Design of a System for Obtaining the Diameter of Tree Trunks Using Depth Vision

Ma Ruiting, Chen Qilin, Li Ming*, Zhu Wenjia, Chen Lina, Wang Runtao, Ma Tao, Wu Jiansheng, Xuan Zijian
School of Electronic and Electrical Engineering, Lingnan Normal University, Zhanjiang, Guangdong, China
**Corresponding author*

Abstract: In order to improve the autonomous navigation and operation efficiency of orchard robots, this study proposes an improved YOLOv5 algorithm to identify the trunk of the fruit tree, and combines with the depth camera for accurate positioning. By combining the SENet attention mechanism module with the residual module in the network, we obtain the improved SE-Res module, which can enhance the extraction of useful feature information and compress useless feature information. After experimental verification, the accuracy of the improved YOLOv5 model is increased by 2.38 percentage points, the recall rate is increased by 0.84 percentage points, the frame rate is reduced by 0.99 frames/s, and the mAP is increased by 0.05 percentage points. Experimental results show that the method can accurately identify and locate fruit trees in the autonomous navigation and fertilization of orchard robots, so as to improve operation efficiency and ensure operation quality, and realize the intellectualization and rationalization of fruit tree fertilization in the orchard environment.

Keywords: YOLOv5 Algorithm; Attention Mechanism; Depth Image; Target Detection; Fruit Tree Trunk; Accuracy

1. Introduction

With the rapid development of computer technology, computer image processing technology continues to make progress. People are increasingly interested in using image acquisition and processing equipment and vision image processing technology to realize fruit tree recognition. Fruit tree identification and positioning technology plays a vital role in improving the efficiency of orchard robots, improving production quality and reducing costs.Early fruit tree recognition techniques mainly rely on traditional image processing methods, such as grayscale processing, edge detection, and threshold segmentation. These methods perform recognition by processing fruit tree images such as denoising, feature extraction, and using rule-based or template matching algorithms. Although these methods have achieved some results under simple fruit tree morphology and background conditions, they perform poorly in the face of complex background, occlusion and deformation. Given the differences in terrain, climate and other conditions in different regions of China, traditional fruit cultivation methods can no longer meet the demand. As a result, deep neural network technology is emerging as required by The Times. It has strong feature sensitivity and feature extraction and recognition ability, and can quickly and accurately complete feature extraction and target recognition, which provides effective support for fruit tree recognition in complex environments.

Based on previous research, this paper improves the YOLOv5 deep learning object detection algorithm. YOLOv5 is an advanced multi-step object detection technology, which is characterized by the ability to determine the best detection results by comparing and analyzing different parameters, so as to identify specific objects more accurately. Its characteristics are as follows: the inference process is fast, the simulation effect is accurate, and the operation cost is less. Therefore, we use YOLOv5 as the four main models (YOLOv5s,YOLOv5m, YOLOv5l, YOLOv5x), of which YOLOv5s as the basis, combined with the attention mechanism, to better meet the real-time requirements, so as to improve the efficiency and accuracy of the system. It greatly improves the performance of the entire system. Such a design improves the accuracy of trunk identification and diameter acquisition.

2. Trunk Detection based on the Improved YOLOv5 Algorithm

2.1 Data Collection

The system uses Intel Realsense D415 camera to take pictures of the tree trunk. The D415 has an excellent field of view, enabling accurate depth capture, and is equipped with an RGB sensor for facial recognition, 3D scanning, and small photography. Combining Intel modules and vision processors, it is a highly innovative design and an excellent choice for manufacturers.

In this paper, the depth data collected by Intel Realsense D415 and the anchor frame identified by YOLOv5 were used to collect the anchor frame information in the depth image, and the minimum depth information near the center of the anchor frame was obtained by bubble sorting and median filtering. It also traverses all pixels in the depth image where the anchor frame is one-third to two-thirds wide. The collected depth image is shown in Figure1 below:

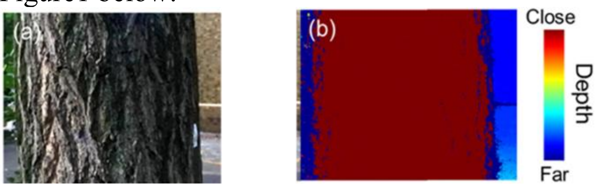

Figure 1. Tree Trunk Depth Image

2.2 SENet Attention Mechanism

The SENet attentional mechanism module allows feature recalculation to improve data validity and reduce the impact of unwanted information. The module consists of three parts: compression layer, excitation layer and proportional layer. Among them, the input characteristic of the compression layer is X, and its size is H'×W'×C', while the output result of the excitation layer is U, and its size is H×W×C. Its structure is shown in Figure2 below:

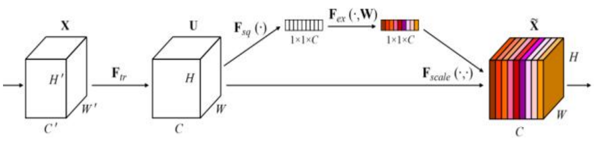

Figure2. Mechanism Diagram of SENet Attention Mechanism

(1)In the compression layer phase, we compress the H×W×C feature graph U into a real number column Z to provide a wider global perception range so that each channel can be related to each other.

$$Z = Squeeze(U) = \frac{1}{W \cdot H}\sum_{i=1}^{W}\sum_{j=1}^{H} U \quad (1)$$

(2)In the excitation layer, we need to perform a nonlinear transformation of the real number column Z to determine the weight of each channel. To achieve this, we need to build a complete grid with two complete layers, the ReLU layer, and two Sigmoid layers. The ReLU function can effectively reduce the number of parameters in the network and control overfitting. In addition, the Sigmoid function determines the weight of each message by limiting the value of each message to a range of 0 to 1.

$$S = Excitation(Z, W) = \sigma[W_2\delta(W_1 Z)] \quad (2)$$

W1 and W2 represent the weights in the first and second fully connected layers, respectively, which can be used to describe the performance and stability of the system. σ represents the Sigmoid activation function ,δ represents the ReLU activation function.

(3)In the proportional layer stage, the real number column S obtained in the excitation layer is multiplicatively weighted in order to transform it into a new channel, thus achieving a redefinition of the feature.

$$\tilde{X} = Scale(U, S) = U \cdot S \quad (3)$$

2.3 SE-Res Residual Module

By combining the SENet and SE-Res modules, we can achieve an efficient arrangement of the model. With two CBLS, we can obtain the input M1 of the model and calculate the final model result by comparing M1 and M2. The SE-Res module is shown in Figure3.

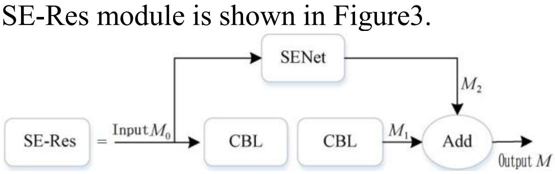

Figure3. The SE-Res Module Structure

2.4 The Enhanced Structure of the Yolov5 Model

Of the four YOLOv5 models, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, the lightest model is YOLOv5s with a weight ratio of only 27. Considering the need for real-time performance, the design chose the YOLOv5s model. The adoption of YOLOv5 significantly enhances positioning accuracy while reducing size. However, because it relies on direct prediction of the target location without prior information, its positioning accuracy is somewhat affected. Moreover, due to strong spatial constraints in its prediction process, it performs suboptimally when dealing with small objects appearing in groups. Therefore, this design chooses to enhance the attention mechanism based on the YOLOv5s basic model.

2.5 The Comparison of Effects Pre- and Post-Improvement

According to Figure4 and Figure5, when SENet is used, the accuracy of the YOLOv5 algorithm is significantly improved, reaching 62.53%, and when SENet is used, this value is further increased, reaching 2.38 percentage points, and even smaller branches can be identified more accurately.

Figure 4. The Initial Recognition Map of YOLOv5

Figure 5. The Improved YOLOv5 Recognition Map

3. Calculation of Trunk Diameter

3.1 Trunk Width Detection

With the depth information gathered by Intel RealSense D415, we can precisely explore the size of the object. The depth camera of the camera can achieve horizontal and vertical resolution of 65°and 40°, respectively, while the RGB camera can achieve FOV values of 69°and 42°. When the color image and depth image are perfectly matched, the FOV value can be increased from 56.7°×44.1°to 56.7°×44.1°. Thus, accurate size detection can be achieved.

$$\theta = \frac{56.7^{\circ}}{c_0}(1 + n_0) \quad (4)$$

Note: n_0 Number of columns of the depth image； c_0 The number of invalid pixels between the current valid pixel and the previous valid pixel

3.2 Calculating Trunk Diameter

The Intel RealSense D415 sensors were placed precisely in front of the mobile robot in the orchard, and as they scanned, they were able to view the cross section of the tree as an accurate cylinder and were able to obtain its curvature, as detailed in Figure6.

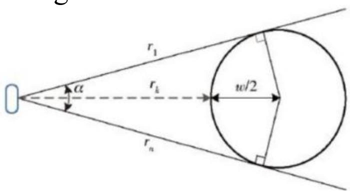

Figure 6. The Camera Scans a Schematic View of the Tree Trunk

From this, the trunk diameter w can be obtained：

$$w = 2r_k \frac{sin(\alpha/2)}{1 - sin(\alpha/2)} \quad (5)$$

Note: $\alpha = (n - 1)$； $r_k = min_{m \in T_i} d_m$

4. System Testing and Analysis

4.1 Comparative Experiments of Different Attention Mechanisms

Taking mainstream attention mechanisms such as CA, CBAM, SENet and ECA into account, we compare and analyze the effect of the new YOLOv5s network structure, and use mean mean precision as the evaluation index to measure the effectiveness of these mechanisms. Experimental results are shown in Table 3 below:

Table 1. Comparison of Different Attention Mechanisms

Model	mAP@0.5	mAP@0.5:0.95
YOLOv5s	89.2	59.0
YOLOv5s+ECA	90.2	59.1
YOLOv5s+CBAM	90.2	59.2
YOLOv5s+SENet	90.4	58.9
YOLOv5s+CA	90.4	59.6

mAP@0.5:0.95 shows mAP over different IoU thresholds (from 0.5 to 0.95 with a 0.05 step). Table 3 shows that after the attention mechanism is added to YOLOv5s, the accuracy of each network is improved to varying degrees compared to the original YOLOv5s network. Among the four attention mechanisms, SENet attention mechanism mAP@0.5 improved significantly (mAP from 89.2% to 90.4%, mAP increased by 1.2 percentage points). Therefore, it was finally decided to add the SENet attention mechanism to YOLOv5s backbone network.

4.2 Total Number of Tree Trunks Identifica-Tion Experiment (as shown in Table 2)

Table 2. Identification Results of the Total Number of Trunks

Algorithm	error	average accuracy rate
Original YOLOv5	5.56%	0.54
Improved YOLOv5	3.33%	0.61

Test conclusion: The recognition rate and average accuracy of the improved YOLOv5 for multiple trees are higher than that of the original YOLOv5, which can achieve the expected goal.

4.3 Trunk Diameter Acquisition Experiment (as shown in Table 3)

Table 3. Results Obtained from Tree Trunk Diameter

Single trunk actual diameter (mm)	Measurement diameter (mm)	error
186	192	+3.22%
186	179	-3.76%
186	182	-2.15%
186	175	-5.38%
186	191	+2.68%

Test conclusion: the error of single tree trunk diameter identification is less than 6%, which can reach the expected standard within the acceptable error range.

5. Conclusion

With the YOLOv5 model, we have developed a new SENet attentional mechanism module that greatly speeds up efficient data collection and processing while also reducing unnecessary data. To this end, we have also combined the SE-Res module for higher precision and better performance. Compared to the YOLOv5 algorithm, before adding SENet module, the accuracy of its algorithm is only 2.38 percentage points, but in this improvement, its tree identification ability has been significantly improved. In addition, the improved algorithm can also better identify small tree trunks, and the recall rate has improved by 0.84 percentage points. However, due to the introduction of the additional attention mechanism module, the frame rate is reduced by 0.99 frames/s. That is, the processing speed is slightly slower. Overall, however, mAP (mean accuracy) improved by 0.05 percentage points, which means that improved overall detection performance is beneficial.

In other words, by introducing the SENet attention mechanism module, both the accuracy and recall rate of tree trunk recognition tasks have been improved based on the YOLOv5 model. Although the processing speed decreased slightly, the overall performance indicator mAP also improved. This improvement brings a significant performance boost to the trunk recognition task.

Variable fertilization belongs to the category of precision agriculture, and the ultimate goal is to distribute nutrients according to the actual distribution of soil, avoiding unnecessary waste and pollution. This technique uses tree diameter detection to control the amount of fertilizer applied, providing important support for the development of modern agriculture. Its emergence will bring us more advanced technology, and promote intelligence, information and large-scale agriculture.

Acknowledgements

This work was supported by Guangdong Province Higher Vocational Education Reform Research and Practice Project "Construction of the Immersive"2+2" Collaborative Educa-tion Curriculum System for Electrical Engineering and Automation" (GDJG2021-368), Ling Shi Jiao Wu [2021] No.165 Document No.34 "Lingnan Normal University - Beijing Sanqing Internet Technology Co., Ltd. Practical Teaching Base", 2019 Campus level Talent Special Project of Lingnan normal university“Research on Sensor Angle Compensation and Noise Reduction Methods for Flow Method Production Measurement”[grant numbers ZL2005, ZL2007], College Student Innovation Program of Lingnan normal university"Fruit tree targeted precision variable fertilization robot based on YOLOv5 deep vision algorithm"[grant numbers X202410579025] and"Design of Spray Device Based on Laser SLAM and Adaptive Kalman Filter Contro" [grant numbers S202410579010].

References

[1] MEI Nan. Research on development countermeasures of Chinese style family farm [D]. Jilin University,2015.

[2] Zhang Jicheng. Research on key technologies of variable fertilization system based on prescription map [D]. Northeast Agricultural University,2013.

[3]GilesDK,SlaughterDC. Precision band spraying with machine vision guidance and adjustable yaw nozzles[J]. Transactions of theAsae, 1997, 40 (1) : 29-36.

[4] NIU Runxin, Zhang Xiangyang, Wang Jie, Zhu Hui, Huang Jian, Chen Zhengwei. Tree-trunk detection algorithm of agricultural robot orchard based on LiDAR [J]. Transactions of the Chinese Society for Agricultural Machinery, 2019,51(11):21-27.

[5] Zhang Yingying, Zhou Jun. Orchard trunk detection based on LiDAR [J]. Journal of China Agricultural University, 2015, 20(05): 249-255. (in Chinese)

[6] LIU Weihong, HE Xiongkui, LIU Yajia, Wu Zhiming, YUAN Changjian, LIU Limin, QI Peng, LI Tian. 3DLiDAR navigation method for orchard rows [J]. Transactions of the Chinese Society of Agricultural Engineering, 201,37(09):165-174.

[7] HAN Zhenhao, Li Jia, Yuan Yanwei, Fang Zhengjing, Zhao Bo, ZHU Licheng. Orchard visual navigation path recognition method based on U-Net network [J]. Transactions of the Chinese Society for Agricultural Machinery, 201,52(01):30-39.

[8] Li Shiwei. Object Detection and License Plate Recognition System based on YOLOv5 algorithm [J]. Electronic Technology and Software Engineering,2022(01):138-141.

[9] GU Baoxing, Liu Qin, Tian Guangzhao, Wang Haiqing, Li He, Xie Shangjie. Tree trunk recognition and localization based on improved YOLOv3 [J]. Transactions of the Chinese Society of Agricultural Engineering,202,38(06):122-129. (in Chinese)

[10] SHEN Yue, Zhuang Zhenzhen, Liu Hui, Jiang Jianbin, OU Mingxiong. Multi-feature tree trunk recognition method based on RealSense depth camera [J]. Transactions of the Chinese Society for Agricultural Machinery, 2002,53(04):304-312.

[11] Wang Han, Liu Haiming, Shao Yuhong. Research on Metal surface defect detection based on improved YOLOv5 algorithm [J/OL]. Mechanical science and technology: 1-6 [2023-10-21]. https://doi.org/10.13433/j.cnki.1003-8728.20230315.

[12] Lu Bibo, Zhou Yun, Li Xiaojun et al. YOLOv5 Personnel Detection Algorithm in Light Coal Mine with Attention mechanism [J]. And coal technology, 2023 (10) : 200-203. The DOI: 10.13301 / j.carol carroll nki. Ct. 2023.10.045.

