

Article

Approach of Dynamic Tracking and Counting for Obscured Citrus in Smart Orchard Based on Machine Vision

Yuliang Feng ^{1,2}, Wei Ma ^{2,*}, Yu Tan ^{1,*}, Hao Yan ¹, Jianping Qian ^{3,*} , Zhiwei Tian ² and Ang Gao ⁴

¹ College of Engineering, China Agricultural University, Beijing 100083, China; fylove@cau.edu.cn (Y.F.); yanhao_re@163.com (H.Y.)

² Institute of Urban Agriculture, Chinese Academy of Agriculture Sciences, Chengdu 610213, China; tianzhiwei@caas.cn

³ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

⁴ College of Mechanical and Electrical Engineering, Shandong Agricultural University, Taian 271018, China; 2019110103@sdau.edu.cn

* Correspondence: mawei03@caas.cn (W.M.); tanyu@cau.edu.cn (Y.T.); qianjianping@caas.cn (J.Q.)

Abstract: The approach of dynamic tracking and counting for obscured citrus based on machine vision is a key element to realizing orchard yield measurement and smart orchard production management. In this study, focusing on citrus images and dynamic videos in a modern planting mode, we proposed the citrus detection and dynamic counting method based on the lightweight target detection network YOLOv7-tiny, Kalman filter tracking, and the Hungarian algorithm. The YOLOv7-tiny model was used to detect the citrus in the video, and the Kalman filter algorithm was used for the predictive tracking of the detected fruits. In order to realize optimal matching, the Hungarian algorithm was improved in terms of Euclidean distance and overlap matching and the two stages life filter was added; finally, the drawing lines counting strategy was proposed. In this study, the detection performance, tracking performance, and counting effect of the algorithms are tested respectively; the results showed that the average detection accuracy of the YOLOv7-tiny model reached 97.23%, the detection accuracy in orchard dynamic detection reached 95.12%, the multi-target tracking accuracy and the precision of the improved dynamic counting algorithm reached 67.14% and 74.65% respectively, which were higher than those of the pre-improvement algorithm, and the average counting accuracy of the improved algorithm reached 81.02%. The method was proposed to effectively help fruit farmers grasp the number of citrus and provide a technical reference for the study of yield measurement in modernized citrus orchards and a scientific decision-making basis for the intelligent management of orchards.

Keywords: smart orchard; automatic detection; target tracking; dynamic counting



Citation: Feng, Y.; Ma, W.; Tan, Y.; Yan, H.; Qian, J.; Tian, Z.; Gao, A. Approach of Dynamic Tracking and Counting for Obscured Citrus in Smart Orchard Based on Machine Vision. *Appl. Sci.* **2024**, *14*, 1136. <https://doi.org/10.3390/app14031136>

Academic Editor: José Miguel Molina Martínez

Received: 9 January 2024

Revised: 24 January 2024

Accepted: 27 January 2024

Published: 29 January 2024



Copyright: © 2024 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

The acquisition of yield information for citrus and other cash crops is an important part of the fine technology system and a key indicator for orchard production management by fruit farmers [1]. It is also an important study direction for field information management in smart orchards. Citrus is one of the most widely planted agricultural products in the world, which is an important pillar of the agricultural economy in many countries, and it is also an important cash fruit crop in Southwest China [2]. With the continuous development of citrus planting technology, efficient cultivation patterns such as dense planting, fine pruning, and citrus bagging have become mainstream [3], which has greatly improved the efficiency of orchard land use. The accurate prediction of citrus production can help to effectively utilize resources and rationally allocate resources such as labor, storage space, and harvesting equipment [4]. However, at present, the number of fruits in intensively cultivated citrus orchards mainly relies on manual sampling and counting estimation,

which is time-consuming and laborious. Therefore, the automatic counting of citrus fruits is essential for obtaining yield information in citrus orchards.

In recent years, automatic counting methods oriented to fruit yield measurement have become the focus of study, but their studies have mainly focused on static models. Mekhalfi et al. [4] used the Viola-Jones based target detection algorithm to estimate the yield by counting kiwifruit images in the field and tested in two large kiwifruit orchards, which resulted in counting errors of 6% and 15%, respectively. Koirala et al. [5] proposed a deep learning model called “MangoYOLO”, trained on daytime images and robust to variations in orchards, varieties, and lighting conditions. It estimated fruit yields in five orchards with errors ranging from 4.6% to 15.2%. Although these related studies achieved good counting performance on static images, the static model is not applicable to the actual production environment, so the detection and tracking of the dynamic model have become a hot spot for future research.

Fast and effective fruit detection in dynamic modeling is the basis for achieving high-precision counting. Xu et al. [6] established a color model for citrus recognition on R-B color metrics using the dynamic thresholding method. Citrus was separated from the background and the detection of single and multiple fruits was carried out under both downlight and backlight conditions, with a high recognition correctness rate. Traditional target detection algorithms usually rely on features such as the color, texture, and shape of the target and combine morphological operations to segment the image to detect the target region [7,8]. However, traditional algorithms often struggle to balance real-time and accuracy, and the generalizability of the findings is limited. In contrast, deep learning-based target detection algorithms have more efficient and accurate data representation, are more generalizable, and are easier to apply to real-world scenarios [9]. Among them, first-order detection algorithms (e.g., YOLO) and second-order detection algorithms (e.g., Faster R-CNN [10]) have faster detection speeds than traditional algorithms and are widely used in fruit detection research. Yan et al. [11] proposed an improved Faster RCNN-based prickly pear fruit recognition method. The region of interest pooling (ROI pooling) in the convolutional neural network is improved to the region of interest calibration (ROI align) of the regional feature aggregation method, which makes the target rectangular box of the prickly pear fruit in the natural environment more accurate. In order to realize real-time detection and counting, the single-stage target detection method represented by the YOLO series has a wider application in the field of fruit detection. Lu et al. [12] proposed a citrus recognition method based on an improved YOLOV3-LITE lightweight neural network, using a combination of hybrid training and migration learning with the training method to improve the generalization ability of the model, which significantly improves the accuracy of citrus fruit detection under the light occlusion of the fruit.

Target tracking localizes the target in subsequent frames based on the position of the tracked target in the initial frame, thus correlating to the same target in the video [13–15]. ZHANG et al. [16] proposed the “OrangeSort” method based on simple online and real-time tracking (SORT) for counting citrus in the field, which reduces the effect of leaf shading on the accuracy of citrus counting results, but the method ignores the appearance characteristics of the target. Xia et al. [17] introduced the Byte algorithm to improve FairMOT’s data association strategy during fruit tracking to predictively track fruits in videos, but its use of depth features only in appearance does not fully represent the target and affects the tracking accuracy. Wojke et al. [18] proposed the DeepSort method which incorporates both motion and appearance information and which can achieve high multi-target tracking accuracy (MOTA) while maintaining real-time speed, and can effectively reduce the impact of the occlusion problem. Therefore, the DeepSort method has reference significance for fruit counting in complex citrus orchard environments.

This study aims to provide an effective dynamic tracking and counting approach for citrus in a complex orchard environment. Additionally, the study introduces the two stages life filter into DeepSORT to improve the algorithm performance. Based on the YOLOv7-tiny deep learning target detection algorithm, the former combines the Kalman filter and the

Hungarian algorithm to realize the fusion of motion and appearance features. The latter significantly reduces the number of missed citrus and improves the accuracy of automatic citrus counting.

2. Material and Method

2.1. Experiment Site

In order to make the collected dataset have practical reference value, pictures and videos of citrus taken from real citrus orchards were used as the experimental dataset in this paper. The experimental orchard is located in Pujiang County, Chengdu City, Sichuan Province, China, and the widely cultivated Kasumi citrus varieties were used as the main study objects. We used a Sony IMX596 camera to capture 2000 citrus images and supplemented the dataset with 200 citrus images from the Jiangxi experiment site, 300 citrus images from the Guangxi experiment site, and 500 citrus images from the Guangdong experiment site in order to increase the generalizability of the dataset. The dataset totaled 3000 citrus images of different orchards with different varieties, and the collection locations and the number of datasets is shown in Figure 1.

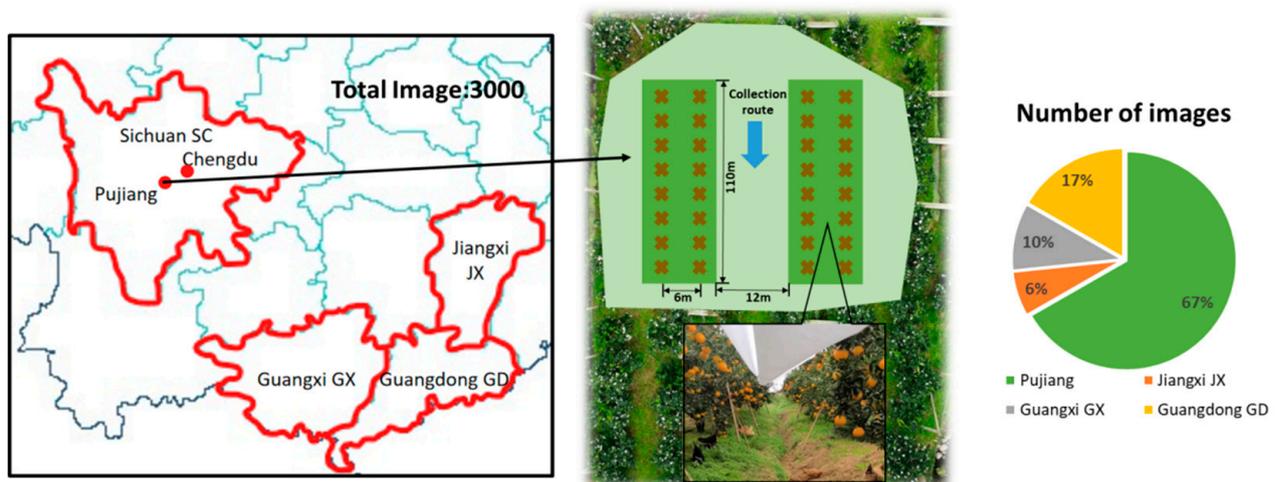


Figure 1. Experiment site and composition of dataset.

2.2. Image Preparation and Processing

The data collection test site was mostly in low-light weather and the number of images of different citrus varieties was limited, so pre-processing was performed to enhance the data. Various methods such as changing the brightness, changing the saturation, and rotating mirror images were used to preprocess the data as needed. In this paper, 3000 citrus images were preprocessed to increase the saturation by 5 times, increase the brightness by 70 pixels, decrease the brightness by 70 pixels, and be flipped vertically as shown in Figure 2. Then, the data collection was supplemented by the public dataset MOT16 [19], the number of citrus images in this experimental dataset was expanded to 15,000, and the training set, validation set and test set were divided according to the ratio of 7:2:1.

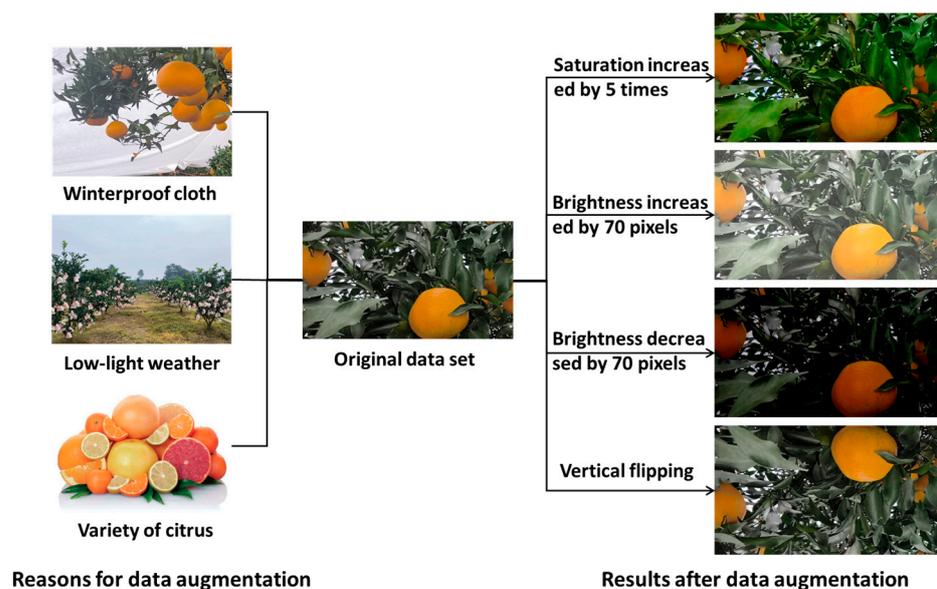


Figure 2. Causes and effects of data augmentation.

2.3. Approach of Citrus Tracking and Counting Based on YOLOv7-Tiny and DeepSORT

Citrus fruit counting is generally accompanied by the bagging of citrus. At this time, citrus has entered the coloring period and there is a clear difference in color with the background, which is conducive to the counting and estimation of yield. The citrus tracking and counting method based on YOLOv7-tiny and DeepSORT proposed in this paper mainly consists of the following steps: (1) target detection: use YOLOv7-tiny network to detect the citrus in the video, and get the target detection frame and the corresponding features; (2) target prediction: use Kalman filtering algorithm to predict the position and state of the citrus in the next frame of the video, and (3) target matching: use the Hungarian algorithm to cascade match and IOU match the predicted tracks and detections in the current frame, which are updated by Kalman filtering and output as tracking results. Among them, the track that fails in target matching is initialized as a new tracking track; in order to reduce the fruit loss and repeated matching errors caused by the new track, the two stages life filter is additionally added to the algorithm in this paper.

The two stages life filter (TTSLF) and three consecutive frames detection are applied in new track. The filtered track enters the target matching session after successful detection and the track is deleted after failure. The overall flow of the tracking and counting method is shown in Figure 3.

In the target detection part, the YOLO algorithm has the advantage of high target detection accuracy, while agricultural robots have high requirements for easy scalability and high accuracy. Therefore, we focus on agricultural robot harvesting research based on algorithm comparison. The YOLOv7-tiny uses fewer convolutional layers and channels than YOLOv7, YOLOX, and other models, the inference speed is faster, suitable for real-time detection of citrus with high real-time requirements, and the model is small in size, occupies less storage space and video memory, and is suitable to be deployed in resource-limited devices. It is also suitable for deployment on resource-limited devices, which is favorable for the development of agricultural robots in citrus orchards. Therefore, YOLOv7-tiny is chosen as the target detection algorithm for citrus.

In the target tracking part, the DeepSORT algorithm [18] added Re-ID appearance features, reducing ID switches by 45% after applying the deep association metric to targets. By considering both motion and appearance in the matching strategy, it achieved good performance at high rates. So, we added a two-stage life filter to the DeepSORT algorithm, which significantly reduced matching failures and improved fruit counting accuracy.

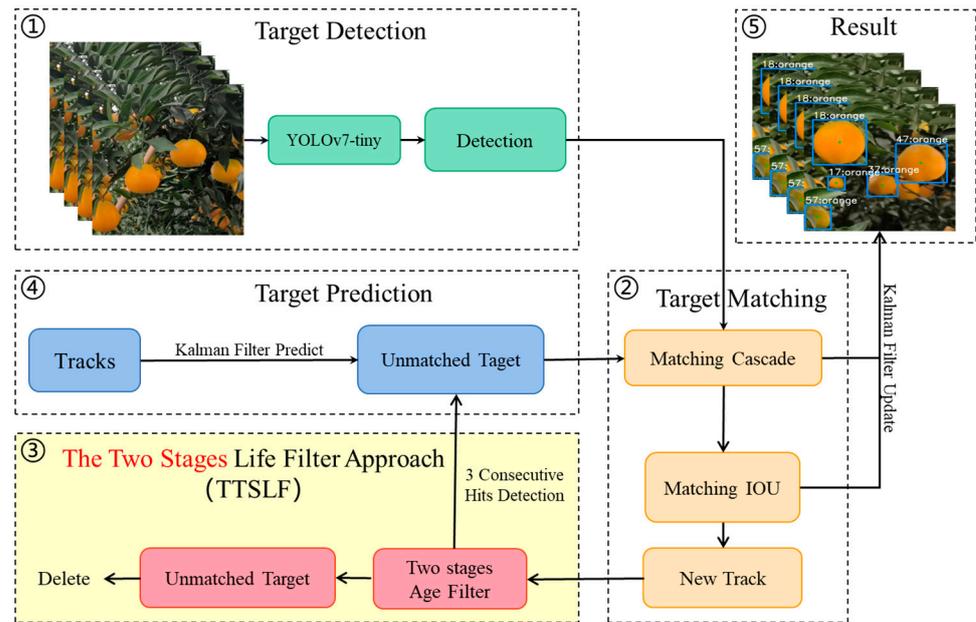


Figure 3. Overall flow chart of citrus tracking and counting method.

2.3.1. Citrus Fruit Detection Based on YOLOv7-Tiny

After investigating and studying the target detection algorithms, we found that the YOLOv7-tiny network used the darknet deep learning framework to achieve end-to-end training of the input image. The YOLO network [20,21] first realizes the input image to 416×416 and converts it into a $416 \times 416 \times 3$ tensor, then the backbone network converts the tensor of the input image into a higher dimensional feature map. The neck converts the feature map extracted by the backbone network into a feature map suitable for target detection. The detection head predicts the location and class of the target in the image. Finally, the detection results are post-processed to remove duplicates and sort the results to obtain the final target detection results.

2.3.2. DeepSORT-Based Citrus Fruit Tracking

The YOLOv7-tiny deep learning model is used to extract the visual features of the targets such that each target has a unique representation in the feature space. Then the DeepSORT [22] serialization optimization algorithm, specifically the SORT (simple online and realtime tracking) algorithm based on the Kalman filter and the Hungarian algorithm is used to achieve the tracking of citrus fruits and maintain the stability and accuracy of tracking.

1. Fruit target prediction based on the Kalman filter algorithm

Kalman filtering (KF) is a recursive filtering method for estimating the state of a system and is commonly used in linear systems in areas such as signal processing, control systems, and target tracking [23,24]. Since the video frame rate of the citrus fruits acquired in this paper is 30 frames/s, the position of citrus fruits varies very little between video sequences (which can be basically regarded as a uniform motion) and it can be assumed that the citrus fruit tracking system is linearly correlated over time. In the prediction phase, the Kalman filter predicts the corresponding citrus tracking position and its covariance matrix based on the detected position of the YOLOv7-tiny network.

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \quad (1)$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \quad (2)$$

where $\hat{x}_{k|k-1}$ is the a priori state estimate of the k th frame, which is the result of the intermediate computation of filtering, i.e., the result of the k th frame predicted based on the optimal estimate of the previous frame ($k - 1$ frame), and the result of the prediction equation; A and B are the parameter matrices of the system; $\hat{x}_{k-1|k-1}$ is the a posteriori state estimate for the $k - 1$ frame, which is one of the results of the filtering; u_{k-1} is the process noise at frame $k - 1$; $P_{k|k-1}$ is the a priori estimated covariance at frame k , which represents the uncertainty of the state and is one of the results of the filtering; and Q is the covariance of the system's process noise, which is the state variable used by the Kalman filter for estimating the discrete-time process.

In the update phase, the Kalman filter optimizes the predicted values obtained in the prediction phase using observations of the current frame to obtain a more accurate new estimate, i.e., updating the fruit tracking position and its covariance matrix based on the matching relationship between the detected and predicted fruits.

$$K_k = P_{k|k-1}H^T (HP_{k|k-1}H^T + R)^{-1} \tag{3}$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k (Z_k - H\hat{x}_{k|k-1}) \tag{4}$$

$$P_{k|k} = (I - K_kH)P_{k|k-1} \tag{5}$$

where K_k is the filter gain matrix, which is the intermediate calculation of the filter; H is the state variable to observation variable conversion matrix, which represents the relationship connecting state and observation. R is the measurement noise covariance; $\hat{x}_{k|k}$ is the number of frames in the k frame of the a posteriori state estimate, which is one of the results of the filtering; Z_k is the measurement value, which is an input to the filtering; $P_{k|k}$ is the a posteriori estimated covariance of the k frame's a posteriori estimated covariance, which is one of the results of the filtering; I is the unit matrix.

2. Fruit Target Matching Based on the Hungarian Algorithm

The Hungarian algorithm is a combinatorial optimization algorithm for solving the task allocation problem in polynomial time [25]. In this paper, after obtaining the video detection results and the prediction results in the citrus video, the Euclidean distance is used to measure the similarity between the Kalman filtered prediction results and the YOLOv7-tiny detection results.

In the citrus target matching stage, the Hungarian algorithm is used to seek the optimal solution for matching the prediction results with the detection results, as in Equation (6), and the successful matching enters the update stage of Kalman filtering.

$$s.t \begin{cases} \sum_{i=1}^m x_{i,j} = 1, i = 1, 2, \dots, m \\ \sum_{j=1}^n x_{i,j} = 1, j = 1, 2, \dots, n \\ x_{i,j} = 0 \text{ or } 1, i(j) = 1, 2, \dots, m(n) \end{cases} \tag{6}$$

where m and n are the number of fruits tracked and number of fruits detected, respectively.

2.3.3. Citrus Drawing Lines Counting Strategy

Citrus fruit trees in the Sichuan and Chongqing regions of China are predominantly evergreen with a high branching rate, robust branching vigor, and abundant foliage. These characteristics contribute to the frequent obscuration of citrus fruits by branches and leaves during the acquisition of citrus counting videos. During the target tracking process, it may lead to the loss of citrus targets or the identity ID of the same citrus target jumping, which is easy to repeat counting or omission (Figure 4).

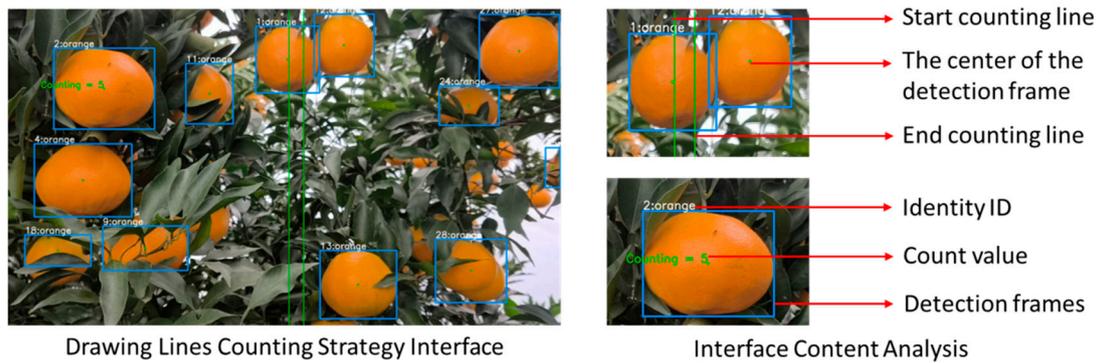


Figure 4. Introduction of drawing lines counting strategy.

3. Experiments and Analysis

3.1. Experiment Design

The training platform consisted of a laptop equipped with a 13th Gen Intel(R) Core (TM) i9-13900HX (2.20 GHz) CPU, 16 GB of RAM, and an NVIDIA RTX 4060 Laptop GPU. Software tools included CUDA 12.0, cuDNN, Pycharm2022, Python 3.8, and OpenCV 4.4.0.

During the citrus target detection experiment, the initial learning rate of the optimal hyperparameters in the network is set to 0.001 according to the grid search algorithm, the weight decay coefficient is 0.0005, and the training strategy adopts the momentum gradient descent algorithm with a momentum term of 0.9, and the number of iterations is set to 50,000. Subsequently, we use the network model obtained from the training to detect the citrus image and citrus video, respectively, and to compare the accuracy of the citrus detection and the accuracy of the citrus video tracking.

For the experimental validation of the citrus counting strategy, we manually counted the 10 citrus videos obtained, and the actual number of citrus fruits in the videos was obtained by averaging the counting results obtained by the researchers. When counting, we played the video frame by frame, first recorded the number of fruits in the first frame, and then recorded the number of newly appeared fruits in the subsequent frames until the end of the video, to get the total number of fruits in the video and completed the manual counting of fruits. Subsequently, based on the algorithm of this paper and the proposed counting strategy, the number of citrus fruits in 10 videos were obtained separately, and the algorithm counting results were compared with the manual counting results to test the citrus video counting accuracy.

3.2. Experimental Evaluation Indicators

3.2.1. Fruits Detection Experiment

Precision (P), recall (R), harmonized average of precision and recall rates ($F1$), average Detection Precision (ADP), and detection time for detecting an image are used to evaluate the performance of the trained network mode. The performance of the obtained network model and the calculation formula is shown below. In addition, video detection accuracy (VDA) was used to represent the ratio of correctly detected targets to all targets by the model, with higher values indicating better model performance.

$$F1 = \frac{2PR}{P + R} \quad (7)$$

$$ADP = \int_0^1 P(R) dR \quad (8)$$

3.2.2. Fruits Tracking Experiment

Fruit ID switch rate (IDS), multiple object tracking accuracy ($MOTA$), and multiple object tracking precision ($MOTP$) are used to evaluate the tracking algorithm performance.

Fruit ID switching rate refers to the ratio of the fruits whose IDs have changed to all counted fruits in the video, and the smaller value means better tracking performance. Multi-target tracking accuracy indicates the performance of the tracking algorithm in maintaining the tracking trajectory and the larger value means the better tracking performance, as shown in Equation (9). Multi-target tracking accuracy refers to the success rate of matching all tracked targets; the higher the accuracy, the better the performance of the algorithm, as shown in Equation (10).

$$MOTA = M/S \quad (9)$$

$$MOTP = M/T \quad (10)$$

where S denotes the total number of fruits tracked in the video, i.e., the number of algorithm counts; M denotes the number of fruits tracked and matched correctly in the video; and T denotes the total number of fruits tracked and matched in the video, i.e., the number of detections.

3.2.3. Fruits Counting Experiment

In fruit counting, this paper adopts an average counting precision A_{cp} as the evaluation index of citrus counting effect. The larger the average counting precision, the more accurate the counting performance of the algorithm in this paper. The formula is:

$$A_{cp} = \frac{\sum_1^n \left(1 - \frac{|S-G|}{G}\right)}{n} \quad (11)$$

where S denotes the number of fruits in the counting video obtained by the algorithm of this paper, and G denotes the number of fruits in the manual counting citrus video, and n denotes the number of videos.

3.3. Results and Analysis

3.3.1. Fruit Detection Performance

In order to validate the detection accuracy of the detection algorithms YOLOv7-tiny on the citrus dataset, four target detection methods, YOLOv5, YOLOv5-tiny, YOLOX, and YOLOv7-tiny, were tested for comparison in this paper, and the test results are showed in Table 1.

Table 1. Test results of several different object detection algorithms on bagged grape dataset.

Methods for Object Detection	Precision	Recall Rate	Average Detection Precision
YOLOv5	96.62%	95.22%	94.17%
YOLOv5-tiny	95.19%	94.00%	94.04%
YOLOX	94.84%	93.24%	89.02%
YOLOv7-tiny	96.81%	96.39%	97.23%

The experiment results show that the precision, recall, and average precision of YOLOv7-tiny are 96.81%, 96.39%, and 97.23%, respectively, which are 1.97 percentage points, 3.15 percentage points, and 8.21 percentage points higher than those of YOLOX, and the precision, recall, and average precision of YOLOv7-tiny are 1.62 percentage points, 2.39 percentage points, and 3.19 percentage points higher than those of YOLOv5-tiny, respectively. Compared with YOLOv5, the precision, recall, and average precision are improved by 0.19 percentage points, 1.17 percentage points, and 3.06 percentage points, respectively. The superior detection performance of YOLOv7-tiny on the citrus dataset was demonstrated.

YOLOv5, YOLOv5-tiny, and YOLOX have unsatisfactory citrus detection mainly due to the following reasons: (1) when two detection targets overlap each other and obscure each other, YOLOv5, YOLOv5-tiny, and YOLOX will recognize the two overlapping citrus

as a single citrus, which leads to errors in tracking counts; (2) the identity ID jumping phenomenon will occur in the detection process of YOLOv5, YOLOv5-tiny, and YOLOX will occur identity ID jumping phenomenon during the detection process, counting the same citrus repeatedly; (3) in the complex environment of citrus orchards, the obscurity of leaves causes YOLOv5, YOLOv5-tiny, and YOLO to fail to detect the obscured citrus.

3.3.2. Fruit Tracking Performance

The proposed citrus tracking and counting algorithm based on YOLOv7-tiny and DeepSORT conducts tracking experiments on citrus videos in the dataset and obtains good tracking results. In the dynamic tracking experiments of citrus videos, fruits are obscured by multiple leaves, resulting in the intermittent detection of fruits, which produces intermittent appearance of detection frames, fragmentation identification, misidentification or identity ID jumps; meanwhile, camera motion-induced motion blur can cause prediction frame matching failures, which significantly affect tracking results. Therefore, we propose a two-stage life filter to improve the success rate of prediction frame matching, reduce fruit leakage, and mitigate identity ID jumping.

During the dynamic tracking of citrus, complex citrus growing environments make multiple conditions, which affect the accuracy of dynamic citrus tracking counting. Those conditions are shown in Table 2. Normal conditions can be easily detected and tracked; fruit overlap identifies two or more citrus as one, resulting in a reduced value counting; fruit obscured by trunk is difficult to be detected and always makes the detection jump; motion blur makes it difficult to detect citrus and result in reduced counts; fruit obscured by branches and leaves makes detection more difficult, which counts a citrus as more than one; fragmentary images result in reduced counts. For those problems arising from obscurity in complex environments, the two stages life filter proposed in this paper effectively improves it. The method reduces the leakage rate and improves the counting accuracy through the re-matching of new trajectories.

Table 2. Multiple conditions of citrus in orchard environment.

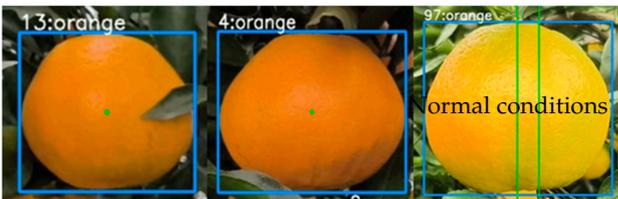
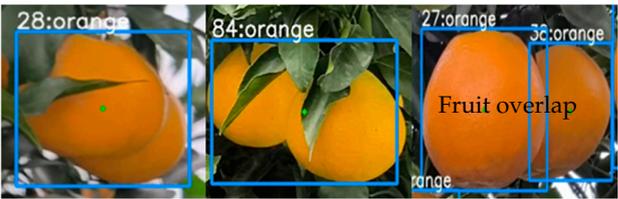
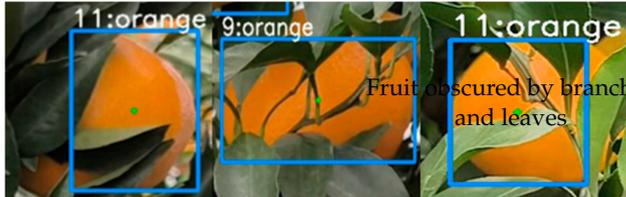
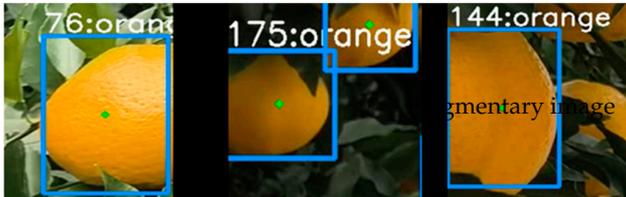
NO	Legend (of a Map, etc.)	Condition	Influence
1		Normal conditions	Detecting and counting targets
2		Fruit overlap	Resulting in a reduced number of counts
3		Fruit obscured by trunk	Making the detection frame jump and reduced counts

Table 2. Cont.

NO	Legend (of a Map, etc.)	Condition	Influence
4		Motion blur	Failing to detect citrus, resulting in reduced counts
5		Fruit obscured by branches and leaves	Counting a citrus as more than one
6		Fragmentary image	Resulting in reduced counts

Ten citrus videos (300–548 frames) were tracked and counted using the pre- and post-improvement tracking and counting algorithms, respectively, and the resulting tracking metrics are shown in Table 3. The MOTA of the improved algorithm is 67.14%, which is 21.78 percentage points higher than that of the pre-improved one. The MOTP is 17.88 percentage points higher, and the tracking credibility has been greatly improved. The high complexity of the citrus orchard environment and the precision of video shooting cause the identity ID of the citrus to jump very easily, which is the main reason for the tracking error. The algorithm improved by adding the two stages life filter reduces the IDSR in the detection result video from 25.97% to 15.63%, which reduces the fruit tracking error to a certain extent.

Table 3. Comparison of tracking results before and after algorithm improvement.

Algorithm Performance		Before Improvement	After Improvement	Results
Video detection accuracy	Normal conditions	95.12%	95.12%	No improvement
	obscured citrus	Same as above	Same as above	No improvement
ID switch rate	Normal conditions	25.97%	15.63%	Improvements
	obscured citrus	Same as above	Same as above	Improvements
Multiple objects tracking accuracy	Normal conditions	45.36%	67.14%	Improvements
	obscured citrus	5% decrease	Not affected	Improvements
Multiple objects tracking precision	Normal conditions	56.77%	74.65%	Improvements
	obscured citrus	7% decrease	Not affected	Improvements

3.3.3. Fruit Counting Accuracy

In order to accurately estimate the number of citrus in the video and realize the counting of citrus to estimate the production, we propose a drawing lines counting strategy to count citrus and conduct experiments on the citrus video dataset. We compare the

manual counting values, the algorithm's counting values before improvement, and the algorithm's counting values after improvement, then validate the effectiveness of the algorithm by fitting the manual counting values and the algorithm's counting values after improvement.

The fitted straight line is shown as Formula (12) and R^2 (The coefficient of determination) is 0.9982, $RMSE$ (Root Mean Square Error) is 32.300.

$$y = 1.0742x + 0.8083 \quad (12)$$

The $RMSE$ (root mean square error) between the number of fruits obtained by the algorithm and the number of fruits obtained by manual counting is 32.300. the counting accuracy of the 10 citrus videos based on Equation (11) was about 80%, with an average accuracy of 81.02%. From the fitting results, the coefficient of determination R^2 value is 0.9982, indicating that the algorithm in this paper has a significant linear correlation between the fruit counting values in modern orchard videos and the real values. From the fitting results, it can be seen that the counting accuracy does not change significantly with the number of citrus in the video, indicating that the method proposed in this paper has good robustness. The comparison of counting values is shown in Figure 5. Figure 5 shows that the algorithm counting values after improvement are closer to the real values of manual counting than the algorithm counting values before improvement.

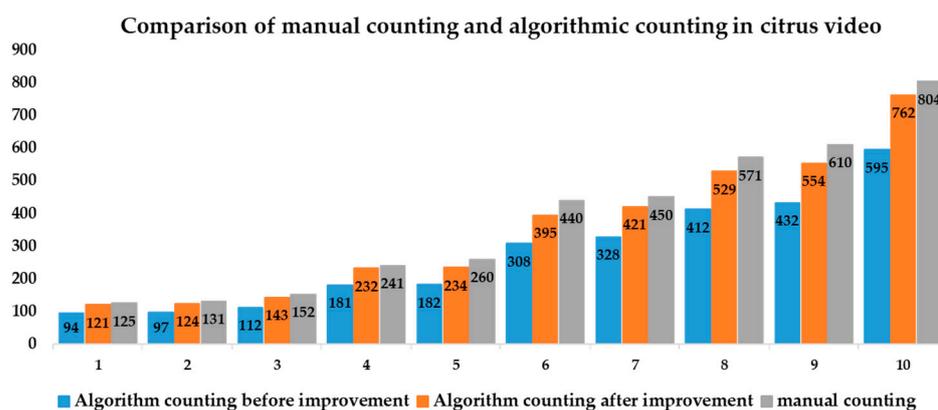


Figure 5. Comparison of manual counting and algorithmic counting in citrus video.

To test whether there is a significant difference between the algorithm counting values after improvement and manual counting value, we used the p -value test of the chi-square test. We take the manual counting value as the target value and the algorithm counting values after improvement as the actual value. Then we formulate the original hypothesis that there is no significant difference between the algorithm counting value and the manual value, and the alternative hypothesis that there is a significant difference between the algorithm counting value and the manual value. Finally, the statistic comes to be tested by the p -value. The p -value is the probability of the chi-square statistic, which indicates the probability that the observed chi-square statistic is greater than or equal to the actual observed chi-square statistic if the original hypothesis holds. If the p -value is less than the significance level, the original hypothesis is rejected and the alternative hypothesis is accepted, i.e., the algorithm counting value is significantly different from the manual value. Otherwise, the original hypothesis is accepted, i.e., there is no significant difference between the algorithm counting value and the manual value.

We calculated a p -value of 0.0133 for the data and in the chi-square test, the significance level for this statistic was chosen to be 0.01. Since the p -value was greater than the significance level, the original hypothesis was accepted that there is no significant difference between the algorithm counting value and the manual value. The result also provides support for the accuracy of the dynamic tracking and counting for obscured citrus in the Smart Orchard Based on Machine Vision.

4. Discussion

Table 4 studies the different fruit detection and counting methods applied to citrus. It compares the task, datasets, and methods and lists the advantages and disadvantages of each to better explain the experimental results obtained in this paper.

Table 4. Studies on different fruit detection and counting methods applied to citrus.

Year	Reference	Task	Dataset	Methods	Pros and Cons
2023	[1]	Detection and counting of citrus fruits in a greenhouse environment.	Smartphone; 7272 from a citrus plantations in Guangzhou city, China.	YOLOv5 model with SE attention mechanism; DeepSORT algorithm	Rapid, high precision but network structure is large
2023	[17]	Detection and automatic estimation of citrus with a lightweight model for edge computing devices	Smartphone; 40 videos from a citrus plantations in Guangxi, China.	CSPDarkNet53+PAFPN network; Improved FairMOT algorithm using Byte algorithm.	Can be used on edge computing equipment
2022	[26]	Counting the number of fruits in a single citrus tree based on static images.	MER-131-U3C Industrial Camera; Dataset from Chongqing, China.	K-Means clustering algorithm; Gradient-based Hough transform;	Have fewer practical applications
2021	[27]	Detection and counting of Kasumi citrus	AT-C0320X2-REV1.1 Camera; 900 photos and some videos from Chengdu city, China.	YOLOv4, CSPResNest50 network; the Recursive Feature Pyramid network.	Rapid, high precision but not as effective as YOLOv7-tiny
2019	[28]	Rapid automatic fruit detection and yield prediction	Smartphone iPhone XR; Dataset from Shiraz Ranch farm in California, USA	Application of MATLAB software (2017b) based on Hough's circle transform method.	Simple, rapid, good robustness but fewer practical application

In this study, we found that due to the complex citrus orchard environment and the biological characteristics of leafy citrus tree, the tracking and counting of citrus were difficult. According to the significant color difference between the citrus fruits and the orchard background, Jiang et al. [26] used the K-Means clustering algorithm and the Gradient-based Hough transform to count the number of fruits in a single citrus tree based on static images. The traditional method is simple but easily affected by the weather and other natural environments, which results in fewer practical applications.

Compared with traditional methods, deep learning often uses trained algorithms for target detection and tracking. Zhuang et al. [1] improved citrus fruits recognition by the incorporation of the SE attention mechanism and combined DeepSORT algorithm with improved YOLOv5. Yi et al. [27] improved the YOLOv4 by using CSPResNest50 network and the Recursive Feature Pyramid network. These YOLO algorithms are better than traditional but not as effective as YOLOv7-tiny.

Considering the practical applications, Xia et al. [17] achieved the detection and automatic estimation of citrus with a lightweight model for edge computing devices and Zhang et al. [28] used the Hough's circle transform method on MATLAB software (v. 2017b) to achieve rapid automatic fruit detection and yield prediction. These approaches are difficult to port to embedded devices and require specialized edge computing devices.

In this study, we proposed a citrus detection and dynamic counting method based on the lightweight target detection network YOLOv7-tiny, the Kalman filter tracking and

the Hungarian algorithm. The YOLOv7 has a fast detection speed, a high target detection accuracy, a lightweight network and easy portability to various embedded devices. The incorporation of the two stages life filter makes DeepSORT have high tracking efficiency and counting accuracy.

5. Conclusions

In this study, the YOLOv7-tiny algorithm and the DeepSORT algorithm are used to detect and track citrus in natural environments, while a citrus fruit counting strategy based on modern citrus plantations is proposed. The YOLOv7-tiny network model was trained using the acquired images of different citrus varieties in different locations, and its average detection precision (ADP) and video average detection accuracy (VDA) values reached 97.23% and 95.12%, and it takes only 0.011s to detect an image. Using this model, the citrus fruits are detected frame by frame in the video dataset, and the pixel coordinates of the citrus are obtained. Based on the Kalman filter to predict the citrus position, and successively based on the Euclidean distance and the intersection over union to measure the similarity between predicted and detected citrus. The Hungarian algorithm is used to match the same fruit inconsecutive video frames, and the prediction frames that failed to be matched are updated to a new trajectory which is re-tracked and matched again after two stages life filter. The multiple object tracking accuracy (MOTA) and multiple object tracking precision (MOTP) are 67.14% and 74.15%, respectively. Finally, fruit counting was achieved by a drawing lines counting strategy, combining the identity ID assigned to each citrus, and with an ACP (average counting precision) of 81.02%. Based on the algorithm and counting strategy of this paper, we obtained the number of citrus in 10 citrus videos, which was processed in combination with the number of manual counts to obtain a coefficient of determination of 0.9982. And the RMSE (root mean square error) is 32.300, which indicates the effectiveness of the drawing lines counting strategy.

In conclusion, the presented approach of dynamic tracking and counting for obscured citrus holds immense potential for smart orchard information monitoring. Its applications include yield estimates for different periods, efficiency enhancement, and inventory management. It will contribute to the development of smart orchards and provide support for the establishment of unmanned standard orchards. At the same time, relevant technologies should also continue to be studied to promote sustainable development and smart agriculture.

Author Contributions: Conceptualization, Y.F., W.M. and Y.T.; Software, Y.F. and A.G.; Validation, Z.T.; Investigation, J.Q. and H.Y.; Resources, Y.F.; Data curation, Y.F.; Writing—original draft preparation, Y.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Institute of Urban Agriculture, Chinese Academy of Agriculture Sciences (grant number: ASTIP2023-34-1UA-10); key projects of the Sichuan Provincial Science and Technology Plan (grant number 2022YFG0147); the Chengdu Local Finance Special Fund Project for NASC (grant number: NASC2022KR08), (grant number: NASC2021KR02) and (grant number: NASC2021KR07).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

1. Zhuang, H.; Zhou, J.; Lin, Y.; Pong, H.; Lin, H. Citrus Fruit Recognition and Counting Study Based on Improved YOLOv5+DeepSort. *South Forum* **2023**, *8*, 9–13.
2. Li, X.; Sun, D.; Liu, H. A Study on China's Citrus Export Growth Based on Ternary Marginal Analysis. *Chin. J. Agric. Resour. Reg. Plan.* **2021**, *42*, 110–118.

3. Sun, J.; Wang, D.; Liu, C.; Wu, X.; Wu, Q.; Cheng, H. Research progress of citrus on-tree storage. *China Fruits* **2021**, *7*, 1–6+10.
4. Mekhalfi, M.L.; Nicolò, C.; Ianniello, I.; Calamita, F.; Goller, R.; Barazzuol, M.; Melgani, F. Vision System for Automatic On-Tree Kiwifruit Counting and Yield Estimation. *Sensors* **2020**, *20*, 4214. [[CrossRef](#)] [[PubMed](#)]
5. Koirala, A.; Walsh, K.B.; Wang, Z.; Mccarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of ‘MangoYOLO’. *Precis. Agric.* **2019**, *20*, 1107–1135. [[CrossRef](#)]
6. Xu, H.; Ye, Z.; Ying, Y. A study on tree citrus recognition based on color information. *Trans. Chin. Soc. Agric. Eng.* **2005**, *21*, 98–101.
7. Xu, L.; Lv, J. Segmentation of Yangmei images based on homomorphic filtering and K-means clustering algorithm. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 202–208.
8. Lv, J.; Wang, F.; Xu, L.; Ma, Z.; Yang, B. A segmentation method of bagged green apple image. *Sci. Hortic.* **2019**, *246*, 411–417. [[CrossRef](#)]
9. Huang, H.; Duan, X.; Huang, X. Research and Improvement of Fruits Detection Based on Deep Learning. *Comput. Eng. Appl.* **2020**, *56*, 127–133.
10. Gou, Y.; Yan, J.; Zhang, F.; Sun, C.; Xu, Y. Research Progress on Vision System and Manipulator of Fruit Picking Robot. *Comput. Eng. Appl.* **2023**, *59*, 13–26.
11. Yan, J.; Zhao, Y.; Zhang, L.; Su, X.; Liu, H.; Zhang, F.; Fan, W.; He, L. Recognition of Rosa roxbunghii in natural environment based on improved Faster RCNN. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 144–151.
12. Lv, S.; Lu, S.; Li, Z.; Hong, T.; Xu, Y.; Wu, B. Orange recognition method using improved YOLOv3-LITE lightweight neural network. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 205–214.
13. Lukežič, A.; Vojíš, T.; ehovin Zajc, L.; Matas, J.; Kristan, M. Discriminative Correlation Filter Tracker with Channel and Spatial Reliability. *Int. J. Comput. Vis.* **2018**, *126*, 671–688. [[CrossRef](#)]
14. Gao, F.; Wu, Z.; Suo, R.; Zhou, Z.; Li, R.; Fu, L.; Zhang, Z. Apple detection and counting using real-time video based on deep learning and object tracking. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 217–224.
15. Lv, J.; Zhang, G.; Liu, Q.; Li, S. Bagged grape yield estimation method using a self-correcting NMS-ByteTrack. *Trans. Chin. Soc. Agric. Eng.* **2023**, *39*, 182–190.
16. Zhang, Y.; Wang, C.; Wang, X.; Zeng, W.; Liu, W. FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking. *Int. J. Comput. Vis.* **2021**, *129*, 3069–3087. [[CrossRef](#)]
17. Xia, X.; Chai, X.; Zhang, N.; Zhou, S.; Sun, Q.; Sun, T. A lightweight fruit load estimation model for edge computing equipment. *Smart Agric.* **2023**, *5*, 1–12.
18. Nicolai, W.; Alex, B.; Dietrich, P. Simple Online and Realtime Tracking with a Deep Association Metric. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.
19. Anton, M.; Leal-Taixe, L.; Reid, I.; Roth, S.; Schindler, K. *MOT16: A Benchmark for Multi-Object Tracking*; Cornell University Library: Ithaca, NY, USA, 2016.
20. Chien-Yao, W.; Bochkovskiy, A.; Hong-Yuan, M.L. *YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors*; Cornell University Library: Ithaca, NY, USA, 2022.
21. Jiang, T.; Zhao, J.; Wang, M. Bird Detection on Power Transmission Lines Based on Improved YOLOv7. *Appl. Sci.* **2023**, *13*, 11940. [[CrossRef](#)]
22. Xie, T.; Yao, X. Smart Logistics Warehouse Moving-Object Tracking Based on YOLOv5 and DeepSORT. *Appl. Sci.* **2023**, *13*, 9895. [[CrossRef](#)]
23. Li, Z.; Yang, X.; Li, L.; Chen, H. Iterated Orthogonal Simplex Cubature Kalman Filter and Its Applications in Target Tracking. *Appl. Sci.* **2024**, *14*, 392. [[CrossRef](#)]
24. Sinopoli, B.; Schenato, L.; Franceschetti, M.; Poolla, K.; Jordan, M.I.; Sastry, S.S. Kalman Filtering with Intermittent Observations. *IEEE Trans. Autom. Control* **2004**, *49*, 1453–1464. [[CrossRef](#)]
25. Mills-Tettey, G.A.; Stentz, A.; Dias, M.B. *The Dynamic Hungarian Algorithm for the Assignment Problem with Changing Costs*; Robotics Institute: Pittsburgh, PA, USA, 2007.
26. Jiang, X. Research on yield prediction of single mature citrus tree based on image recognition. *Agric. Mach. Agron.* **2022**, *12*, 66–68.
27. Yi, S.; Li, J.; Zhang, P.; Wang, D. Detecting and counting of spring-see citrus using YOLOv4 network model and recursive fusion of features. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 161–169.
28. Zhang, X.; Ma, R.; Wu, K.; Huang, Z.; Wang, J. Fast Detection and Yield Estimation of Ripe Citrus Fruit Based on Machine Vision. *Guangdong Agric. Sci.* **2019**, *46*, 156–161.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.