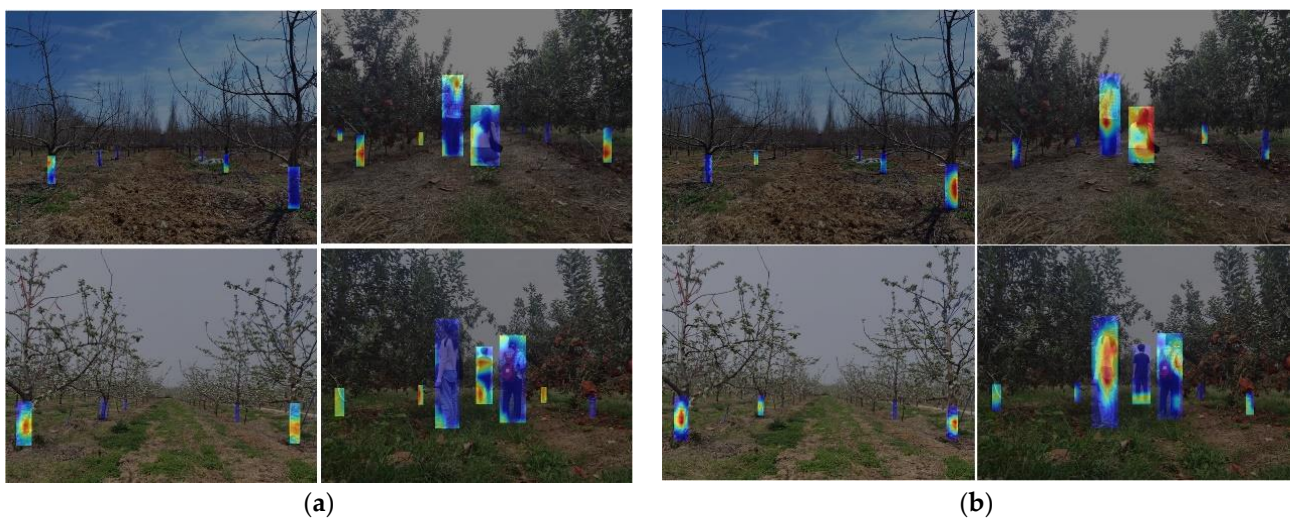


Figure 9. Visualization based on Grad-CAM technology. (a) the YOLOv5s model. (b) the improved YOLOv5s model.

3.3. Model Detection Performance Comparison among Four Seasons

The detection performance of the improved YOLOv5s model is compared among four seasons. The mAP% index of the tree trunk and obstacle detection for spring, summer, autumn and winter in the test set is given in Table 3. When the season stage is among branches and leaves growing to ripening, the tree trunk has a clearly differentiable color with the branches and leaves, so the detection performance in spring, summer and autumn are better. However, in winter, there are only branches and trunks, the branches and leaves have fallen off, so the difference between the characteristics of trunks and the background is not obvious, with a lower detection accuracy obtained. Figure 10 is an illustration of detection performance on four seasons.

Table 3. The mean average precision of target recognition in different season.

Season	Spring	Summer	Autumn	Winter
mAP/%	95.61	98.37	96.53	89.61

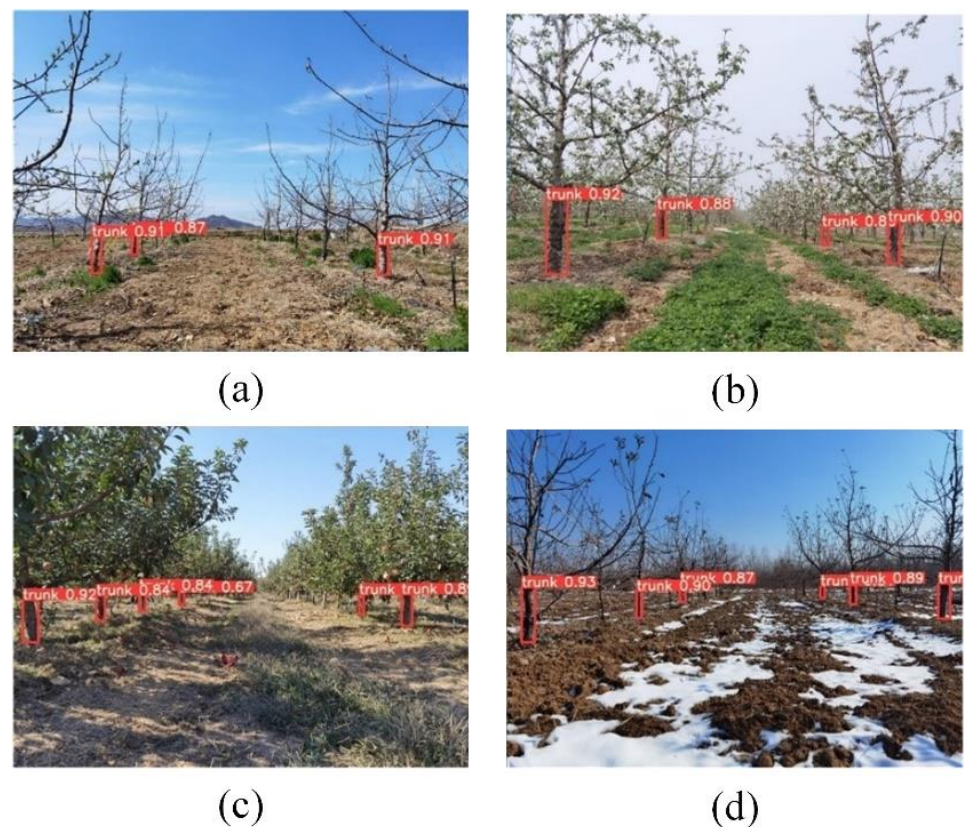


Figure 10. The detection performance illustration of the improved YOLOv5s for different seasons. (a) spring, (b) summer, (c) autumn and (d) winter.

Through the results, there is the following analysis in the article. The meaning here is that the leaves fall in winter, the color of the branches is close to the color of the trunk, and the thicker branches in the back row may cause interference to the trunk detection. Figure 11 shows the data of winter and spring respectively. There is no difference between the trees in the images of spring and winter, with almost no leaves, but the soil after snow in winter is wet and is dry in spring. Therefore, the color of soil in winter is darker than in spring. That is to say, the soil and the back branches are both close to the trunk color in winter, which together bring interference to the detection of the trunk, so the detection effect in winter is poor.

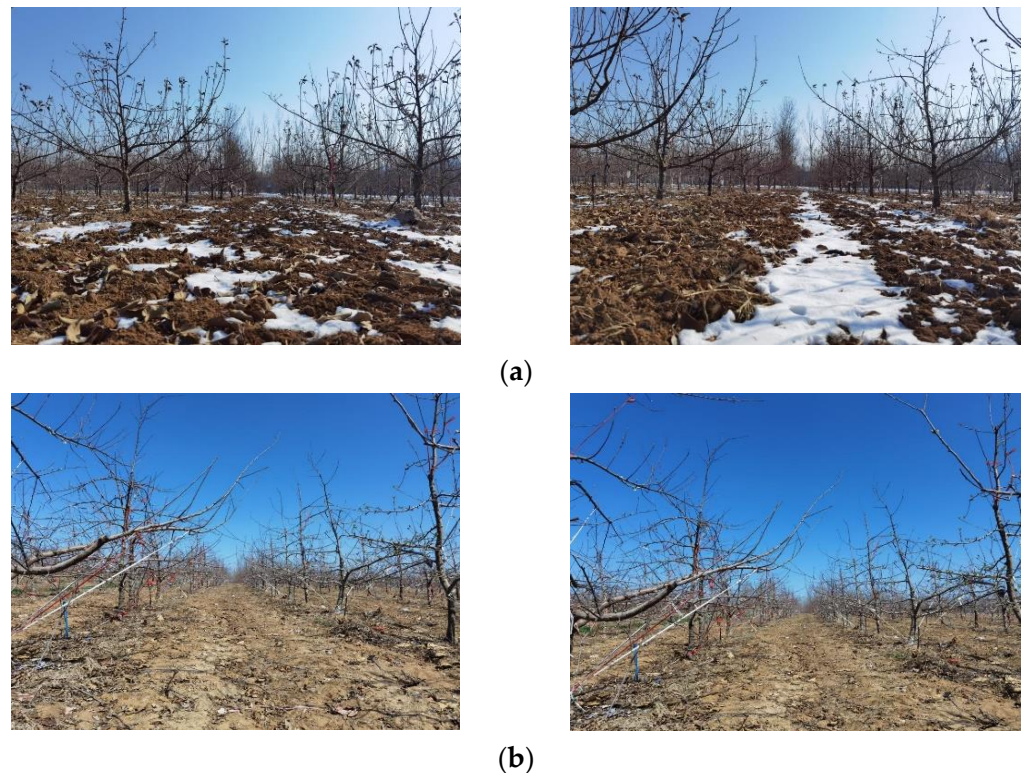


Figure 11. Data collected in winter and spring. (a) Winter, (b) Spring.

3.4. Comparison under Different Lighting Conditions

The tree trunk and obstacle detection indices for different lighting conditions in the test set are shown in Table 4, results show that the mAP of the front-light condition is 0.03 higher than that of the back-light condition, with the mAP value of the tree trunk and the obstacle person detection in the front-light condition is 0.037 and 0.024 higher than that of the back-light condition, respectively. Thus, the model works a little better under front-light than the back-light condition.

Table 4. Model detection performance under different lighting conditions.

Class	P/%		R/%		mAP/%	
	Front-Light	Back-Light	Front-Light	Back-Light	Front-Light	Back-Light
All	90.50	90.40	90.50	83.30	94.20	91.20
Trunk	88.00	86.20	82.20	72.30	89.50	85.80
Obstacle	93.00	94.70	98.70	94.30	99.00	96.60

3.5. In Field Detection on a Carrier Platform

The orchard carrying platform is a crawler type movement vehicle, with experimental computer and image acquisition camera. This vehicle also has power, relay, control modules and so on, the detailed configuration is listed in Table 5. The overall working framework is shown in Figure 12. The camera is used to collect the actual orchard environment video, and the improved YOLOv5s model implemented on the computer can detect and recognize the moving targets. The traveling speed of the orchard carrier platform is 0–1.4m/s, thus the detection speed of the proposed model is suitable for the intelligent equipment of apple orchard. Since human is indispensable in the actual orchard environment, the real-time detection of human as a kind of obstacles can effectively avoid the interference of obstacles in the working condition.

Table 5. Experimental platform specifications.

Configuration	Parameter
Power	2 × 48 V
Power inverter	HS-08
Camera	OPENMV4 H7 PLUS
Control module	STM32F103ZET6
Relay	MY4N-J
AC contactor	NXC-40

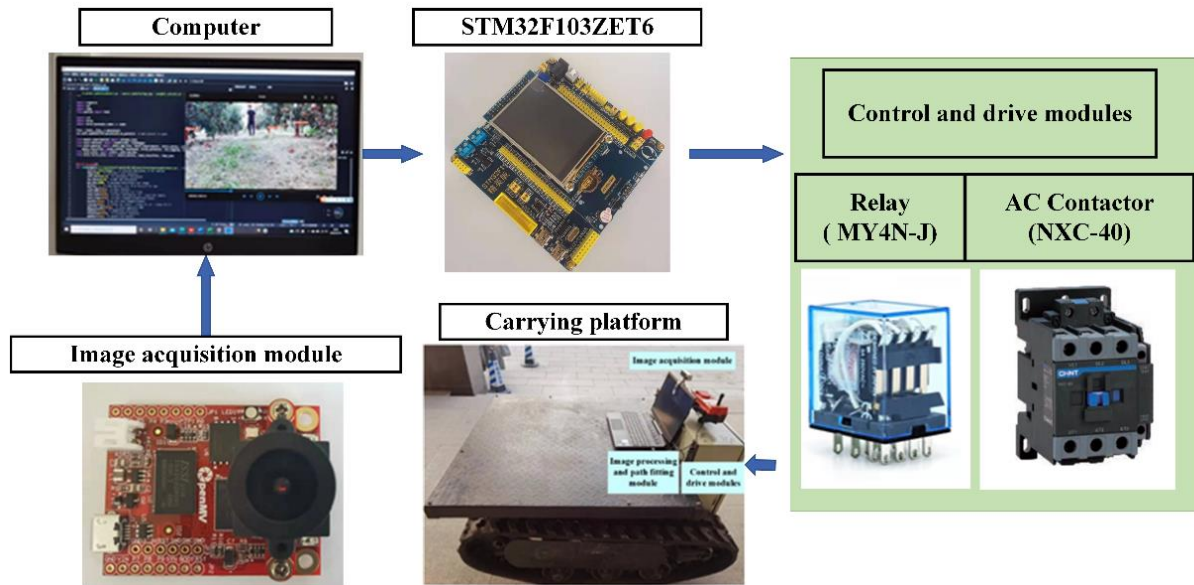


Figure 12. The overall working framework.

Figure 13 shows the real-time detection of tree trunks and obstacles in a video. Since the target is moving relative to the camera and the detection of moving target is more difficult than that of static target. In the video, the differentiation degree between tree trunks and background and obstacles is obvious, and the trunks and obstacles are correctly detected, which fully embodies the superiority of the improved YOLOv5s model.



Figure 13. Actual orchard environment video.

4. Discussion

This paper compares the proposed improved YOLOv5s model with the existing trunk detection methods, including the comparison of the detected targets in the orchard environment, whether the detection results are considered for the application on the path, and the influence of different seasons on the detection results.

Li et al. proposed a target detection method based on YOLOv3 that only detected typical obstacles (such as humans, cement columns and utility poles), without combining with path problems, and seasonal effects were not considered [2]. Yang et al. proposed a method called yolov3_trunk_model (Y3TM) to detect trunks, streetlight and telephone poles in the forest [21]. The study was conducted in a forest rather than an orchard environment, and the effects of seasonal changes were not considered, and no further study of the routing problem was done. Chen et al. presented a novel tree trunk detection method based on the histograms of oriented gradient (HOG) and support vector machine (SVM) and reduced the error of mobile robot localization identification by fusing multiple cameras, ultrasonic sensors, and moving average filters in an orange orchard [7]. Through the combination of tree trunk detection and position data, the mobile robot was guaranteed to travel, but the influence of orchard seasonal changes on tree trunk detection was not considered.

In the actual orchard operation, the requirements on detection speed and time varies according to the job type. The driving speed of the orchard crawled carrier platform is 0–1.4 m/s, and the scheme is proposed to be suitable for the intelligent production of the landing orchard [31]. In current studies about orchard path detection, Aguiar et al. proposed the Single Shot Multibox Detector model with average precision value of 84.16% and inference time of 23.14 ms [32]. Badeka et al. used Faster R-CNN and two YOLO versions with average precision of 73.2% and execution time of 29.6 ms [33]. In terms of network selection, the YOLOv5s model could balance speed and accuracy, and better detection performance is achieved after improvement. The mAP value of the proposed model is 95.2%, and the detection time is 33 ms. Compared with current studies, the improvement in model size and detection time by the improved YOLOv5s model is meaningful.

The method proposed in this paper is to monitor the tree trunk in real time. Due to the serious occlusion of the tree trunk in the back row of the collected images, the back rows can easily be mistaken as branches, so only the first two rows of tree trunks are annotated, it is not possible to perform non-real-time path calculation at the entrance of the orchard. Non-real-time computation is more efficient, and the algorithm needs to be further improved to increase the detection accuracy of the trunk in the back row [34]. Through comparative analysis, the method proposed in this paper considers the influence of different seasons on the detection results, and uses the tree trunk detection results to determine the traveling direction of the carrying platform, so the analysis is more comprehensive. The comparison results are shown in Table 6.

Table 6. The proposed method is compared with existing methods.

Source	Method	Object	Orchard Season Considerations	Result
Li et al. [2]	YOLOv3	humans, cement columns and utility poles	No	mAP: 88.64%
Yang et al. [21]	YOLOv3	trunk, telephone poles and streetlight	No	Recall: >93.00%
Chen et al. [7]	ultrasonic sensors	trunk	No	Recall: 92.14%
Proposed method	YOLOv5s	trunk, obstacle and determine the direction	Yes	mAP: 95.20%

5. Conclusions

To improve the detection accuracy of tree trunks and obstacles and meet the real-time requirements of the carrier platform in the orchard operation, in this paper, we propose a

method based on improved YOLOv5s. After adding SE module in YOLOv5s network, the K-means clustering algorithm is used for training and pruning operation, and the target recognition model is obtained after fine-tuning. The improved model reduces 13.6 MB smaller than the original YOLOv5s model, the average detection accuracy 1.30% higher, and the average detection time is 2 ms shorter. The Grad-CAM technology is used to visualize the network before and after improvement, and the improved network can focus on the target well. Finally, the trained model is applied in a real orchard environment, and the trunks and obstacles are detected correctly. The centric coordinates of recognition frame obtained after the trunk detection are fitted by the least square method to obtain the end of the path, and the forward direction of the carrier platform is determined. This will provide technical reference for orchard path planning.

Author Contributions: D.Z. and G.W. collected data; D.Z. and Y.S. designed hardware; Y.S. finished hardware emulation; D.Z. and Y.Z. analyzed the data; D.Z., Y.Z. and L.Z. wrote the paper; F.S. and L.Z. drew pictures for this paper; F.S., Y.Y., L.Z. and S.C. reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Tianjin Science and Technology Planning Project in 2020 (20YDTPJC00940), and the Shandong modern agricultural technology system (SDAIT-18-06).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All data are presented in this article in the form of figures and tables.

Acknowledgments: The authors would like to thank the editors and all the reviewers who participated in the review.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Clark, B.; Jones, G.D.; Kendall, H.; Taylor, J.; Cao, Y.; Li, W.; Zhao, C.; Chen, J.; Yang, G.; Chen, L.; et al. A proposed framework for accelerating technology trajectories in agriculture: A case study in China. *Front. Agric. Sci. Eng.* **2018**, *5*, 485–498. [[CrossRef](#)]
- Li, Y.; Li, M.; Qi, J.; Zhou, D.; Liu, K. Detection of typical obstacles in orchards based on deep convolutional neural network. *Comput. Electron. Agric.* **2021**, *181*, 105932. [[CrossRef](#)]
- Harshe, K.D.; Gode, N.P.; Mangtani, P.P.; Patel, N.R. A review on orchard vehicles for obstacle detection. *Int. J. Electr. Electron. Data Commun.* **2013**, *1*, 69.
- Malavazi, F.B.P.; Guyonneau, R.; Fasquel, J.B.; Lagrange, S.; Mercier, F. LiDAR-only based navigation algorithm for an autonomous agricultural robot. *Comput. Electron. Agric.* **2018**, *154*, 71–79. [[CrossRef](#)]
- Gu, C.; Zhai, C.; Wang, X.; Wang, S. CMPC: An Innovative Lidar-Based Method to Estimate Tree Canopy Meshing-Profile Volumes for Orchard Target-Oriented Spray. *Sensors* **2021**, *21*, 4252. [[CrossRef](#)] [[PubMed](#)]
- Kolb, A.; Meaclem, C.; Chen, X.Q.; Parker, R.; Milne, B. Tree trunk detection system using LiDAR for a semi-autonomous tree felling robot. In Proceedings of the 2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA), Auckland, New Zealand, 15–17 June 2015; pp. 84–89.
- Chen, X.; Wang, S.; Zhang, B.; Liang, L. Multi-feature fusion tree trunk detection and orchard mobile robot localization using camera/ultrasonic sensors. *Comput. Electron. Agric.* **2018**, *147*, 91–108. [[CrossRef](#)]
- Vodacek, A.; Hoffman, M.J.; Chen, B.; Uzkent, B. Feature Matching With an Adaptive Optical Sensor in a Ground Target Tracking System. *IEEE Sens. J.* **2015**, *15*, 510–519.
- Zhang, X.; Karkee, M.; Zhang, Q.; Whiting, M.D. Computer vision-based tree trunk and branch identification and shaking points detection in Dense-Foliage canopy for automated harvesting of apples. *J. Field Robot.* **2021**, *58*, 476–493. [[CrossRef](#)]
- Wang, L.; Lan, Y.; Zhang, Y.; Zhang, H.; Tahir, M.N.; Ou, S.; Liu, X.; Chen, P. Applications and Prospects of Agricultural Unmanned Aerial Vehicle Obstacle Avoidance Technology in China. *Sensors* **2019**, *19*, 642. [[CrossRef](#)] [[PubMed](#)]
- Shalal, N.; Low, T.; Mccarthy, C.; Hancock, N. A preliminary evaluation of vision and laser sensing for tree trunk detection and orchard mapping. In Proceedings of the Australasian Conference on Robotics and Automation (ACRA 2013), University of New South Wales, Sydney, Australia, 2–4 December 2013; pp. 80–89.
- Freitas, G.; Hamner, B.; Bergerman, M.; Singh, S. A practical obstacle detection system for autonomous orchard vehicles. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Algarve, Portugal, 7–12 October 2012; pp. 3391–3398.
- Bietresato, M.; Carabin, G.; Vidoni, R.; Gasparetto, A.; Mazzetto, F. Evaluation of a LiDAR-based 3D-stereoscopic vision system for crop-monitoring applications. *Comput. Electron. Agric.* **2016**, *124*, 1–13. [[CrossRef](#)]

14. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [[CrossRef](#)]
15. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
16. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)] [[PubMed](#)]
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
20. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
21. Yang, T.; Zhou, S.; Xu, A. Rapid Image Detection of Tree Trunks Using a Convolutional Neural Network and Transfer Learning. *IAENG Int. J. Comput. Sci.* **2021**, *48*, 257–265.
22. Zhang, Q.; Karkee, M.; Tabb, A. The use of agricultural robots in orchard management. *arXiv* **2019**, arXiv:1907.13114.
23. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 618–626.
24. Zhang, C.; Yong, L.; Chen, Y.; Zhang, S.; Ge, L.; Wang, S.; Li, W. A Rubber-Tapping Robot Forest Navigation and Information Collection System Based on 2D LiDAR and a Gyroscope. *Sensors* **2019**, *19*, 2136. [[CrossRef](#)]
25. Zhang, X.; Li, X.; Zhang, B.; Zhou, J.; Tian, G.; Xiong, Y.; Gu, B. Automated robust crop-row detection in maize fields based on position clustering algorithm and shortest path method. *Comput. Electron. Agric.* **2018**, *154*, 165–175. [[CrossRef](#)]
26. Chandra, A.L.; Desai, S.V.; Guo, W.; Balasubramanian, V.N. Computer Vision with Deep Learning for Plant Phenotyping in Agriculture: A Survey. *arXiv* **2020**, arXiv:2006.11391.
27. Jie, H.; Li, S.; Gang, S.; Albanie, S. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023.
28. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [[CrossRef](#)]
29. Li, Y.; Iida, M.; Suyama, T.; Suguri, M.; Masuda, R. Implementation of deep-learning algorithm for obstacle detection and collision avoidance for robotic harvester. *Comput. Electron. Agric.* **2020**, *174*, 105499. [[CrossRef](#)]
30. Moghadam, P.; Starzyk, J.A.; Wijesoma, W.S. Fast Vanishing-Point Detection in Unstructured Environments. *IEEE Trans. Image Process.* **2012**, *21*, 425–430. [[CrossRef](#)]
31. Zhou, Y.; Yang, Y.; Zhang, B.; Wen, X.; Yue, X.; Chen, L. Autonomous detection of crop rows based on adaptive multi-ROI in maize fields. *Int. J. Agric. Biol. Eng.* **2021**, *014*, 217–225. [[CrossRef](#)]
32. Aguiar, A.S.; Monteiro, N.N.; Santos, F.N.D.; Pires, E.J.S.; Silva, D.; Sousa, A.J.; Boaventura-Cunha, J. Bringing semantics to the vineyard: An approach on deep learning-based vine trunk detection. *Agriculture* **2021**, *11*, 131. [[CrossRef](#)]
33. Badeka, E.; Kalampokas, T.; Vrochidou, E.; Tziridis, K.; Papakostas, G.; Pachidis, T.; Kaburlasos, V. Real-time vineyard trunk detection for a grapes harvesting robot via deep learning. In Proceedings of the Thirteenth International Conference on Machine Vision, Rome, Italy, 2–6 November 2020; Osten, W., Nikolaev, D.P., Zhou, J., Eds.; International Society for Optics and Photonics (SPIE): Bellingham, WA, USA, 2021; Volume 11605, pp. 394–400.
34. Zhao, W.; Wang, X.; Qi, B.; Runge, T. Ground-Level Mapping and Navigating for Agriculture Based on IoT and Computer Vision. *IEEE Access* **2020**, *8*, 221975–221985. [[CrossRef](#)]