# Precise Positioning of Picking Grapes

Duanfeng Liu, Gaili Gao*

College of Engineering, China Agricultural University, Beijing 100083, China

*Corresponding Author Email: ggl1965@126.com

**Abstract.** Aiming at the problem that it is difficult to determine the picking location for the picking robot, precise positioning of picking grapes based on Mask-RCNN network and Cannylines detector is proposed in this paper. Firstly, the grape images under shading, direct sunlight and backlight conditions are collected, and grape images are marked with masks to make grape data sets. Then, based on the Mask-RCNN network training data set, the trained model is used to identify the grapes, and the box of framed grape cluster and the mask approximately equal to the size of grape cluster are obtained. Then, the AOI of main stem and centroid of grape cluster are calculated according to the obtained box and mask, the obtained AOI is equalized. Finally, the Cannylines detector based on Bilateral filtering is used to detect the line segment of main stem in the AOI, and the picking location is obtained according to the minimum distance between the centroid and the line segment. 150 images from the test set were used for the experiment, with 50 images selected from each type of light, including shading, direct sunlight and backlight. Experiments show that the accuracies of picking positioning under the three types of light conditions, are 0.900, 0.920, and 0.860, respectively. This method can provide precise picking location information for the grape picking robot.

**Keywords:** Grape; Mask RCNN network; Cannylines detector; Picking location.

## 1. Introduction

Fresh grapes have thin skin, tender meat and large volume. Many factors need to be considered in the picking process, which increases the difficulty of picking. The manual picking efficiency is low and the cost is high. Therefore, it is of great practical significance to study automatic grape picking.

Image processing and machine vision technology are widely used in grape image recognition. Early researchers mainly recognized grape clusters based on traditional image segmentation methods. Yang et al. took |G-R| + |G-B| chromatic aberration image as the segmented image [1]; Liu et al. used Otsu algorithm to segment grapes in H and V channels [2]. Traditional image recognition methods are easily disturbed by the environment, and the recognition accuracy of the algorithm is unstable. The application deep learning algorithm improves the recognition accuracy of grapes. Dolezel et al. used feedforward multilayer neural network to classify grapes [3]; Liu et al. used K-means clustering algorithm to segment grape images [4]; Cecotti et al. used Alexnet, Googlenet, Resnet and other networks to identify grapes [5]; Franczyk et al. proposed to combine Resnet neural network classifier model with multi-layer perceptron (MLP) to form an ExtResnet classifier to identify grape [6].

The shape of the grapes is irregular and each grape is scattered, and fresh grapes are easy to be damaged, so it is difficult to grab the fruit directly, so another key problem in picking the grape is to determine the exact picking location for each bunch of grapes. Xiong et al. analyzed clusters of grapes as a "flexible rod-hinge-rigid rod-mass ball " composite cluster model to determine the picking location [7]; Lei et al. carried out threshold segmentation and Harris corner detection on the fruit stem [8]; Ning et al. used Mask-RCNN network and regional growth algorithm to identify the fruit stem [9].

In order to realize the automatic picking grape, the identification and location of grape clusters based on the Mask-RCNN network are proposed in this study, which use Cannylines detector to detect the line segment on the grape stem, set the conditions according to the position between the grape stem and grape, and then calculate the picking location of the stem.

## 2. Grape recognition

### 2.1 Data set collection and production

The images of the grape clusters were collected at the Jiubaihe Grape Plantation in Xiqing District, Tianjin. The grapes were photographed according to the three illumination conditions of shading, direct sunlight and backlight. The grape variety is Red Globe Grape, and the total number of samples is 260. The images were captured by iPhone. The captured images are compressed, and the resolution of the compressed images is 704 × 832 pixels. The configuration of the computer processing images: the GPU is NVIDIA GeForce GTX 950M 4GB; Memory is 12GB. The Labelme software is used to annotate RGB images to obtain mask images, as shown in Fig. 1. In order to be close to the real environment, the data enhancement technology is used for the RGB and mask images of the grape cluster to expand the sample size. After data enhancement, the total sample data is 900. The sample data is randomly divided into training set and test set according to the ratio of 5:1.

a. RGB image b. Mask image

**Fig. 1** RGB and mask image of grapes

### 2.2 Grape identification based on Mask-RCNN

In this study, Mask-RCNN network is used to segment grapes, and its network structure is shown in Fig. 2. The information in the grape images is more complicated, including the ground, grape leaves, vines, and supporting frames, etc., and different light will affect the color and texture of the grapes, and the shape and size of the grapes vary greatly. The resnet101 network has a deep network depth, which can sequentially extract low-level features in grape images, such as color, edges, etc., and high-level features, such as grapes, branches and leaves, etc. The FPN network integrates the features of the bottom and high layers, and can fully extract the multi-layer information in the feature map. Therefore, the resnet101+FPN structure is selected as the backbone of the Mask-RCNN network. The Region Proposal Network (RPN) uses the proposal frame to select the grape cluster from the background without classification. The proposal framed region is transferred to RoIAlign and processed into a fixed-size feature map. Finally, the interest classifier is used to classify grapes specifically, the bounding box regressor adjusts the framed region again, and the mask generation network divides the grape cluster to generate a mask consistent with the shape and size of the grape cluster. The final output image has a box for the mask of the selected grape cluster.

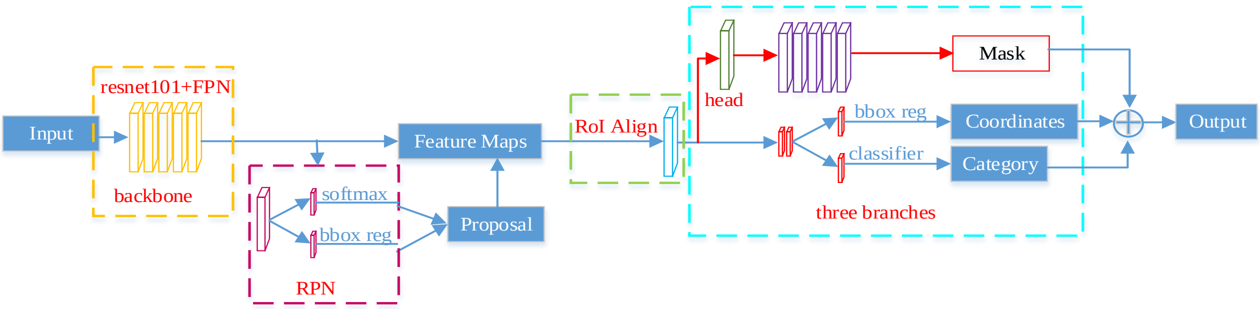



**Fig. 2** Mask-RCNN network

# 3. Precise positioning of picking grapes

## 3.1 Area of interest (AOI) for main stem of grape cluster

The grape cluster generally grows vertically due to its gravity, and the main stem of the grape cluster is located above its center of mass. In order to reduce the amount of calculation, the main stem area of interest is set above the grape cluster according to the box size of the grape framed region. When the main stem is accurately positioned, only the image processing is performed on the area of interest that is calculated as follows.

$$ROI_x = a, \quad x_1 + \frac{1}{7}(x_2 - x_1) \le a \le x_1 + \frac{6}{7}(x_2 - x_1) \tag{1}$$

$$ROI_y = b, \quad y_1 - \frac{1}{15}(y_2 - y_1) \le b \le y_1 + 10 \tag{2}$$

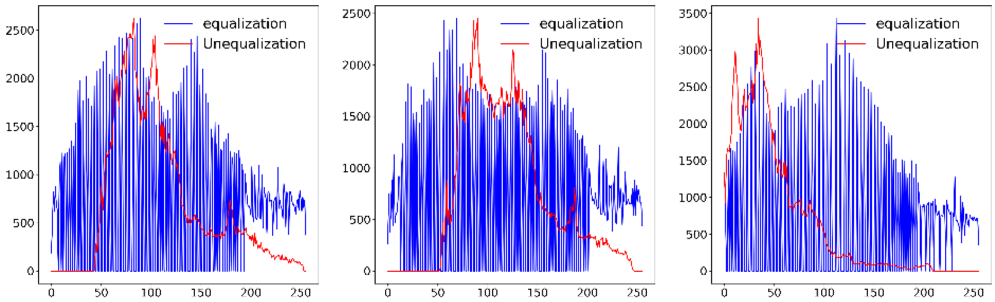
Where $(x_1, y_1)$ is the coordinates of the upper-left corner of the box, $(x_2, y_2)$ is the coordinates of the lower-right corner of the box, $ROI_x$ is the area of interest in the x-coordinate direction, and $ROI_y$ is one in the y-coordinate direction.

## 3.2 Precise positioning for main stem of grape cluster

### 3.2.1 Image preprocessing for the AOI

In addition to the main stem of the grape cluster, other objects, such as the branches, leaves and ground, are also contained in the AOI, which interfere with the identification of the main stem. In order to obtain the more accurate edge of the main stem, the RGB image of the AOI (Fig. 3a) is equalized to enhance the contrast between the main stem and other objects. That is, the R, G, and B channels of the image are respectively equalized, the color components in each channel are redistributed, and the color distribution width of each channel increases (Fig. 3b). The three processed channels are reconstructed into an RGB image (Fig. 3c), and the edge information of the main stem in the AOI is enhanced to prepare for the subsequent edge extraction.

a. Original ROI image




b. R, G, B component histogram

c. Equalized ROI image

**Fig. 3** Image equalization

### 3.2.2 Edge extraction for main stem of grape cluster

The edge extraction of the main stem of the grape cluster is mainly to obtain its edge contour, but the grape branches, the veins of the grape leaves and the ground in the AOI will interfere with the extraction. In this paper, Cannylines detector [10] is used to detect the picking location on the main stem of the grape cluster. It consists of the nonparametric Cannypf edge detector and line segment detector. The first step of Cannypf edge detector is to smooth the image using Gaussian filter, which only considers the spatial distribution and ignores the difference of pixel values. In the process of image smoothing, the edges in the image become blurred. In order to smooth out the useless texture in the image without losing the edge and detail information of the main stem, a bilateral filter is used to smooth the image. Its kernel (equation (3)) is generated by two functions of the spatial domain (equation (4)) and the range domain kernels (equation (5)), which nonlinearly combines the gray information and spatial information of the image. It can filter the noise and smooth the image while maintaining the image edge. After the edge image is obtained by CannyPF edge detector, the line segment detector performs line segment detection on the image to obtain the final line segment.

Where $(x, y)$ is the pixel coordinate, $(x_c, y_c)$ is the center coordinate, $\sigma_d$ is the standard deviation of the spatial domain, $gray(x, y)$ is the gray value corresponding to the pixel point, $gray(x_c, y_c)$ is the gray value corresponding to the center point, $\sigma_r$ is the standard deviation of the range domain.

The Cannylines detector based on the bilateral filter is used to detect the line segment of the main stem in the AOI, and the endpoint coordinates of the line segment are recorded. As can be seen from Fig. 4, the detected edge of the AOI smoothed by bilateral filter is clearer than that by Gaussian filter, which can better highlight the main stem in the image. Moreover, the continuity of edge is enhanced, and the number of useless line segments detected is less, which can reduce the number of line segments of picking location in the later calculation.

$$G(x, y, x_c, y_c) = G_d(x, y, x_c, y_c) * G_r(x, y, x_c, y_c) \tag{3}$$

$$G_d(x, y, x_c, y_c) = \exp(-\frac{(x-x_c)^2+(y-y_c)^2}{2\sigma_d^2}) \tag{4}$$

$$G_r(x, y, x_c, y_c) = \exp(-\frac{\|gray(x,y)-gray(x_c,y_c)\|^2}{2\sigma_r^2}) \tag{5}$$


**Fig. 4** Edge extraction and line segment detection

### 3.3 Picking location for main stem of grape cluster

The main stem of the grape cluster is approximately cylindrical, and the line segments detected on the main stem should be approximately parallel to each other. The grapes grow vertically due to its gravity. Theoretically, the extension of the line segment on the main stem should pass through the centroid, but in practice because of the interference of other factors, it is offset from the centroid by a certain distance.

The calculation of the picking location is divided into three steps: (1) The equations and slopes of the line segments on the main stem are calculated, and the line segments whose absolute values of the slopes are less than √3 are removed; (2) The line segments that are approximately parallel to each other are selected, that is, if the absolute value of the slope difference between the line segments is within [0,0.5], the line segments are considered to be approximately parallel and retained. (3) The distances between the centroid and the retained line segments are calculated according to the distance equation from point to line so that the line segment with the smallest distance is used as one where the picking location is located, and the midpoint of the line segment is the picking location.

## 4. Results and Discussion

The Mask-RCNN deep learning network based on transfer learning is used to train the grape cluster data set to obtain the training model. The weight obtained by training on the coco data set is initialized into the Mask-RCNN network, and then the established training data set is input into the Mask-RCNN network for training. When the model is trained, the initial learning rate, momentum and weight decay are set to 0.002, 0.9 and 0.0001, respectively. To verify the effectiveness of picking location, 150 images from the test set were used for the experiment, with 50 images selected from each type of light, including shading, direct sunlight and backlight.

Fig. 5 shows the identification and picking location of the grape cluster under three types of illumination. Under shading condition (Fig. 5a), there is under-segmentation at the edge of the grape cluster, but the area of the unsegmented grape is small (Fig. 5b). The box of framed grapes is close to the circumscribed rectangle of the grape cluster, and the picking location (the yellow dot in Fig. 5c) is located in the main stem area. Under the direct sunlight condition (Fig. 5d), most of the area of grapes is identified, and the upper boundary of the box of the selected grapes is slightly higher than the grape cluster (Fig. 5e), and the picking location (the yellow dot in Fig. 5f) is also accurately located in the main stem area. Under backlight condition (Fig. 5g), about one-third of the grape at the top of the grape cluster is not recognized, with fewer main stem in the area of interest for obtaining the main stem (Fig. 5 h, i), but picking location (the yellow dot in Fig. 5i) can still be identified.

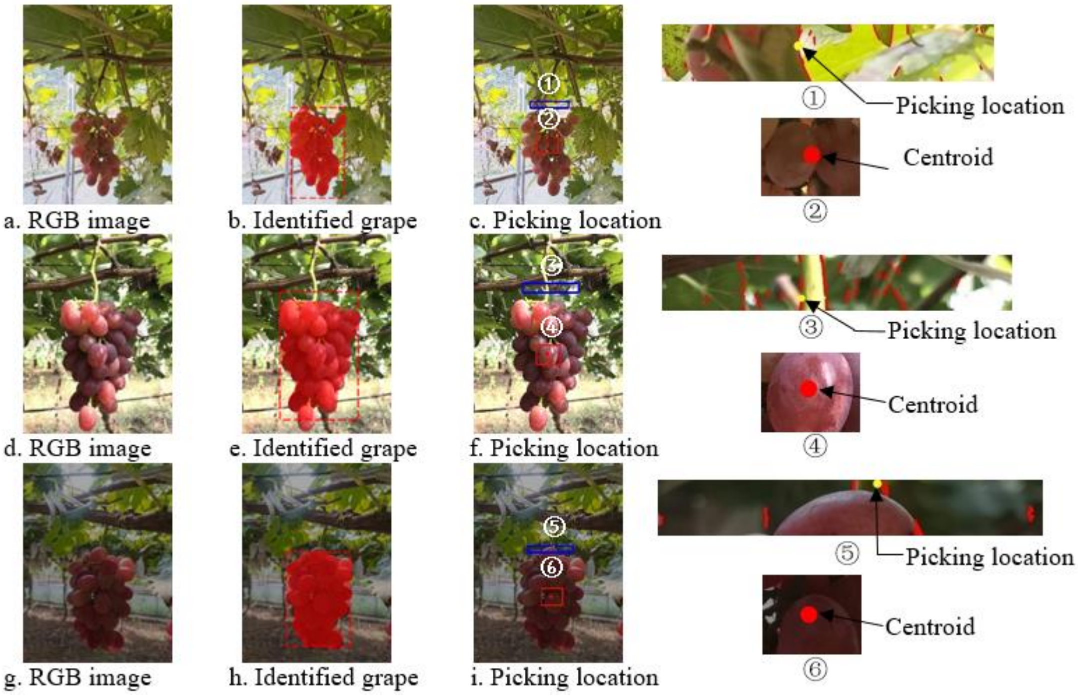



**Fig. 5** Grape cluster identification and picking location

**Table 1** Experimental results

| Types of light | Total number of images | mAP | Accurate location | Failed Location | Accuracy of location |
|---|---|---|---|---|---|
| shading | 50 | 0.912 | 45 | 5 | 0.900 |
| direct sunlight | 50 | 0.874 | 46 | 4 | 0.920 |
| backlight | 50 | 0.921 | 43 | 7 | 0.860 |

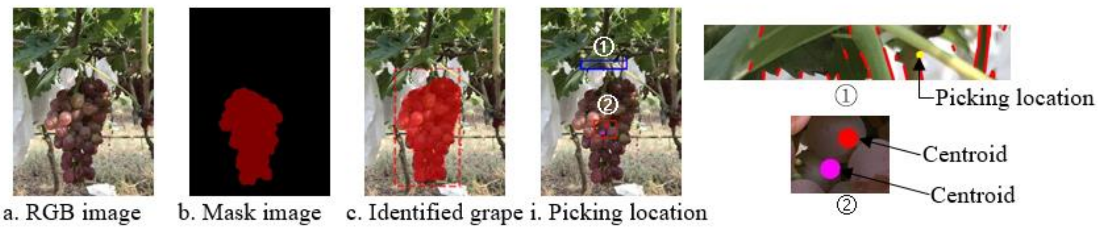



**Fig. 6** Failed picking positioning

In this paper, the average precision (AP) defined in Pascal VOC challenge [11] is used to evaluate the trained model. The evaluation data are shown in Table 1, where mAP is the mean value of AP values of various categories. It can be seen from the data in Table 1 that the mask RCNN model based on transfer learning has good recognition effect, in which the mAP value under backlight is the highest, with a value of 0.921. The accuracy of picking location under shading and direct sunlight conditions is higher, which are 0.900 and 0.920 respectively. However, a small number of grape images cannot be accurately identified. As shown in Fig. 6, the grape clusters in the image are overlapped, and the overlapped grape clusters are identified as a target, the mask area and contour obtained are far from the mask label. The upper boundary of the box of the selected grapes is much higher than the grapes. The calculated AOI is far away from the grape cluster. The centroid (red dot) obtained from the grape mask acquired from the Mask RCNN network recognition image does not coincide with one (purple dot) obtained from the mask label, which ultimately resulted in the obtained picking location not being located in the main stem area. The main reasons for the failure of picking location are as follows:

(1) wrong segmentation occurs when the model recognizes the grape clusters;

(2) In the AOI, the stems are staggered, with more vines and leaves;

(3) The grape cluster or main stem are badly covered.

## 5. Conclusions

(1) The Mask-RCNN deep learning network based on transfer learning is used to train the grape cluster data set to obtain the training model, and also get the box and mask of the selected grape cluster to realize the positioning of the grape cluster. The mAP values for identifying grapes under shading, direct sunlight and backlight conditions are 0.912, 0.874 and 0.921, respectively.

(2) The Cannylines detector based on bilateral filter extracts the edge of the main stem from the image of the main stem. According to the position relationship between the main stem and centroid and the approximate parallel characteristics of the main stem edges, the picking position is accurately obtained. Experiments show that the accuracies of picking positioning under shading, direct sunlight and backlight conditions are 0.900, 0.920, and 0.860, respectively.

## References

[1] Qinghua Yang, Can Liu, Yi Xun, et al. Target Recognition for Grape Bagging Robot. Transactions of the Chinese Society for Agricultural Machinery. Vol. 44 (2013) NO 08, p:234-239.

[2] Scarlett Liu, Mark Whitty. Automatic grape bunch detection in vineyards with an SVM classifier. Journal of Applied Logic. Vol. 13 (2015) NO 04, p: 643-653.

[3] Petr Dolezel, Pavel Skrabanek, Lumir Gago. Detection of grapes in natural environment using feedforward neural network as a classifier. 2016 SAI Computing Conference (SAI). Londo, 13 July 2016, p:1330-1334.

[4] Ping Liu, Yanjun Zhu, Tongxun Zhang, et al. Algorithm for recognition and image segmentation of overlapping grape cluster in natural environment. Transactions of the Chinese Society of Agricultural Engineering. Vol. 36 (2020) NO 06, p: 161-169.

[5] Hubert Cecotti, Agustin Rivera, Majid Farhadloo, et al. Grape detection with convolutional neural networks. Expert Systems with Applications. Vol. 159 (2020), p: 113588.

[6] Bogdan Franczyk, Marcin Hernes, Adrianna Kozierkiewicz, et al. Deep learning for grape variety recognition. Procedia Computer Science. Vol. 176 (2020), p: 1211-1220.

[7] Juntao Xiong, Zhiliang He, Linyue Tang, et al. Visual Localization of Disturbed Grape Picking point in Non-structural Environment. Transactions of the Chinese Society for Agricultural Machinery. Vol. 48 (2017) NO 04, p: 29-33+81.

[8] Wangxiong Lei, Jun Lu. Visual positioning method for picking point of grape picking robot [J]. Jiangsu Journal of Agricultural Sciences. Vol. 36 (2020) NO 04, p: 1015-1021.

[9] Zhengtong Ning, Lufeng Luo, Jiaxin Liao, et al. Recognition and the optimal picking point location of grape stems based on deep learning. Transactions of the Chinese Society of Agricultural Engineering. Vol. 37 (2021) NO 09, p: 222-229.

[10] Xiaohu Lu, Jian Yao, Kai Li, et al. CannyLines: A parameter-free line segment detector. 2015 IEEE International Conference on Image Processing. Canada, 10 December 2015, pp:507-511.

[11] Mark Everingham, Luc Van Gool, Christopher K. I. Williams, et al. The Pascal Visual Object Classes (VOC) Challenge. International Journal of Computer Vision. Vol. 88 (2010) NO 02, p: 303-338.