Based on Machine Vision Fruit Target Detection Method

Guanghui Yin, Yuanmin Xie*, Juntong Yun*, Hang Chen and Chuanjie Yang
Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China
Email: 2811221986@qq.com, xieyuanmin@wust.edu.cn
*Corresponding author

Abstract. Fruit detection and location is the key technology of fruit picking automation. In this paper, aiming at the problem of fruit detection and location, fruits growing in natural environment were studied based on the Faster R-CNN target detection model. Experimental verification shows that the detection and recognition of fruit based on Faster R-CNN model can achieve higher detection accuracy and speed.

1. Introduction
China is a large agricultural country; fruit planting occupies a considerable part in agricultural production. In order to improve the labor efficiency of fruit picking and reduce the production cost, it has become an imminent problem to realize the automation of fruit picking process, and the realization of fruit identification, detection and location is one of the key technologies [1]. Target recognition and location of fruit growing in natural environment is to obtain target position information of fruit by computer vision technology and transmit the obtained position information to the fruit picker, so as to achieve accurate fruit picking by controlling the end of the fruit picker.

In recent years, many scholars have studied the identification, detection and localization of fruits. Ji et al. [2] studied the method based on region growth and color feature segmentation, and recognized the extracted apple color and shape features through support vector machine. Si et al. [3] located the pixel position of apple by RGB color channel color difference threshold and color difference ratio, and determined the two-dimensional information of fruit by random ring method. Lu et al. [4] proposed a segmentation technique based on fused chromatic aberration information and normalized RGB model to identify citrus, but the influence of illumination was not excluded in the preliminary segmentation stage. Suchet et al. [5] proposed two feature extraction algorithms of multi-scale multi-layer perceptron and convolutional neural network to segment apple to achieve detection and recognition. Fu longsheng et al. [6] used the LeNet model of parameter optimization and structure reduction to carry out independent feature learning of kiwifruit in the field environment, but failed to achieve good effects on shielding and overlapping fruits.

At present, the accuracy and speed of fruit detection and recognition need to be improved. And the regional convolutional neural network (Faster R-CNN) has a good performance in the accuracy and speed of target detection. Taking oranges as the research object, this paper uses Faster R-CNN to detect and identify fruits from multiple pictures collected in the natural environment.

2. Target Detection Method Based on Faster R-CNN
Convolutional Neural Networks (CNN) is a kind of Neural network with Convolutional computation and deep structure. At present, with the proposal of deep learning theory and the upgrade of numerical calculation equipment, convolutional neural network develops rapidly and is widely used in machine
vision, natural language processing, digital image processing and other research fields. It is an efficient artificial neural network [7].

Faster R-CNN is a target detection model based on candidate regions, which is mainly composed of three parts: candidate region selection, image feature extraction and target region classification. Faster R-CNN is to improve Fast R-CNN and directly use the convolutional neural network to obtain the candidate region [8]. That is suppose there are two convolutional neural networks. One is the region generation network, and each candidate region in the image can be obtained; the other is the classification of candidate regions and the border regression network.

The Faster R-CNN network model is mainly composed of three parts: feature extraction network, regional suggestion network and classification location network, as shown in figure 1. The feature extraction network is used to extract and detect the features to form the convolution feature map [9]. Three kinds of convolutional neural networks are usually used: small network ZF, medium network VGGM and large network VGG16. It is suggested that region proposal network (RPN) adopt full convolutional network to obtain the regions of interest and transfer these regions to the full connection layer. The classification and positioning network uses the pooling layer to process the region of interest, extract the image features of candidate boxes of RPN network output, and transmit them to the subsequent full connection layer to achieve the target classification and positioning.

![Figure 1. Faster R-CNN Network Model Framework](image)

There will be two losses in the training process of the Faster R-CNN model, the RPN network carries out the regional suggestion stage and the prediction stage [10]. The RPN network loss function consists of two parts: (1) classification loss function, which is used to describe whether the proposed region is the target region; (2) regression loss function, used to describe the gap between the border of the proposed region and the border of the real target region. The RPN loss function is shown in formula (1).

$$L\left(\{p_i\},\{t_i\}\right) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*)$$

(1)

Where, $p_i$ represents the probability that the anchor $i$ rectangle box is the target; $p_i^*$ represents the probability that the anchor $i$ is the real target: if the anchor $i$ is the real target, then $p_i^*$ is 1; otherwise, it is 0; $t_i$ is a vector composed of four parameters of the network output; $t_i^*$ represents a vector composed of four parameters of GT as the anchor of the target.

As shown in formula (1), the regression loss function $L_{reg}$ is multiplied by $p_i^*$ before it, indicating that the regression loss function exists only when the third anchor is the real target. The border regression loss function is used to correct the coordinates of the anchor rectangular box and the real rectangular box, and the deviation calculation formula of the coordinates and dimensions is shown in formula (2).
\[ t_s = \frac{(x - x_a)}{w_a}, \quad t_s^* = \frac{(y - y_a)}{h_a} \]
\[ t_w = \log\left(\frac{w}{w_a}\right), \quad t_h = \log\left(\frac{h}{h_a}\right) \]
\[ t_s^* = \frac{(x^* - x_a)}{w_a}, \quad t_y^* = \frac{(y^* - y_a)}{h_a} \]
\[ t_w^* = \log\left(\frac{w^*}{w_a}\right), \quad t_h^* = \log\left(\frac{h^*}{h_a}\right) \] (2)

Where \( x, y, w, \) and \( h \) denote the box's center coordinates and its width and height. Variables \( x, x_a, \) and \( x^* \) are for the predicted box, anchor box, and groundtruth box respectively (likewise for \( y, w, h \)).

### 3. Experiment and Result Analysis

#### 3.1. Experimental Process

Image of mature oranges in the natural environment were collected to form the experimental data set, which contained 1000 different photos. The data set is divided into training verification set and test set in a ratio of 7:3, and the training set is divided into training set and verification set in a ratio of 7:3. The training set is used for model training and parameter adjustment, and the test set is used for model effect test. At the same time, the images of the training set were labeled, and the targets in the samples were labeled as oranges. Partial labeling results were shown in figure 2.

![Figure 2. The Sample Labeling Diagram](image)

In the experiment, the parameters of the Faster R-CNN model were set as follows: the candidate region was set to 100, the batch_size was set to 24, the training times was set to 200,000, and the training stopped after the model converges.

Some text.

#### 3.2. Experimental Results and Analysis

The test set was substituted into the trained Faster R-CNN model for testing, and the average detection accuracy was 82.4%, and part of the test results were shown in figure 3. The total amount of data collected is too small, and the sample type may be insufficient, which may lead to low classification accuracy.
4. Conclusion
Fruit identification and location is the core problem of fruit picking automation. In this paper, oranges growing in the natural environment were detected and identified based on the Faster R-CNN model to achieve higher detection accuracy and speed. But there are also many problems, which need to be solved in the follow-up research. The fruit environment under the natural environment is complex, how to achieve accurate and efficient positioning of fruit and improve the precision of multi-target detection by partial fruit contour is a need for further research. Tables should have only horizontal rules and no vertical ones. Generally, only three rules should be used: one at the top of the table, one at the bottom, and one to separate the entries from the column headings. Table rules should be 0.5 points wide.

5. Acknowledgements
This work was supported by Training Project of Innovation and Entrepreneurship Training Program for college students in Hubei Province (201810488054). The research reported in the paper was also supported by Wuhan University of Science and Technology.

6. References
[1] Edan Y, Han S, Kondo N. Automation in Agriculture[M]. New York: Springer Publishing Company, 2009.
[2] Ji Wei, Zhao Dean, Cheng Fengyi, et al. Automatic recognition vision system guided for apple harvesting robot[J]. Computers & Electrical Engineering, 2012, 38(5): 1186–1195.
[3] Si Yongsheng, Liu Gang, Feng Juan. Location of apples in trees using stereoscopic vision[J]. Computers and Electronics in Agriculture, 2015, 112: 68–74.
[4] Lu Jun, Nong Sang. Detecting citrus fruits and occlusion recovery under natural illumination conditions[J]. Computers and Electronics in Agriculture, 2015, 110: 121–130.
[5] Bargoti S, Underwood J P. Image segmentation for fruit detection and yield estimation in apple orchards[J]. Journal of Field Robotics, 2017, 34(6): 1039–1060.
[6] Fu Longsheng, Feng Yali, ElkamilTola, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(2): 205–211.
[7] Tara N,Brian Kingsbury, George Saon, et al. Deep Convolutional Neural Networks for Large-scale Speech Tasks [J]. Neural Networks, 2015, 64:39–48.
[8] Ren S ,He K,Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J], arXiv:1506.01497v3, 2016.
[9] Abhinav Shrivastava, Abhinav Gupta. Contextual Priming and Feedback for Faster R-CNN[C]. European Conference on Computer Vision. Springer International Publishing, 2016. DOI:10.1007/978-3-319-46448-0_20

[10] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real–Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6):1137-1149.