



Article A Method of Grasping Detection for Kiwifruit Harvesting Robot Based on Deep Learning

Li Ma¹, Zhi He¹, Yutao Zhu¹, Liangsheng Jia¹, Yinchu Wang¹, Xinting Ding¹ and Yongjie Cui^{1,2,3,*}

- ¹ College of Mechanical and Electrical Engineering, Northwest A&F University, Yangling 712100, China
- ² Key Laboratory of Agricultural Internet of Things, Ministry of Agriculture and Rural Affairs, Yangling 712100, China
- ³ Shaanxi Key Laboratory of Agricultural Information Perception and Intelligent Service, Yangling 712100, China
- * Correspondence: agriculturalrobot@nwafu.edu.cn

Abstract: Kiwifruit harvesting with robotics can be troublesome due to the clustering feature. The gripper of the end effector will easily cause unstable fruit grasping, or the bending and separation action will interfere with the neighboring fruit because of an inappropriate grasping angle, which will further affect the success rate. Therefore, predicting the correct grasping angle for each fruit can guide the gripper to safely approach, grasp, bend and separate the fruit. To improve the grasping rate and harvesting success rate, this study proposed a grasping detection method for a kiwifruit harvesting robot based on the GG-CNN2. Based on the vertical downward growth characteristics of kiwifruit, the grasping configuration of the manipulator was defined. The clustered kiwifruit was mainly divided into single fruit, linear cluster, and other cluster, and the grasping dataset included depth images, color images, and grasping labels. The GG-CNN2 was improved based on focal loss to prevent the algorithm from generating the optimal grasping configuration in the background or at the edge of the fruit. The performance test of the grasping detection network and the verification test of robotic picking were carried out in orchards. The results showed that the number of parameters of GG-CNN2 was 66.7 k, the average image calculation speed was 58 ms, and the average grasping detection accuracy was 76.0%, which ensures the grasping detection can run in real time. The verification test results indicated that the manipulator combined with the position information provided by the target detection network YOLO v4 and the grasping angle provided by the grasping detection network GG-CNN2 could achieve a harvesting success rate of 88.7% and a fruit drop rate of 4.8%; the average picking time was 6.5 s. Compared with the method in which the target detection network only provides fruit position information, this method presented the advantages of harvesting rate and fruit drop rate when harvesting linear clusters, especially other cluster, and the picking time was slightly increased. Therefore, the grasping detection method proposed in this study is suitable for near-neighbor multi-kiwifruit picking, and it can improve the success rate of robotic harvesting.

Keywords: kiwifruit; harvesting robot; grasping angle; GG-CNN; deep learning

1. Introduction

Kiwifruit has an average vitamin C content of 70 mg per 100 g, and is considered a highly nutritious product [1]. China is the origin of kiwifruit and the largest kiwifruit producer in the world. The planting area in 2020 was 1.85×10^5 hectares, and the yield was 2.23 million tons [2]. Kiwifruit orchards need to be carefully managed throughout the year. In particular, fruit harvesting in autumn is labor-intensive work, accounting for more than 25% of the production costs [3]. In order to overcome the growing labor shortage, the development of efficient and adaptable kiwifruit harvesting robots has become a research hotspot [4–7].

Fruit target detection is one of the important steps to perform robot harvesting. Robots working in orchards with complex lighting conditions require reliable information from



Citation: Ma, L.; He, Z.; Zhu, Y.; Jia, L.; Wang, Y.; Ding, X.; Cui, Y. A Method of Grasping Detection for Kiwifruit Harvesting Robot Based on Deep Learning. *Agronomy* **2022**, *12*, 3096. https://doi.org/10.3390/ agronomy12123096

Academic Editor: Chao Chen

Received: 27 October 2022 Accepted: 4 December 2022 Published: 7 December 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). visual perception systems. In recent years, due to the strong adaptability of deep-learning technology to orchard scenes, it has been widely used in the field of agriculture [8]. The kiwifruit target-detection method based on deep learning has been widely studied by scholars. The mAP (mean average precision) is up to 93%, and the average time for single image processing is 34 ms, which basically meets the accuracy and speed requirements of the kiwifruit target-detection task [9–12]. However, the success rate of the most advanced kiwifruit harvesting robot is less than 80% [5,13], and the two main problems are fruit retention and fruit dropping. Fruit retention occurs because of unsuccessful grasping or unstable grasping causing the fruit to slip out of the gripper. At the same time, approximately 87% of the kiwifruit is distributed in clusters in canopies [14] and the gripper will interfere with the adjacent fruit due to the improper bending direction, which causes the fruit to drop. The reason for these two problems can be attributed to the improper grasping pose of the gripper [13]. However, the target detection only obtains the fruit position information, and

rate and harvesting success rate. The shape of the kiwifruit's bottom is circular, and the fruit axis is approximately vertically downward. Therefore, the six-dimensional position and pose of the fruit can be reduced to four dimensions. Since the manipulator approaches and grasps the fruit vertically from bottom to top, and the growth height of the fruit in the cluster is not significantly different, the grasping process is regarded as planar grasping. Grasp-detection methods are divided into traditional methods [15], point-cloud segmentation methods [16,17], and deep-learning methods [18,19]. GG-CNN (generative grasping convolutional neural network) is a deep-learning-based planar grasping-detection algorithm. Compared with other algorithms, this algorithm is several orders of magnitude smaller, and achieves better performance in cluttered scenes. However, the GG-CNN network has a simple structure and takes a single channel depth image as the input; it only uses the depth prior knowledge and discards the color and other advanced prior knowledge, which makes it difficult for the algorithm to effectively learn the significant features related to grasping. The authors achieved a grasping rate of 84% on a group of unknown objects with adversarial geometric shapes and a grasping rate of 94% on household items [19]. Therefore, we consider adopting GG-CNN2 for transfer learning and applying the algorithm to grasping detection for kiwifruit harvesting.

does not involve the grasping pose of the manipulator. Therefore, we consider combining the target-detection method and the grasping-detection method to improve the grasping

In this study, the clustered kiwifruit was mainly divided into three categories: single fruit, linear cluster, and other cluster. The grasping-detection network GG-CNN2 was used to predict the grasping angle of the gripper. The rest of this paper is structured as follows: Section 2 introduces the definition of the grasping configuration, image acquisition, grasping dataset, GG-CNN2 network architecture, and the improvement in loss function based on focal loss. Section 3 analyzes and discusses the network training results, grasping prediction results, and robotic-harvesting verification test results. Finally, Section 4 outlines the conclusions obtained from this work.

2. Materials and Methods

2.1. Description of Kiwifruit in Orchard

The kiwifruit orchard is scaffold-cultivated in Meixian County, Shaanxi Province, China (108.00° E, 34.13° N). The average row width is 4 m, and the average plant space is 3 m. The branches and leaves form a dense canopy by fixing and extending the branches with steel wires. The fruits are naturally drooping, and distributed in the spatial range of 1.5 m–1.8 m above the ground [20,21]. Figure 1 shows the distribution characteristics of kiwifruit. The outline and calyx characteristics of kiwifruit are obvious. The fruits grow in clusters and are adjacent to each other. The clusters include single fruit, linear cluster, and other cluster, and the number of fruits in a single cluster is approximately 2–10 [14,22]. A single fruit is defined as one fruit with no adjacent fruit around it. A linear cluster is defined as the number of fruits being more than or equal to two, with the fruits distributed

in chains; each fruit has at most two adjacent fruits, and there is only one adjacent fruit at the beginning and end of the chain. Approximately 87% of the fruits within the canopy only have two adjacent fruits [14]. The other cluster is defined as the number of fruits being more than or equal to four, with the fruits distributed in an irregular regional shape, and with some fruits having more than or equal to three adjacent fruits.



Figure 1. Distribution characteristics of kiwifruit.

2.2. Description of the Grasping Pose of Manipulator

2.2.1. Grasping Pose

The world coordinate system {*W*} is set as the robot base coordinate system {*B*}. The positioning information of the target fruit (x_k , y_k , z_k) is obtained by the deep-learning algorithm YOLO v4. The robotic arm moves to the canopy underside corresponding to the axis of the target fruit based on the target pose represented by Equation (1), as shown in Figure 2a.

$$E_1 = \left[x_k, y_k, 500 \text{ mm}, Rx_1 = \frac{\pi}{2}, Ry_1 = 0, Rz_1 = \frac{\pi}{2}\right]^{\mathrm{T}}$$
(1)

Robotic arm grasping pose $E_2 = [x_k, y_k, z_k, R_x, R_y, R_z]^T$ is equivalent to the transformation matrix of the end effector coordinate system {*E*} relative to the robot base coordinate system {*B*}. T_E^B can be expressed as Equation (2) [23].

$$T_E^B = T_K^B T_E^K R_{E2}^{E1} \tag{2}$$

where T_K^B is the transformation matrix of the fruit coordinate system {*K*} relative to the robot base coordinate system {*B*}, T_E^K is the transformation matrix of the end-effector coordinate system {*E*} relative to the fruit coordinate system {*K*}, and R_{E2}^{E1} is the rotation matrix of the end effector around the *y*-axis of its own coordinate system {*E*}. T_K^B can be calculated by combining the internal- and the external-parameters transformation matrices of the camera [7]. T_E^K can be expressed as Equation (3).

$$I_E^K = \begin{bmatrix} R(z, \frac{\pi}{2})R(y, 0)R(x, \frac{\pi}{2}) & \begin{bmatrix} 0 & d & 0 \end{bmatrix}^T \\ 0_{1\times 3} & 1 \end{bmatrix}$$
(3)

$$d = z_k - 500 \text{ mm} \tag{4}$$

where *d* is the grasping distance (mm). Therefore, the problem of grasping pose detection can be transformed into finding the optimal rotation matrix R_{E2}^{E1} ; it can be expressed as in Equation (5).

$$R_{E2}^{E1} = \begin{bmatrix} r_{y}(\theta) & 0_{3\times 1} \\ 0_{1\times 3} & 1 \end{bmatrix}$$
(5)

where θ is the grasping angle (°), $r_y(\theta)$ is the Rodrigues-transformed 3 × 3 rotation matrix rotated by θ around the *y* axis. As shown in Figure 2b, the effective range of the grasping Euler angle *Rz* is [0, 2 π]. Since the gripper is a two-finger gripper with rotational symmetry on the central axis (*y* axis), the value range of the Euler angle *Rz* is [0, π] \cup [$-\pi$, 0]. At the same time, due to the initial Euler angle being *Rz*₁, the Euler angle *Rz* can be expressed as Equation (6).

$$Rz = \begin{cases} Rz_1 + \theta, Rz \in [0, \pi] \\ \theta - Rz_1, Rz \in [-\pi, 0] \end{cases}, -\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$$
(6)



Figure 2. Schematic diagram of grasping pose: (a) coordinate transformation, (b) grasping path.

Robotic-arm grasping pose E_2 can be expressed as Equation (7).

$$E_{2} = \left[x_{k}, y_{k}, z_{k}, \frac{\pi}{2}, 0, Rz\right]^{\mathrm{T}}$$
(7)

Based on the above analysis, the detection problem of the grasping pose is finally transformed into the calculation of the grasping angle θ .

2.2.2. Grasping Angle

In Figure 3, the planar grasping configuration *g* is defined as Equation (8).

$$g = \{q, u, v, \theta, w\}$$
(8)

where *q* represents the grasping quality, (u, v) represents the grasping point of the pixel coordinates, and the grasping angle θ is defined as the angle between the opening–closing direction of the gripper (green line) and the horizontal axis of the camera (sky blue line), and *w* represents the grasping width in the image. Since the gripper spacing can adapt to the maximum diameter of the kiwifruit [6], we do not make strict requirements for the

prediction accuracy of the grasping width. In Figure 3, the black fruit area indicates that the gripper cannot touch other fruits except the current fruit, and the white background area is regarded as the free area, which represents the area that the gripper can reach. The bending normal vector is perpendicular to the opening–closing direction of the gripper. The gripper rotates by 60° to be able to safely separate the fruit from the stalk [6].



Figure 3. Schematic diagram of grasping-angle definition.

2.3. Image Acquisition

The images of kiwifruit clusters were collected at the kiwifruit experimental station in Meixian County, Shaanxi Province, China, during the daytime from August to October in 2022, as shown in Figure 4. The images were acquired by a depth camera (RealSense D435i, Intel Corporation, Santa Clara, CA, USA). The image resolution was set to 640×480 . The depth camera was placed approximately 30 cm below the canopy for image acquisition from bottom to top. The color and depth images were saved in PNG and TIFF formats. Backlighting did not affect the quality of the depth images of kiwifruit. The effective filling rate of the fruit area was above 95% [24]. A total of 360 original images were obtained, including 50 single-fruit images, 220 linear-cluster images, and 90 other-cluster images. All images were collected at different locations to ensure that there were no overlapping regions in the images.





2.4. Grasping Datasets

The Connell grasping dataset can be used for the training of the grasping-detection network [25], which is mainly used for the two-finger gripper or sucker. In this paper, we refer to the AFFGA-Net annotation method to construct a grasping dataset [26], including depth images, color images, and grasping labels in MAT format. Figure 5 shows the visualized results of the grasping labels. According to the kiwifruit distribution characteristics of each cluster, a specific positive-sample labeling method was used. The grasping label of a single fruit (referred to as SF) is shown in Figure 5a. The blue area in the figure is the grasping point, corresponding to the kiwifruit calyx area, and the green circles indicate that the fruit can be safely grasped by the gripper at any grasping angle. The grasping label of a linear cluster (referred to LC) is shown in Figure 5b. Since the fruits of the linear cluster are distributed in a chain shape, there is only one adjacent fruit along the chain. The green lines in the figure represent the opening–closing direction of the gripper. Each grasping point on the blue line corresponds to a grasping angle indicated by a green line. The grasping label of other cluster (referred to OC) is shown in Figure 5c. Since there are three or more fruits adjacent to the central fruit, there is no continuous free area of approximately 180 degrees, and it is difficult for the gripper to approach the fruit at a safe grasping angle; therefore, there may be unlabeled fruits in the other cluster. For the peripheral fruit of the other cluster, there are two fruits adjacent to it, and the blue line is approximately perpendicular to the line connecting the two adjacent fruits calyxes. The grasping dataset was divided into the training set and the test set at a ratio of 4:1. In order to expand the number of samples in the training set, the images and labels were simultaneously enhanced by scaling, rotating, and flipping.







Figure 5. Visualization results of the grasping labels: (**a**) SF, (**b**) LC, (**c**) OC.

2.5. Grasping Detection Network

2.5.1. Network Structure

The grasping configuration is predicted based on the grasping-detection network GG-CNN2 [19], and the network architecture is shown in Figure 6. First, the depth image is scaled to 300 × 300 pixels and sent to the network. Then, the image feature extraction is performed by stacking four standard convolutions of different sizes and two maximum pooling to generate a low-resolution feature map. Then, the feature map is restored in the scale space by stacking two bilinear interpolation up-sampling and standard convolutions. Finally, the maps of three-channel grasping pose G_{θ} are output, including the map of grasping quality Q_{θ} , the map of grasping width W_{θ} and the map of grasping angle Φ_{θ} . The map of grasping quality Q_{θ} describes the grasping feasibility of each pixel in the depth image. The closer the value is to 1, the higher the grasping quality and the darker the color appears in figure. The method for generating the optimal grasping configuration g^* is based on the heatmap maximum value strategy [27], in which the position parameters of g^* depend on the peak-point coordinates of Q_{θ} , and the angle and width parameters are the peak-point coordinates of Φ_{θ} and W_{θ} , respectively. The formula is defined as follows:

$$g^* = \max_{Q_{\theta}} G_{\theta} = \{q, u, v, \theta, w\}$$
(9)



Figure 6. Network architecture for GG-CNN2.

Normalization of input features can speed up convergence of the model. The grasping width divides the maximum width value of 250 pixels. The cosine and sine prediction maps of the grasping angle are obtained by linear regression, and then Φ_{θ} is obtained by solving Equation (10) [19].

$$\Phi_{\theta} = \frac{1}{2} \arctan \frac{\sin(2\Phi_{\theta})}{\cos(2\Phi_{\theta})} \tag{10}$$

In order to prevent the network from generating the optimal grasping configuration in the background or at the edge of the fruit due to the imbalance of positive and negative samples, the original loss function of mean squared error (MSE) is improved based on the focal loss [28] with binary cross entropy (BCE) to improve the learning efficiency and generalization ability of the network. Predicting the grasping region is a binary classification problem. The sigmoid function is used to normalize the prediction results, and the focal loss is used to calculate. The grasping quality loss L_{qua} is defined as Equation (11).

$$L_{qua} = -\frac{1}{N} \sum_{n=0}^{N} \left[(1-\alpha) \cdot y_q^n \cdot \left(p_q^n \right)^{\gamma} \cdot \log\left(p_q^n \right) + \alpha \cdot \left(1 - y_q^n \right) \cdot \left(1 - p_q^n \right)^{\gamma} \cdot \log(1 - p_q^n) \right]$$
(11)

where *N* is the size of the feature map, p_q^n is the predicted probability, y_q^n is the sample label, α is the balance factor, and γ is the regulatory factor. Predicting the grasping angle is a regression problem. First, the sigmoid function is used to normalize the output of the angle head, and then the BCE function is used to calculate the loss. The grasping angle loss L_{ang} is defined as Equation (12).

$$L_{ang} = -\frac{1}{N} \sum_{n=0}^{N} [y_l^n \cdot \log(p_l^n) + (1 - y_l^n) \cdot \log(1 - p_l^n)]$$
(12)

where p_l^n is the predicted probability, and y_l^n is the sample label. Predicting the grasping width is a regression problem. The BCE function is used to calculate the loss, and the grasping width loss L_{wid} is defined as Equation (13) [26].

$$L_{wid} = -\frac{1}{N} \sum_{n=0}^{N} [y_w^n \cdot \log(p_w^n) + (1 - y_w^n) \cdot \log(1 - p_w^n)]$$
(13)

where is p_w^n the predicted grasping width, and y_w^n is the sample label. In order to balance the loss of each branch, the total loss is used to optimize the network by calculating the loss of the output of each head, and the multi-task loss L_{total} is defined as Equation (14).

$$L_{total} = L_{qua} + L_{ang} + L_{wid} \tag{14}$$

2.5.2. Evaluation and Hyperparameters

In this paper, the Jaccard index of the grasping rectangle [29] is used to determine whether the grasping estimation is effective. Specifically, the grasping prediction needs to meet two conditions at the same time: (1) the difference between the predicted grasping angle and the labeled grasping angle is less than 15°; and (2) the Jaccard index of the predicted grasping frame and the true grasping frame is not lower than 0.25. The Jaccard index is calculated by Equation (15).

$$J(G_P, G_T) = \frac{G_P \cap G_T}{G_P \cup G_T}$$
(15)

where G_P represents the area of the predicted grasping frame, G_T represents the area of the true grasping frame, $G_P \cap G_T$ represents the intersection, and $G_P \cup G_T$ represents the union. The accuracy of the test set data is used as the evaluation index, and the accuracy is calculated according to Equation (16).

$$accuracy = \frac{N_{correct}}{N_{total}} \times 100\%$$
(16)

The network is implemented based on the PyTorch deep-learning framework. The operating environment is Ubuntu 16.04, CPU, AMD Ryzen 7 pro 4750U with Radeon Graphics. The network uses the Adam optimization function. The initial learning rate is set to 0.001, the weight attenuation coefficient is set to 0.01, and the batch size is set to 2; a total of 2000 epochs were trained.

3. Results and Analysis

3.1. Network Training Results

The network training process data was downloaded from the TensorBoard. Figure 7a shows that the curve of the loss function gradually decays with the number of iterations. The loss function decays quickly in the early iteration stage, then it starts to converge and stabilizes at approximately 1.8 after 1000 iterations of training. Figure 7b shows the graspable curve gradually increases with the number of iterations, and the graspable converges to approximately 80%. This curve indicates that the GG-CNN2 network can effectively predict the grasping configuration and the generalization ability is gradually improved.



Figure 7. Training evaluation results: (a) loss curve; (b) graspable curve.

3.2. Grasping Detection Results

In order to evaluate the generalization ability of the algorithm in the orchard scenario, we take the clustered kiwifruit scenarios at random locations in the orchard as the test environment, and carry out the detection test of the grasping configuration based on

the grasping-detection algorithm. Figure 8 shows the results of the grasping detection. The results of grasping angle were annotated in the figure, which is the most important parameter in grasping configuration. The results show that for different fruit-distribution scenarios, the grasping algorithm can generate an optimal grasping configuration with the highest grasping quality while meeting the requirements. Although the background in the figure shows different lighting conditions—some features of the fruit are lost due to the backlight—the network can still rely on the depth prior information provided by the depth image to complete the prediction of the grasping configuration.



Figure 8. Test results of grasping detection.

Figure 9 shows the process of detecting the grasping angle. The depth image generated candidate grasping areas through the grasping-detection network (Figure 9b), then the candidate grasping angles were selected corresponding to the grasping-quality peak pixels in the region (Figure 9c); the final grasping angle was selected based on the principle of maximum grasping quality (Figure 9d). As the fruit depth information on the left side of the depth image (Figure 9a) was incomplete, the fruit was not detected as a candidate grasping area.



Figure 9. Detection process of the grasping angle: (a) depth image; (b) candidate grasping areas; (c) candidate grasping angles; (d) grasping angle.

Figure 10 shows the case of a false-positive prediction. As the leaf outline around the fruit in the depth image is clear, the shape is approximately circular, and the leaf depth value is close to the fruit, which leads to a false-positive prediction. Therefore, the grasping prediction will be affected by the interference of leaves and the depth filling rate in the actual orchard environment if it only depends on the depth image.





Figure 10. Detection process of the grasping angle.

Table 1 shows the performance results of the gasping-detection network in different scenarios. The results show that the number of parameters of GG-CNN2 was 66.7 k, the average image calculation speed was 58 ms, and the average grasping-detection accuracy was 76.0%, which ensures the grasping detection can run in real time. The algorithm shows better grasping-prediction ability for single fruit and linear cluster compared with other cluster. It can complete the grasping-prediction task for most fruits. At the same time, the lightweight feature of the network is deployed in the portable graphics processing unit, which can realize the application in the real scene.

Table 1. Performance results of grasp network.

Algorithm	Parameters	Clusters	Samples	Average Accuracy	Speed (ms)
GG-CNN2	66.7 k	SF LC OC	25 25 25	80.3% 77.7% 70.0%	58

3.3. Verification Test of Robotic Picking

In order to verify whether the robot can improve the fruit-grasping rate and harvesting success rate under the condition of combining the information of the position and grasping angle of the target fruit, a picking experiment was conducted in the Yangling International Kiwifruit Innovation and Entrepreneursnip Park's kiwifruit orchard trellis-cultivation environment.

3.3.1. Overall Structure

As shown in Figure 11, the overall structure of the kiwifruit picking robot consists of five parts: robotic arm, end effector, vision system, fruit-collection device, and mobile platform. The robotic arm (UR5, Universal Robots, Odense, Denmark) is a multi-joint robotic arm with the characteristics of being lightweight and having high flexibility. The robotic arm is composed of six rotating joints, with a repeatability of ± 0.1 mm, a working radius of 850 mm, and an effective working load of 5 kg. The end effector is composed of two 3D-printed lightweight grippers, photoelectric sensors and pneumatic components. The inner curved surface of the grippers is designed to adapt to the shape of the kiwifruit, thereby reducing fruit damage during the picking process. The total weight of the end effector is 3.5 kg, and the separation force between the stalk and fruit is 3–10 N [6], which meets the requirement that the effective load of the robotic arm be less than 5 kg. The vision system includes an RGB-D camera (RealSense D435i, Intel, Santa Clara, CA, USA) and an image-processing unit (Jetson Nano, NVIDIA, Santa Clara, CA, USA). The camera detects

and locates the target kiwifruit in a bottom-up direction through the arrangement of the eyes on the hand [10]. The fruit collection device includes a bellows and a box, and the harvested fruits slide into the box by the buffering effect of the bellows. The mobile platform (Safari-880T, Guoxing Intelligent Technology, Shenzhen, China) is a crawler chassis with good trafficability in the orchard.



Figure 11. Structure of harvesting robot.

3.3.2. Control System

The picking-robot control system is developed based on the ROS-MoveIt (Robot Operation System Motion Planning Framework) [30], as shown in Figure 12. The RGB-D camera captures fruit color images and depth images and transmits the images to the image-processing unit. The image-processing unit first performs fruit target detection and grasping detection based on the deep-learning model, and then obtains the pose information of the target fruit relative to the robot base coordinate system based on the internal and external parameter matrices of the camera. The fruit-pose information is sequentially published in the form of topics and the robotic-arm control node subscribes to the topic. The rapidly exploring random trees (RRT) algorithm in the Open Motion Planning Library (OMPL) is used for path planning. The inverse kinematics solution is solved by calling the inverse solver IKFast to form the dynamic trajectory of the robotic-arm kinematics group and drive the robotic arm to arrive at the target pose. After the robotic arm completes the current target-fruit picking task, the image-processing node updates the fruit-pose information until all fruit-picking tasks are completed.

Figure 12. Schematic diagram of harvesting-robot control system based on ROS-Movelt.

3.3.3. Test Method

We implemented two methods for harvesting clustered kiwifruit. Method I is the original method, which is described as follows: the fruit target detection is performed based on the YOLO v4 network, and then the picking order is determined according to the principle of the shortest spatial distance; the manipulator combines the current-pose and fruit-position information to perform motion planning to complete all the fruit-grasping and picking tasks one by one. Method II is specifically described as follows: fruit target detection and grasping detection are performed based on the YOLO v4 network and the GG-CNN2 network, respectively, and then the manipulator performs motion planning under the condition of combining the information of the fruit position and the grasping angle, and, finally, the manipulator completes all fruit-picking tasks one by one. During the test of Method II, the robot removes the fruit associated with the optimal grasping configuration from the scene after each prediction, and the fruit is removed one by one, finally forming the picking sequence. The manipulator picking the fruit includes three steps. First, the manipulator receives the instruction to move to the canopy underside corresponding to the axis of the target fruit. Then, the end effector moves vertically upward to the fruit positioning point, and the photoelectric sensor signal controls the gripper to close to complete the fruit grasping. Finally, the separation of the fruit and the peduncle is completed by rotating the wrist joint of the robotic arm to a certain angle. During the second step of the picking, the manipulator adjusts the gripper according to the predicted grasping angle to safely approach and grasp the target fruit. In this test, the kiwifruits located in different positions in the canopy were randomly selected, and clustered fruits. Such as single fruit, two-fruit linear cluster, three-fruit linear cluster, and other cluster, were tested. The number of fruits picked in the fruit box and the number of fruits unseparated from the branches and the number of fruits dropped on the ground were counted. The harvesting success rate and the fruit drop rate were calculated. In addition, a phone timer was used to record the total time from the initial position of the end effector to the end of a cluster being picked, and the total time was divided by the number of kiwifruits in each cluster; the average value of several groups of mean times was taken as the average picking time.

3.3.4. Results and Analysis

Figure 13 shows the picking process of the manipulator in the kiwifruit orchard. As shown in Figure 13a, the deep learning based target detection network obtains the position information of all fruits in the color image. Figure 13b shows changes in acquired depth images. As shown in Figure 13c, the deep-learning-based grasping detection network obtains the grasping angle information corresponding to the current fruit depth image with

the highest grasping quality. Figure 13d shows that the manipulator combines the target position and pose information to complete motion planning and executes grasping. As the fruits were separated and dropped into the box along the bellows, the distribution characteristics of the fruits in the depth image also change accordingly. Therefore, the grasping network needs to evaluate the grasping quality and grasping angle of the remaining fruits in the current depth image, and determine the next fruit to be grasped.

Figure 13. Robotic picking process of the manipulator in the kiwifruit orchard: (**a**) target detection; (**b**) depth images; (**c**) grasping angle; (**d**) manipulator performs grasping.

The robotic-picking test results are shown in Table 2. The single fruit, linear cluster, and other cluster were used to perform grasping tests 10, 25, and 27 times, respectively. The results show that the grasping rate of single fruit and linear cluster are both higher than other cluster. This is because the free area around the other cluster is relatively small. Method II, combining with target-detection and grasping-detection information, can achieve a fruit-harvesting success rate of 88.7% and a fruit drop rate of 4.8%. Compared with Method I, the harvesting success rate increased by 8.1%, and the fruit drop rate decreased by 4.9%; the average picking time was 6.5 s, which was a slight increase. There was no obvious difference in the grasping rate between two methods for single-fruit picking, but for the linear cluster, especially the other cluster, there was an obvious difference, indicating that Method II would be effective when the robot was facing the clustered fruit-picking tasks. It is a safe method which can effectively improve the success rate and drop rate of clustered fruit. In addition, the small number of the fruits left on the branches was mainly due to unsuccessful detection caused by environmental factors such as leaf occlusion and backlighting; the picking sequence of clustered fruit is also an important influencing factor.

Table 2. The results of robotic picking test in kiwifruit orchard.

Method _	(Grasping Rate			Dropped	Harvesting	Average Picking
	SF	LC	OC	· · · · · · · · · · · · · · · · · · ·		Success Rate	Time (s)
Method I	9/10 90.0%	21/25 84.0%	20/27 74.1%	6 9.7%	6 9.7%	80.6%	5.8
Method II	9/10 90.0%	23/25 92.0%	23/27 85.2%	4 6.5%	3 4.8%	88.7%	6.5

Several kinds of fruit and vegetable picking robots using multi-joint manipulators were compared and analyzed, as shown in Table 3. For greenhouse vegetables, such as

tomato and sweet-pepper picking robots, the picking efficiency is relatively lower than for fruit picking robots. Most of these robots need to cut stalks and pick a single target selectively with requirements on positioning accuracy. The harvesting rate of our kiwifruit picking robot is 80.6%, and the picking time is 5.8 s. However, there is an obvious difference between robot and manual picking in operating efficiency; we need to further optimize the perceptual and planning algorithms.

Table 3. Comparison of different fruit harvesting robots.

	Objects	Harvesting Rate	Picking Time (s)
Ola Ringdahl, et al. [31]	sweet pepper	61%	24
Hiroaki Yaguchi, et al. [32]	tomato	60%	23
Pengbo Wang, et al. [33]	cherry tomato	72%	14
Abhisesh Silwal, et al. [34]	apple	84%	6.0
Ours	kiwifruit	80.6%	5.8

4. Conclusions

- (1) In this study, a grasping-detection method for a kiwifruit harvesting robot was proposed based on the GG-CNN2, which enables the gripper to safely and effectively grasp the clustered fruits and avoid the interference of the bending action on the neighboring fruits. We mainly divided the clustered kiwifruit into three types, including single fruit, linear cluster, and other cluster.
- (2) The performance test results of the grasping-detection network showed that the number of parameters of the GG-CNN2 was 66.7 k, the average image calculation speed was 58 ms, and the average accuracy was 76.0%, which ensures that the grasping prediction can complete the most tasks and run in real-time.
- (3) The verification test results of robotic picking showed that the manipulator combined with the position information provided by the target-detection network YOLO v4 and the grasping angle provided by the grasping-detection network GG-CNN2 achieved a harvesting success rate of 88.7% and a fruit drop rate of 4.8%; the average picking time was 6.5 s. Compared with the method which was only based on the target-detection information, the harvesting success rate of this method was increased by 8.1%, and the fruit drop rate was decreased by 4.9%; the picking time was slightly increased. The grasping-detection method is suitable for near-neighbor multi-kiwifruit picking.

Author Contributions: Conceptualization, L.M. and Y.Z.; methodology, L.M. and Z.H.; software, X.D.; validation, L.M. and Y.W.; formal analysis, Z.H. and X.D.; investigation, L.J.; resources, Y.W.; data curation, L.M.; writing—original draft preparation, L.M.; writing—review and editing, L.M. and Z.H.; visualization, L.J. and Y.W.; supervision, Y.C. and Z.H.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 31971805.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data are presented in this article in the form of figures and tables.

Acknowledgments: This study was conducted in the College of Mechanical and Electronic Engineering, Northwest A&F University. We thank the Yangling International Kiwifruit Innovation and Entrepreneursnip Park for providing kiwifruit orchard to support the robotic harvesting test.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

- Baranowska, W.E.; Dominik, S. Characteristics and pro-health properties of mini kiwi (*Actinidia arguta*). *Hortic. Environ. Biotechnol.* 2019, 60, 217–225. [CrossRef]
- Production of Kiwi (Fruit) by Countries. UN Food and Agriculture Organization. 2020. Available online: https://www.fao.org/ faostat/en/#data (accessed on 10 August 2022).
- García-Quiroga, M.; Nunes-Damaceno, M.; Gómez-López, M.; Arbones-Maciñeira, E.; Muñoz-Ferreiro, N.; Vázquez-Odériz, M.L.; Romero-Rodríguez, M.A. Kiwifruit in syrup: Consumer acceptance, purchase intention and influence of processing and storage time on physicochemical and sensory characteristics. *Food Bioprocess Technol.* 2015, *8*, 2268–2278. [CrossRef]
- Williams, H.A.M.; Jones, M.H.; Nejati, M.; Seabright, M.J.; Bell, J.; Penhall, N.D.; Barnett, J.J.; Duke, M.D.; Scarfe, A.J.; Ahn, H.S.; et al. Robotic kiwifruit harvesting using machine vision, convolutional neural networks, and robotic arms. *Biosyst. Eng.* 2019, 181, 140–156. [CrossRef]
- 5. Williams, H.; Ting, C.; Nejati, M.; Jones, M.H.; Penhall, N.; Lim, J.Y.; Seabright, M.; Bell, J.; Ahn, H.S.; Scarfe, A.; et al. Improvements to and large-scale evaluation of a robotic kiwifruit harvester. *J. Field Robot.* **2020**, *37*, 187–201. [CrossRef]
- 6. Mu, L.; Cui, G.; Liu, Y.; Cui, Y.; Fu, L.; Gejima, Y. Design and simulation of an integrated end-effector for picking kiwifruit by robot. *Inf. Process. Agric.* 2020, 7, 58–71. [CrossRef]
- Cui, Y.; Ma, L.; He, Z.; Zhu, Y.; Wang, Y.; Li, K. Design and Experiment of Dual Manipulators Parallel Harvesting Platform for Kiwifruit Based on Optimal Space. *Trans. CSAM* 2022, 53, 132–143.
- 8. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. Comput. Electron. Agric. 2018, 147, 70–90. [CrossRef]
- 9. Fu, L.; Feng, Y.; Elkamil, T.; Liu, Z.; Li, R.; Cui, Y. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks. *Trans. CSAE* 2018, 34, 205–211. [CrossRef]
- 10. Mu, L.; Gao, Z.; Cui, Y.; Li, K.; Liu, H.; Fu, L. Kiwifruit Detection of Far-view and Occluded Fruit Based on Improved AlexNet. *Trans. CSAM* 2019, *50*, 24–34. [CrossRef]
- 11. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; Al-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* **2021**, *22*, 754–776. [CrossRef]
- 12. Suo, R.; Gao, F.; Zhou, Z.; Fu, L.; Song, Z.; Dhupia, J.; Li, R.; Cui, Y. Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking. *Comput. Electron. Agric.* **2021**, *182*, 106052. [CrossRef]
- 13. Au, C.K.; Redstall, M.; Duke, M.; Kuang, Y.C.; Lim, S.H. Obtaining the effective gripper dimensions for a kiwifruit harvesting robot using kinematic calibration procedures. *Ind. Robot Int. J. Robot. Res. Appl.* **2021**, *49*, 865–876. [CrossRef]
- 14. Fu, L.; Tola, E.; Al-Mallahi, A.; Li, R.; Cui, Y. A novel image processing algorithm to separate linearly clustered kiwifruits. *Biosyst. Eng.* **2019**, *183*, 184–195. [CrossRef]
- 15. Liu, M.-Y.; Tuzel, O.; Veeraraghavan, A.; Taguchi, Y.; Marks, T.K.; Chellappa, R. Fast object localization and pose estimation in heavy clutter for robotic bin picking. *Int. J. Robot. Res.* **2012**, *31*, 951–973. [CrossRef]
- Wang, X.; Kang, H.; Zhou, H.; Au, W.; Chen, C. Geometry-aware fruit grasping estimation for robotic harvesting in apple orchards. Comput. Electron. Agric. 2022, 193, 106716. [CrossRef]
- 17. Gao, R.; Zhou, Q.; Cao, S.; Jiang, Q. An Algorithm for Calculating Apple Picking Direction Based on 3D Vision. *Agriculture* **2022**, 12, 1170. [CrossRef]
- Ni, P.; Zhang, W.; Bai, W.; Lin, M.; Cao, Q. A new approach based on two-stream cnns for novel objects grasping in clutter. J. Intell. Robot. Syst. 2019, 1, 161–177. [CrossRef]
- 19. Morrison, D.; Corke, P.; Leitner, J. Learning robust, real-time, reactive robotic grasping. *Int. J. Robot. Res.* **2020**, *39*, 183–201. [CrossRef]
- 20. Cui, Y.; Su, S.; Wang, X.; Tian, Y.; Li, P.; Zhang, F. Recognition and Feature Extraction of Kiwifruit in Natural Environment Based on Machine Vision. *Trans. CSAM* **2013**, *44*, 247–252. [CrossRef]
- 21. Fu, L.; Zhang, F.; Gejima, Y.; Li, Z.; Wang, B.; Cui, Y. Development and Experiment of End-effector for Kiwifruit Harvesting Robot. *Trans. CSAM* 2015, 46, 1–8. [CrossRef]
- 22. Mu, L. Full Field of View Information Perception and Integrated Picking Method for Kiwifruit Harvesting Robot; Northwest A&F University: Xianyang, China, 2019.
- 23. Xu, J.; Liu, N.; Li, D.; Lin, L.; Wang, G. A Grasping Poses Detection Algorithm for Industrial WorkpiecesBased on Grasping Cluster and Collision Voxels. *Robot* 2022, 44, 153–166. [CrossRef]
- Xiao, Z.; Zhou, M.; Yuan, H.; Liu, Y.; Fan, C.; Cheng, M. Influence Analysis of Light Intensity on Kinect v2 Depth Measurement Accuracy. *Trans. CSAM* 2021, 52, 108–117. [CrossRef]
- 25. Lenz, I.; Lee, H.; Saxena, A. Deep learning for detecting robotic grasps. Int. J. Robot. Res. 2015, 34, 705–724. [CrossRef]
- 26. Wang, D.; Liu, C.; Chang, F.; Li, N.; Li, G. High-Performance Pixel-Level Grasp Detection Based on Adaptive Grasping and Grasp-Aware Network. *IEEE Trans. Ind. Electron.* **2022**, *69*, 11611–11621. [CrossRef]
- Zeng, A.; Song, S.; Yu, K.T.; Donlon, E.; Hogan, F.R.; Bauza, M.; Ma, D.; Taylor, O.; Liu, M.; Romo, E.; et al. Robotic Pick-and-Place of Novel Objects in Clutter with MultiAffordance Grasping and Cross-Domain Image Matching. *Int. J. Robot. Res.* 2017, 41, 690–705. [CrossRef]
- 28. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Adn Mach. Intell.* 2017, 99, 2999–3007.

- Jiang, Y.; Moseson, S.; Saxena, A. Efficient grasping from rgbd images: Learning using a new rectangle representation. In Proceedings of the 2011 IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 3304–3311. [CrossRef]
- Hernandez-Mendez, S.; Maldonado-Mendez, C.; Marin-Hernandez, A.; Rios-Figueroa, H.V.; Vazquez-Leal, H.; Palacios-Hernandez, E.R. Design and implementation of a robotic arm using ROS and MoveIt! In Proceedings of the International Autumn Meeting on Power, Electronics and Computing (ROPEC), Ixtapa, Mexico, 8–10 November 2017. [CrossRef]
- 31. Arad, B.; Balendonck, J.; Barth, R.; BenShahar, O.; Edan, Y.; Hellström, T.; Hemming, J.; Kurtser, P.; Ringdahl, O.; Tielen, T.; et al. Development of a sweet pepper harvesting robot. *J. Field Robot.* **2020**, *37*, 1027–1039. [CrossRef]
- 32. Yaguchi, H.; Nagahama, K.; Hasegawa, T.; Inaba, M. Development of an autonomous tomato harvesting robot with rotational plucking gripper. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 652–657. [CrossRef]
- 33. Rong, J.; Wang, P.; Wang, T.; Hu, L.; Yuan, T. Fruit pose recognition and directional orderly grasping strategies for tomato harvesting robots. *Comput. Electron. Agric.* **2022**, 202, 107430. [CrossRef]
- Silwal, A.; Davidson, J.R.; Karkee, M.; Mo, C.; Zhang, Q.; Lewis, K. Design, integration, and field evaluation of a robotic apple harvester. J. Field Robot. 2017, 34, 1140–1159. [CrossRef]