Information acquisition of anterior branches of fruit based on Mask R-CNN

Wang Zhuo¹, Xiong Longye¹, Liao Haishen¹, Kang Xilong¹, Yang Changhui¹²*

¹College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China
²College of Mechanical Engineering, Xi’an Jiaotong University, Xi’an 710049, China
*Corresponding author’s e-mail: yangchanghui@cqut.edu.cn

Abstract: In order to obtain the deflection information of the anterior branch to enable the end effector to perform citrus picking in the correct posture, this paper proposes a method based on Mask R-CNN model and centroid estimation for pre-fruit branch deflection information acquisition. The method marks the front of the fruit branches by means of a rectangular-like marking method, and completes the identification of the pre-fruit branches with random posture. The ROI is obtained by the discrete mask obtained for the model, The ROI area is equally divided into two sub-areas for centroid extraction, At last, The two centroids obtained are used to calculate the deflection angle of the fruiting branches. The experimental results show that the recognition accuracy of the model under the test set is 96.44%, the average recall rate is 80.71%, the average deviation of the deflection angle of the pre-fruit branches is 7.4°, and the maximum deviation of the angle is 12°.

1. Introduction
With the development of intelligent agricultural machinery in China, research on fruit picking robots has made great progress [1-5]. However, under natural conditions, the growth of the branches before the fruit is random, the shape is different, and the fruits are close to the branches. If the distribution of these branches is not considered, the end effector may be blocked by the branches when the robot is picking. This situation will lead to the failure of picking. Therefore, research on how to obtain pre-fruit branch information is very valuable for the development of picking robots.

At present, there are many target extraction methods in the field of image processing [6-11]. Zhang Fugui [12] segmented the fruit trees by analyzing the RGB image components and then thresholding them. The segmentation accuracy is 79.67%. Ji Wei [13] proposed an apple branching threshold segmentation method based on adaptive equalization histogram. The correct segmentation rate of the method is 94%. Cai Jianrong [14] extracted some skeletons of citrus trees through morphological and regional refinement steps, and then repaired the fruit tree skeleton by Hough straight line transformation and occlusion of the skeleton skeleton restoration. The correct recognition rate was 67.3%. He Leiying [15] carried out branch reconstruction of leafless walnut trees by double contour synchronous tracking method, but the applicable scene of this method is too simple. Tabb Amy [16] used superpixels to determine the background low-texture area, and then segmented the fruit branches. The segmentation method is only suitable for simple backgrounds. Amatya Suraj [17] detected the cherry stems by clustering and geometric methods. The overall detection accuracy was 93.8%. However, this method only realized the identification of a single branch, but did not carry out the
branches with bifurcation. Detection and consideration. Majeed Yaqoob [18] used convolutional neural networks and Kinect V2 cameras to achieve a segmentation accuracy of 92% and 93% for apple trees, respectively. This method only studies the unobstructed apple branches and does not consider the presence of occlusion. Shalal Nagham [19] used a camera and a laser scanner to detect the trunk of a fruit tree with a detection accuracy of 96.64%. This method can only detect the coarser main trunk and cannot identify the subdivided fork. Zhang Qin [20] tested apple tree branches by RCNN and fitting method. The average recall rate and accuracy were 91.5% and 85.5%. The method is still for single branches and unobstructed branches. For branches with bifurcation and occlusion, no studies have been conducted.

In view of the above situation, this paper takes the citrus tree in the natural scene as the research object, and obtains accurate branch information by identifying multiple branches and occluded branches in front of the fruit, providing basis for robot path planning and obstacle avoidance.

2.Materials and Methods

2.1. Graphic gathering
The vision system of this study consists of a BB2-08S2C-60 binocular camera (resolution 1024×768) produced by Point Gray, a 1394 image acquisition card, and a mobile workstation, as shown in Figure 1. The width of the experimental orchard road is about 200cm, and the robot walks along the middle line of the road. The horizontal distance between the end of the arm and the center of the crown is about 68cm, and the horizontal distance between the end of the arm and the camera is 60cm. The calculation distance of the camera is about 90.69cm. The image acquisition distance is maintained at 80 cm to 100 cm. The collection location is located in Jinguoyuan, Beibei District, Chongqing, and the citrus orchard of Chongqing University of Technology. The shooting time is December 2016 and January 2019, and the weather is cloudy.

![Diagram of the picking robot](image)

Note: L1 indicates the horizontal distance between the end of the arm and the camera, L2 indicates the horizontal distance between the end of the arm and the center of the crown, and L3 indicates the horizontal distance between the camera and the center of the canopy.

2.2. Practice data set
Because the growth pattern of fruit trees is random and the shape of branches is different, the image threshold segmentation method is not effective, especially for the environment with strong outdoor illumination changes. Therefore, this paper uses the deep learning method to identify the randomness of the posture of the branches. At the same time, the shape characteristics of the branches were analyzed. In this paper, the branches were labeled with a rectangular shape, and the citrus and pre-fruit branches were marked with the labeling tool Labelme[21], as shown in Fig. 2. The citrus (Orange) is abbreviated as O, branches. (Branch) class is referred to as B.
2.3. Identification model test
The total sample size of this experiment was 150 pictures, of which 100 pictures were used as training sets and 50 pictures were used as test sets. The experimental environment is: CPU is Intel i7 7800X, single 11G GPU NVIDIA 1080Ti, memory 32G PC; software environment: Ubuntu16.04, deep learning framework Tensorflow1.4.

Mask R-CNN [22] is an improved model of Faster R-CNN [23]. Both models are used for target recognition and localization. Faster R-CNN selects the identified objects by a rectangular box. Mask R-CNN is Based on Faster R-CNN, a mask branch is introduced to realize the segmentation of the object of the frame selection. In order to realize the segmentation of the pixel level, the paper first obtains the preliminary recognition result through the Mask R-CNN model, and then uses the obtained mask. Carry out the next step.

Using 100 pictures in the Mask R-CNN model, the final model recognition effect is shown in Figure 3. Table 1 shows the recall rate and accuracy of different models for different types of objects, and the average recall rate of citrus. It is 82.09%, the average accuracy is 93.15%, the average recall rate of branches is 80.71%, and the average accuracy is 96.44%.

| Marking type | Recall (%) | Precision (%) |
|--------------|------------|---------------|
| Citrus       | 82.09      | 93.15         |
| Branch       | 80.71      | 96.44         |

2.4. Estimation of the slope of the fruit
As shown in FIG. 4, the actuator can cut the stalk of the yaw angle $|\beta| \leq 30^\circ$ when picking, FIG. 4a is a physical diagram of the end effector, and FIG. 4b and FIG. 4c are schematic diagrams of the actuator structure. When the branches are parallel to the running plane of the actuator, if the picking is
carried out in the manner shown in Fig. 4b, the actuator will not reach the picking point due to the obstruction of the front branches, and when the actuator is picked in the manner shown in Fig. 4c, it can be avoided. The branches before the opening of the fruit accurately reach the picking point. Therefore, it is important to obtain the stem slope information to guide the actuator to pick in the correct posture.

![Figure 4. shows the different arrangement of branches](image)

After the Mask R-CNN model is identified, the output information includes the category information of the recognition target, the frame information, and the mask segmentation area. However, for the fruit branches, the information obtained by the model detection cannot obtain the slope of the stem. The acquired border is a positive border, and the obtained mask is a closed area, and the slope of the branch cannot be obtained.

In order to solve this problem, the paper firstly takes the obtained trunk-like frame as the sensation area, then divides the area into two areas on the Y-axis direction, and finally estimates the inclination of the branches by calculating the slope of the line where the center of mass of the left and right areas is located. The experimental results are shown in Fig. 5, in which Fig. 5a is the model recognition effect, Fig. 5b is the screening target mask, and Fig. 5c is the interest segmentation effect.

![Figure 5. Get the ROI and segment](image)

For a 2D continuous image \( f(x, y) (x \geq 0, y \geq 0) \), the Moment \( m_{pq} \) and Center distance \( O_{pq} \) definition is as shown in equations (1) and (2):

\[
m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy
\]

\[
O_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy
\]

For discrete digital images, Equation (1)(2) can be changed as shown in Equation (3)(4) below:

\[
m_{pq} = \sum_{j=1}^{N} \sum_{i=1}^{N} i^p j^q f(i, j)
\]

\[
O_{pq} = \sum_{j=1}^{N} \sum_{i=1}^{N} (i - i_c)^p (j - j_c)^q f(i, j)
\]

Where \( (i_c, j_c) \) is the centroid coordinate, the specific formula is as shown in the following formula (5) and formula (6):
Through the equations (5) and (6), the mask centroids in the two divided graphs are obtained. As shown in FIG. 6, FIG. 6a is the acquired two centroids, and FIG. 6b is the centroid connection.

\[ i_c = \frac{m_{10}}{m_{00}} \quad \text{(5)} \]
\[ j_c = \frac{m_{01}}{m_{00}} \quad \text{(6)} \]

Figure 6. Obtaining the slope of the fruiting branch

3. Experimental results and analysis

By extracting the two centroids of the left and right regions after the mask segmentation along the Y-axis direction, the stem tilt angle and the midpoint information of the branches can be obtained, and the obtained stem tilt angle and the midpoint information can be used to accurately locate the branches. According to the result of the positioning, the actuator is guided to pick in a reasonable posture, and the experimental result is shown in FIG. 7.

![Image of experimental results and analysis](image)

Figure 7. Results of the front branch identification

In order to verify the accuracy of the slope estimation of the proposed method, this paper compares the straight line slope \( k_1 \) with the artificial value and the straight line slope \( k_2 \) obtained by the method. Statistics of 50 sets of deviation angles \( \alpha = \arctan |k_1 - k_2| \times \frac{180}{\pi} \), The statistical results are shown in Fig. 8. The branch angle calculated by the method in this paper is artificially fitted with a maximum deviation of 12° from the angle of the branches, The average deviation is 7.4°, Meet the end effector handle slew angle \( |\beta| \leq 30^\circ \).
6

Figure 8. Calculating the angular deviation statistics of branches and artificially fitting branches

4.Conclusion

(1) Aiming at the growth mode of the anterior branches, a rectangular labeling method was proposed, and two training data sets of citrus and stem were constructed. The trained Mask R-CNN model realized the accurate identification of citrus and pre-fruit branches. The average recognition accuracy rate of citrus was 93.15%, the average recall rate was 82.09%, the average recognition accuracy of branches was 96.44%, and the average recall rate was 80.71%.

(2) Accurate acquisition of the slope and midpoint information is achieved by extracting the centroid. The average deflection of the deflection angle estimated by the centroid is 7.4° and the maximum deviation is 12°, which can meet the actual picking requirements.

Acknowledgments

This research is funded by the Special Project of Key Technological Innovation (No.CSTC2015zdcy-ztzx7003) and research program of basic research and frontier Technology (No.CSTC2016jcyjA0444) received from Chongqing Science and Technology Commission.

References

[1] Liu J.Z.(2017) Research progress analysis of robotic harvesting technologies in greenhouse[J].Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(12):1-18.

[2] Wang D.D., Song H.B., He D.G.(2017) Research advance on vision system of apple picking robot[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 33(10):59-69.

[3] Wang Y., Xu H.B., Zhang M., et al.(2018) Design and Experiment of Bite-model End-effector for Citrus Harvesting by Simulating with Mouth of Snake[J]. Transactions of the Chinese Society for Agricultural Machinery, 49(10):54-64.

[4] Williams H.A.M., Jones M.H, Nejati M., et al.(2019) Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms[J]. Biosystems Engineering,181:140-156.

[5] Xu C.L., Wang Q., Chen H., et al.(2017) Design and simulation of artificial limb picking robot based on somatosensory interaction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 33(Supp.1):49-55.

[6] Wang J., Zhou Q., Yin A.J.(2018) Self-adaptive segmentation method of cotton in natural scene by combining improved Otsu with ELM algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 34(14):173-180.
[7] Xu S.Y., Peng C.L., Chen K., Wang L.Q., et al.(2018) Measurement Method of Wheat Stalks Cross Section Parameters Based on Sector Ring Region Image Segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 49(4): 53–59.

[8] Cao J.N.(2012) Review on Image Segmentation Based on Entropy[J]. Pattern Recognition and Artificial Intelligence. 25(6):958-971.

[9] Li J, Luo W, Wang Z, et al.(2019) Early detection of decay on apples using hyperspectral reflectance imaging combining both principal component analysis and improved watershed segmentation method[J]. Postharvest Biology and Technology, 149:235-246.

[10] Jiao L, Jiao Y, Wu J.(2018) An image segmentation method based on network clustering model[J]. Physica A: Statistical Mechanics and its Applications, 490:1532-1542.

[11] Russell B.C., Torralba A., Murphy K.P., et al.(2008) LabelMe: A Database and Web-Based Tool for Image Annotation[J]. International Journal of Computer Vision, 77(1-3):157-173.

[12] Zhang F., Yuan K., Shen M.M., et al.(2014) Different methods of fruiter image segmentation in variable-rate spray[J]. Journal of Chinese Agricultural Mechanization, 35(5).

[13] Ji W, Qian Z, Xu B, et al.(2016) Apple tree branch segmentation from images with small gray-level difference for agricultural harvesting robot[J]. Optik - International Journal for Light and Electron Optics,127(23):11173-11182.

[14] Cai J.R., Zhou X.J., Wang F., et al.(2009) Obstacle Identification of Citrus Harvesting Robot[J]. Transactions of the Chinese Society for Agricultural Machinery, 40(11): 171–175.

[15] He L.Y., Wu C.Y., Du X.Q.(2014) Fruit tree extraction based on simultaneous tracking of two edges for 3D reconstruction[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 30(7): 182–189.

[16] Tabb A, Medeiros H.(2018) Automatic segmentation of trees in dynamic outdoor environments[J]. Computers in Industry, 98:90-99.

[17] Amatya S, Karkee M.(2016) Integration of visible branch sections and cherry clusters for detecting cherry tree branches in dense foliage canopies[J]. Biosystems Engineering,149:72-81.

[18] Majeed Y, Zhang J, Zhang Q, et al.(2018) Apple Tree Trunk and Branch Segmentation for Automatic Trellis Training Using Convolutional Neural Network Based Semantic Segmentation[J]. IFAC Papers OnLine,51(17):75-80.

[19] Shalal N, Low T, McCarty C, et al.(2015) Orchard mapping and mobile robot localisation using on-board camera and laser scanner data fusion – Part A: Tree detection[J]. Computers and Electronics in Agriculture, 119:254-266.

[20] Zhang Q, Zhang X, Zhang J, et al.(2018) Branch detection for apple trees trained in fruiting wall architecture using depth features and Regions-Convolutional Neural Network (R-CNN)[J]. Computers and Electronics in Agriculture, 155:386-393.

[21] Russell B.C., Torralba A, Murphy K.P., et al.(2008) LabelMe: A Database and Web-Based Tool for Image Annotation[J]. International Journal of Computer Vision, 77(1-3):157-173.

[22] Chen X, Wang S, Zhang B, et al.(2018) Multi-feature fusion tree trunk detection and orchard mobile robot localization using camera/ultrasonic sensors[J]. Computers and Electronics in Agriculture, 147:91-108.

[23] Zhang L., Lin L., Liang X., et al.(2016) Is faster R-CNN doing well for pedestrian detection?[J]. arXiv:1607.07032 [cs.CV].

[24] Wang B., Zhi Q.C., Geng G.H., et al.(2004) Computation of Center of Mass for Gray Level Image Based on Differential Moments Factor[J], (10):1360-1365.