Article

An Autonomous Fruit and Vegetable Harvester with a Low-Cost Gripper Using a 3D Sensor

Tan Zhang, Zhenhai Huang, Weijie You, Jiatao Lin, Xiaolong Tang and Hui Huang *

Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University, Shenzhen 510000, China; tan.zhang@utoronto.ca (T.Z.); zhenhaihuang0@gmail.com (Z.H.); youweijiee@gmail.com (W.Y.); zeup300@gmail.com (J.L.); tangtxl80@gmail.com (X.T.)
* Correspondence: huihuang@szu.edu.cn

check for updates

Abstract: Reliable and robust systems to detect and harvest fruits and vegetables in unstructured environments are crucial for harvesting robots. In this paper, we propose an autonomous system that harvests most types of crops with peduncles. A geometric approach is first applied to obtain the cutting points of the peduncle based on the fruit bounding box, for which we have adapted the model of the state-of-the-art object detector named Mask Region-based Convolutional Neural Network (Mask R-CNN). We designed a novel gripper that simultaneously clamps and cuts the peduncles of crops without contacting the flesh. We have conducted experiments with a robotic manipulator to evaluate the effectiveness of the proposed harvesting system in being able to efficiently harvest most crops in real laboratory environments.

Keywords: harvesting robot; gripper; segmentation; cutting point detection

1. Introduction

Manual harvesting of fruits and vegetables is a laborious, slow, and time-consuming task in food production [1]. Automatic harvesting has many benefits over manual harvesting, e.g., managing the crops in a short period of time, reduced labor involvement, higher quality, and better control over environmental effects. These potential benefits have inspired the wide use of agricultural robots to harvest horticultural crops (the term crops is used indistinctly for fruits and vegetables throughout the paper, unless otherwise indicated) over the past two decades [2]. An autonomous harvesting robot usually has three subsystems: a vision system for detecting crops, an arm for motion delivery, and an end-effector for detaching the crop from its plant without damaging the crop.

However, recent surveys of horticultural crop robots [3,4] have shown that the performance of automated harvesting has not improved significantly despite the tremendous advances in sensor and machine intelligence. First, cutting at the peduncle leads to higher success rates for crop detachment, and detachment at the peduncle reduces the risk of damaging the flesh or other stems of the plant and maximizes the storage life. However, peduncle detection is still a challenging step due to varying lighting conditions and occlusion of leaves or other crops [5], as well as similar colors of peduncle and leaves [6,7]. Second, existing manipulation tools that realize both grasping and cutting functions usually require a method of additional detachment [8,9], which increases the cost of the entire system.

In this paper, we present a well-generalized harvesting system that is invariant and robust to different crops in an unstructured environment aimed at addressing these challenges. The major contributions of this paper are the following: (i) development of a low-cost gripper (Supplementary Materials) that simultaneously clamps and cuts peduncles in which the clamper and cutter share the same power-source drive, thus maximizing the storage life of crops and reducing the complexity and cost; (ii) development of a geometric method to detect the cutting point of

the crop peduncle based on the Mask Region-based Convolutional Neural Network (Mask R-CNN) algorithm [10].

The rest of this paper is organized as follows: In Section 2, we introduce related work. In Section 3, we describe the design of the gripper and provide a mechanical analysis. In Section 4, we present the approach of cutting-point detection. In Section 5, we describe an experimental robotic system to demonstrate the harvesting process for artificial and real fruits and vegetables. Conclusions and future work are given in Section 6.

2. Related Work

During the last few decades, many systems have been developed for the autonomous harvesting of soft crops, ranging from cucumber [11] and tomato-harvesting robots [12] to sweet-pepper [2,13] and strawberry-picking apparatus [14,15]. The work in [16] demonstrated a sweet-pepper-harvesting robot that achieved a success rate of 6% in unstructured environments. This work highlighted the difficulty and complexity of the harvesting problem. The key research challenges include manipulation tools along with the harvesting process, and perception of target fruits and vegetables with the cutting points.

2.1. Harvesting Tools

To date, researchers have developed several types of end-effectors for autonomous harvesting, such as suction devices [17], swallow devices [18], and scissor-like cutters [15]. The key component of an autonomous harvester is the end-effector that grasps and cuts the crop. A common approach for harvesting robots is to use a suction gripping mechanism [16,19,20] that picks up fruit by pushing the gripper onto the crop to generate air to draw a piece of crop into the harvesting end and through the tubular body to the discharge end. These types of grippers do not touch the crops during the cutting process, but physical contact occurs when they are inhaled and slipped into the container, potentially causing damage to the crops, especially to fragile crops. Additionally, the use of additional pump equipment increases the complexity and weight of the robotic system, resulting in difficulties for compact and high-power density required in autonomous mobile robots [21,22]. Some soft plants, such as cucumbers and sweet peppers, must be cut from the plant and thus require an additional detachment mechanism and corresponding actuation system. The scissor-like grippers [16,23] that grasp the peduncles or crops and cut peduncles can be used for such soft plants. However, these grippers are usually designed at the expense of size. A custom harvesting tool in [16] can simultaneously envelop a sweet pepper and cuts the peduncle with a hinged jaw mechanism; however, size constraints restrict its use for crops with irregular shapes [24].

To overcome these disadvantages, we have designed a novel gripper that simultaneously clamps and cuts the peduncles of crops. In addition, the clamping and detachment mechanism share the same power-source drive rather than using separate drives. The benefits of sharing the same power source for several independent functions are detailed in our previous work [25,26]. Since the gripper clamps the peduncle, it can be used for most fruits that are harvested by cutting the peduncle. Since the gripper does not touch the flesh of the crop, it can maximize the storage life and market value of the crops. As the clamping and cutting occurs simultaneously at the peduncle, the design of the gripper is relatively simple and therefore cost-effective.

2.2. Perception

Detection in automated harvesting determines the location of each crop and selects appropriate cutting points [1,27,28] using RGB cameras [29,30] or 3D sensors [13,15]. The 3D localization allows a cutting point and a grasp to be determined. Currently, some work has been done to detect the peduncle based on the color information using RGB cameras [6,7,29,30]; however, such cameras are not capable of discriminating between peduncles, leaves, and crops if they are in the same color [7]. The work in [31] proposed a dynamic thresholding to detect apples in viable lighting conditions, and they further used a dynamic adaptive thresholding algorithm [32] for fruit detection using a small set of training

images. The work in [27] facilitates detection of peduncles of sweet peppers using multi-spectral imagery; however, the accuracy is too low to be of practical use. Recent work in deep neural networks has led to the development of a state-of-the-art object detector, such as Faster Region-based CNN (Faster R-CNN) [33] and Mask Region-based CNN (Mask R-CNN) [10]. The work in [34] uses a Faster R-CNN model to detect different fruits, and it achieves better performance with many fruit images; however, the detection of peduncles and cutting points has not been addressed in this work. To address these shortcomings, we use a Mask R-CNN model and a geometric feature to detect the crops and cutting points of the peduncle.

3. System Overview and the Gripper Mechanism

3.1. The Harvesting Robot

The proposed harvesting robot has two main modules: a detection module for detecting an appropriate cutting point on the peduncle, and a picking module for manipulating crops (see Figure 1). Specifically, the detection module determines the appropriate cutting points on the peduncle of each crop. The picking module controls the robot to reach the cutting point, clamps the peduncle, and cuts the peduncle. The clamping and cutting actions are realized by the newly designed gripper.
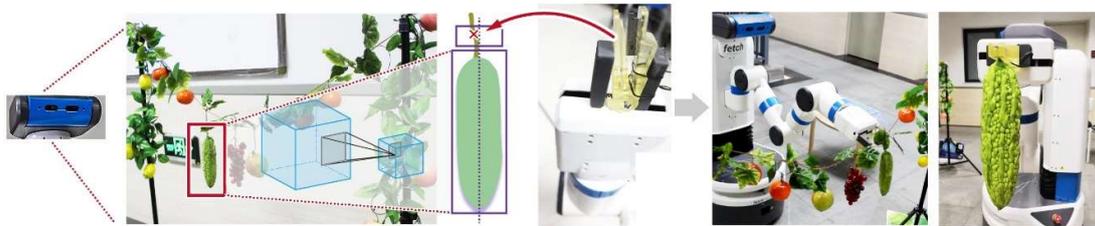

Figure 1. Architecture of the proposed harvesting robot.

3.2. The Gripper Mechanism

The proposed gripper is illustrated in Figure 2 from a 3D point of view. Inspired by skilled workers who use fingers to gently hold the peduncles and use their fingernails to cut the peduncles, this gripper is designed with both clamping and cutting functions with three key parts only: a top plate with a cutting blade, a middle plate to press the peduncle, and a bottom plate to hold the peduncle. Two torsion springs and an axle are included.
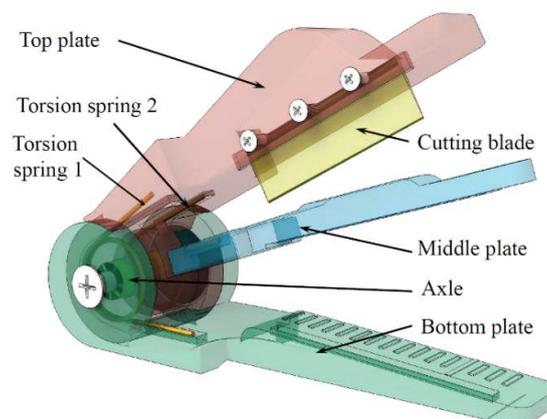

Figure 2. The 3D view of the proposed gripper [35,36].

The clamping and cutting processes work as follows:

1. Clamping the peduncle: pressing the top plate, the middle plate contacts the bottom plate based on "*torsion spring 2*", and thus the peduncle is clamped, as shown in Figure 3b.

2. Cutting the peduncle: continuing to press the top plate, the cutting blade contacts the peduncle on the bottom plate, thereby cutting the peduncle, as shown in Figure 3c.

After releasing the top plate, the top and middle plates are automatically restored to the rest position (Figure 3a) by "*torsion springs 1 and 2*", respectively.
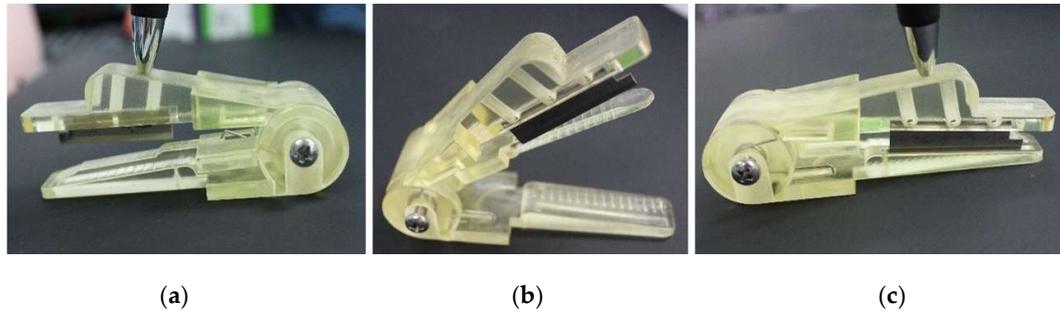

(a)　　　　　　　　　　　　**(b)**　　　　　　　　　　　　**(c)**

Figure 3. Cutting process: (**a**) rest position; (**b**) clamping state; (**c**) cutting state.

3.3. Force Analysis of the Gripper

As shown in Figure 4, when the cutting blade touches the bottom plate, the cutting force on the peduncle is calculated using the following equations:

$$M_{cut} = M_{push} - M_{spring\_1} - M_{spring\_2} \tag{1}$$

$$M_{spring\_1} = k_1 \cdot \theta_1 \tag{2}$$

$$M_{spring\_2} = k_2 \cdot \theta_2 \tag{3}$$

where M_{cut}, M_{push}, M_{spring_1}, M_{spring_2} are the moments of cutting force, pushing force, spring 1 and spring 2. k_1 and k_2 are the torsion coefficients of "torsion springs 1 and 2", respectively. θ_1 and θ_2 are the angles of deflection from the equilibrium position of "torsion springs 1 and 2", respectively.

"*Torsion spring 1*" is the spring between the top plate and bottom plate, while "*torsion spring 2*" is the spring between the top plate and middle plate; see also Figures 2 and 4d. With the moment M_{cut}, the cutting force F_{cut} can be described as:

$$M_{cut} = F_{cut} \cdot L_{cut} \tag{4}$$

where F_{cut} is the force applied on the peduncle from the cutting blade, L_{cut} is the length between the axis of the coil and the line of the cutting force, see Figure 4b. P is the contacting point between the cutting blade and the peduncle.

Thus, F_{cut} is derived from Equations (1)–(4) as follows:

$$F_{cut} = \frac{M_{push} - k_1 \cdot \theta_1 - k_2 \cdot \theta_2}{L_{cut}} \tag{5}$$

For further investigating the cutting force on peduncle, we analyzed the stress:

$$F_{cut} = \tau_{cut} \cdot A \tag{6}$$

$$\tau_{cut} = \frac{M_{push} - k_1 \cdot \theta_1 - k_2 \cdot \theta_2}{A \cdot L_{cut}} \tag{7}$$

where τ_{cut} is the stress, and A is the surface area of the peduncle. The two variables determine if the peduncle can be cut off. Equation (7) is derived from Equations (5) and (6).

As shown in Figure 5, when pressing the top plate and the middle plate touches the bottom plate, "*torsion spring 2*" creates the pressing force on the middle plate, and thus generates force on the bottom plate. This force ensures that the gripper clamps the peduncle before the cutting blade touches the peduncle:

$$M_{clamp} = M_{spring\_2} \tag{8}$$

$$M_{clamp} = F_{clamp} \cdot L_2 \tag{9}$$

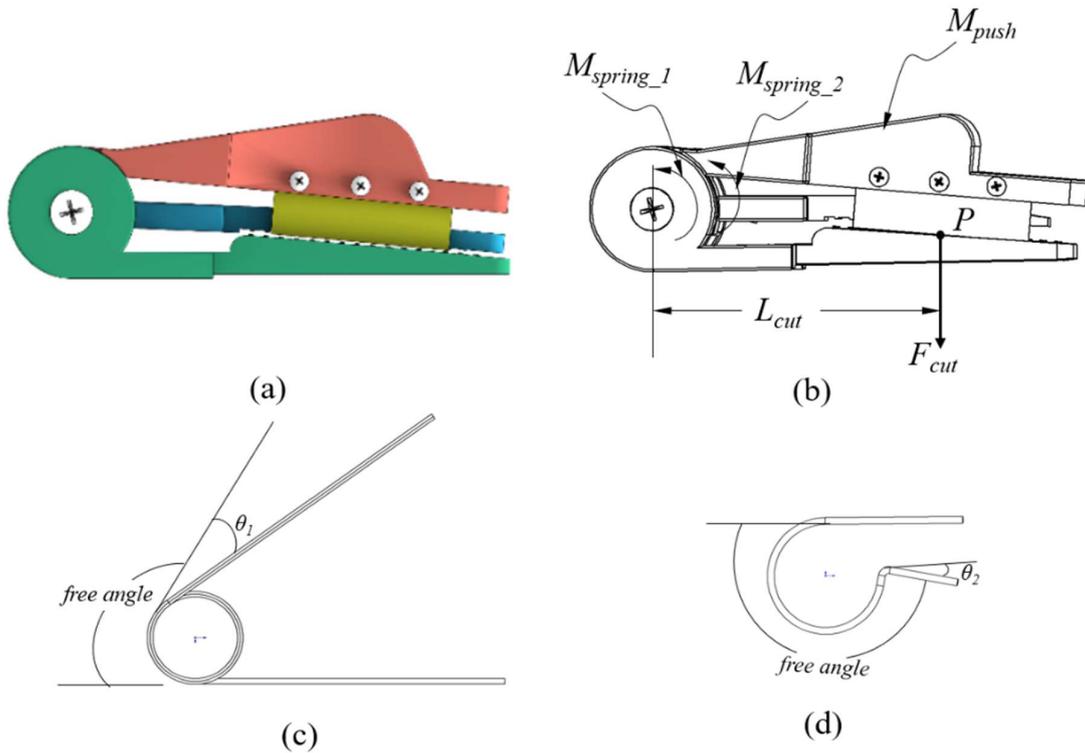$$M_{spring\_2} = k_2 \cdot \theta_2 \tag{10}$$

Figure 4. Force diagram of the cutting blade touching the bottom plate. (**a**) The diagram of the gripper. The force analysis of the gripper (**b**), *Torsion spring* 1 (**c**), and *Torsion spring* 2 (**d**).
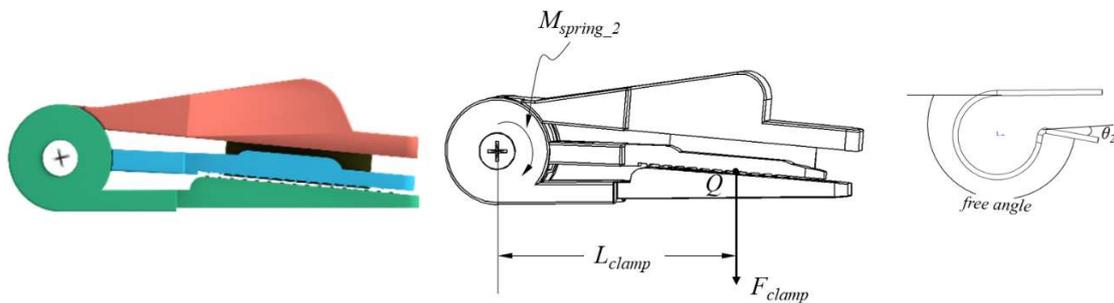

Figure 5. Force diagram of the middle plate touching the bottom plate.

After the peduncle is cut, the crop is detached from the peduncle if the clamping force is smaller than the gravity of the crop. To ensure that the crop is clamped, the minimum clamping force overcoming the gravity is expressed as:

$$F_{clamp} \geq \frac{m \cdot g}{\mu} \tag{11}$$

where *m* is the weight of the crop, *g* the gravity, and μ the coefficient of friction for a crop on the gripper plate surface. Based on the Equations (8)–(11), we obtain the minimum coefficient of "*torsion spring 2*":

$$k_2 \geq \frac{m \cdot g \cdot L_{clamp}}{\mu \cdot \theta_2} \tag{12}$$

4. Perception and Control

4.1. Cutting-Point Detection

The approach to detecting the cutting points of the peduncles consists of two steps: (i) the pixel area of crops is obtained using deep convolutional neural networks; (ii) the edge images of fruits are extracted, and then a geometric model is established to obtain the cutting points. Here, we obtain the external rectangle size of each pixel area to determine the region containing the peduncle for each crop. The architecture of the cutting-point-detection approach is shown in Figure 6.

Accurate object detection requires detecting not only which objects are in a scene, but also where they are located. Accurate region proposal algorithms thus play significant roles in the object-detection task. Mask R-CNN [10] makes use of edge information to generate region proposals by using the Region Proposal Network (RPN). Mask R-CNN uses color images to perform general object detection in two stages: region proposal, which proposes candidate object bounding boxes, and a region classifier that predicts the class and box offset with a binary mask for each region of interest (ROI). To train Mask R-CNN for our task, we perform fine-tuning [37], which requires labeled bounding-box information for each of the classes to be trained. After extracting the bounding box for each fruit, the geometric information of each region (e.g., external rectangle) is extracted by geometric morphological calculation. The cutting region of the interest of the targeted crop can thus be determined. A RGB-D camera like the Kinect provides RGB images along with per-pixel depth information in real time. With the depth sensor, the 2D position of the cutting point is mapped to the depth map, and thus the 3D information of the cutting point (x,y,z) is obtained.
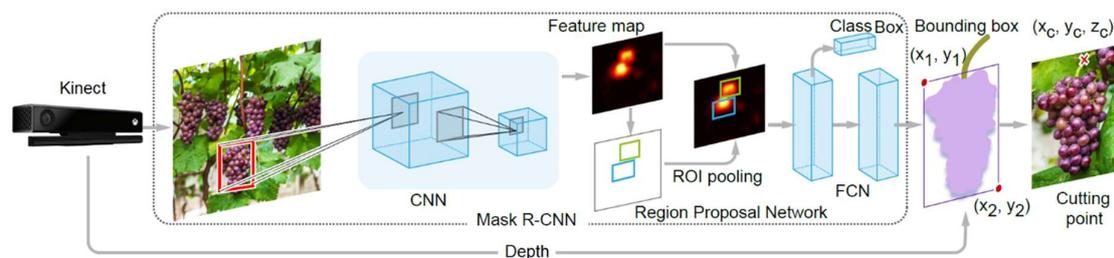
Figure 6. Illustration of the vision system that detects and localizes the crop and its cutting point.

The peduncle is usually located above the fruit. Thus, we set the ROI of the peduncle above the fruit, as shown in Figure 7.

The directions of the crops can be downward or tilted. We also adapted the minimum bounding rectangle to Mask R-CNN, thus to obtain the bounding boxes for crops in tilted direction. Other cases such as the direction of the crop parallel to the camera or the peduncle being "occluded" by its flesh will be discussed in the future work.

We then obtain the bounding box of the fruit, i.e., a top point coordinate (x_t, y_t), and a bottom point coordinate (x_b, y_b). The coordinates of the cutting point (x_c, y_c) can be determined as follows:

$$Roi_L = 0.5 \cdot L_{max} \tag{13}$$

$$Roi_H = 0.5 \cdot H_{max} \tag{14}$$

$$L_{max} = |x_b - x_t| \tag{15}$$

$$H_{max} = |y_b - y_t| \tag{16}$$

$$x_c = 0.5(x_b - x_t) \pm 0.5Roi_L \tag{17}$$
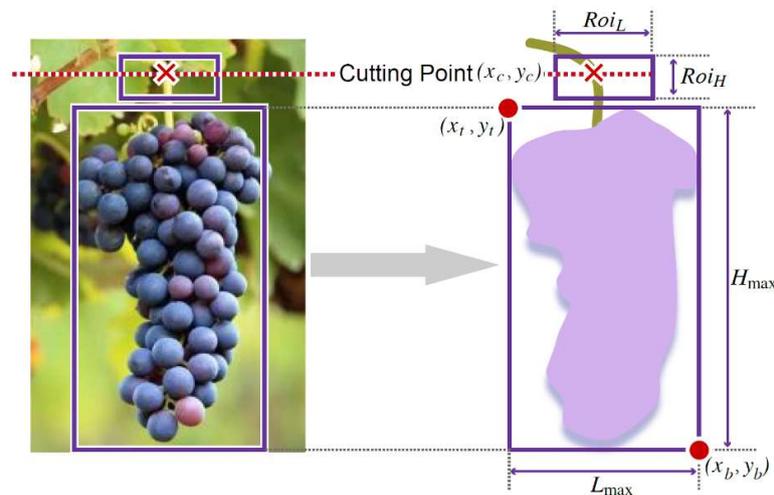
$$y_c = y_t - 0.5Roi_H \tag{18}$$

Figure 7. A schematic diagram of the cutting-point-detection approach.

4.2. System Control

A common method of motion planning for autonomous crop harvesting is open-loop planning [16,38]. Existing work on manipulation planning usually requires some combination of precise models, including manipulator kinematics, dynamics, interaction forces, and joint positions [39]. The reliance on manipulator models makes transferring solutions to a new robot configuration challenging and time-consuming. This is particularly true when a model is difficult to define, as in custom harvesting robots. In automated harvesting, a manipulator is usually customized to a special crop; thus, they are not as accurate as modern manipulators and it is relatively difficult to obtain accurate models of the robots. In addition, requiring high position sensing becomes challenging in manipulators given the space constraints.

In this paper, we demonstrate a learning-based kinematic control policy that maps target positions to joint actions directly, using convolutionally neural network (CNN) [40]. This approach makes no assumptions regarding the robot's configuration, kinematic, and dynamic model, and has no dependency on absolute position sensing available on-board. The inverse-kinematics function is learned using CNNs, which also considers self-collision avoidance, and errors caused by manufacturing and/or assembly tolerances.

As shown in Figure 8, the network comprises three convolutional layers and three fully connected layers. The input is the spatial position of the end-effector, and the output is the displacement of all of the joints. The fully connected layers will serve as a classifier on top of the features extracted from the convolutional layers and they will assign a probability for the joint motion being what the algorithm predicts it is.

We collected data by sending a random configuration request $q = (q_1, q_2, \ldots, q_n)$ to a robot with n joints and observing the resulting end-effector position $x = (x_1, x_2, \ldots, x_n)$ in 3D space. n is the number of joints. This results in a kinematic database with a certain number of trails:

$$X = (< q_1, x_1 >, < q_2, x_2 >, \ldots, < q_n, x_n >) \tag{19}$$

We use the neural network shown in Figure 8 to estimate the inverse kinematics from the collected data. Then we generate a valid configuration q_n for a given target position x_n. q_n and x_n are the n_{th} set of training data. We use this approach for planning to reach a cutting point of the peduncle.
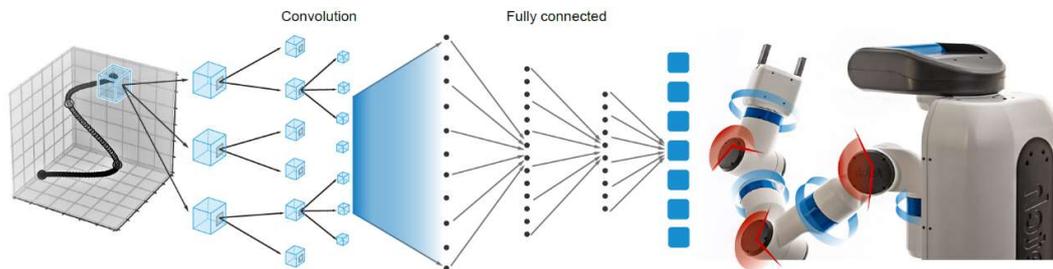

Figure 8. Network architecture for the kinematic control of a robotic manipulator.

5. Hardware Implementation

The following experiments aim at evaluating the general performance of the harvesting system. As shown in Figure 9, the proposed gripper was mounted on the end-effector of the Fetch robot (see Figure 9c). The Fetch robot consists of seven rotational joints, a gripper, a mobile base, and a depth camera. The depth camera was used to detect and localize the crops. The robot is based on the open source robot operating system, ROS [41]. The transformation between the image coordinate frame and the robot coordinate frame are obtained with ROS packages. The three plates and the shaft of the proposed gripper are created using Polylactic Acid (PLA) with a MakerBot Replicator 2 printing machine. The length, width, and height of modules are taken as 11, 5, and 9 cm, respectively.

Since the Fetch robot is not suitable for working in the field, the experiments were conducted in a laboratory environment. In order to verify the effectiveness of the proposed system in identifying and harvesting a variety of crops, we use both real and plastic ones, as shown in Figure 9a,b. Two experiments are presented in the following section aimed at validating the perception system and the harvesting system. First, we present the accuracy of the detection of crops and the cutting points. Second, we present the experiment of the full harvesting platform, demonstrating the harvesting performance of the final integrated system.
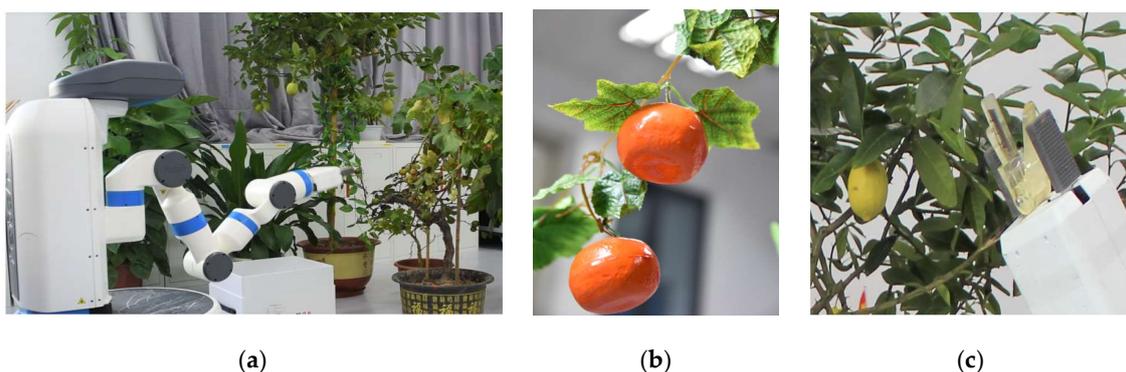

(**a**) (**b**) (**c**)

Figure 9. Setup for the experiments. (**a**) The experiment setup; (**b**) The plastic crops with plastic peduncles hung on a rope; (**c**) The proposed gripper mounted on the end-effector of the Fetch robot.

5.1. Detection Results

Seven types of crops were tested, including lemons, grapes, common figs, bitter melon, eggplants, apples, and oranges. For each type of crop, we used 60 images for training and noises were introduced in the images. The number of images was inspired by Sa et al. [34]. We collected images from Google and set labelled bounding box information for each crop. The images taken with the Kinect sensor

were used for testing only. In this paper, we used 15 testing images for each type of crop. These images are taken from different viewpoints. We have one plant for each type of the real fruits.

An example is shown in Figure 10 where the real crops were on the plants (see Figure 9a). For a convenient harvesting, all the plastic crops were hung on a rope, as shown in Figure 9b. We had three attempts and the crops were repositioned the crops every time. Therefore, we used 27 plastic crops (six apples, three lemons, six oranges, three bitter melons, three clusters of grapes, and six eggplants) and 31 real crops (seven grape bunches, 14 lemons, and 10 common figs).
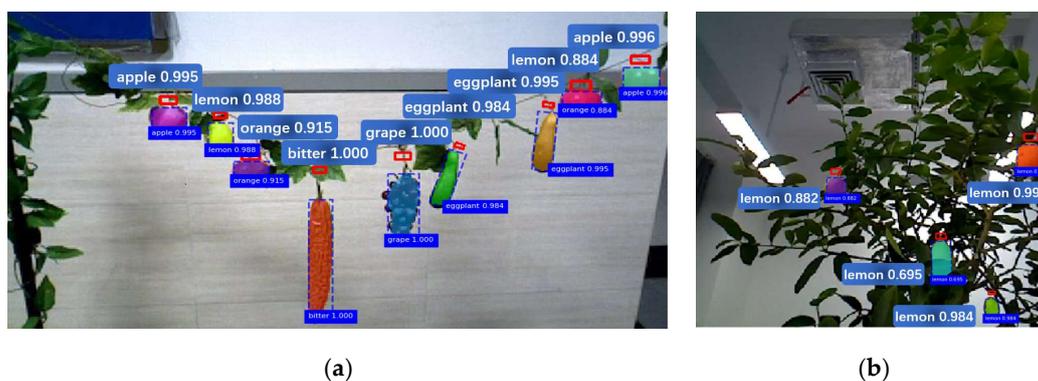

(**a**) (**b**)

Figure 10. Results of the cutting-point-detection for some plastic (**a**) and real crops (**b**).

We utilize the precision-curve [42] as the evaluation metric for the detection of cutting points of lemons. The performance of the proposed detection method and the baseline method Conditional Random Field (CRF) [43] was compared. Figure 11 shows that the two methods have similar performance. It can be seen that there is a gradual linear decrease in precision with a large increase in recall for the proposed approach.
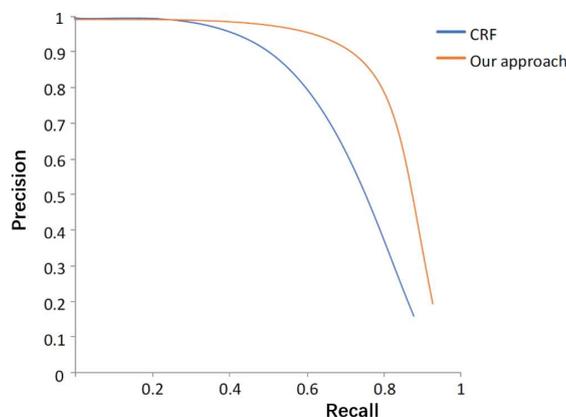

Figure 11. Precision-recall curve for detecting lemons using the CRF-based approach and the proposed approach.

Some of the detection results for real and plastic crops were shown in Figure 10. The results show that the detection success rates for plastic and real crops are 93% (25/27) and 87% (27/31), respectively. The detection success rates for the cutting points for plastic and real crops are 89% (24/27) and 71% (22/31), respectively. The result can be seen in Figure 12.
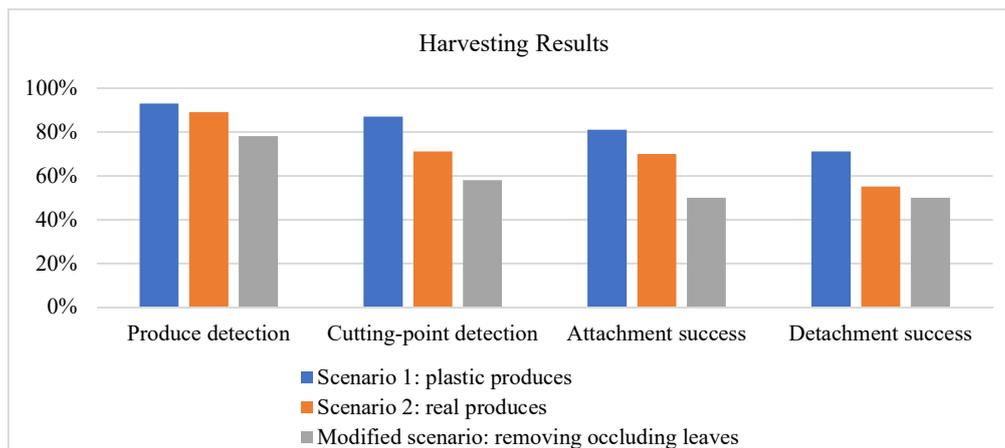
Figure 12. Harvesting rates for autonomous harvesting experiment.

5.2. Autonomous Harvesting

Motion planning is performed for autonomous attachment and detachment. The attachment trajectory starts at a fixed offset from the target fruit or vegetable determined by the bounding box. Then, the trajectory makes a linear movement towards the cutting point. The whole motion was computed using the control policy that maps target positions to joint actions introduced in Section 4.2. Self-collisions of the robot were included within the motion planning framework to ensure the robot safely planned. Once the attachment trajectory has been executed attaching the peduncle, the end-effector is moved horizontally to the robot from the cutting point.

A successful harvest consists of a successful detection of the cutting point and a successful clamping and cutting of the peduncle. The crop that was successfully picked was replaced with a new one hung on the rope. We repositioned the crop for repicking if it was not harvested successfully. For real crops, the robot harvests each crop once, including seven grape bunches, 14 lemons, and 10 common figs. Similar to plastic fruits, reaching the crop's cutting point is counted as a successful attachment, and clamping and cutting of the peduncle is counted as a successful detachment. Results show that most plastic fruits and vegetables were successfully harvested. The result is shown in Figure 12. The detailed success and failure cases can be seen in Table 1. The attachment success, i.e., the success rate of the robot's end-effector reaching the cutting point is 81% and 61% for plastic and real crops, respectively. The detachment success rates are 67% and 52%, respectively.

Table 1. Harvesting results for plastic and real crops.

Success Cases	Plastic Crops							Real Crops			
	Apple (6)	Lemon (3)	Orange (6)	Bitter melon (3)	Grapes (3)	Eggplants (6)	Total	Grapes (7)	Melons (14)	Common figs (10)	Total
Attachment	5	2	5	3	2	5	81%	5	11	6	71%
Detachment	5	2	4	1	2	5	70%	4	8	5	55%

Figure 13 is the video frames showing the robot reaching a cluster of plastic, cutting the peduncle, and placing them onto the desired place. Figure 14 is the video frames showing the robot harvesting real lemons, grapes and common figs in the laboratory environment. A video of the robotic harvester demonstrating the final experiment by performing autonomous harvesting of all crops is available at http://youtu.be/TLmMTR3Nbsw.

For cases where the crop's cutting point is occluded by leaves, the robot can remove the leaves to make the crop's cutting point visible. First, the overlap between crop's cutting area and the leaf is calculated to see if it exceeds a given threshold. Then, the robot detects the leaf's cutting point and cuts it using the same method as that for cutting crops. Figure 15 shows a modified scenario in which the robot removed leaves that occluded the fruits. We made four attempts and failed twice.

The figure shows that the leaf stuck onto the cutting blade and it did not detach even if the gripper is open.

Figure 13. Video frames showing the robot reaching a cluster of plastic grapes, cutting the peduncle, and placing them onto a table.

Figure 14. Video frames showing robot harvests real lemons (top), grapes (middle), and common figs (bottom).
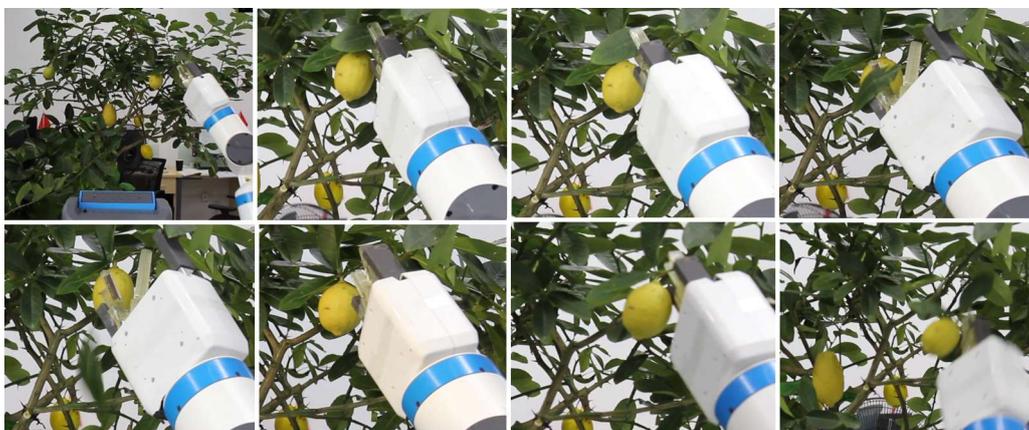

Figure 15. Video frames for the modified scenario in which the robot picked an occluded lemon (bottom) by removing an occluding leaf (top).

5.3. Discussion

A failure analysis was conducted to understand what were the major causes of harvesting failures within the system proposed, and thus to improve the system in the future work. The mode of failures includes: (1) fruits or vegetables not detected or detected inaccurately due to occlusions; (2) cutting point not detected due to the irregular shapes of the crops; (3) peduncle not attached due to motion planning failure; (4) peduncle partially cut since the cutting tool stuck into the peduncle; (5) crop dropped since the gripper does not clamp the crop tightly; (6) detachment fails since the crops or leaves stays on the blade when the gripper is open.

Figure 11 shows the success rates for three scenarios: Scenario 1 with plastic crops, Scenario 2 with real crops, and a modified scenario in which the robot picked crops by removing occlusions. The detection of the fruits and the cutting-points, attachment, and detachment harvest success rate were included. The detachment success rate reflects the overall harvesting performance.

The results showed that Scenario 1 with plastic crops has a higher harvest success. The most frequent failure mode was (2) cutting point not detected and (5), occurring 30% harvest failure. For instance, the two attachment failures (orange 1 and eggplant 1) occurred in the first nine crops. For orange 1, the cutting blade missed the peduncle as the peduncle is short and the cutting blade missed it. Eggplant 1 was irregularly shaped, which resulted in a poor cutting-pose estimation. We then adjusted the directions of the orange and eggplant for the following two rounds, and there was no detection error. The lower detachment success rate is most likely attributed to the grapes and bitter melon since they have a larger weight and fall.

The harvest success rate (55%) was lower for Scenario 2 with real crops. The most frequent failure modes were (4), and (5) which represented a significant portion of the failure cases and showed that attachment and detachment process is challenging due to the lower stiffness of the cutting mechanism, as well as the sophisticated environment. For instance, the common figs have shorter peduncles and the gripper is thick and wider, resulting in a detachment failure. The grape peduncle is thicker and difficult to be cut off.

The results also showed a lower harvest success rate (50%) for the modified scenario. The most failure modes were (2) and (6). For instance, the leaves of the lemon tree did not fall when the gripper is open since the leaves were thin and light. The cutting points of the leaves were difficult to be detected due to the random poses. Thus, the leaves may not be removed completely once. In the future work, we will improve the detection of the occluded crops and the leaves, as well as the robot grasping pose.

6. Conclusions and Future Work

In this paper, we developed a harvesting robot with a low-cost gripper and proposed a cutting-point detection method for autonomous harvesting. The performance of the proposed harvesting robot was evaluated using a physical robot in different scenarios. We demonstrated the effectiveness of the harvesting robot on different types of crops. System performance results showed that our system can detect the cutting point, cutting the fruit at the appropriate position of the peduncle. Most failures occurred in the attachment stage due to irregular shapes. This problem was avoided by adjusting the directions of crops. In the future, we plan to devise a new approach to detect the cutting pose for irregularly shaped crops.

There are some detachment issues related to the fabrication of the gripper. For instance, the crop did not detach from the peduncle when the gripper is open, while the clamped crop falls due to gravity. Thus, we are working to improve the gripper using metal materials with a better cutting blade, so the gripper would be sufficiently stiff to successfully clamp different types of crops, including one with a thicker peduncle.

In this paper, we mainly focused on the cases where the directions of the crops are downward or tilted. The hypothesis posed in this paper is that a robot capable of reaching targets or leaves without considering obstacles. Human approaches a fruit by pushing a leaf aside or going through the stems without causing any damage to the plant since a damage can seriously hamper plant growth

and affect future production of fruit. Inspired by this, we demonstrated how a robot removes leaves to reach the target crop. However, the proposed crop picking process by removing leaves is still in an ideal situation. In the future, we will improve the detection of obstacles, such as unripe crops, leaves and stems. For cases where the direction of a crop parallel to the camera or the peduncle being "occluded" by its flesh, the robot needs to rotate the camera to the next best view [44] to make the cut point of the crop visible. A method to determine the next best view with an optimal grasp [45] based on occlusion information will be considered in the future work.

For the case where stems are viewed as obstacles which are not avoided, we are currently working on developing a vision-guided motion planning method, using a deep reinforcement learning algorithm. This approach allows the robot to reach the desired fruits and avoid obstacles using current visual information. The aforementioned next-best-view-based approach will be combined with the motion planning. Furthermore, we aim to explore the estimation of the locations of crops in three dimensions and predict the occlusion order between objects [46].

As our current robot is inappropriate for use outside, we have not tested our harvesting robot in the field. Our next step is to test more types of crops in the field using a UR5 robot and, in addition, consider the autonomous reconstruction of unknown scenes [47].

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

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kapach, K.; Barnea, E.; Mairon, R.; Edan, Y.; Ben-Shahar, O. Computer vision for fruit harvesting robots-state of the art and challenges ahead. *Int. J. Comput. Vis. Robot.* **2012**, *3*, 4–34. [CrossRef]
2. Bac, C.W.; Hemming, J.; Tuijl, B.A.J.V.; Barth, R.; Wais, E.; Henten, E.J.V. Performance evaluation of a harvesting robot for sweet pepper. *J. Field Robot.* **2017**, *34*, 6. [CrossRef]
3. Bac, C.W.; van Henten, E.J.; Hemming, J.; Edan, Y. Harvesting robots for high-value crops: State-of-the-art review and challenges ahead. *J. Field Robot.* **2014**, *31*, 888–911. [CrossRef]
4. Kamilaris, A.; Prenafeta-Boldu, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
5. Arad, B.; Kurtser, P.; Barnea, E.; Harel, B.; Edan, Y.; Ben-Shahar, O. Controlled lighting and illumination-independent target detection for real-time cost-efficient applications the case study of sweet pepper robotic harvesting. *Sensors* **2019**, *19*, 1390. [CrossRef] [PubMed]
6. Sa, I.; Lehnert, C.; English, A.; McCool, C.; Dayoub, F.; Upcroft, B.; Perez, T. Peduncle detection of sweet pepper for autonomous crop harvesting—Combined color and 3-d information. *IEEE Robot. Autom. Lett.* **2017**, *2*, 765–772. [CrossRef]
7. Luo, L.; Tang, Y.; Lu, Q.; Chen, X.; Zhang, P.; Zou, X. A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard. *Comput. Ind.* **2018**, *99*, 130–139. [CrossRef]
8. Eizicovits, D.; van Tuijl, B.; Berman, S.; Edan, Y. Integration of perception capabilities in gripper design using graspability maps. *Biosyst. Eng.* **2016**, *146*, 98–113. [CrossRef]
9. Eizicovits, D.; Berman, S. Efficient sensory-grounded grasp pose quality mapping for gripper design and online grasp planning. *Robot. Auton. Syst.* **2014**, *62*, 1208–1219. [CrossRef]
10. He, K.; Gkioxari, G.; Dolla'r, P.; Girshick, R. Mask r-cnn. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, 2961–2969.

11. van Henten, E.J.; Slot, D.A.V.; Cwj, H.; van Willigenburg, L.G. Optimal manipulator design for a cucumber harvesting robot. *Comput. Electron. Agric.* **2009**, *65*, 247–257. [CrossRef]

12. Chiu, Y.C.; Yang, P.Y.; Chen, S. Development of the end-effector of a picking robot for greenhouse-grown tomatoes. *Appl. Eng. Agric.* **2013**, *29*, 1001–1009.

13. Lehnert, C.; English, A.; McCool, C.S.; Tow, A.M.W.; Perez, T. Autonomous sweet pepper harvesting for protected cropping systems. *IEEE Robot. Autom. Lett.* **2017**, *2*, 872–879. [CrossRef]

14. Yamamoto, S.; Hayashi, S.; Yoshida, H.; Kobayashi, K. Development of a stationary robotic strawberry harvester with picking mechanism that approaches target fruit from below (part 2)—Construction of the machine's optical system. *J. Jpn. Soc. Agric. Mach.* **2010**, *72*, 133–142.

15. Hayashi, S.; Shigematsu, K.; Yamamoto, S.; Kobayashi, K.; Kohno, Y.; Kamata, J.; Kurita, M. "Evaluation of a strawberry-harvesting robot in a field test. *Biosyst. Eng.* **2010**, *105*, 160–171. [CrossRef]

16. Hemming, J.; Bac, C.; van Tuijl, B.; Barth, R.; Bontsema, J.; Pekkeriet, E.; van Henten, E. A robot for harvesting sweet-pepper in greenhouses. In Proceedings of the International Conference of Agricultural Engineering, Zurich, Switzerland, 6–10 July 2014.

17. Hayashi, S.; Takahashi, K.; Yamamoto, S.; Saito, S.; Komeda, T. Gentle handling of strawberries using a suction device. *Biosyst. Eng.* **2011**, *109*, 348–356. [CrossRef]

18. Dimeas, F.; Sako, D.V. Design and fuzzy control of a robotic gripper for efficient strawberry harvesting. *Robotica* **2015**, *33*, 1085–1098. [CrossRef]

19. Bontsema, J.; Hemming, J.; Pekkeriet, E. Crops: High tech agricultural robots. In Proceedings of the International Conference of Agricultural Engineering, Zurich, Switzerland, 6–10 July 2014.

20. Baeten, J.; Donn'e, K.; Boedrij, S.; Beckers, W.; Claesen, E. Autonomous fruit picking machine: A robotic apple harvester. *Springer Tracts Adv. Robot.* **2007**, *42*, 531–539.

21. Haibin, Y.; Cheng, K.; Junfeng, L.; Guilin, Y. Modeling of grasping force for a soft robotic gripper with variable stiffness. *Mech. Mach. Theory* **2018**, *128*, 254–274. [CrossRef]

22. Mantriota, G. Theoretical model of the grasp with vacuum gripper. *Mech. Mach. Theory* **2007**, *42*, 2–17. [CrossRef]

23. Han, K.-S.; Kim, S.-C.; Lee, Y.B.; Kim, S.C.; Im, D.H.; Choi, H.K.; Hwang, H. Strawberry harvesting robot for bench-type cultivation. *J. Biosyst. Eng.* **2012**, *37*, 65–74. [CrossRef]

24. Tian, G.; Zhou, J.; Gu, B. Slipping detection and control in gripping fruits and vegetables for agricultural robot. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 45–51. [CrossRef]

25. Zhang, T.; Zhang, W.; Gupta, M.M. An underactuated self-reconfigurable robot and the reconfiguration evolution. *Mech. Mach. Theory* **2018**, *124*, 248–258. [CrossRef]

26. Zhang, T.; Zhang, W.; Gupta, M. A novel docking system for modular self-reconfigurable robots. *Robotics* **2017**, *6*, 25. [CrossRef]

27. Bac, C.; Hemming, J.; van Henten, E. Robust pixel-based classification of obstacles for robotic harvesting of sweet-pepper. *Comput. Electron. Agric.* **2013**, *96*, 148–162. [CrossRef]

28. McCool, C.; Sa, I.; Dayoub, F.; Lehnert, C.; Perez, T.; Upcroft, B. Visual detection of occluded crop: For automated harvesting. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 16–21 May 2017; pp. 2506–2512.

29. Blasco, J.; Aleixos, N.; Molto, E. Machine vision system for automatic quality grading of fruit. *Biosyst. Eng.* **2003**, *85*, 415–423. [CrossRef]

30. Ruiz, L.A.; Molt'o, E.; Juste, F.; Pla, F.; Valiente, R. Location and characterization of the stem–calyx area on oranges by computer vision. *J. Agric. Eng. Res.* **1996**, *64*, 165–172. [CrossRef]

31. Zemmour, E.; Kurtser, P.; Edan, Y. Dynamic thresholding algorithm for robotic apple detection. In Proceedings of the 2017 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), Coimbra, Portugal, 26–28 April 2017; pp. 240–246.

32. Zemmour, E.; Kurtser, P.; Edan, Y. Automatic parameter tuning for adaptive thresholding in fruit detection. *Sensors* **2019**, *19*, 2130. [CrossRef]

33. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, UK, 2015; pp. 91–99.

34. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. Deep-fruits: A fruit detection system using deep neural networks. *Sensors* **2016**, *16*, 1222. [CrossRef]

35. Zhang, T.; You, W.; Huang, Z.; Lin, J.; Huang, H. A Universal Harvesting Gripper. Utility Patent. Chinese Patent CN209435819, 27 September 2019. Available online: http://www.soopat.com/Patent/201821836172 (accessed on 27 September 2019).

36. Zhang, T.; You, W.; Huang, Z.; Lin, J.; Huang, H. A Harvesting Gripper for Fruits and Vegetables. Design Patent, Chinese Patent CN305102083, 9 April 2019. Available online: http://www.soopat.com/Patent/201830631197 (accessed on 9 April 2019).

37. Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; Darrell, T. Decaf: A deep convolutional activation feature for generic visual recognition. In Proceedings of the International Conference on Machine Learning, Berkeley, CA, USA, 21-26 June 2014; pp. 647–655.

38. Baur, J.; Schu¨tz, C.; Pfaff, J.; Buschmann, T.; Ulbrich, H. Path planning for a fruit picking manipulator. In Proceedings of the International Conference of Agricultural Engineering, Garching, Germany, 6–10 June 2014.

39. Katyal, K.D.; Staley, E.W.; Johannes, M.S.; Wang, I.-J.; Reiter, A.; Burlina, P. In-hand robotic manipulation via deep reinforcement learning. In Proceedings of the 30th Conference on Neural Information Processing Systems, Barcelona, Spain, 5–8 December 2016.

40. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, UK; Lake Tahoe, NV, USA, 2012; pp. 1097–1105.

41. Quigley, M.; Conley, K.; Gerkey, B.; Faust, J.; Foote, T.; Leibs, J.; Wheeler, R.; Ng, A.Y. Ros: An open-source robot operating system. In Proceedings of the ICRA Workshop on Open Source Software, Kobe, Japan, 12–17 May 2009; pp. 1–6.

42. Davis, J.; Goadrich, M. The relationship between precision-recall and roc curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240.

43. Tseng, H.; Chang, P.; Andrew, G.; Jurafsky, D.; Manning, C. A conditional random field word segmenter for sighan bakeoff 2005. In Proceedings of the fourth SIGHAN Workshop on Chinese language Processing, Jeju Island, Korea, 14–15 October 2005; pp. 168–171.

44. Wu, C.; Zeng, R.; Pan, J.; Wang, C.C.; Liu, Y.-J. Plant phenotyping by deep-learning-based planner for multi-robots. *IEEE Robot. Autom. Lett.* **2019**, *4*, 3113–3120. [CrossRef]

45. Barth, R.; Hemming, J.; van Henten, E.J. Angle estimation between plant parts for grasp optimisation in harvest robots. *Biosyst. Eng.* **2019**, *183*, 26–46. [CrossRef]

46. Lin, D.; Chen, G.; Cohen-Or, D.; Heng, P.-A.; Huang, H. Cascaded feature network for semantic segmentation of rgb-d images. In Proceedings of the IEEE International Conference on Computer Vision, Honolulu, HI, USA, 21–26 July 2017; pp. 1311–1319.

47. Xu, K.; Zheng, L.; Yan, Z.; Yan, G.; Zhang, E.; Niessner, M.; Deussen, O.; Cohen-Or, D.; Huang, H. Autonomous reconstruction of unknown indoor scenes guided by time-varying tensor fields. *ACM Trans. Graph.* **2017**, *36*, 1–15. [CrossRef]