Article

Research on the Relative Position Detection Method between Orchard Robots and Fruit Tree Rows

Baoxing Gu [1], Qin Liu [2], Yi Gao [1], Guangzhao Tian [1,3,*], Baohua Zhang [4], Haiqing Wang [1] and He Li [1]

1   College of Engineering, Nanjing Agricultural University, Nanjing 210031, China; gbx@njau.edu.cn (B.G.); 9203011507@stu.njau.edu.cn (Y.G.); whq_nj@126.com (H.W.); lihe@njau.edu.cn (H.L.)
2   School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China; 230229544@seu.edu.cn
3   SUNWAY-AI Technology (Changzhou) Co., Ltd., Changzhou 213161, China
4   College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China; bhzhang@njau.edu.cn
*   Correspondence: tgz@njau.edu.cn

Abstract: The relative position of the orchard robot to the rows of fruit trees is an important parameter for achieving autonomous navigation. The current methods for estimating the position parameters between rows of orchard robots obtain low parameter accuracy. To address this problem, this paper proposes a machine vision-based method for detecting the relative position of orchard robots and fruit tree rows. First, the fruit tree trunk is identified based on the improved YOLOv4 model; second, the camera coordinates of the tree trunk are calculated using the principle of binocular camera triangulation, and the ground projection coordinates of the tree trunk are obtained through coordinate conversion; finally, the midpoints of the projection coordinates of different sides are combined, the navigation path is obtained by linear fitting with the least squares method, and the position parameters of the orchard robot are obtained through calculation. The experimental results show that the average accuracy and average recall rate of the improved YOLOv4 model for fruit tree trunk detection are 5.92% and 7.91% higher, respectively, than those of the original YOLOv4 model. The average errors of heading angle and lateral deviation estimates obtained based on the method in this paper are 0.57° and 0.02 m. The method can accurately calculate heading angle and lateral deviation values at different positions between rows and provide a reference for the autonomous visual navigation of orchard robots.

Keywords: orchard robot; autonomous navigation; positional parameters; machine vision; YOLO

1. Introduction

With the rapid development of sensor technology and computer technology, robotic autonomous navigation technology has been applied in orchards on a large scale [1–3]. Autonomous navigation technology is the key technology to realize the intelligence of agricultural equipment in orchards, which is conducive to reducing the work intensity of operators, improving work efficiency, and enhancing the quality of operations [4–6].

Autonomous navigation technology mainly perceives the dynamic environment around the vehicle through a variety of sensors to plan and navigate the path and complete operations, such as orchard weeding [7,8], furrowing and fertilizing [9–11], and picking [12,13]. Commonly used navigation methods are classified as Global Navigation Satellite Systems (GNSS)/Global Positioning Systems (GPS) [14,15], Light Detection and Ranging (LIDAR) [16,17], vision sensors [18,19], and multi-sensor fusion navigation [20]. Wei Shuang et al. [21] proposed a pure tracking model based on the GNSS autonomous navigation path search method for agricultural machines and pre-sighting point search. In straight-line navigation, the root mean square error of navigation is 3.79, 4.28, and 5.39 cm when the speed is 0.8, 1.0, and 1.2 m/s, respectively. Luo Xiwen et al. [22] developed an automatic navigation control system based on DGPS; the maximum error of this system is

less than 15 cm, and the average error is less than 3 cm in straight-line navigation when the tractor is at a speed of 0.8 m/s. However, when the fruit tree branches and leaves are luxuriant, the canopy is closed, and the environment between the rows of the tree is almost in the semi-closed state. Satellite signals are then blocked by the canopy, leading to certain limitations of the positioning and navigation method based on GNSS/GPS positioning and navigation methods subject to certain limitations. Liu Weihong et al. [23] proposed an orchard inter-row navigation method based on 3D LiDAR, the method in the real pear orchard with mobile robot tracking of tree rows at a speed of 1.35 m/s. The maximum lateral deviation obtained is less than 22.1 cm. 3D LiDAR and subject to weather conditions. It is expensive, greatly increasing the cost of inputs in orchard environments, and therefore difficult to use universally. Han Zhenhao et al. [24] proposed an orchard visual navigation path recognition method based on the U-Net network, which uses a camera to collect images and semantically segment the road information in the image to calculate the navigation path. Using this method to drive in the orchard road with a width of about 3.1 m, the average distance error ratio is 1.4%, which is about 4.4 cm, and the average processing time of a single frame of image is 0.154 s.

In recent years, deep learning has made breakthrough progress in image classification research and has also driven the rapid development of target visual detection [25–27]. Deep learning-based target detection models include two-stage target detection models and single-stage target detection models [28,29]. Two-stage target detection models have higher detection accuracy but slower detection speed; therefore, most real-time detection tasks currently use single-stage target detection models. Typical single-stage detection models include the YOLO [30–32] family of models and the SSD [33] model. Xie Shuang et al. [34] used an improved YOLOv8 model for tea recognition, which combines deformable convolution, an attention mechanism, and improved spatial pyramid pooling, thus enhancing the model's ability to learn complex target invariance, reducing the interference of irrelevant factors, and achieving multi-feature fusion, which improves the detection accuracy. Wang et al. [35] constructed a model that fuses YOLO v5s and the attention mechanism of convolutional neural network model YOLO_CBAM for the detection of spiny calyx lobelia weed. They devised a method for slicing high-resolution images. This method constructs the dataset by calculating the overlap rate to reduce the possibility of loss of details due to compression of high-resolution images during training, and the final accuracy is 92.72%. Tian et al. [36,37] successfully achieved the detection of grape maturity and weed identification and localization through the improved YOLOv4 algorithm. The experimental results show that this method has high accuracy.

By summarizing the work of relevant researchers, we can find that when working, orchard robots or agricultural robots mostly need navigation satellite systems for absolute positioning. However, in the actual working process, due to the obstruction of the tree canopy, the orchard robot may not be able to receive satellite signals normally. At this time, the orchard robot needs to obtain its relative pose with the tree row through its own sensors in real time. In this way, the orchard robot can navigate autonomously in the tree row.

This paper proposes a method based on a binocular camera to determine the positional parameters of an orchard robot. When the orchard robot advances between rows, the binocular camera acquires images and transmits them to the improved YOLOv4 model for fruit tree trunk detection, obtains the camera coordinates of the trunks, and then obtains the ground projection coordinates of each trunk after coordinate conversion. The ground projection coordinates of different sides of the orchard robot are combined to take the midpoint, the least squares method is used to fit a straight line to the midpoint to obtain the navigation path, and the heading angle and lateral deviation values of the orchard robot are obtained through calculation.

2. Materials and Methods

In this paper, the proposed method for estimating the positional parameters between rows of orchard robots consists of four parts: fruit tree trunk detection, fruit tree trunk

localization, navigation path calculation, and calculation of positional parameters. The detection of fruit tree trunks is based on the improved YOLOv4 model to detect the trunks on both sides of the orchard robot; the positioning of fruit tree trunks is based on the parallax principle after the binocular camera acquires the image to obtain the camera coordinates of the fruit tree trunks; the calculation of the navigation path is based on the camera coordinates of the fruit tree trunks obtained from the previous step through the coordinate transformation to obtain the ground projection coordinates, the combination of projection coordinates of the different sides of the orchard robot to take the mid-point; and the mid-point is calculated using the least-squares method. The least squares method is used to fit the midpoint to a straight line; the attitude parameters are obtained by calculating the straight line relationship between the fitted straight line and the vehicle base coordinate system.

2.1. Hardware Composition

In this study, the electric-driven tracked orchard operation platform designed by the College of Engineering of Nanjing Agricultural University is used as the mobile carrier, which is driven by two 48 V DC servo motors; the motors are driven and controlled by one dual-channel servo motor driver; the upper computer is the New Creation Cloud Embedded Industrial Controller with an external display screen; STM32 is selected as the lower computer to control the servo motor driver of the mobile carrier; the vision sensor uses the ZED2i (polarized version) high-definition camera; the camera's focal length f is 1.8 mm, the base distance b is 120 mm, and the camera captures the left and right images with a resolution of 1280 × 720. The overall scheme of the system is shown in Figure 1:
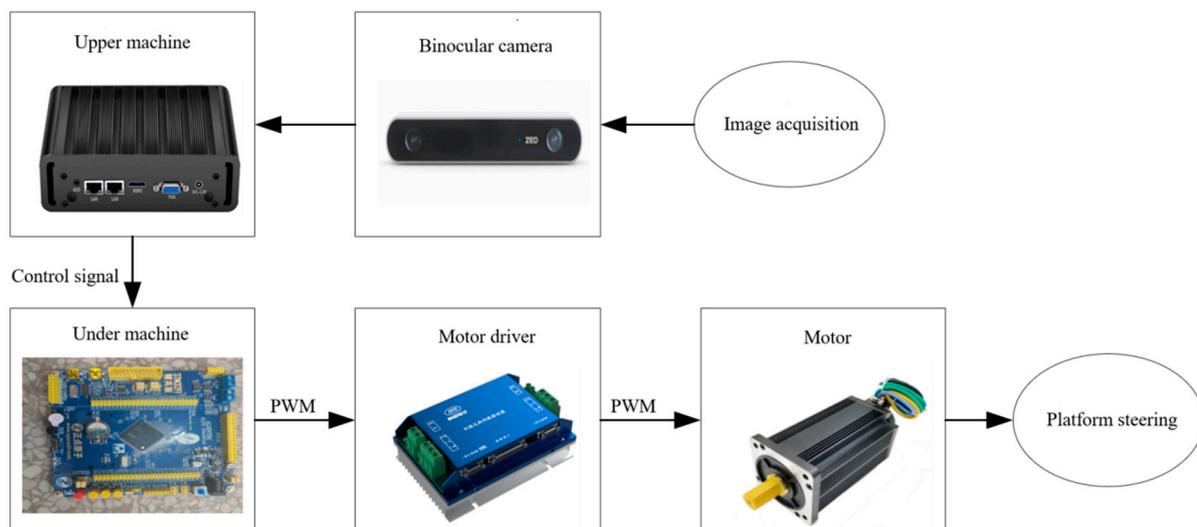

Figure 1. Entire design of the system.

2.2. Technological Route

The technical flow of the machine vision-based inter-row position estimation and navigation method for orchard robots at night is shown in Figure 2.

2.3. Improvements to the YOLOv4 Algorithm

2.3.1. YOLOv4 Algorithm

The YOLO family of algorithms is one of the most advanced target detection algorithms in the world. The YOLOv4 algorithm uses the CSPDarknet53 as the backbone network for feature extraction. It uses the CBM module (consisting of Convolution, Normalization, and Mish activation functions) and the CBL module (consisting of Convolution, normalization, and LeakyReLU activation functions) for feature extraction. It also joins the Spatial Pyramid Pooling Networks (SPPNet) and Path Aggregation Network (PANet)

modules for feature extraction while adding two network modules, SPPNet and PANet, so that the target detection accuracy of the YOLOv4 algorithm is better than that of the YOLOv3 algorithm. Additionally, the YOLOv4 algorithm adds image enhancement during training, thus further expanding the dataset.
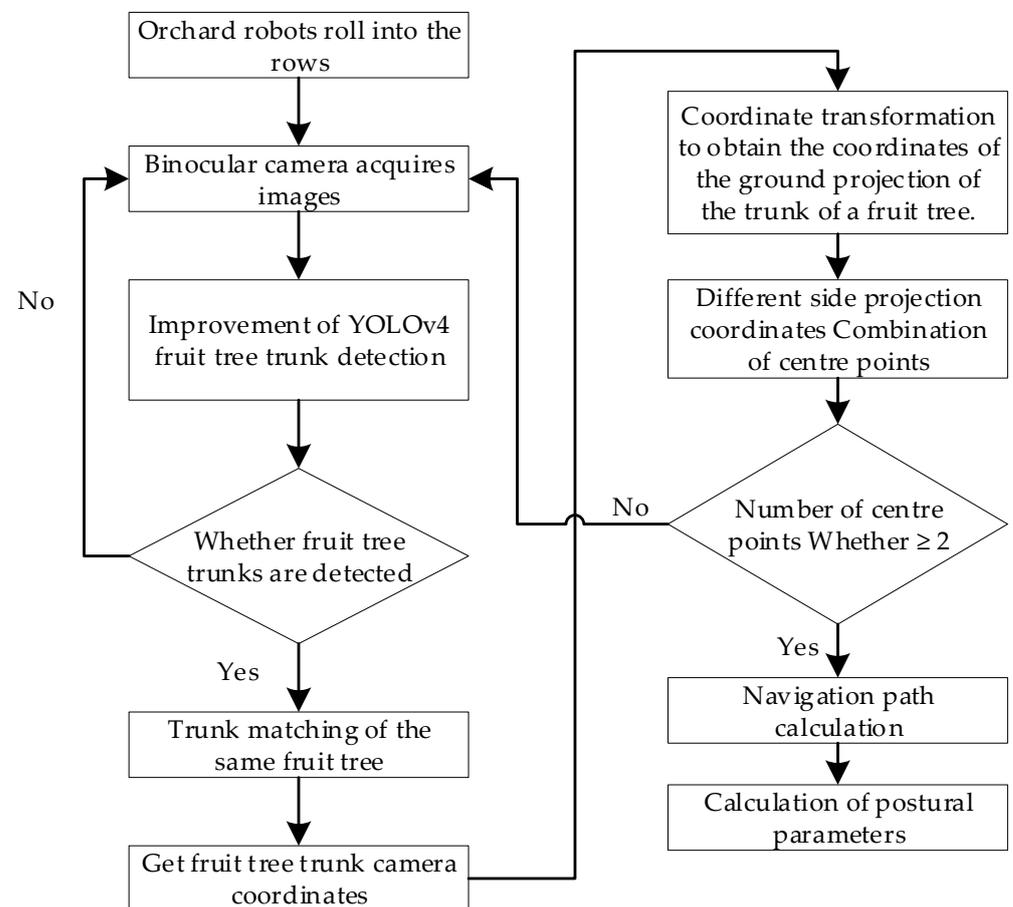
Figure 2. Technical process of estimating pose parameters between trees.

2.3.2. Improvement of the YOLOv4 Algorithm

In this paper, the YOLOv4 algorithm is used to detect fruit tree trunks. When the orchard robot travels between rows, the environment is relatively complex, and the detection accuracy of fruit tree trunks is an important prerequisite to ensure that the orchard robot can safely navigate. Therefore, it is necessary to carry out certain optimizations of the YOLOv4 algorithm so as to improve the detection accuracy of fruit tree trunks.

In this paper, we introduce an efficient channel attention mechanism ECA module, as shown in Figure 3, where $k = 5$ is used as an example. The ECA module maintains the dimensionality of the channel, performs a global average pooling (GAP) operation on the channel, and generates the channel weights using a Sigmoid activation function (σ) after one-dimensional convolution. Finally, it multiplies the channel weights with the original feature layer one by one to obtain a new feature layer, where the convolution kernel scale k can be adaptively determined and is proportional to the channel dimension. Therefore, if the channel dimension C is known, then the convolutional kernel scale k can be obtained by calculating Equation (1). The ECA module adds little computational effort to the algorithm and improves performance in all aspects.

$$k = \psi(C) = \left| \frac{\log_2(C)}{a'} + \frac{b'}{a'} \right|_{\text{odd}} \tag{1}$$

where $|*|_{\text{odd}}$ denotes the nearest odd number to $|*|$, and a' and b' are function coefficients. In this paper, a' takes the value of 2, and b' takes the value of 1.

In the YOLOv4 algorithm, the input dataset outputs three kinds of feature layers— $52 \times 52 \times 256$, $26 \times 26 \times 512$. and $13 \times 13 \times 1024$—after feature extraction by the backbone feature extraction network CSPDarknet53. Here, the $52 \times 52 \times 256$ and $26 \times 26 \times 512$ are input into the PANet module after one convolution, and the $13 \times 13 \times 1024$ feature layer is input into PANet after channel stacking and three convolutions after the SPP module. In this paper, the channel attention mechanism ECA module is added to these three feature layers before channel stacking and five convolution operations in PANet. In the PANet module, there are two up-samplings of the feature layer. This paper adds an ECA module after up-sampling and performs an enhanced feature extraction operation on the feature layer after up-sampling. This paper, therefore, adds five ECA modules to the original YOLOv4 algorithm and obtains the optimized new algorithm ECA5-YOLOv4 algorithm. The network model of the ECA5-YOLOv4 algorithm is shown in Figure 4.
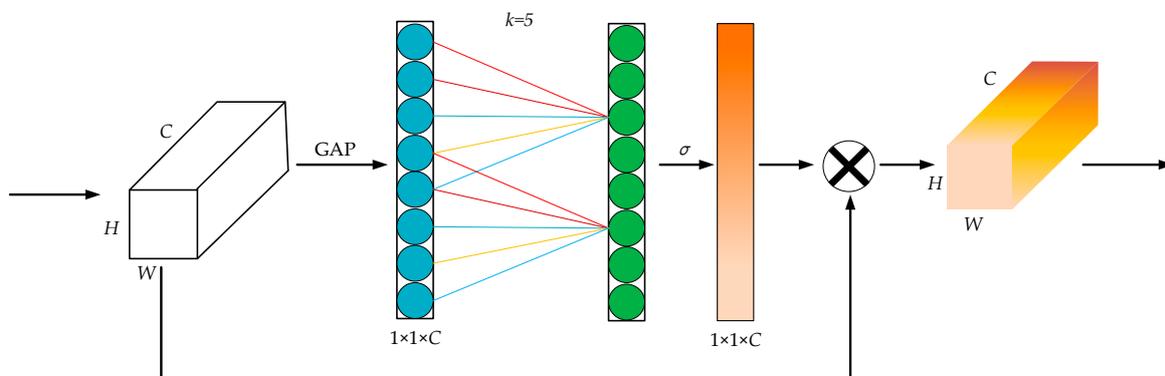

Figure 3. The ECA module. Note: C, W, and H denote the feature channel size, width, and height of the feature map, respectively; σ denotes the Sigmoid activation function.
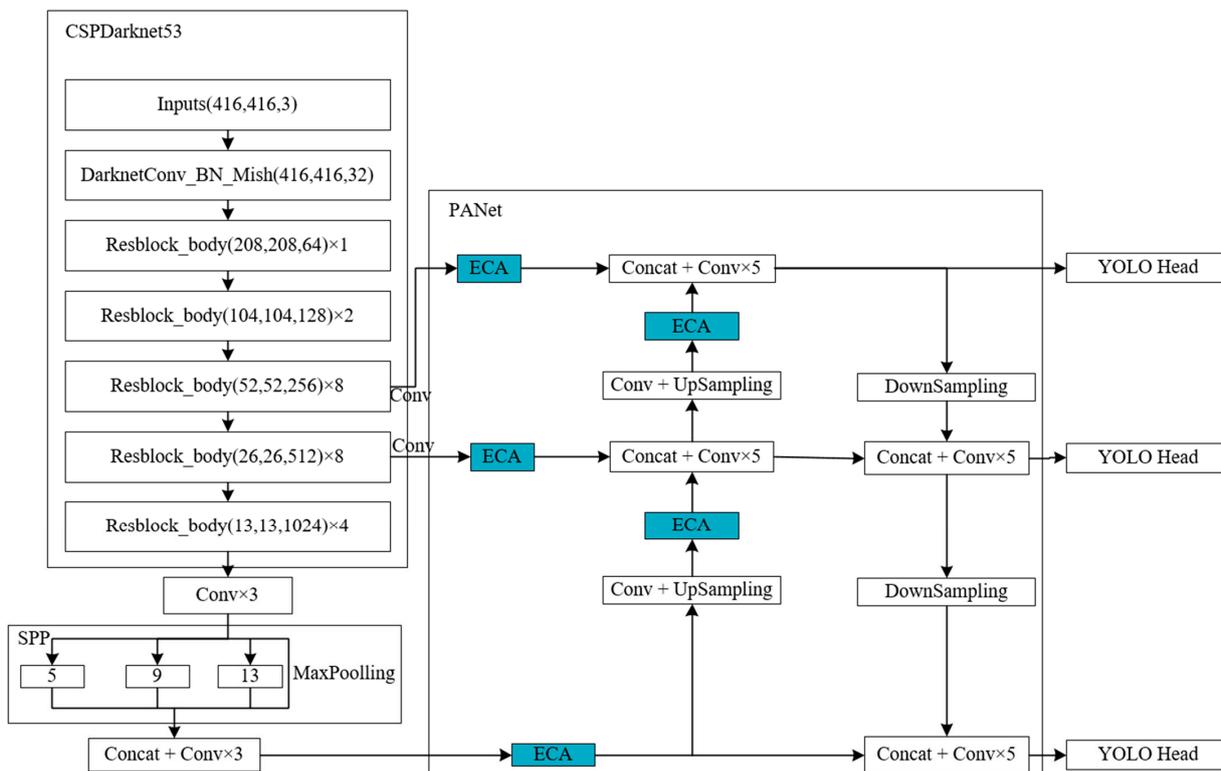

Figure 4. The network model of the ECA5-YOLOv4 algorithm.

2.4. Fruit Tree Trunk Positioning

2.4.1. Fruit Tree Trunk Camera Coordinates

Fruit tree trunks are detected by improving the YOLOv4 model on the left and right images captured by the binocular camera. After detecting the fruit tree trunks, the fruit tree trunks are matched, and when the same fruit tree trunks in the left and right images are matched successfully, the pixel parallax D and image parallax d of the same fruit tree trunks in the left and right images are obtained; according to the principle of triangulation of the binocular camera, the camera coordinates of the fruit tree trunks are obtained (X_C, Y_C, Z_C). The method in this paper is to project the ground so the Y_C coordinates do not need to be solved to simplify the calculation process. The pixel coordinates of the geometric center of the fruit tree trunk in the left image are (u_L, v_L), and the image coordinates are (x_L, y_L); the pixel coordinates of the geometric center of the fruit tree trunk in the right image are (u_R, v_R), and the image coordinates are (x_R, y_R):

$$D = (u_L - u_R) \tag{2}$$

$$d = (x_L - x_R) \tag{3}$$

Binocular camera left and right camera optical centers were expressed by O_L, O_R, with O_L as the origin, horizontally to the right for the axis X_L positive direction, vertically down for the axis Y_L positive direction, and horizontally forward for the positive direction of the Z_L, to establish the camera coordinate system O_L-$X_L Y_L Z_L$. f indicates the focal length of the binocular camera in mm; b indicates the baseline of the left and right cameras in mm. The structure of the binocular camera is shown in Figure 5.
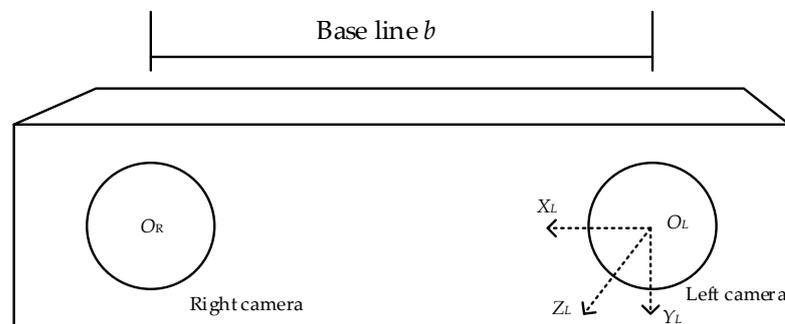

Figure 5. Structure of binocular camera.

The principle of conversion of pixel coordinates to image coordinates and the similarity triangle relationship between image coordinates and camera coordinates is rendered as follows:

$$X_C = \frac{Z_C \cdot (u_L - u_0)}{f_x} \tag{4}$$

where f_x denotes the pixel coordinate system u-axis scale factor, respectively, and u_0 denotes the amount of lateral translation of the image coordinate system origin in the pixel coordinate system.

According to the triangulation method,

$$Z_C = \frac{f \cdot b}{d} = \frac{f \cdot b}{x_L - x_R} \tag{5}$$

The horizontal coordinates of the geometric center of the fruit tree trunk in the pixel coordinate system, u_L, u_R in the left and right images, respectively:

$$\begin{cases} u_L = \frac{f \cdot x_L}{f_x} + u_0 \\ u_R = \frac{f \cdot x_R}{f_x} + u_0 \end{cases} \tag{6}$$

From Equations (5) and (6), the depth Z_C is obtained as follows

$$Z_C = \frac{f_x \cdot b}{u_L - u_R} \tag{7}$$

Substituting Equation (7) into Equation (4), there is:

$$X_C = \frac{b \cdot (u_L - u_0)}{u_L - u_R} \tag{8}$$

The coordinates (Xc, Zc) of the fruit tree trunk in the planar coordinate system $X_L O_L Z_L$ are obtained from Equations (7) and (8) where, b, u_0, f_x are camera internal parameters, which can be obtained by camera calibration.

2.4.2. Fruit Tree Trunk Coordinate Conversion

Using the coordinate system and conversion relationship shown in Figure 6, this paper will provide the orchard operation platform on the ground projection for the ground coordinate system origin O: the binocular camera installed in the front of the vehicle in the center; binocular camera geometric center of the projection of the ground for the point F; $\overrightarrow{OF}$ for the ground coordinate system axis Z positive direction; horizontally to the right for the axis of the X-positive direction; and the establishment of the ground coordinate system XOZ. The ground coordinate system can be obtained by the camera coordinate system in the planar coordinate system $X_L O_L Z_L$. The translation of X and Z direction is X_0 and Z_0, respectively. When the camera coordinates are converted to ground coordinates, the Y-axis direction is not to be calculated. The fruit tree trunk in the plane coordinate system $X_L O_L Z_L$ coordinates for (X_C, Z_C), and, in the ground, coordinate system coordinates for (X_g, Z_g); the conversion relationship is shown in Equation (9).

$$\begin{cases} X_g = X_C - X_0 \\ Z_g = Z_C + Z_0 \end{cases} \tag{9}$$
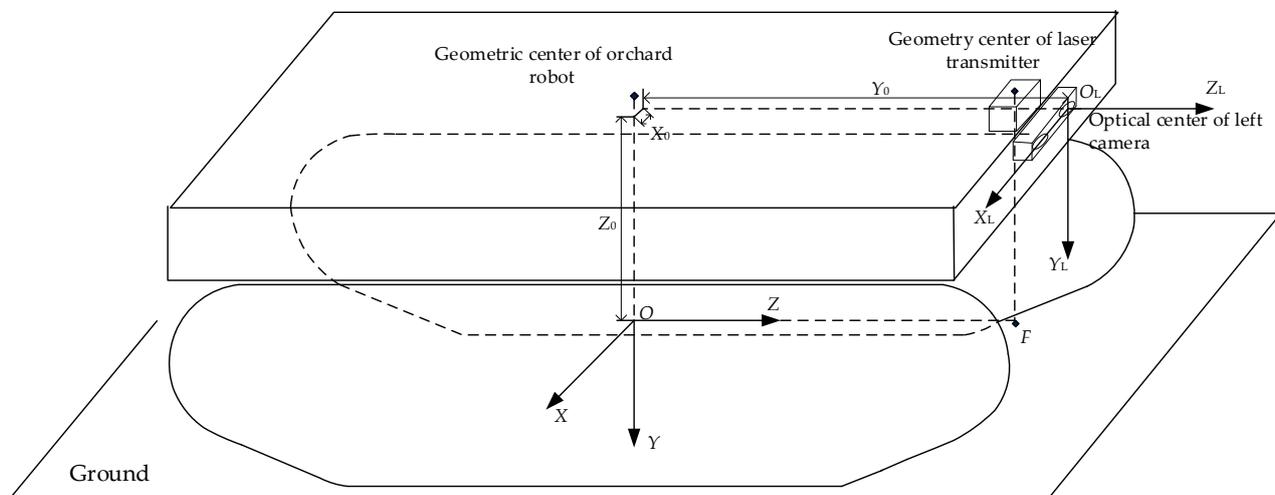

Figure 6. Coordinate system and transformation relations.

2.5. Calculation of Navigation Path and Attitude Parameters

When the orchard robot is traveling between rows, the binocular camera is used to locate the tree trunks. Then, the coordinate conversion is used to obtain the sitting mark of a plurality of fruit tree trunks in the ground coordinate system as $G_i(X_{gi}, Z_{gi})$. Among them, the coordinate point of $G_i(X_{gi}, Z_{gi}) < 0$ is regarded as being located on the left side of the orchard operation vehicle, and at this time, the left-side sitting mark is $G_{Lj}(X_{glj}, Z_{glj})$. The coordinate point of $X_{gi} \geq 0$ is regarded as being located on the right side of the orchard operation vehicle, and at this time, the right-side sitting mark is $G_{Rk}(X_{grk}, Z_{grk})$.

The combination of points on the left and the right is taken to be the midpoint, with a total of $j \times k$ midpoints, denoted $C_{j,k}(X_{j,k}, Z_{j,k})$.

$$\begin{cases} X_{j,k} = \frac{X_{glj}+X_{grk}}{2} \\ Z_{j,k} = \frac{Z_{glj}+Z_{grk}}{2} \end{cases} \tag{10}$$

2.5.1. Navigation Path Calculation

The obtained $j \times k$ midpoints are fitted with a straight line using the least squares method, and the calculated straight line is the navigation path, noting that the fitted straight line is $Z = a_1 X + a_0$. The calculation process is as follows:

$$\begin{bmatrix} j \times k & \sum\limits_{j=1,k=1}^{j,k} X_{j,k} \\ \sum\limits_{j=1,k=1}^{j,k} X_{j,k} & \sum\limits_{j=1,k=1}^{j,k} X_{j,k}^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1,k=1}^{j,k} Z_{j,k} \\ \sum\limits_{j=1,k=1}^{j,k} X_{j,k} Z_{j,k} \end{bmatrix} \tag{11}$$

Find a_0 and a_1 from Equation (11):

$$\begin{cases} a_0 = \dfrac{\sum\limits_{j=1,k=1}^{j,k} X_{j,k}^2 \sum\limits_{j=1,k=1}^{j,k} Z_{j,k} - \sum\limits_{j=1,k=1}^{j,k} X_{j,k} \sum\limits_{j=1,k=1}^{j,k} X_{j,k} Z_{j,k}}{(j \times k) \cdot \sum\limits_{j=1,k=1}^{j,k} X_{j,k}^2 - \left(\sum\limits_{j=1,k=1}^{j,k} X_{j,k} \right)^2} \\[6mm] a_1 = \dfrac{\sum\limits_{j=1,k=1}^{j,k} X_{j,k} Z_{j,k} - \sum\limits_{j=1,k=1}^{j,k} X_{j,k} \sum\limits_{j=1,k=1}^{j,k} Z_{j,k}}{\sum\limits_{j=1,k=1}^{j,k} X_{j,k}^2 - \left(\sum\limits_{j=1,k=1}^{j,k} X_{j,k} \right)^2} \end{cases} \tag{12}$$

2.5.2. Calculation of Postural Parameters

The positional relationship of the orchard operation platform when traveling between rows is shown in Figure 7, and the traveling direction of the vehicle is the same as the positive direction of the Z-axis, which can be expressed by $X = 0$. Therefore, the heading angle φ of the vehicle can be obtained by calculating the angle between $X = 0$ and $Z = a_1 X + a_0$; the lateral deviation λ of the vehicle can be obtained by calculating the perpendicular distance from the origin O to the fitted straight line $Z = a_1 X + a_0$. The calculation formulas are shown in (13) and (14):

$$\varphi = 90° - |\arctan a_1| \tag{13}$$

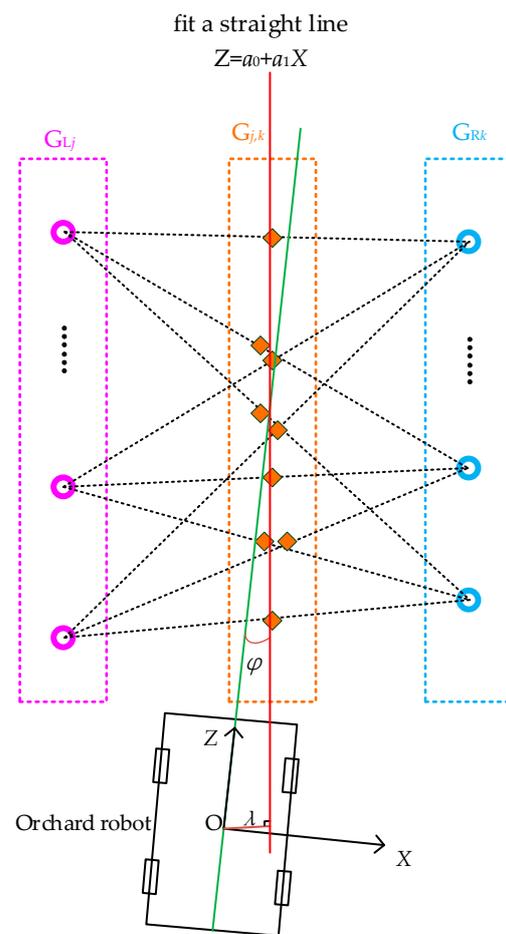$$\lambda = \frac{|a_0|}{\sqrt{1+a_1^2}} \tag{14}$$

Figure 7. Driving position relationship between rows.

3. Results

3.1. Data Acquisition and Model Training

The experimental dataset collection site was located in the College of Engineering, Nanjing Agricultural University, and 700 fruit tree trunk images were collected. Data enhancement was performed on the images by adding noise and flipping to expand the dataset to 1500 sheets, including 1200 sheets for the training set and 300 sheets for the validation set. The fruit tree trunk part was labeled using the LabelMe tool with the category information "tree."

In order to verify the higher accuracy of the ECA5-YOLOv4 algorithm, this paper replaces the five ECA modules in the model with the new attention mechanism modules, SENet module, and CBAM module. The original YOLOv4 algorithm, ECA5-YOLOv4 algorithm, SENet5-YOLOv4 algorithm, and CBAM5-YOLOv4 algorithm were trained for 500 generations using the same dataset, respectively. The loss functions of the training models are shown in Figure 8.

Training platform: Intel(R) Xeon(R) E5 2689 2.60 GHz CPU, 32 G RAM, NVIDIA GeForce GTX 1070 8 G graphics card.

The evaluation metrics of training results in this experiment include Precision (*P*), Recall (*R*), and Frame rate. *P* and *R* are calculated as:

$$P = \frac{T_P}{T_P + F_P} \times 100\% \tag{15}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{16}$$

where T_P denotes the number of correctly detected fruit tree trunks in the picture, F_P denotes the number of detection errors in the picture, and F_N denotes the number of missed targets in the picture.
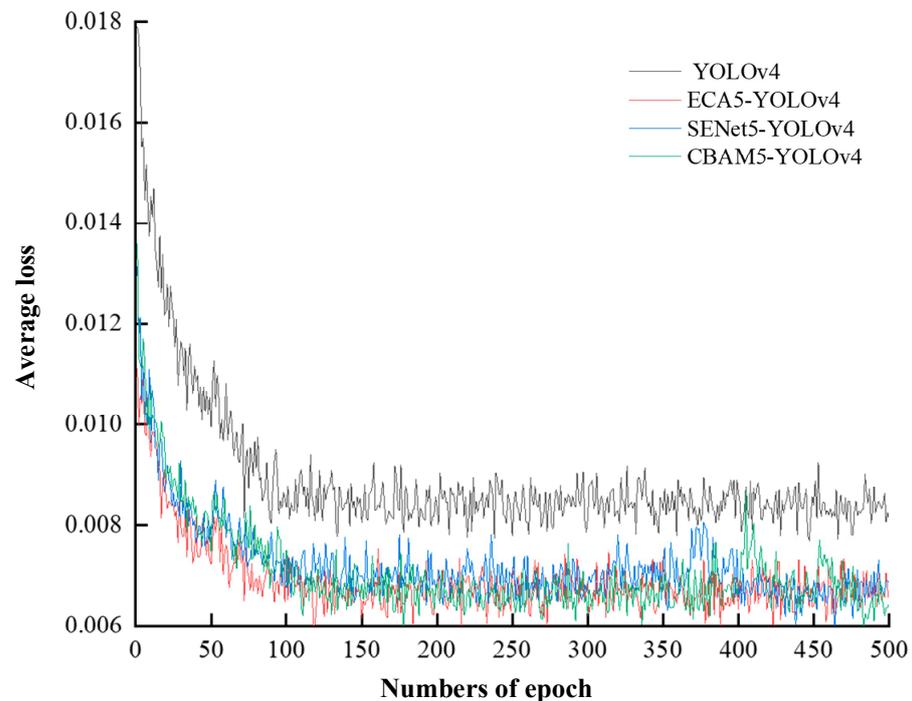

Figure 8. The loss function of model training.

Table 1 shows the training results of the original YOLOv4 algorithm, the ECA5-YOLOv4 algorithm, the SENet5-YOLOv4 algorithm, and the CBAM5-YOLOv4 algorithm.

Table 1. Results of model training.

Model	Precision/%	Recall/%	Frame Rate/(Frame·s^{-1})
YOLOv4	91.13	87.51	15.47
ECA5-YOLOv4	97.05	95.42	17.59
SENet5-YOLOv4	94.25	91.01	18.03
CBAM5-YOLOv4	96.83	95.14	17.50

From the result analysis, the ECA5-YOLOv4 algorithm, SENet5-YOLOv4 algorithm, and CBAM5-YOLOv4 algorithm have shown improvements in precision rate, recall rate, and frame rate compared to the original YOLOv4 algorithm. Among them, the ECA5-YOLOv4 algorithm has the highest precision rate and recall rate of 97.05% and 95.42%, respectively, which are improvements of 5.92%, 2.8%, and 0.22% in precision rate, and 7.91%, 4.41%, and 0.28% in recall rate when compared to the YOLOv4 algorithm, SENet5-YOLOv4 algorithm, and CBAM5-YOLOv4 algorithm, respectively. The SENet5-YOLOv4 algorithm has the highest frame rate, which is 2.56 fps, 0.44 fps, and 0.53 fps higher than the YOLOv4 algorithm, ECA5-YOLOv4 algorithm, and CBAM5-YOLOv4 algorithm, respectively.

3.2. Posture Parameter Determination Test

3.2.1. Binocular Camera Internal Reference Measurement

In this experiment, a ZED2i (polarized version) HD camera was used; the focal length f of the binocular camera was 1.8 mm, the base distance b was 120 mm, and the resolution of the camera grabbing the left and right images was 1280 × 720. The internal parameters of the camera were calibrated using the Software Development Kit (SDK 4.0) that comes

with the binocular camera. According to the needs of fruit tree trunk localization, u0 and result were obtained as 645.25 pixels and 529.88 pixels, respectively.

3.2.2. Experimental Design and Evaluation Indicators

This experiment was carried out in the College of Engineering of Nanjing Agricultural University, and the orchard robot was driven into the rows of fruit trees and stopped at any position in the rows, at which time the real values of the heading angle and lateral deviation of the orchard robot were measured and recorded as φ_t, λ_t; the values measured by the method in this paper are estimated values and recorded as φ_e, λ_e. The difference between the real values of the heading angle and lateral deviation and the estimated values are the error values and recorded as E_φ, E_λ, which were taken as the evaluation indexes of the present experiment. The valuation index is calculated as follows:

$$E_\varphi = |\varphi_t - \varphi_e| \tag{17}$$

$$E_\lambda = |\lambda_t - \lambda_e| \tag{18}$$

3.2.3. Experimental Results and Analysis

The fruit tree trunks are detected by different models, and the positional parameters of the orchard robot are also calculated. The fruit tree trunk results are shown in Figure 9, and the calculation results are shown in Table 2. The ECA5-YOLOv4 image at Site 1 is enlarged separately, and the effect is shown in Figure 10. In Figure 9, the actual heading angle and lateral deviation of the orchard robot from Position 1 to Position 3 are 150.2° and 0.53 m, 158.5° and 0.48 m, and 160.4° and 0.52 m, respectively. From Table 2, it can be seen that the best estimation of the positional parameters of the orchard robot was made by the ECA5-YOLOv4 model, with the mean values of the errors of heading angle and lateral deviation being 0.57° and 0.02 m. The results were compared with those of the original YOLOv4, SENet5-YOLOv4, and CBAM5-YOLOv4 models; the mean values of errors were reduced by 1.50° and 0.01 m, 0.86° and 0.03 m, and 0.60° and 0.03 m, respectively.
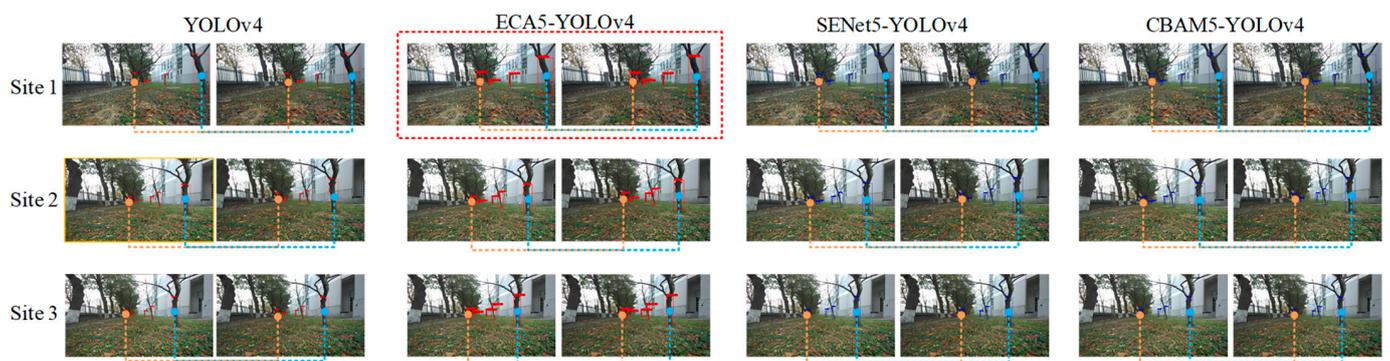

Figure 9. The results of tree trunk detection based on different models.

Table 2. Test results of pose parameters determination.

| Site	YOLOv4		ECA5-YOLOv4		SENet5-YOLOv4		CBAM5-YOLOv4	
	Heading Angle φ_e/°	Lateral Deviation λ_e/m	Heading Angle φ_e/°	Lateral Deviation λ_e/m	Heading Angle φ_e/°	Lateral Deviation λ_e/m	Heading Angle φ_e/°	Lateral Deviation λ_e/m
1	152.9	0.48	151.1	0.50	151.9	0.46	152.2	0.48
2	156.2	0.45	158.0	0.49	156.7	0.52	157.5	0.45
3	161.6	0.50	160.7	0.55	159.6	0.48	159.9	0.46
mean value of error	2.07	0.03	0.57	0.02	1.43	0.05	1.17	0.05
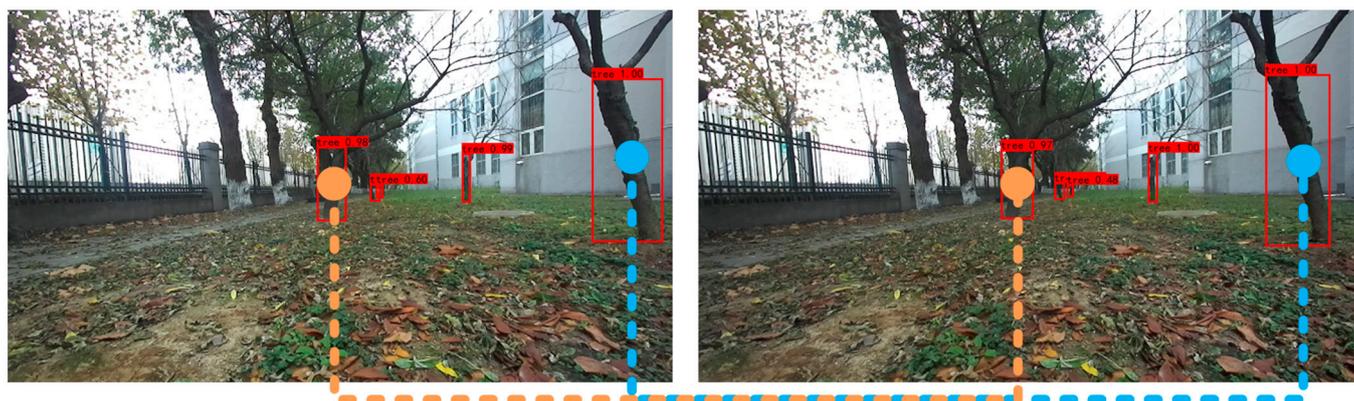
Figure 10. Enlarged ECA5-YOLOv4 image.

4. Discussion

In this study, based on the YOLOv4 model, the ECA5-YOLOv4 model is obtained by introducing five ECA attention mechanism modules into the PANet module of the model, and the ECA5-YOLOv4 model can efficiently and accurately detect the trunks of fruit trees, which can provide a guarantee for obtaining the positional parameters of the orchard robot.

(1) Compared with the original YOLOv4, SENet5-YOLOv4, and CBAM5-YOLOv4 models, the accuracy of the ECA5-YOLOv4 model for fruit tree trunks improved by 5.92%, 2.8%, and 0.22%, respectively;

(2) The estimation method of inter-row position parameters of the orchard robot proposed in this paper obtains the mean values of the errors of heading angle and lateral deviation as 0.57° and 0.02 m, with low errors, which can provide a theoretical basis for the orchard robot to navigate between the rows of fruit trees.

References

1. Zhao, C.J. The development current situation and future outlook of smart agriculture. *J. South China Agric. Univ.* **2021**, *42*, 1–7.
2. Hu, J.T.; Gao, L.; Bai, X.P.; Li, T.C.; Liu, X.G. Research progress in automatic navigation technology of Agricultural machinery. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 1–10.
3. Zhou, J.; He, Y.Q. Research progress in agricultural machinery navigation path planning. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 1–14.
4. Qi, F.; Zhu, M.; Zhou, X.Q.; Wei, X.M. Analysis of the relationship between agricultural engineering and China's agricultural modernization. *Trans. Chin. Soc. Agric. Eng.* **2015**, *31*, 1–10.
5. Luo, X.W.; Liao, J.; Hu, L.; Zang, Y.; Zhou, Z.Y. Improve the level of agricultural mechanization to promote sustainable agricultural development. *Trans. Chin. Soc. Agric. Eng.* **2016**, *32*, 1–11.

6.   Guo, C.Y.; Zhang, S.; Zhao, J.; Chen, J. Research on autonomous navigation system of orchard agricultural vehicles based on RTK-BDS. *J. Agric. Mech. Res.* **2020**, *42*, 254–259.

7.   Li, X.J.; Wang, P.F.; Yang, X.; Li, J.P.; Liu, H.J.; Yang, X.V. Design and test of orchard ridge mower. *J. Chin. Agric. Mech.* **2019**, *40*, 47–52.

8.   Zhou, Z.K.; Hu, J.N.; Zhang, C.Y.; Yang, J.M.; Ren, Z.H. Design of Obstacle Avoidance Lawn Mower Control System Based on Image and LiDAR. *J. Agric. Mech. Res.* **2022**, *44*, 80–84.

9.   Song, C.C.; Zhou, Z.Y.; Wang, G.B.; Wang, X.W.; Zang, Y. Optimization of groove wheel structure parameters of groove wheel fertilizer ejector of fertilization UAV. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 1–10.

10. Song, Y.P.; Zhang, Z.H.; Fan, G.Q.; Gao, D.S.; Zhang, H.M.; Zhang, X.H. Research status and development trend of orchard ditch fertilization machinery in China. *J. Chin. Agric. Mech.* **2019**, *40*, 7–12.

11. Zhang, H.J.; Xu, C.B.; Liu, S.X.; Jiang, H.; Zhang, C.F.; Wang, J.X. Design and Experiment of Double Row Ditching Fertilizer Applicator with Automatic Depth Adjustment. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 62–72.

12. Hu, G.R.; Kong, W.Y.; Qi, C.; Zhang, S.; Bu, L.X.; Zhou, J.G.; Chen, J. Optimization of navigation path for mobile picking Robot in orchard Environment. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 175–184.

13. Li, T.; Qiu, Q.; Zhao, C.J.; Xie, F. Task planning of multi-arm picking robot in dwarf dense planting orchard. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 1–10.

14. Han, J.; Park, C.; Park, Y.; Kwon, J.H. Preliminary results of the development of a single-frequency GNSS RTK-based autonomous driving system for a speed sprayer. *J. Sens.* **2019**, *2019*, 4687819. [CrossRef]

15. Yue, S.; Zhu, H.P.; Ozkan, H.E. Development of a variable-rate sprayer with laser scanning sensor to synchronize spray outputs to tree structures. *Trans. ASABE* **2012**, *55*, 773–781.

16. Zhang, Y.Y.; Zhou, J. Detection of orchard trunk based on lidar. *J. China Agric. Univ.* **2015**, *20*, 249–255.

17. Niu, R.X.; Zhang, X.Y.; Wang, J.; Zhu, H.; Huang, J.; Chen, Z.W. A trunk detection algorithm based on agricultural robot orchard. *Trans. Chin. Soc. Agric. Mach.* **2020**, *51*, 21–27.

18. Bi, S.; Wang, Y.H. Orchard robot visual navigation interline pose estimation and fruit tree target positioning method. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 16–26.

19. Cai, S.P.; Sun, Z.M.; Liu, H.; Wu, H.X.; Zhang, Z.Z. Real-time detection method of orchard obstacles based on the modified YOLOv4. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 36–43.

20. Zhong, Y.; Xue, M.Q.; Yuan, H.L. Design of GNSS/INS combined navigation System. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 40–46.

21. Wei, S.; Li, S.C.; Zhang, M.; Ji, Y.H.; Xiang, M.; Li, M.Z. Agricultural machinery automatic navigation path search and steering control based on GNSS. *Trans. Chin. Soc. Agric. Eng.* **2017**, *33*, 70–77.

22. Luo, X.W.; Zhang, Z.G.; Zhao, Z.X.; Chen, B.; Hu, L.; Wu, X.P. DGPS automatic navigation control system of Dongfanghong X-804 tractor. *Trans. Chin. Soc. Agric. Eng.* **2009**, *25*, 139–145.

23. Liu, W.H.; He, X.K.; Liu, Y.J.; Wu, Z.M.; Yuan, C.J.; Liu, L.M.; Qi, P.; Li, T. 3D LiDAR navigation method between orchard rows. *Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 165–174.

24. Han, Z.H.; Li, J.; Yuan, Y.W.; Fang, X.F.; Zhao, B.; Zhu, L.C. Path recognition of orchard visual navigation based on U-Network. *Trans. Chin. Soc. Agric. Mach.* **2021**, *52*, 30–39.

25. Feng, C.Y.; Nie, G.H.; Naveed, Q.N.; Potrich, E.; Sankaran, K.S.; Kaur, A.; Sammy, F. Optimization of sorting robot control system based on deep learning and machine vision. *Math. Probl. Eng.* **2022**, *2022*, 5458703. [CrossRef]

26. Nasirahmadi, A.; Sturm, B.; Edwards, S.; Jeppsson, K.; Olsson, A.C.; Müller, S.; Hensel, O. Deep Learning and Machine Vision Approaches for Posture Detection of Individual Pigs. *Sensors* **2019**, *19*, 3738. [CrossRef]

27. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

28. Ju, M.; Hu, Q. Brain-inspired filtering Network for small infrared target detection. *Multimed. Tools Appl.* **2023**, *82*, 28405–28426.

29. Qian, H.; Wang, H.; Feng, S.; Yan, S. FESSD:SSD target detection based on feature fusion and feature enhancement. *J. Real-Time Image Process.* **2023**, *20*, 2. [CrossRef]

30. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017.

31. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.

32. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.

33. 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 Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Amsterdam, The Netherlands, 2016.

34. Xie, S.; Sun, H. Tea-YOLOv8s: A tea bud detection model based on deep learning and computer vision. *Sensors* **2023**, *23*, 6576. [CrossRef] [PubMed]

35. Wang, Q.; Cheng, M.; Huang, S.; Cai, Z.; Zhang, J.; Yuan, H. A deep learning approach incorporating YOLO v5 and attention mechanisms for field real-time detection of the invasive weed *Solanum rostratum* dunal seedlings. *Comput. Electron. Agric.* **2022**, *199*, 107194. [CrossRef]

36. Zhao, J.W.; Tian, G.Z.; Qiu, C.; Gu, B.X.; Zheng, K.; Liu, Q. Weed detection in potato fields based on improved YOLOv4: Optimal speed and accuracy of weed detection in potato fields. *Electronics* **2022**, *11*, 3709. [CrossRef]
37. Qiu, C.; Tian, G.Z.; Zhao, J.W.; Liu, Q.; Xie, S.J.; Zheng, K. Grape maturity detection and visual pre-positioning based on improved YOLOv4. *Electronics* **2022**, *11*, 2677. [CrossRef]