# A Comprehensive Review of the Research of the "Eye–Brain–Hand" Harvesting System in Smart Agriculture

Wanteng Ji [1,2,3,†], Xianhao Huang [1,2,3,†] ⓘ, Shubo Wang [1,2,4,*] ⓘ and Xiongkui He [1,2,4,*] ⓘ

1   Centre for Chemicals Application Technology, China Agricultural University, Beijing 100193, China;
    jwt17863978031@163.com (W.J.); xianhao@cau.edu.cn (X.H.)
2   College of Agricultural Unmanned System, China Agricultural University, Beijing 100193, China
3   College of Engineering, China Agricultural University, Beijing100083, China
4   College of Science, China Agricultural University, Beijing 100193, China
*   Correspondence: shubowang@qdu.edu.cn (S.W.); xiongkui@cau.edu.cn (X.H.)
†   These authors contributed equally to this work.

**Abstract:** Smart agricultural harvesting robots' vision recognition, control decision, and mechanical hand modules all resemble the human eye, brain, and hand, respectively.  To enable automatic and precise picking of target fruits and vegetables, the system makes use of cutting-edge sensor technology, machine vision algorithms, and intelligent control and decision methods.  This paper provides a comprehensive review of international research advancements in the "eye–brain–hand" harvesting systems within the context of smart agriculture, encompassing aspects of mechanical hand devices, visual recognition systems, and intelligent decision systems.  Then, the key technologies used in the current research are reviewed, including image processing, object detection and tracking, machine learning, deep learning, etc. In addition, this paper explores the application of the system to different crops and environmental conditions and analyzes its advantages and challenges. Finally, the challenges and prospects for the research on picking robots in the future are presented, including further optimization of the algorithm and improvement of flexibility and reliability of mechanical devices. To sum up, the "eye–brain–hand" picking system in intelligent agriculture has great potential to improve the efficiency and quality of crop picking and reduce labor pressure, and it is expected to be widely used in agricultural production.

**Keywords:** smart agriculture; visual recognition; decision control; end-effector; harvesting robots; research review

## 1. Introduction

In agricultural production, harvest is one of the most important links.  It directly relates to the quality of product harvest.  However, traditional manual picking methods have problems such as high cost, low efficiency, and labor shortage, which seriously restrict the development of agricultural production. Therefore, it is urgent and necessary to study the technology of fruit and vegetable-picking robots, which is an important way to solve this problem. The visual recognition system, decision control system, and end-effector system are the key contents of the research on harvesting robot technology, and their technological maturity directly affects the harvesting effect of the robot. Intelligent harvesting system refers to a system that utilizes modern technology to achieve autonomous harvesting by the steps of recognition, decision, control, and grasping during the agricultural harvesting process. The intelligent picking system has the characteristics of efficiency, precision, and reliability, which can greatly improve agricultural production efficiency, reduce the labor burden of fruit farmers, and improve the quality and safety of agricultural production [1–8]. The literature distribution of intelligent harvesting systems for different crops is shown in Figure 1.
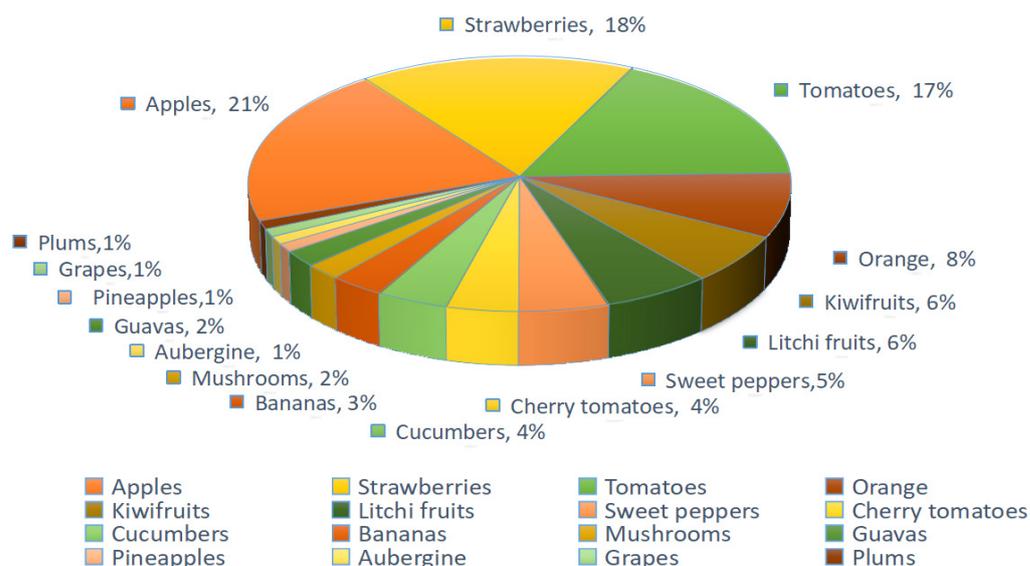
**Figure 1.** Distribution of the referenced literature per crop type.

The unmanned agricultural harvesting system mainly consists of three parts: a recognition module; a decision control module; and an end-effector module [9–11]. Therefore, this review will start from the recognition, decision control, and end-effector gripper of unmanned agricultural harvesting systems and analyze the current situation of intelligent harvesting systems. Among them, the recognition module is a key part of the system to obtain crop information, which mainly uses technologies such as machine vision and lidar. By using RGB [12,13], depth cameras, various sensors, or other devices, information about picked objects is obtained based on target features, feature fusion, or deep learning, including position, shape, size, maturity, surrounding environment, etc. [14–20]. To enhance the precision and accuracy of image segmentation, various image preprocessing techniques have been utilized, facilitating superior decision and control in subsequent stages [21–23]. The decision and control module serves as a crucial component in autonomous decision and control during the harvesting process. During the decision and control process, a broad range of network models and optimization algorithms [24–26], including Support Vector Machine (SVM) [27,28], decision trees, deep learning, genetic algorithms, and particle swarm optimization, are extensively employed [9,29–33]. These models and algorithms provide intelligent decision control adapted to different harvesting situations, leading to optimal harvesting outcomes. The end-effector module, a key hardware component responsible for automated fruit picking, influences the harvesting results directly through its design type, harvesting method, and size dimensions. Researchers categorize the end-effector grippers into four major types: negative pressure adsorption end-effectors; shearing-style end-effectors; negative pressure adsorption end-effectors; cavity retrieval end-effectors; and flexible grasping end-effectors. [10,34–45]. The selection of an appropriate end-effector gripper is determined by the physical characteristics of the fruits, such as their types, sizes, and the hardness of their peels. These three modules, encompassing vision, manipulation, and decision capabilities, work synergistically within the integrated eye–brain–hand system to accomplish the harvesting task.

In the context of the continuous advancement of artificial intelligence and robotics, unmanned systems for agricultural harvesting are gaining increasing attention as a novel method of crop collection. Within these agricultural harvesting robots, the perception system, decision-control system, and end-effectors play pivotal roles, serving as the essential components for achieving automated harvesting. Harvesting robots equipped with eye–brain–hand-integrated systems are not only characterized by high precision, efficiency, and reliability, but they are also adaptable to various harvesting environments, demonstrating a broad range of prospective applications.

Studying agricultural harvesting robots is not only an important measure to adapt to agricultural modernization and market demand but also a crucial step in solving labor shortages and improving the quality of agricultural products. Through continuous innovation and improvement of technology, agricultural production can be automated, refined, and intelligent, making positive contributions to the sustainable development of global agriculture. This article summarizes the important research progress of over 120 agricultural harvesting robots in the past 6 years. This research focuses on the "eye–brain–hand" of robots and comprehensively analyzes the key role of robots in the agricultural product-picking process. This article delves into various technological breakthroughs, including high-precision visual perception technology, intelligent path-planning and decision-making methods, and advanced end-effector design. By summarizing the research results of these aspects, this review provides insight into the field of agricultural picking robots and provides important guidance and enlightenment for future research and application. These contributions will have a profound impact on promoting the improvement of agricultural production efficiency, solving labor problems, and ensuring the quality of agricultural products. This article mainly discusses the three main parts of agricultural harvesting robots, namely "eye, brain, and hand", based on the "structure–activity relationship method". In these three parts, more detailed descriptions are provided according to the "time sequence method" and the "current situation countermeasure method". In the fifth part, the challenges and prospects in the field of harvesting robots are summarized according to the "current situation countermeasure method". The outline of this article is shown in Figure 2. In this paper, Section 2 introduces common target perception hardware systems, perception methods, and image preprocessing techniques. Hardware systems can be categorized into active vision, passive vision, and applications combining various sensors. The perception methods are primarily based on three aspects: target features; feature fusion; and deep learning. Section 3 primarily elaborates on decision strategies and control methods, encompassing regional division and task allocation, active and passive obstacle avoidance strategies, path planning based on various technologies, and numerous control methods. Section 4 presents the various end-effector mechanisms and evaluation metrics for harvesting robots. Section 5 presents challenges and prospects for agricultural harvesting robots. Finally, a summary of this article is provided in Section 6.
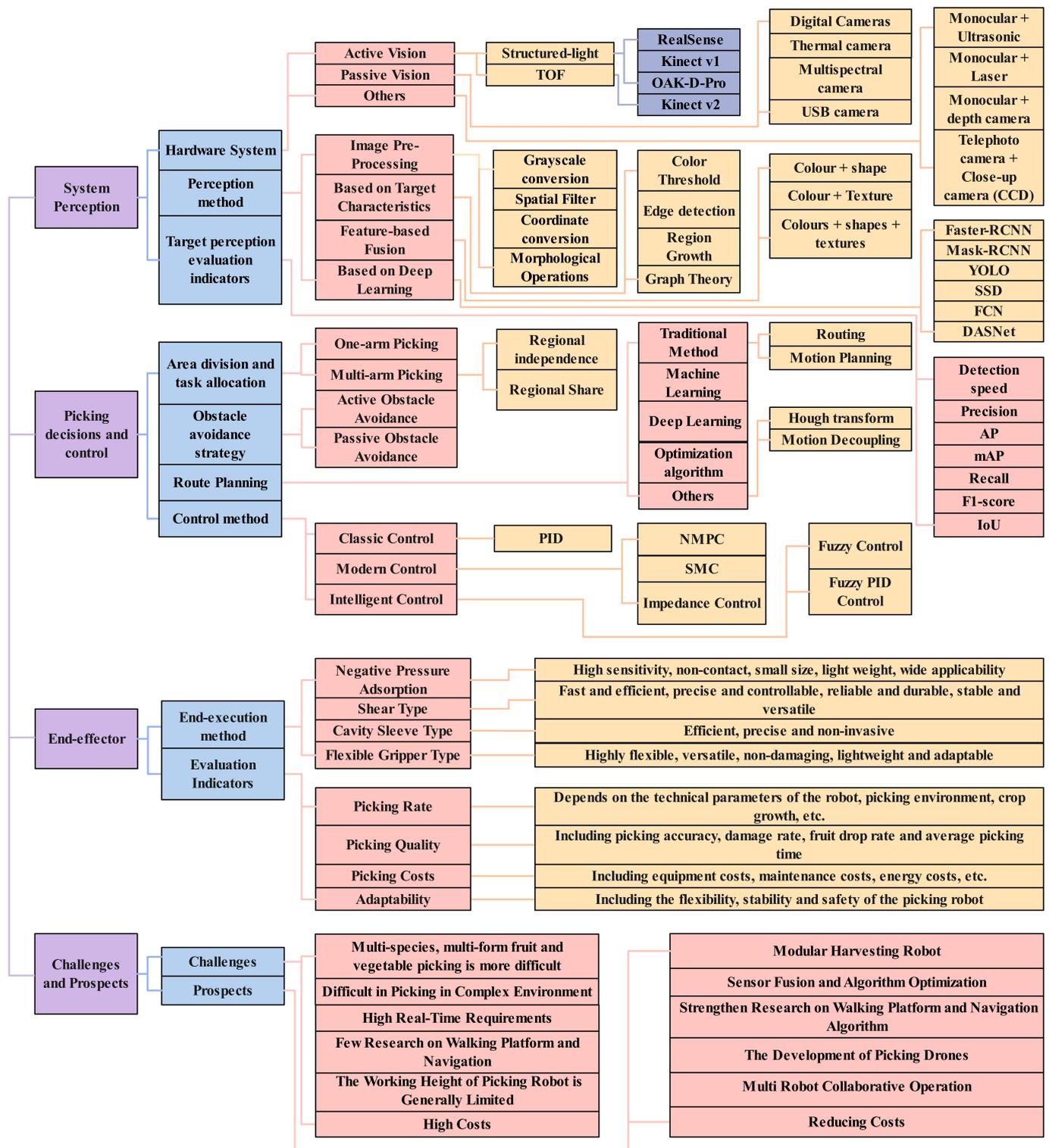
**Figure 2.** Outline of the article.

## 2. Intelligent Harvesting "Eye" System

### 2.1. Perception Hardware System

Considering the intelligent point of view of unmanned system, it is necessary to see accurately in order to achieve better expected results. Therefore, the system perception as the first part of the picking system has been widely paid attention to both at home and abroad. At present, the perception methods of agricultural picking robots mainly include

binocular vision, LiDAR, and the combination of monocular cameras and other sensors. Due to cost constraints, the LiDAR-based perception method is less applied, and the other two perception methods are mainly described here.

### 2.1.1. Object Perception Based on Binocular Vision

Binocular vision measurement is similar to the stereo perception of the human eye; it uses two cameras to image the object to be measured from different angles, based on the stereoscopic parallax of the corresponding points in the two images, combined with the principle of triangulation to realize the acquisition of 3D information of the object. Binocular vision measurement techniques can be divided into two types based on whether the light source is actively projected during the detection process, active vision and passive vision, both of which are often used in hand–eye harvesting. The classification, characteristics, and examples of active and passive visual cameras are shown in Table 1.

(a) Active vision

At present, the active vision technology based on Structured light is popular in the market for binocular cameras, which has the advantages of strong anti-interference ability, a wide application environment, and a mature technical scheme. Therefore, active visual cameras—especially those from the RealSense series from Intel, the Kinect series from Microsoft, and the OAK-D-Pro from OpenAI—are more frequently employed in the actual harvesting of fruits and vegetables. The guava harvesting robot designed by Lin et al. [46] used a Kinect V2 camera as a visual sensor, which consists of an RGB color camera, an infrared camera, and an IR light source. The IR light source actively projects the near-infrared spectrum, which forms random reflection spots when illuminated on an object. These spots will be received by infrared cameras to read the depth information of the object. Based on this Time of Flight (TOF) principle, it obtains depth images of targets between 0.5 and 4.5 m. In addition, the camera is inexpensive and stable. Due to its many advantages, the Kinect-V2 depth camera was also equipped by Ning et al. [47] on a pepper-picking robot to realize pepper picking in a greenhouse planting environment. In order to determine the location of the kiwifruit in an RGB image, extract coordinates, and locate it, Mu et al. [48] also used this camera as their machine vision equipment in the construction of a kiwifruit harvesting robot. Different from the above research, Zhang et al. [49] equipped the cherry tomato picking robot with Intel's RealSense D415 depth camera and installed it on the side of the base of the mechanical arm to ensure that it is not blocked during tomato harvesting. Similarly, Yu et al. [50] used the DF810-HD depth camera to provide visual support to the picking robot during tomato picking. Considering the height of the greenhouse and the growth of tomato plants, and in order to meet the demand for real-time control, the camera is placed on a liftable platform and fixed, which effectively increases the sensing range while reducing the calculation amount of the visual system. For the harvesting of short, densely planted apples in a greenhouse environment, Li et al. [51] chose to combine a RealSense D455 depth camera with a multi-Cartesian mechanical arm and used a task planning algorithm to achieve accurate identification and effective harvesting of short, densely planted apples. Similarly, Kang et al. [52] used the RealSense D-435 depth camera as the hardware support of the visual system of the apple-picking robot and applied the depth neural network DAS-Net to identify the fruit. The experiment showed that the visual system has high precision in fruit detection and segmentation. However, the setting of a single camera is difficult to adapt to the complex and changeable working environment. In order to adapt to an unstructured harvesting environment, Sarabu et al. [53] designed a double-arm harvesting robot consisting of a grab arm and a search arm for the same apple harvesting task, and each arm was equipped with an RGB-D depth camera (Hand-Eye). The camera on the grab arm is used to locate the picking apple in the field of view. The camera on the search arm can detect the target outside the dead zone of the grab arm camera and quickly plan a clear and suitable picking path in combination with relevant algorithms. Moreover, multi-camera and multi-view method can solve the problem of overlapping and blocking fruit detection to a certain extent. Gong et al. [54] provided us with a new

idea when designing a greenhouse tomato-picking robot, which is to improve the image segmentation accuracy in a complex environment through multi-source fusion images. RGB, depth, and infrared images are acquired by the Kinect v2 camera, fused to obtain RGB-D-I images, and target segmentation accuracy is improved by 7.6% in combination with the extended Mask-RCNN network. Although active vision cameras can be used in a wide range of scenarios, they are, in some cases, influenced by environmental factors. To achieve better sensing in the night-time environment, Fu et al. [55] equipped four 850 Lumens LED lights on their apple-picking robot platform, and the Kinect V2 camera was installed in the center of the four LED lights, which provided a bright and stable night-time working environment for the picking platform. Although the presence of an active light source can enhance the depth camera's perception in low-light conditions, it remains challenging to maintain optimum and consistent perception in situations with strong light or varying light intensity. To solve this problem, Xiong [56] applied a U-shaped straw-picking robot to a structured greenhouse-picking environment, with two independent picking systems on both sides of the arched structure and a real D435 depth camera for the visual system. The design of a U-shaped full shielding frame structure can greatly reduce or even eliminate the impact of ambient light changes on the quality of acquired images and greatly improve the detection and positioning accuracy of the visual system without any specific correction algorithms.

(b) Passive Vision

Passive vision technology uses a pure RGB camera and binocular parallax principle to detect and locate the target. Compared with active vision, passive vision is widely used in scenes with bright vision and wide space due to its simple structure and low power consumption. Although color is the most intuitive feature to distinguish the target fruit from the background, color-based methods of identification are susceptible to factors such as varying light. To this end, we can reduce the impact of light by means of relevant algorithms. Lv et al. [57] used Sony's Cyber-shot color camera to provide RGB images for their apple-picking robot and then corrected the images of apples affected by external light with an adaptive Gamma algorithm, thus greatly improving the image segmentation accuracy. In addition, the shape-based recognition method is not easily affected by changing lighting. Wang et al. [58] reduced the minimum relative error of Apple, ranging up to 0.96%, by combining a linear fusion detection algorithm and AD-Census matching algorithm based on CMOS binocular camera. However, the apple-picking robot designed by Yu et al. [59] is based on a binocular camera combined with color threshold and edge detection for target identification, and the success rate of apple identification is up to 82.5%. Similarly, Yang et al. [44] also used an RGB camera to identify the color and texture of the target in the Hangzhou chrysanthemum-picking robot. After eliminating the noise with the bilateral filter, the color and texture characteristics of the image were extracted through the RGB value and the grayscale paragenesis matrix of the image and then input into the Least Squares Support Vector Machine model (LS-SVM). The segmentation time of the trained model for the Hangzhou chrysanthemum was as low as 0.7 s. In addition, during tomato fruit picking, Zhou et al. [60] used a variable baseline USB binocular camera (HNY-CV-002), combined with the identification method of circular Hough transform and RGB color space, to achieve efficient picking of the target tomato. In addition, Jin et al. [61] applied depth learning technology to the binocular camera to identify the target fruit in order to better realize the perception of tomato fruit and achieved good results. Similar techniques, such as those discussed above for active vision, are also used for passive binocular vision in order to optimize the recognition impact. Ye et al. [62] installed the Micro-vision MV-VD120SC industrial camera on the end-effector of the litchi-picking robot and planned auxiliary target pickup points for the robot. After the end-effector arrived at the auxiliary target pickup point from the initial point, it would perform environmental perception again and plan the motion between the target pickup points. This strategy can avoid interference between the end-effector and obstacles around the target as much as possible while compensating for visual errors and improving positioning accuracy.

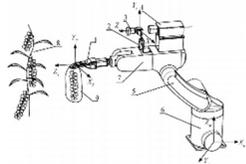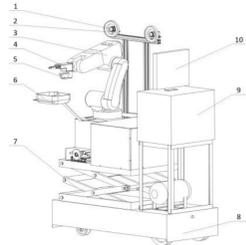**Table 1.** Characteristics of Active and Passive Vision Technology and Representative Camera Examples.

| Active/Passive Vision | Type | Advantages | Disadvantages | Representative Cameras | |
|---|---|---|---|---|---|
| Active Vision | Structured-light | More mature and easier to miniaturize<br>Low power consumption<br>Can be used at night<br>High accuracy and resolution within a certain range | Easily disturbed by ambient light<br>The accuracy deteriorates as the detection distance increases | RealSenseD435i |  |
| | | | | Kinect v1 |  |
| | | | | OAK-D-Pro |  |
| | TOF | Long detection distance<br>Less interference by ambient light | High requirements for equipment<br>High power consumption<br>Low edge accuracy<br>Lower frame rate and resolution | Kinect v2 |  |
| | | | | PMD CamCube 3.0 |  |
| Passive Vision | - | Low hardware requirements and low cost<br>Suitable for both indoor and outdoor use | Very sensitive to ambient light<br>Not for monotonous scenes that lack texture<br>Calculations are more complex | Digital Cameras |  |
| | | | | Thermal camera |  |
| | | | | Multispectral camera |  |

2.1.2. Target Perception Based on Multi-Sensor Combination

Due to the low accuracy and poor fault tolerance of a single sensor, the perception strategy based on multi-source sensor fusion is widely used in the fields of autonomous driving and industrial robots. Similarly, this technology has also been borrowed by agricultural robot experts for fruit's 3D perception. At present, there are several combinations of multi-source sensors used for fruit perception: monocular + ultrasound, monocular + laser, and monocular + depth camera. The relevant descriptions of different types of multi-sensor combination perception are shown in Table 2.

Oktarina et al. [63] used a combination of a Pi camera placed on a robotic arm and an ultrasonic sensor HC-SR04 to achieve recognition and positioning of red and green tomatoes. Lower-cost network cameras with appropriate resolution offer color image information of the target, and ultrasonic sensors offer depth information of the target fruits. A new visual unit was created by Feng et al. [64] for his cherry tomato-picking robot. A monocular camera and a laser sensor are both mounted on the manipulator arm. The target tomato is sensed and recognized using an RGB camera, and its distance from the vision system is calculated using a laser sensor. Using a suitable combination of the two sensors, corresponding algorithms, and a shear end actuator, harvesting success can be increased by 83%. In contrast to the examples above, Sepulveda et al. [65] created a dual-arm eggplant harvesting robot with a visual system made up of two monocular cameras, specifically, the SR4000 depth measurement capture camera and the Prosilica GC2450C color camera. The former provides high-resolution color images, while the latter provides depth information of images. In addition, in order to identify and locate strawberries more accurately, Feng et al. [66] creatively developed a vision system combining far and near. After the far-sighted unit acquires a larger field-of-view image, the robotic arm, carrying the end-effector and the close-range camera, approaches one by one from the left side to the right side to sense and pick the ripe fruits again.

**Table 2.** Comparison of different forms of multi-sensor combination perception.

| Applied Crops | Perception | Sensors | Characteristic | Structure | Effect | Ref. |
|---|---|---|---|---|---|---|
| Red tomato, Green tomato | Monocular RGB +Ultrasonic | Pi camera (mobile) + ultrasonic sensor HC-SR04 (mobile) | Simple method, low cost, and adaptable to the limited computing resources of microcontrollers |  | Picking time Red: 4.932 s Green: 5.276 s | [63] |
| Cherry tomato | Monocular RGB + Laser | FL3-U type RGB camera + LY-LDS-61 type laser sensor | Simple structure, accurate distance measurement, avoiding obstacle obstruction |  | Harvesting success rate: 83% | [64,65] |
| Eggplant | Monocular RGB + depth camera (TOF) | Prosilica GC2450C RGB camera + Mesa SwissRanger SR4000 depth measurement capture camera | High precision and sensitivity |  | Sensing time: 0.81 s | [65] |
| Strawberry | Far-field monocular RGB + Close up monocular RGB | 8 mm lens 1280 × 976 pixels 1/3 inch CCD telephoto camera (fixed) + 5 mm lens 640 × 480 pixels 1/2 inch CCD close-up camera (mobile) | Global and local image information can be obtained simultaneously and possible occlusions can be avoided, but it takes longer. |  | Harvesting success rate: 84% Average harvest time: 10.7 s | [66] |

*2.2. Target Perception Methods*

2.2.1. Image Preprocessing Methods

The fruit and vegetable-picking robot must assess the image of the target fruit based on the target features or the trained neural network model after acquiring the original image through the system perception hardware. However, in order to eliminate background noise, recover real information, enhance detectability, and simplify data, the original image is generally preprocessed first. Commonly used preprocessing methods mainly include grayscale transformation (contrast enhancement, contrast compression, gamma correction, etc.), spatial filtering (Gaussian filtering, mean filtering, median filtering, edge detection, etc.), coordinate transformation (translation, mirroring, rotating), morphological operations (erosion, dilation, open operations, closed operations), and so on.

During image recognition, Xiong et al. [67] first converted RGB images into HSV color saturation images and then judged whether the adaptive color threshold was reached and determined which strawberries could be picked. Moreover, Feng et al. [64] improved the image quality and region delineation accuracy by using the R-G color model to enhance the color features of the images acquired from the camera and then determining the candidate regions of ripe tomato bunches based on the column pixel grey scale statistics. In addition, in order to ensure the stability of obtaining basic information about the target, besides the addition of an external light source, as mentioned earlier, such problems can also be solved by using lighting balance algorithms to preprocess the original image. Zhuang et al. [68] proposed an iterative Retinex algorithm based on the weighted intensity of the fruit region in RGB color images, which can adaptively improve images with poor light distribution, and more than 97% of pixels within the litchi region were correctly segmented after light compensation. In order to enhance the segmentation accuracy and success rate of oil palm fruit under complex backgrounds, different illumination, and different fruit maturity, Huang et al. [69] transformed RGB color space into Lab color space and then obtained a region of interest (ROI) containing oil palm fruit by using Otsu algorithm and morphological operation. After the ROI image in the color space is converted to a Grayscale and smoothed by a Gaussian filter, the target at the edge of the image can be clearly detected. These image preprocessing methods are not only commonly used in traditional image analysis techniques based on target features but also widely used in deep learning techniques. These methods are generally used to reasonably augment a limited dataset by stretching, scaling, rotating, panning, and contrast adjustment as a way to achieve data augmentation and to improve the accuracy and robustness of the neural network model. In the greenhouse of the Guangdong Academy of Agricultural Sciences, Ning et al. [47,70] designed a sweet pepper-picking robot that collected 400 images of 9882 sweet peppers from multiple angles in a variety of weather conditions with a depth camera. In order to provide the YOLO-V4-CBAM model with a sufficient training set and to improve the model detection accuracy, the training set was augmented with data using exposure, blurring, mirroring, and rotation, and 1500 images were obtained, totaling 33,780 sweet peppers. In terms of noise reduction, Mao et al. [70], in order to overcome the interference of complex backgrounds, such as soil, hay, and irrigation pipeline in the cucumber image, the original image of the cucumber was processed under G component to filter out the objects with large color difference in the background. The image is then smoothed using a 3×3 median filter and segmented using the Otsu algorithm to obtain a preliminary denoised background image. After that, MSER (Maximum Stable Extreme Region) is used to further eliminate leaf noise, which enables deep learning to extract cucumber features from complex backgrounds more easily.

2.2.2. Perception Methods Based on Target Features

The traditional techniques for image segmentation based on target features are mainly color thresholding-based, edge detection-based, region growing-based, and graph theory-based. Empirical thresholds and adaptive thresholds are the two types of threshold segmentation technologies that are utilized in real applications. Empirical thresholds are more

frequently used and can be adjusted to meet production needs. In the automation process, an adaptive threshold is utilized more frequently, and adaptive algorithms typically have to select an adaptive threshold. To achieve real-time detection of strawberries, Xiong et al. [71] used a simple color thresholding algorithm based on RGB channels with faster processing speed to detect strawberries. At the same time, in order to remove the noisy pixels and fill the holes, morphological opening and closing operations are performed on the original RGB image based on erosion and expansion of the binary image and objects that are too far or too close to the robot are removed by depth filtering of the depth image. Feng et al. [64] used the R-G color model to enhance the difference between the target fruit and the background by analyzing the color features of the images captured by the RGB camera and selecting the candidate region of the saint fruit from the R-G image based on the gray statistics of the column pixels. Finally, the fruits were recognized using the CogPMAlignTool in the Cognex Vision Pro image processing class library. In litchi picking, the localization of the picking point has always been an important part of orchard operations by picking robots, but the localization accuracy of the picking point is easily affected by unstructured growing environments, such as light intensity variations. In order to eliminate the effect of illumination variations, Zhuang et al. [68] improved the illuminance distribution of weakly illuminated images by employing an adaptive iterative Retinex algorithm while keeping the illuminance distribution of well-illuminated images unchanged. The stem is segmented, and noise is filtered using the histogram of intensity distribution after the litchi region has been divided up into RGB color space. Finally, the location of the picking point was determined based on the connection and positional relationship between the segmented litchi and the stem. Although the segmentation method based on color thresholding has been widely used in image segmentation, its shortcomings are also obvious. This method is only applicable to targets whose colors differ significantly from the image background and whose ripe fruits have a relatively single color, and it fails in the face of fruits and vegetables whose ripe fruits are similar in color to the surrounding environment or have multiple colors. In order to achieve effective segmentation of oil palm fruits with various shapes and colors, Septiarini et al. [69] used an edge detection method widely used for fruit segmentation—Canny detection. In order to reduce the noise interference and improve the image quality, Gaussian smoothing is used to connect the small discontinuities in the image before Canny detection. Then, morphological extension, filling, and reconstruction were carried out, and two morphological operations—opening and closing—were used to correct misclassification. A comparison of image segmentation methods based on different target features is shown in Table 3.

**Table 3.** Comparison of image segmentation methods based on different target features.

| Splitting Technology | Applied Crops | Description | Advantages | Disadvantages | Applicable Environment | Examples | Ref. |
|---|---|---|---|---|---|---|---|
| Color Threshold | Strawberry Cherry Tomato Litchi | One or several thresholds to classify the grayscale histogram, grayscale values in the same category belong to the same object | Most commonly used, simple, fast and efficient calculation | Cannot effectively segment targets with little difference in grayscale values and overlap, more sensitive to noise | Applicable when the difference between image background and target features is obvious | Otsu K-means clustering Maximum entropy method | [64,68,71] |
| Edge detection | Oil palm fruit | Different images have different grayscale, and there are generally distinct edges at the boundary, so use this feature for image segmentation | Faster retrieval and better detection of different image edges | More sensitive to noise, conflicts between noise immunity and detection accuracy | Applicable when low noise, large difference in edge features between different regions | Canny Sobel Robert Prewitt Laplaeian | [69] |
| Regional Growth | Eggplant Kiwifruit Chili Guava | Divide the image into different segmentation regions according to the similarity criterion | It has better area characteristics and overcomes the disadvantage of continuous segmentation area that exists in other methods | Prone to over-segmentation | Applicable when a more definite structural division of the area is required | Meyer Watershed Method, Adams Seed Area Growth Method, Gonzalez Regional Split Merge Method | [46,65,72] |
| Graph Theory | - | The essence is to remove specific edges and divide the graph into several subgraphs to achieve segmentation. | Suitable for a wide range of target shapes | Longer operation time | - | Graph Cuts Grab Cuts Random Walk | - |

### 2.2.3. Feature Fusion-Based Perception Methods

The previous section summarized some methods for object detection based on single features, such as color, texture, and shape, and listed some application examples. However, the above methods often do not perform well enough when encountering complex working environments, and in order to solve this problem, researchers have proposed a detection method based on multi-target feature fusion. This method can integrate the sensing advantages between different features and effectively improve the detection accuracy and robustness of the target-sensing system under complex working conditions.

Kiwifruit image recognition is a bit difficult because of the interference of occlusion and overlapping; in order to solve this problem, many scholars use a feature fusion-based approach for effective perception. Liu et al. [72] also proposed a more complete set of methods for this purpose. In the image processing stage, after converting the RGB color space to HSV, frequency domain filtering and homomorphic filtering techniques are used to eliminate a large amount of noise in the original image and to improve the contrast. Then, the images of kiwifruit were segmented in three stages by combining the Otsu algorithm, the regional growth method, and a dynamic fast identification algorithm. Yang et al. [44] proposed an image segmentation algorithm based on LS-SVM for the visual detection and location of Hangzhou white chrysanthemum. The color and texture features in the RGB space are input to the LS-SVM model after being de-noised by a bilateral filter. The experiment showed that the trained model could effectively separate images of Hangzhou white chrysanthemums from the front, back, and shadow illumination, with an accuracy of more than 90% and a segmentation time of only 0.7 s.

Generally, the above method can only identify and detect one kind of fruit, but it is not applicable to other kinds of fruit and vegetables. Is there any algorithm that can detect multiple types of fruits and vegetables at the same time? To solve this problem, Lin et al. [46] proposed a novel detection method. This technique uses an SVM classifier based on angle, color, and shape characteristics to detect spherical or cylindrical fruit that is common in natural environments. It integrates a clustering algorithm based on region growth and a three-dimensional shape detection algorithm based on m-estimated sample conformance (MSAC). The experiment demonstrates that, for pepper, eggplant, and guava, the algorithm's detection accuracy is 0.866, 0.888, and 0.866, respectively, and that the average detection time for a single fruit is 1.41 s, 4.07 s, and 4.70 s. Similarly, Sepulveda et al. [65] proposed an image segmentation algorithm composed of support vector machine (SVM), watershed transformation, and point cloud extraction for eggplant picking under complex working conditions. The supervised training of the Cubic SVM support vector machine is carried out according to the color characteristics of different scene elements. The trained classifier can identify and segment eggplants in most cases, and the watershed transformation can effectively segment eggplant images in the overlapping state. The above methods improve the perception of the mature target fruit by means of feature fusion but do not detect and analyze the quality of the fruit. If inefficient picking can be prevented by recognizing the rotten and damaged fruits, the picking quality of the picking robot can be enhanced to some amount. In this respect, Kurpaska et al. [73] conducted some research and proposed a method to detect and judge the quality of strawberries based on texture, color, and contour shape analysis. This method uses the analysis of geological samples based on color and texture analysis to detect and analyze the quality of strawberries. The experiment showed that the comprehensive detection method can effectively distinguish different quality strawberries.

### 2.2.4. Perception Methods Based on Deep Learning

Deep learning is a new research direction in the field of machine learning, first proposed by Hinton et al. in 2006. The neural network in deep learning can be divided into three layers: the input layer; the hidden layer; and the output layer. After the input layer obtains the input image information, it passes the information to the hidden layer for feature extraction, and finally, the hidden layer outputs the model results. The working principle

of the neural network presented in fully connected form is shown in Figure 3. Compared with the traditional shallow learning structures such as support vector machine (SVM) and artificial neural network (ANN), deep learning can extract the hidden features in the image and automatically learn to obtain the hierarchical feature representation (as shown in Figure 4a), which is more conducive to the classification or feature visualization. As the amount of training data increases, the advantages of deep learning models become more and more obvious (as shown in Figure 4b). In addition, deep learning has the flexibility to choose the number of network layers according to the designer's needs. Based on the above advantages, in recent years, deep learning has been widely used in the target detection of fruits, vegetables, and other crops with good results.
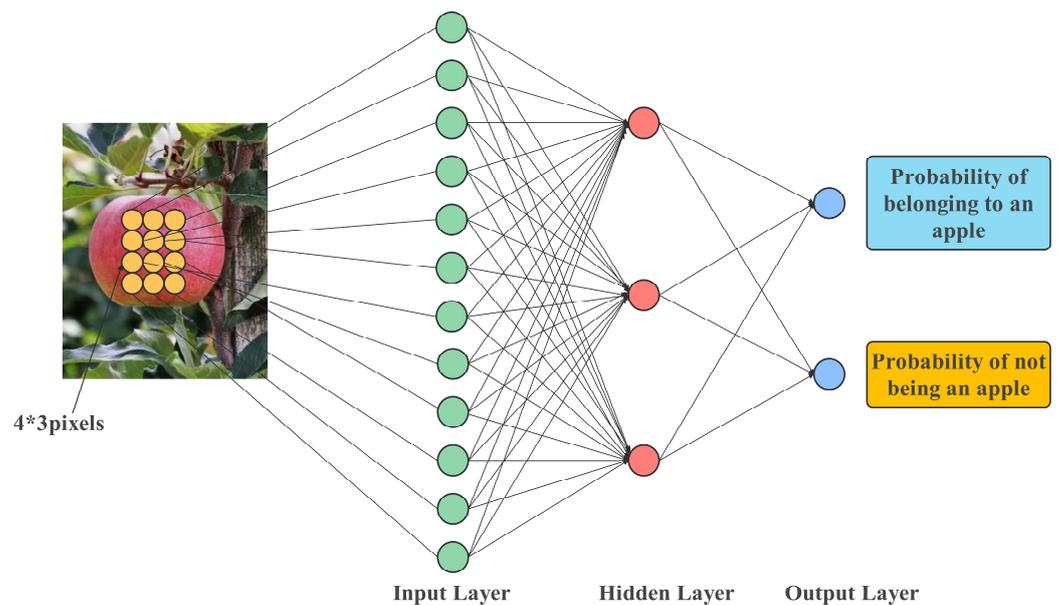


**Figure 3.** Working diagram of neural network in fully connected form.

Unlike traditional manual feature-based detection algorithms (VJ, HOG, DPM), there are many detection algorithms in deep learning, which can be roughly divided into one-stage detection algorithms and two-stage detection algorithms according to the detection stage (as shown in Figure 5). Among them, two-stage detection algorithms mainly include RCNN, Faster R-CNN, Mask R-CNN, and so on. This type of algorithm performs target detection by first generating a pre-selected box that may contain the object to be detected (Proposal box) and then completing the identification and localization of the target after further detection based on the characteristics of the object. This kind of algorithm was quickly developed in the early stages of the application of deep learning technology because it has high detection precision and accuracy, but it also has the drawback of slower detection speed and is time-consuming. Unlike two-stage detection algorithms, single-stage detection algorithms, such as YOLO and SSD, do not require a region candidate network (PRN) and can directly extract features in the network to predict object classification and location, which is characterized by a one-step process and faster detection speed. The Yolo series of algorithms can reach 200 fps, much higher than the 5 fps of the two-stage algorithm Mask R-CNN, which is especially suitable for mobile platforms, but its detection accuracy is a bit poorer than that of algorithms such as Faster-RCNN. Table 4 compares and analyzes various network models used by different researchers.

(a)



(b)
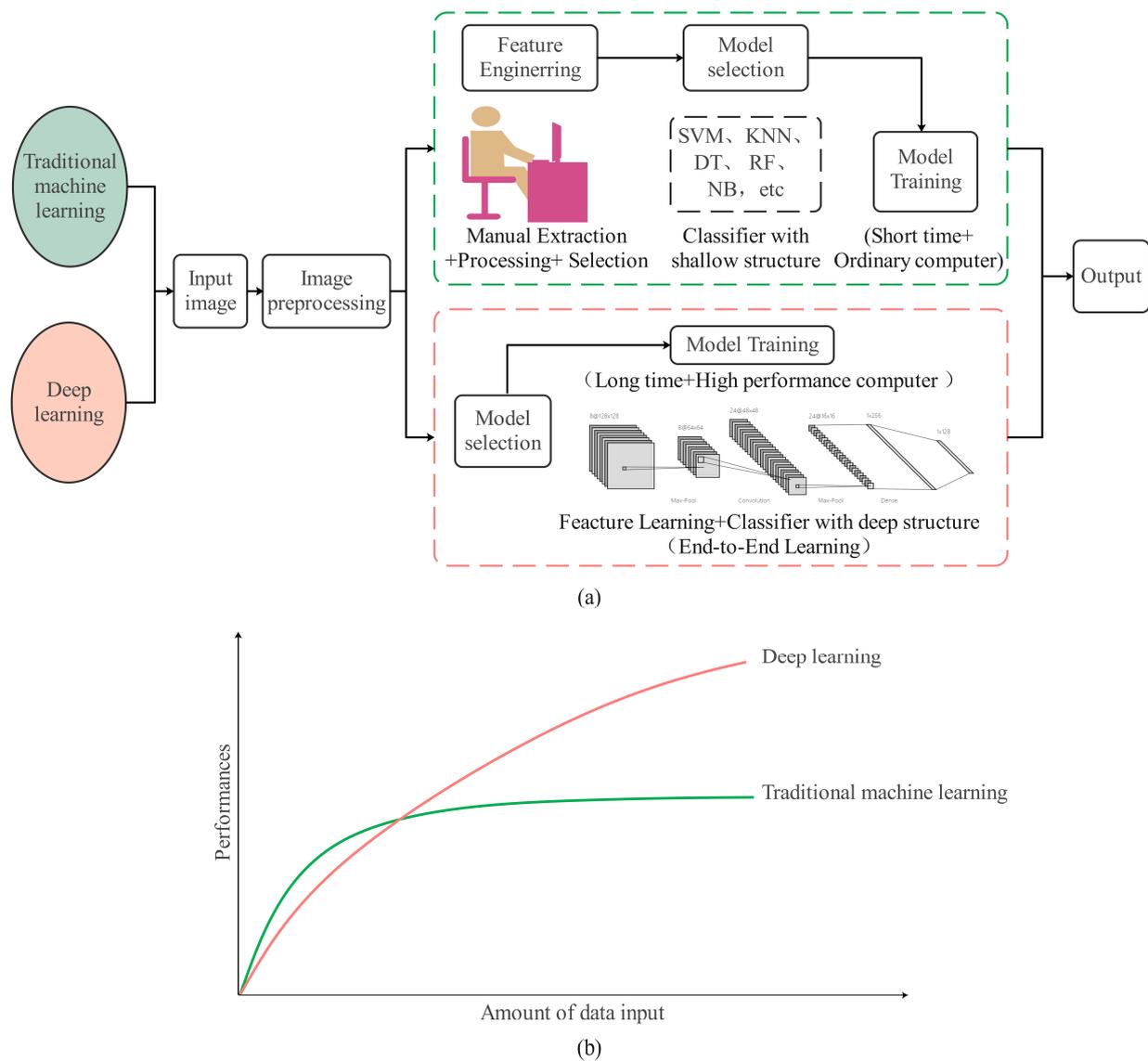
**Figure 4.** Comparison of Process and Performance between Traditional Machine Learning and Deep Learning in Image Processing: (**a**) Comparison of Image Processing Processes between Traditional Machine Learning and Deep Learning; (**b**) The relationship between the performance of traditional machine learning and deep learning and the amount of data input.

(a) Faster R-CNN

Faster R-CNN is a more classical two-stage target detection network, a detection algorithm that was proposed in 2015 after R-CNN and Fast RCNN. Architecturally, it consists of two main networks, Fast R-CNN and RPN (Regional Proposal Network). Compared with the previous two, Faster -RCNN integrates feature extraction, proposal extraction, bounding box, regression, and classification in a single network, which significantly improves the detection speed and greatly improves the comprehensive performance. Mu et al. [48] used Faster R-CNN for kiwifruit recognition, where color and depth images acquired from a Kinectv2 camera were fed into a convolutional neural network, and the neural network was used to detect and extract the coordinates of kiwifruit. The picking robot applying this network model showed an extremely high picking success rate of 94.2% in an orchard test containing 240 samples, with an average picking time of 4–5 s. Similarly, Fu et al. [55] selected two network structures (ZFNet and VGG16) based on Faster-RCNN for apple picking and used the network of two structures to detect the Original-RGB and Foreground-RGB

images acquired from Kinectv2. The experimental results showed that the VGG16 network has the highest average detection accuracy (AP) of 0.893 for Foreground-RGB images.
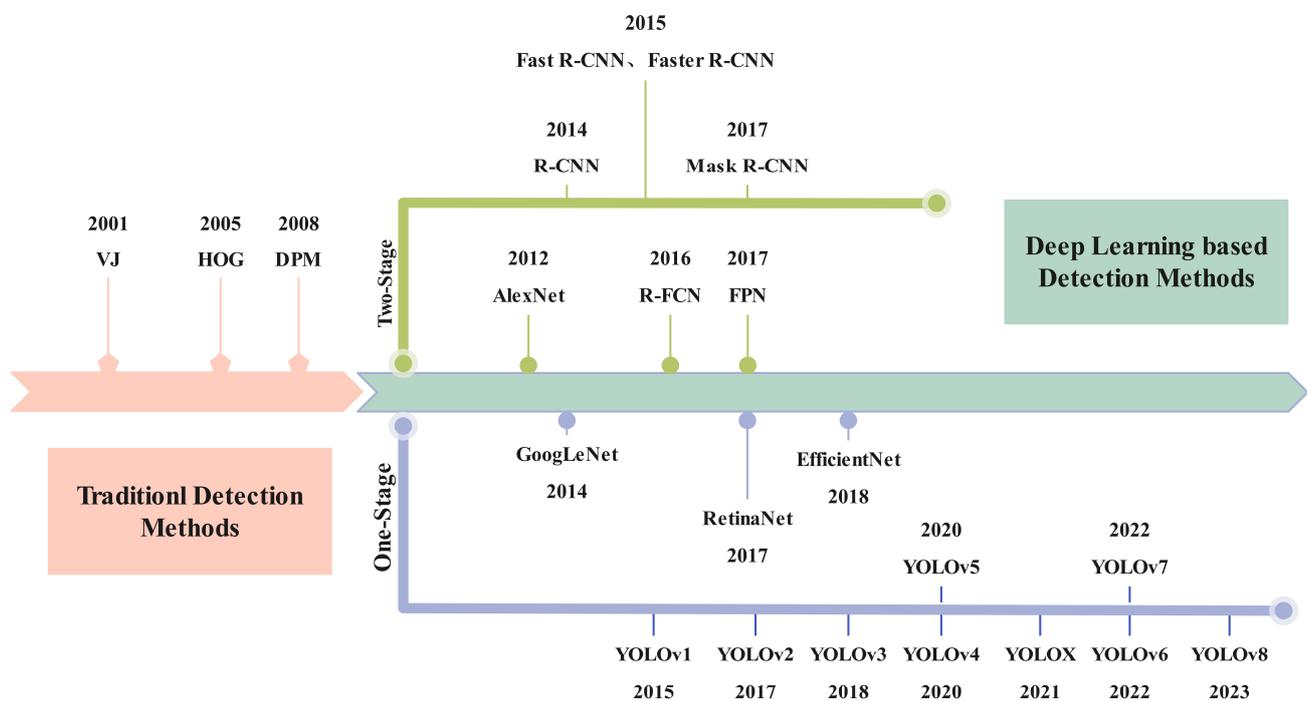


**Figure 5.** Target detection technology timeline.

(b) Mask-RCNN

Mask-RCNN is another classical deep learning network after Faster R-CNN in the two-stage network, which is based on Faster R-CNN with a fully convolutional Mask Prediction Branch added to the Head layer. The ROI Pooling is improved, and ROI Align is proposed, which solves the problem of twice region mismatch caused by rounding in ROI Pooling in Faster R-CNN. Different from Faster R-CNN, which uses VGG as the skeleton network, Master R-CNN uses ResNet50 or ResNet101 as the skeleton network. Combined with the FCN network structure, four modes can be formed, namely, ResNet50, ResNet101, ResNet50 + FPN, and ResNet101 + FPN. The ROI generation method, the selection of RPs, and the selection of RPs to be projected onto the feature map will be different for different combinations, and the size of the feature maps into the Head layer will also be different so that the researchers can choose flexibly according to their needs. Compared with Faster R-CNN, Mask-RCNN is able to simultaneously achieve target detection, target classification, and pixel-level target segmentation by combining object detection and semantic segmentation. Yu et al. [74] used Mask-RCNN as the detection network of the vision module in order to improve the target detection performance of a strawberry-picking robot and chose Resnet50 as the skeleton network, which was combined with a feature pyramid network (FPN) architecture for feature acquisition of target strawberries. The target detection experiments showed that the average detection accuracy (AP) of the trained model is 95.78%, which is particularly effective for strawberry detection under complex growth states such as changing light intensity, overlap, and occlusion. Similarly, in order to better detect tomatoes that are in an overlapping state with smooth texture and uniform color, Gong et al. [54] used Mask-RCNN, which has a better performance in dealing with overlapping targets, as the basic network and used RGB-D-I fused images as the training set. The test results showed that the target segmentation accuracy is improved by 7.6% over the RGB-based mask R-CNN using the extended Mask R-CNN model trained with fused images. In addition, in order to solve the problem of fruit recognition and localization under different occlusion states, Yang et al. [75] purposely proposed a citrus fruit and

branch recognition model based on Mask R-CNN. While constructing a training dataset including multiple complex conditions, a segmentation labeling method is proposed for irregular branches. The experiments showed that the average detection accuracies of the trained model for fruits and branches were 88.15% and 96.27%, respectively, and the average measurement errors for citrus transverse, longitudinal, and branch diameters were 2.52 mm, 2.29 mm, and 1.17 mm, respectively.

(c) YOLO

YOLO series networks belong to one-stage representative networks. Unlike Faster R-CNN and Mask-RCNN, YOLO does not have an RPN network structure and combines object classification and object localization (bounding box) into a regression problem in the detection process. Different from R-CNN's "Look twice" (candidate box extraction and classification), YOLO only needs to Look Once, so the detection speed of the YOLO network is much faster than the two-stage network of the R-CNN series.

In view of the advantages of YOLO series networks and in order to detect banana fruits quickly and accurately in a complex orchard environment, Fu et al. [76] proposed a banana fruit detection method based on YOLOv4. Experimental results showed that the detection rate of the algorithm was 99.29%, the average detection time was 0.171 s, and the AP value was 0.9995. Similarly, in order to meet the identification and positioning requirements of litchi fruits and stems in the nighttime environment, Liang et al. [77] proposed a litchi fruit detection method based on YOLOv3. Under high, medium, and low brightness conditions, the mean Average Precision (mAP) of the model for fruit detection is 96.43%, and the average detection time is 0.026 s. For the segmentation of stem, the accuracy is 95.54%, and the average segmentation time is 0.071 s. In order to verify whether different classification patterns will affect the detection effect of the kiwi detection model, Suo et al. [78] collected and classified 1160 kiwi images according to picking strategies and occlusion conditions and input them into two network models of YOLOv4 and YOLOv3 for training and testing. The experimental results showed that labeling and classifying the data set as detailed as possible can effectively improve the detection accuracy of the network model. Ning et al. [47] used YOLO-V4-CBAM based on YOLO-V4 to identify and locate sweet peppers in dense planting environments so as to improve the recognition and positioning accuracy of sweet pepper-picking robots for multi-target fruits in complex planting environments. Experimental results showed that the F1-score of the proposed method for sweet pepper in a dense planting environment is 91.84%, which is 9.14% higher than that of YOLO-V4, and the positioning accuracy is 89.55%. On the basis of previous research, Xiong et al. [56] combined YOLOv4, Deep SORT, and color threshold to develop a faster and more accurate vision system for strawberry real-time detection, tracking, and positioning. Field experiments showed that the picking success rate of strawberry picking robots using the new system is 62.4%, which is 36.8% higher than before. Similarly, Yu et al. [79] designed a fruit pose estimator called R-YOLO for their new strawberry ridge-harvesting robot. This model is based on YOLOv3 and uses the lightweight network Mobilenet-V1 as the backbone network for feature extraction, which improves the running speed of the model. Tests showed that the model has an average recognition rate of 94.43%, and the processing speed of a single image is 3.6 times faster than YOLOV3. Xu et al. [80] proposed a green mango detection model Light-YOLOv3 based on YOLOv3 for picking green mangoes under different lighting and occlusion environments. This model uses a lightweight unit based on the green mango's color, texture, and shape features to replace the Resnet unit in YOLOv3 and combines the MSCA (Multiscale context aggregation) module to concatenate and predict multi-layer features, which effectively improves the detection effect of green mangoes. Similarly, in order to solve the problem of tomato detection in complex scenes and adapt to embedded devices, Xu et al. [81] proposed a fast detection method based on YOLOv3-tiny. The new model uses improved depth-wise separable convolution and residual structure to replace the standard convolutional network, which increases the depth of the network and greatly reduces the number of Flops. Experiments showed that the f1-score of the new model is 12% higher than that of YOLOv3-tiny, and the

detection speed reaches 25 frames per second. In addition, in order to solve the problem of information loss and insufficient semantic feature extraction of small targets in the process of network transmission of Yolov3, Chen et al. [82] proposed an improved Yolov3 cherry tomato detection algorithm YOLOv3-DPN based on DPNs. The improved algorithm can extract richer semantic features of small targets and reduce the information loss in the propagation process. It is worth mentioning that many groups have performed a lot of research on how to better detect objects, but very little attention has been paid to the problem of phasing objects. To this end, Wang et al. [83] proposed a multi-stage strawberry fruit detection method based on Detailed Semantic Enhancement (DSE-YOLO) on the basis of YOLOv3. This model includes the DSE module, EBCE, and DEMSE loss functions, which solve the problem of foreground class imbalance of the original model and can distinguish different stages of fruits with higher accuracy while better detecting small fruits. Aiming at the problems of low accuracy and poor robustness in traditional green pepper detection methods, Li et al. [84] proposed an improved green pepper object detection algorithm based on Yolov4_tiny. The algorithm was based on the backbone network in the classical object detection model and introduced adaptive feature fusion and feature attention mechanism. It improves the recognition accuracy of green pepper small targets and ensures classification accuracy. Similarly, according to the characteristics of the small shape and dense growth of plums, Wang et al. [85] proposed an improved version of YOLOv4 lightweight model based on YOLOv4. This model uses Darknet53 generation MobilenetV3 on the backbone network and uses Depthwise Separable Convolution (DSC) to replace the standard convolution so as to lightweight the model. At the same time, the $152 \times 152$ feature layer is introduced to improve the target extraction ability in the dense state. Experiments showed that the model has a higher Mean Average Precision (mAP) than YOLOv4, YOLOv4-Tiny, and MobileNet-SSD. The size is 77.85% smaller than YOLOv4, and the detection speed is 112% faster than YOLOv4. At present, most apple detection algorithms cannot distinguish apples occluded by branches from apples occluded by other apples, which is highly likely to cause damage to the target apple, the robotic arm, and the end-effector during the picking process. In order to solve this problem, Yan et al. [86] proposed an apple detection algorithm based on improved YOLOv5s. Experimental results showed that the algorithm can effectively distinguish between pickable and non-pickable apples. Compared with the classical model, the proposed method effectively improves the mAP while compressing the model size, and the average detection time of a single image is only 0.015 s, which can meet the needs of real-time detection.

(d) SSD

SSD is also a one-stage network; unlike the YOLO series, the SSD network has different scales and aspect ratios of Prior boxes, which allows for the use of different sizes of feature maps for the detection of targets of various scales. Qian et al. [87] proposed an SSD-based method for accurate and real-time mushroom detection and location and optimized the backbone network in the original SSD model to improve the real-time detection performance in the embedded device. The model performs well in tests, with an F1 score of 0.951 and an average localization error of 2.43 mm for mushrooms.

(e) FCN

FCN is the pioneering work of deep learning for semantic segmentation. Compared with CNN, FCN replaces the fully connected layer with a convolutional layer and solves the problem of smaller image size due to convolution and pooling by using up-sampling to recover the image size. FCN does not include a full convolution network with a full connection layer, but it can adapt to target input of any size. Its convolution layer can refine the output results as much as possible, and FCN combined with the jump structure of different depth layer results can also ensure robustness and accuracy. In order to achieve collision-free automatic picking of guava, Lin et al. [12] used a Full Convolutional Network (FCN) for the segmentation of guava color images, and the experimental results showed that the average accuracy of the FCN model for the fruit class is 0.893, and the IOU is 0.806, which indicates that the model is able to be able to segment the guava fruits very

well. Unlike Lin et al., in order to improve the accuracy and increase the efficiency of the vision system of the picking robot, Liu et al. [88] combined deep learning algorithms with machine vision and proposed a novel detection algorithm R-FCN combining region-proposed network (RPN) and full convolutional neural network (FCN). The algorithm utilizes FCN to convolve the input image to achieve pixel-level feature extraction and uses RPN to generate multiple candidate frames on the feature map after the convolution operation to effectively separate the foreground and background of the image. In the identification test of apples and oranges, the detection accuracy of the algorithm reaches 97.66% and 96.50%, and the identification accuracy of large fruit bananas reaches 82.30%.

In addition to the above common network models, Li et al. [89] proposed a semantic segmentation method based on Deeplabv3 to segment the fruit, branches, and background in RGB images in order to adapt to the complex growth environment of litchi and detect and locate the fruit branches of multiple litchi clusters. The experiment showed that the extraction accuracy of the test set is 83.33%, and the mean intersection over union (MIOU) is 79.46%, which has a good segmentation effect. Similarly, during the picking process of litchi, in order to better detect the branches and avoid them from damaging the picking robot, Peng et al. [90] used the DeepLabV3+ semantic segmentation model based on the Xception_65 feature extraction network for target detection of litchi. The experimental results showed that the model has an MIoU of 0.765, which is an improvement of 0.144 over the original DeepLabV3+ model, as well as a stronger robustness. Likewise, for fruit and branch segmentation of apples, Kang et al. [52] used the DASNet network model. The f1 score and IoU of the model for fruit detection and segmentation accuracy were 0.871 and 0.862, respectively, according to the test results in the lab and in the orchard setting, indicating that the model was able to precisely and successfully detect and segment orchard apples.

**Table 4.** Comparison of different object detection network models based on DL.

| | Model Based on | Applied Crops | Data | | | Evaluation Indicators | | Feature | Ref |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | Training Sets | Testing Sets | Detection Speed | Others | | |
| One-Stage | YOLOV3 | Litchi | 545 | - | - | 26 ms | mAP: 96.43% | The detection speed is faster than Faster RCNN and SSD, enabling real-time detection | [77] |
| | Light-YOLOv3 | Green Mane | 500 | - | - | 192 fps (5.21 ms) | FLOPs: 10.12 BN Volume: 44 MB F1-score: 97.7% | The problem of insufficient location and semantic information in YOLOv3 prediction feature maps is solved, and the operation speed is improved by 5 times | [80] |
| | YOLOv3-tiny | Tomato | - | 5500 | - | 25 fps (40 ms) | F1-score: 91.92% | Adapts to detection in complex environments and to embedded devices | [81] |
| | Yolov3-DPN | Virgin fruit | 1825 | 1460 | 365 | 58 ms | Precision Light changes: 93.54% Fruit shading: 94.59% F1-score: 94.18% | Richer semantic features of small targets can be extracted and information loss in the propagation process can be reduced | [82] |
| | R-YOLO | Strawberry | 2000 | 1900 | 100 | 56 ms | Precision: 94.43% Recall: 93.46% | Detection speed is 3.6 times faster than YOLOv3, with good real-time performance | [79] |
| | DSE-YOLO | Strawberry | 21,921 | 14,614 | 7307 | 18.2 fps (55 ms) | mAP: 86.58% F1-score: 81.59% | Better detection of small fruits and more accurate differentiation of different stages of fruits | [83] |
| | YOLOv4 | Kiwifruit | 1160 | 928 | 232 | 25.5 ms | mAP: 91.9%. | More detailed classification of the dataset can improve the detection of YOLOv4 | [78] |
| | YOLOV4-CBAM | Sweet Pepper | - | - | 100 | - | Positioning accuracy: 89.55%. F1-score: 91.84% | Compared to YOLO-V4, YOLO-V4- cbam has a higher F1 score | [47] |
| | Deep sort-YOLOv4 | Strawberry | - | - | - | - | Cluster picking success rate: 62.4% | The cluster selecting success rate increased by 36.8% from the previous rate to 62.4%. | [56] |
| | YOLOv4 | Banana | 1164 | 835 | 120 Validation set (Vs): 209 | 171 ms | Detection rate: 99.29% AP: 0.9995 | - | [76] |
| | Improved-YOLOv4 | Plum | 1890 | 1512 | 378 | 42.55 fps (23.5 ms) | mAP: 88.56% | 77.85% size compression and 112% faster detection than YOLOv4 | [85] |
| | Improved-Yolov4_tiny | Green Pepper | 1500 | 1355 | 145 | 89 fps (11.24 ms) | AP: 95.11%, Precision: 96.91% Recall: 93.85% | It can ensure real-time production and can effectively improve the detection of difficult samples of green pepper. | [84] |
| | Improved-YOLOv5s | Apple | 1214 | 1014 | 200 | 66.7 fps (15 ms) | Recall: 91.48% Precision: 83.83% mAP: 86.75% F1-score: 87.49% | It can effectively identify apples that are obscured by leaves and branches | [86] |
| | SSD | Mushroom | 4300 | 4000 | 300 | | F1-score: 0.951 | - | [87] |

**Table 4.** *Cont.*

| | Model Based on | Applied Crops | Data | | | Evaluation Indicators | | Feature | Ref |
|---|---|---|---|---|---|---|---|---|---|
| | | | Total | Training Sets | Testing Sets | Detection Speed | Others | | |
| Two-Stage | Faster-RCNN | Apple | 800 | 560 | 120 Vs: 120 | 181 ms | AP: 0.893 | The VGG16 foreground- rgb image has an AP of up to 0.893, allowing for almost real-time monitoring | [55] |
| | Mask-RCNN | Tomato | - | - | 500 | 456 ms | Iou: 0.916 | The segmentation accuracy is effectively improved by the model trained based on RGB-D-I fused images | [54] |
| | Mask-RCNN | Strawberry | 2000 | 1900 | 100 | 8 fps (125 ms) | MIou: 89.85% AP: 95.78% Recall: 95.41% | - | [74] |
| | Mask-RCNN | Citrus | - | 1000 | - | - | MAP Fruits: 88.15% Branches: 96.27% | It can effectively detect citrus and tree branches at the same time, and can plan pick-up paths and perform reasonable obstacle avoidance. | [75] |
| | FCN | Guava | 437 | 350 | 87 | 565 ms | Mean Accuracy0.893 IOU: 0.806 | - | [12] |
| | R-FCN | Apple Orange Banana | 160,000 | 80,000 | 40,000 Vs: 40,000 | | Accuracy Apple: 97.66% Orange: 96.50% Banana: 82.30% | Better robustness in real-world engineering | [88] |
| | Deeplabv3 | Litchi | - | - | 90 | 464 ms | Precision: 83.33% IOU: 79.46% | - | [89] |
| | DeepLabV3+ | Litchi | 65,625 | 50,000 | 15,625 | - | MIoU: 0.765 | MIoU improves 0.144 over the original DeepLabV3+ model, while having stronger robustness and higher detection accuracy | [90] |
| | DASNet | - | 1277 | 567 | 560 Vs: 150 | 477 ms | Precision: 0.88 F1-score: 0.871 Recall: 0.868 IoU: 0.862 | - | [52] |

## 3. Intelligent Harvesting "Brain" System

The picking decision and control of fruit and vegetable picking robots are key to ensuring the normal work and efficient picking of the robots. On the one hand, the design of the picking strategy needs to carry out a picking feasibility analysis according to the characteristics of the target fruits, maturity degree, growth environment, and other factors, and combine the hardware facilities of the picking platform, such as robotic arm, end-effector, sensor, to formulate a reasonable picking route and picking mode. On the other hand, picking control needs to realize the accurate positioning of the robot and accurate control of the motion of the manipulator, avoid damage to fruits and vegetables, ensure the picking efficiency and speed, and combine the actual scene and picking strategy for real-time adjustment and optimization. Therefore, reasonable picking strategy and accurate picking control are necessary conditions to ensure the efficient and stable operation of fruit and vegetable picking robots and are also one of the key technologies to realize agricultural production automation.

The decision of picking time and the location of the target fruit are mainly completed by the visual perception system, which is described in detail in Section 2. This section mainly focuses on region division and task allocation, obstacle avoidance strategies, path planning, and control methods.

### 3.1. Spatial Partitioning and Task Allocation

Based on the number of different robotic arms, we divide the region division and task allocation strategy into single-arm harvesting and multi-arm harvesting. The different strategies adopted by the researchers are shown in Table 5.

### 3.1.1. Single Mechanical Arm Harvesting

Single robotic arm picking is a common picking mode at present. It has high picking flexibility, strong picking consistency and stability, and can be used with different end-effectors to complete the picking of various fruits, vegetables, and flowers, which can better adapt to diverse agricultural picking needs. In terms of the division of the working area of a single robotic arm, Zhang et al. [49] divided the picking space into several vertical bar subspaces according to the growth characteristics of tomatoes and screened out invalid subspaces by calculating whether there was enough space volume between adjacent branch obstacles to carry the string of tomatoes with claws. This method can effectively solve the problem of a difficult return journey caused by volume increase after successful harvesting.
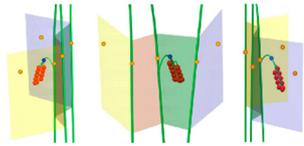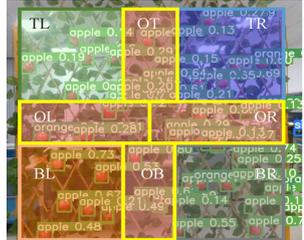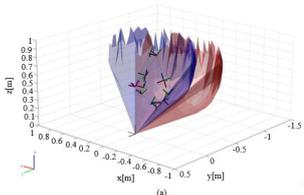
### 3.1.2. Multi-Mechanical Arm Harvesting

Compared with single-arm picking, multi-arm cooperative picking can effectively shorten the picking time, improve the picking efficiency, and better adapt to the complex and changing unstructured picking environment, more suitable for different types, different shapes, and different sizes of crop picking, with stronger adaptability and flexibility. Divided by working area, multi-robot cooperative picking can be categorized into two picking strategies: regional independent and regional shared.

Regional independence: The region-independent strategy means that each robotic arm is individually responsible for a completely independent picking region, that there is no cross overlap between the sub-regions, and each robotic arm is only responsible for picking the target fruits in each sub-region, and this kind of task allocation can avoid the collision interference between multiple arms, and the requirements for the control system are also relatively low. Xiong et al. [67] used a low-cost dual Cartesian robotic arm in their preliminary study of strawberry picking, in which the two robotic arms had completely independent working partitions during picking, and each sub-area was divided into left and right half-areas, and the two arms started picking from the left half-area or the right half-area at the same time according to the density of the target strawberries, which ensured that there was a sufficient safety distance between the arms and avoided possible collisions. In the next study, Xiong et al. [91] put a new type of U-shaped arch-picking robot into

the application; two three-degree-of-freedom robotic arms with non-contact fixtures were installed on both sides of the arch, which were responsible for picking strawberries in the left and right regions, respectively, which was a more complete independent region division, completely solving the problem of collision between multiple arms and at the same time, could effectively reduce the complexity of the control system.

**Table 5.** Space planning and task allocation of harvesting robots.

| Applied Crops | Classification Type | Mechanical Arm | Feature | Sketch Map | Ref. |
|---|---|---|---|---|---|
| Tomato | - | Single | Sieve out invalid subspace, solve the problem of difficult return, improve work efficiency |  | [49] |
| Strawberry | Regional independence | Double | Sufficient safety distance to avoid collision of two robotic arms |  | [67] |
| Strawberry | | Double | The picking area is completely independent, completely solving the collision problem |  | [91] |
| Apple | Regional sharing | Quadra | Effectively solving the problem of double picking and missed picking |  | [51] |
| Eggplant | | Double | Avoiding collisions between robotic arms, effectively shortening picking time |  | [65] |

Regional sharing: The region-shared strategy means that multiple robotic arms are jointly responsible for a large picking area, which is divided into several small sub-areas, with shared overlap between the sub-areas; each robotic arm works independently in the sub-area it is responsible for, and neighboring robotic arms collaborate with each other in the shared area. Under this strategy, each robotic arm collaborates with each other, which can effectively avoid the occurrence of repeated picking and missed picking.

In a dwarfed and densely planted environment, Li et al. [51] used a four-armed robot for the collaborative picking of target apples and planned work partitions for each robotic arm. In addition to the exclusive picking area of the four robotic arms, there are four overlapping picking areas between each neighboring robotic arm. However, to reduce the

amount of computation and control difficulty, at most one robotic arm is allowed to enter the overlapping picking area at the same time, and the whole picking task is categorized as an asynchronous overlapped multiple traveling salesman problem, which can effectively shorten the traversal time. In contrast, to solve the eggplant picking problem in an occluded environment, Sepulveda et al. [65] designed a dual-arm cooperative picking robot. This picking platform can not only simultaneously pick target fruits within the respective working range of the two arms but also pick occluded targets by cooperative operation in the shared area of the two arms. Experiments showed that its average harvesting success rate is as high as 91.67%.

*3.2. Obstacle Avoidance Strategies*
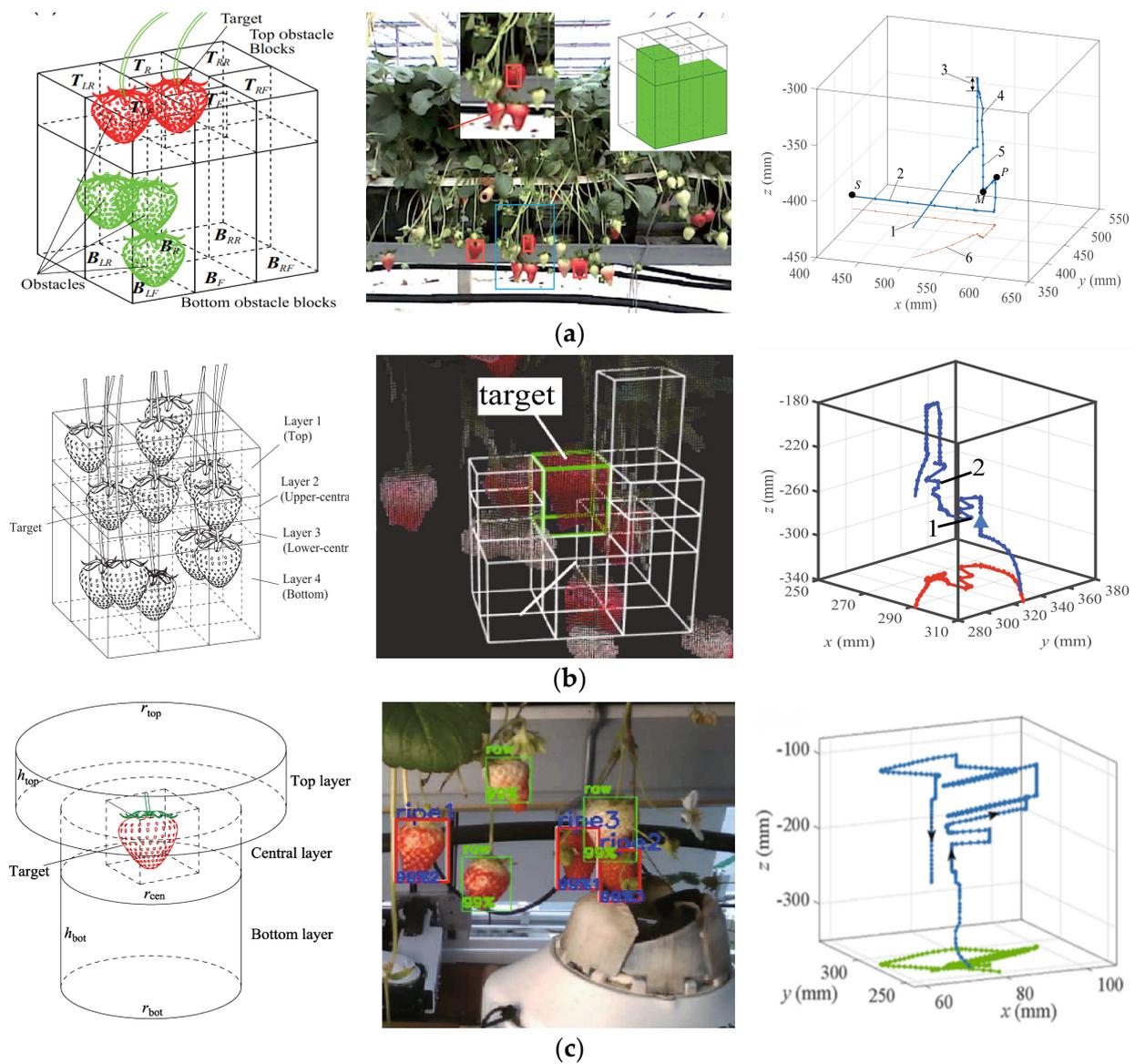
### 3.2.1. Passive Obstacle Avoidance Strategies

Passive obstacle avoidance is the most common and widely used obstacle avoidance strategy. It mainly refers to taking some passive measures to avoid collision or conflict when planning the path, considering the obstacles that the robot or unmanned aerial vehicle may encounter when performing the task. It is mainly realized in the path-planning stage by modeling the surrounding environment and adding real-time obstacle avoidance factors to obtain a smooth route without collision.

Considering the obstacle avoidance problem after tomato bunch picking, Zhang et al. [49] proposed a real-time motion path-planning algorithm (OPS) based on spatial segmentation. This method can plan an effective picking subspace for the robotic arm in advance based on the position information of the environment and tomato bunches and to avoid exploring the path in the invalid subspace. In addition, the OPS algorithm can adjust the end attitude of the robotic arm in real time, according to the relative position between the obstacle and the robotic arm, to realize obstacle avoidance. Experiments showed that the picking time of a single bunch of tomatoes by this method is 12.51 s, and the picking success rate is close to 100%.

### 3.2.2. Active Obstacle Avoidance Strategies

In the actual picking process, especially in complex unstructured environments, dense foliage or compact fruit distribution will make the passive obstacle avoidance "bypassing" strategy fail and then will need some more complex active strategies to solve this problem. Unlike passive obstacle avoidance, active obstacle avoidance can be used to "push away" obstacles through a series of complex sequential movements or multi-arm coordination, which is more suitable for target picking under dense shade. To solve the problem of eggplant picking in an occluded environment, Sepulveda et al. [65] used a strategy of pushing away obstacles with one arm and picking with the other arm in their dual-arm picking robot. Experiments showed that the robot had a high success rate of 81.25% in pushing away from obstacles, which is an effective active obstacle avoidance strategy. In addition, regarding obstacle avoidance techniques for strawberry picking in structured growing environments, Xiong et al. [56,67,91] performed extensive research. In their previous work, to determine the number and location of obstacles around the target, Xiong et al. [67] set up a simpler region of interest (ROI) around the target strawberries. This region divides the obstacles into two layers, top and bottom, with six sub-parts in each layer, which is combined with a simple linear operation to push away the possible obstacles at the top and bottom of the target (as shown in Figure 6a). However, for long-stalked strawberries such as "Murano", a single linear push would be ineffective when there are multiple neighboring obstacles around the target, so Xiong et al. [91] added zigzag pushes in the upward and horizontal directions to the original linear push strategy. In addition, a handheld drag operation (in-hand drag) that can avoid accidentally swallowing the upper obstacle is proposed, and a more complex four-layer structure ROI is set around the target (as shown in Figure 6b), which can better solve the obstacle avoidance problem in complex environments. However, it is not reasonable to measure the presence and number of obstacles by the sub-blocks with point cloud information in the region of interest; for this

reason, the research team redefined the layout of the ROI area [56] and used the push–drag maneuvers to accurately separate obstacles based on their exact location (as shown in Figure 6c). In addition, to obtain the information of the obstacles after dragging in time, the middle and top layers use continuous "look and move" for real-time sensing and determine a new round of push–drag operation. The experiments showed that under the premise of constant picking speed, the cluster picking success rate of the improved method reaches 62.4%, which is 36.8% higher than the previous one.



**Figure 6.** Three Different Active Obstacle Avoidance Strategies: (**a**) a simple linear obstacle avoidance strategy (Ref. [67], 2020, Xiong, Y.); (**b**) a zigzag obstacle avoidance strategy (Ref. [91], 2020, Xiong, Y.); (**c**) obstacle avoidance strategy with new ROI regions and continuous "look and move" (Ref. [56], 2021, Xiong, Y.).

*3.3. Path-Planning Techniques*

3.3.1. Classic Path-Planning Algorithms

The task area needs to be modeled before the path is planned out, and the various obstacle information in the task area is obtained through modeling, and the optimal path for the whole area is planned on this basis. Classical path-planning algorithms include global path-planning algorithms and local path-planning algorithms, and common global

planning algorithms include Dijkstra's algorithm, A* algorithm, RRT algorithm, and so on. Since global path planning needs to consider more factors, such as obstacles, work area size, time, etc., it is time-consuming and not easy to cope with the dynamically changing environment. So Sarabu et al. [53] adopted an improved RRT-based algorithm, RRT-Connect, for apple-picking path planning in complex environments. Preliminary experiments showed that this method achieves good results without complex optimization. Moreover, in the apple-picking process, Kang et al. [52] used the Octrees algorithm to preprocess and model the surrounding environment and searched the optimal path through eight subspaces for picking. Compared with other methods, Octrees is more advantageous in terms of storage efficiency.

### 3.3.2. Machine Learning-Based Path-Planning Algorithms

Although dynamic path planning in classical path planning can achieve real-time adjustment of the picking path according to the surrounding environment, it can only avoid individual obstacles and cannot achieve global optimization, while machine learning-based path planning can effectively solve this drawback. The machine learning-based path-planning method applies shallow neural networks or decision trees and other models to path planning; this method requires a large amount of data input and training through the learning of historical data and can be based on the prediction of the results to determine the next action, for complex environments and tasks can achieve good results.

In order to push away the occluded fruits and reach the specified location to pick the target strawberry successfully in a dense planting environment, Mghames et al. [92] proposed a path-planning algorithm known as Interactive Probabilistic Motion Proto-Principle (I-ProMP) and experimentally verified the starting validity, which is well-suited to be used for solving the problem of obstacle avoidance and path planning in the three-dimensional space, and the computation time is very short, which is about 100 ms.

### 3.3.3. Deep Learning-Based Path-Planning Algorithms

Through deeper neural networks, deep learning-based path planning can learn and forecast the best routes, which is capable of handling more complex environments and tasks, as well as being adaptive and efficient compared to machine learning. Common fruit and vegetable recognition and path-planning methods generally suffer from poor recognition robustness and difficulty in generating collision-free picking paths in dense and complex environments; for this reason, Ning et al. [47] proposed an algorithm for sweet pepper recognition and picking sequence planning—AYDY algorithm, which combines the improved YOLOV4 detection algorithm, an improved DPC algorithm with an anti-collision picking sequence method that introduces a winner-takes-all strategy. The experimental results showed that the AYDY algorithm can effectively shorten the traversal path and picking time and enhance robustness, and the collision-free harvesting success rate is as high as 90.04% compared with the traditional sequential and random traversal algorithms. Similarly, Yang et al. [75] utilized an integrated system developed based on a Masked Region Convolutional Neural Network (Mask R-CNN) and Branch Segment Merging algorithm, which can efficiently plan reasonable collision-free harvesting paths for the harvesting robots while detecting citrus and tree branches.

### 3.3.4. Optimization Algorithm-Based Path-Planning Strategies

The path-planning method based on an optimization algorithm is a kind of optimization algorithm to find an optimal path that satisfies the constraints under the conditions of a given starting point and end point. This method usually transforms the path-planning problem into an optimization problem and obtains the optimal path by finding the optimal solution in multiple aspects, such as time, energy consumption, and pulsation.

Sepulveda et al. [65] adopted a stochastic trajectory optimization algorithm (STOMP) to deal with the path-planning problem in the picking process in the design of the two-armed eggplant picking platform, which generates an optimal picking path based on the

workspace, the position of the fruits, and the configuration of the arms and determines the sequence of motions required to grasp and separate the eggplants. This algorithm performs a global search while avoiding the problem that traditional algorithms need to traverse the entire search space by random sampling, which greatly reduces the computational complexity. Similarly, to improve the operational efficiency of a multi-mechanical arm of a collaborative picking robot in a dwarf-dense planting environment, Li et al. [51] generalized the multi-mechanical arm-picking problem with overlapped domains into an asynchronous overlapped multiple traveling salesman problem and solved it optimally, based on genetic algorithm. Experiments showed that the task-planning method based on a genetic algorithm reduces the job traversal time dramatically relative to the random traversal method and the sequential planning method and effectively improves the operational efficiency under the premise of ensuring that each robotic arm does not conflict. In addition, to address the drawbacks of traditional path-planning algorithms that are time-consuming, as well as to solve the problem of low picking success rate caused by the collision between robotic arms and branches in unstructured environments, Ye et al. [62] obtained collision-free picking poses during litchi picking by an improved adaptive weight particle swarm optimization (APSO) algorithm and used an optimization algorithm based on the Bi-RRT algorithm (AtBi-RRT) to quickly determine the appropriate collision-free picking path. Simulation results showed that the average computation time of the At Bi-RRT algorithm is 3.71 s shorter than the TRRT algorithm.

In the past decades, many teams have focused on visual perception and path planning but neglected the research on motion planning; yet, stable motion planning is crucial for the realization of efficient and lossless picking. In order to achieve stable, efficient, and lossless harvesting of apples, Cao et al. [93] proposed an improved multi-objective particle swarm optimization algorithm (GMOPSO). The algorithm combines the methods of variation operator, annealing factor, and feedback mechanism to optimize the motion trajectory in terms of time, energy consumption, and pulsation so as to accelerate the convergence speed while satisfying the stable motion and avoiding the local optimal solution, and finally realize the optimal trajectory of the robotic arm. Tests showed that the picking platform optimized by the GMOPSO algorithm can effectively achieve stable, efficient, and lossless picking, and its average picking time is 25.5 s, with a success rate of 96.67%.

In addition to the above four common path-planning methods, there are some other path-planning and trajectory optimization methods. For example, in the greenhouse cucumber picking process, Chen et al. [94] used an improved prediction point Hough transform to quickly and accurately fit the path of a cucumber picking robot to obtain a smoother and easier-to-handle path. Aiming at the shortcomings of the traditional Hough transform in terms of large traversal angle range, wide intersection detection range, and being time-consuming, this method makes relevant improvements in three aspects: traversal angle range; intersection detection range; and fitting accuracy. Experiments showed that this method is more time-saving than the traditional Hough transform while having higher accuracy and better robustness. For another example, Colucci et al. [95] proposed a simplified motion planning algorithm based on motion decoupling for precise agricultural applications, which can simplify the complex motion planning problem into a series of simple sub-problems, significantly reduce the computational cost and, thus, improve the efficiency and accuracy of motion planning.

### 3.4. Control Methods

In this section, we divide control methods into two categories for description: traditional control methods (classical and modern control); and intelligent control methods. The comparison of the advantages and disadvantages of specific control strategies is shown in Table 6.

**Table 6.** Comparison of the characteristics of different control methods.

| Control Type | Control Method | Applied Crops | Mechanical Arm | Advantages | Disadvantages | Ref. |
|---|---|---|---|---|---|---|
| Classic Control | PID | Strawberry | Single | Simple to implement Easy to adapt Fast response time Good stability | Sensitive to noise Difficult to adjust parameters Unable to handle non-linear systems Unable to handle time-varying systems | [71] |
| | | Eggplant | Double | | | [65] |
| Modern Control | NMPC | - | Double | Wide applicability Robustness Optimizable for multiple objectives Can handle constraints | Large calculation volume Difficult to adjust parameters High impact of model error Poor stability | [96] |
| | Impedance Control | Apple | Single | Wide adaptability High robustness High control accuracy Flexible interaction possible | Large calculation volume Difficult parameter adjustment High requirements for sensors Not very stable | [97] |
| | SMC | Famous Tea | Single | Robust Rapid response | High-frequency oscillation Complexity of nonlinear design | [56] |
| Intelligent Control | Fuzzy Control | Wolfberry | - | Robustness Wide adaptability Adjustable control effects Flexible knowledge representation | Large calculation volume Difficult parameter adjustment Unstable control effect | [33] |
| | Fuzzy PID Control | Wool | - | Robust Flexible fuzzy rules Easy operation | Computationally complex Poor interpretability Difficulty in choosing parameters | [98] |

### 3.4.1. Classical and Modern Control Methods

Traditional control methods refer to control methods based on mathematical models and control theories, specifically PID control, state feedback control, optimal control, and so on. These methods are normally based on accurate mathematical models through modeling and analysis of the system and designing controllers to achieve stable control of the system. They are characterized by good stability and controllability.

Among them, PID control is a common classical control method that can calculate the error of the robot by measuring parameters such as position, speed, and acceleration of the robot and adjusting the control parameters of the robot according to the error so that it can better control the motion trajectory and posture. Sepulveda et al. [65] used a PID controller to receive the trajectory points generated by the STOMP planning algorithm containing information and provided motion execution commands to the picking robot. This information includes the positions, velocities, and accelerations of all the joints of both arms, as well as the start point of the next trajectory path. In addition, Xiong et al. [71] used PID control in an earlier study of a strawberry-picking robot and combined it with information obtained from a vision system to move the arm to the optimal cutting position.

In addition, nonlinear model predictive control (NMPC) belongs to a kind of optimal control in modern control. It is an advanced control strategy that uses a nonlinear model to describe the system and predict the future behavior of the system and computes a series of control inputs so that the system achieves optimal control under certain constraints. NMPC is usually a better choice in some control problems that require higher accuracy. In order

to better solve the picking problem of a multi-arm robot in an orchard, Flécher et al. [96] proposed a VPC strategy combining NMPC and IBVS (Image-Based Visual Servoing), by which different end-effectors are controlled to approach the specified target fruits. Simulation experiments showed that the control strategy can enable the multi-arm robot to perform multiple tasks effectively in a shared space.

Impedance control, which also belongs to modern control, is a control method based on the relationship between force and position, which can realize the control of force and position of the robots when interacting with the environment. In the apple harvesting process, to reduce the damage to apples by the picking robots, Ji et al. [97] proposed an adaptive impedance control method based on impedance control, which can adaptively adjust the impedance parameters to adapt to different environments and tasks, so that the end-effector can grasp apples quickly, stably, and with a low overshoot even when the environmental stiffness and position are not clear.

Sliding mode control (SMC) is a nonlinear control technique; its main idea is to introduce a specific switching function on a sliding surface so that the system state slides rapidly on this sliding surface and remains on it, which helps realize the robust control and anti-disturbance ability of the system. On the automatic picking platform of famous and high-quality tea, Zhou et al. [56] designed and optimized the control strategy of robotic arm picking based on sliding mode control, which effectively suppressed the vibration phenomenon of the sliding mode surface during rapid convergence. In the testing process, it also showed a high picking success rate and integrity rate.

### 3.4.2. Intelligent Control Methods

Intelligent control methods are control methods based on artificial intelligence techniques, such as fuzzy control, neural network control, and genetic algorithm control. These methods are usually based on a data-driven approach, where the controller is designed to realize the intelligent control of the system by learning and analyzing the data of the system. Compared with traditional control methods, intelligent control methods are characterized by high adaptability and good robustness.

Fuzzy control is a control method based on fuzzy logic, which realizes control by establishing a fuzzy rule base and a fuzzy inference mechanism and is able to deal with the fuzzy or uncertainty problems of the system with strong adaptability and robustness. Therefore, to solve the problem of accurate navigation in the unstructured Goji berries environment, Ma et al. [33] used the fuzzy control method to control the navigation of a Goji berry-picking robots, and the experimental results showed that this method could effectively reduce the influence of environmental variables on the picking platform and improve the robustness of the control system. In order to better study the influence of various factors on the effect of cotton picking, Wang et al. [98] developed a cotton-picking measurement and control system, which is a fuzzy PID control system that integrates classical PID control with fuzzy control. This system can realize the continuous adjustability of cotton-picking speed, conveyor belt speed, and fan speed, and its research results can provide support for the optimization of the picking mechanism.

### 4. Intelligent Picking "Hand" System

Harvesting robots typically employ end-effectors to accomplish the task of harvesting. The end-effector serves as a crucial component of harvesting robots responsible for executing specific tasks, for instance, tasks such as picking, transporting, or assembling. In the context of harvesting robots, the principal function of the end-effectors is to facilitate the efficacious picking of crops while concurrently preserving the integrity of the plants. In terms of end-effectors in harvesting robots, commonly employed types include negative-pressure adsorption, shearing-style, cavity-retrieval, and flexible grasping mechanisms. This section will provide an in-depth discussion of these diverse types of end-effectors.

### 4.1. End-Effector Modes of Operation

The modes of operation for end-effectors in agricultural harvesting robots typically encompass four methods: negative-pressure adsorption; shearing; cavity retrieval; and flexible grasping. Negative-pressure adsorption end-effectors, as depicted in Figure 7a, principally utilize the principle of negative pressure adsorption to adhere the crops onto the robot's end-effector, after which they are harvested via the robotic arm or other components. Shearing end-effectors, as illustrated in Figure 7b, predominantly employ a clamping method akin to scissors, severing the crops from their branches or stems. Cavity retrieval end-effectors, as demonstrated in Figure 7c, function by extending the cavity retrieval device into the crop, using the robotic arm and leveraging the air pressure difference to secure the crop within the cavity, followed by its extraction. Flexible grasping end-effectors, as portrayed in Figure 7d, leverage the properties of flexible materials, enabling the robotic arm to drive the grasper in securing the crop, thereby accomplishing the harvesting task. Its advantage lies in its suitability for fruits and vegetables of various shapes and sizes, with the capability of adopting different grasper shapes and sizes for different crops. The following sections will provide detailed insights into the research developments and applications of these four distinct types of end-effectors.



| (a) | (b) | (c) | (d) |
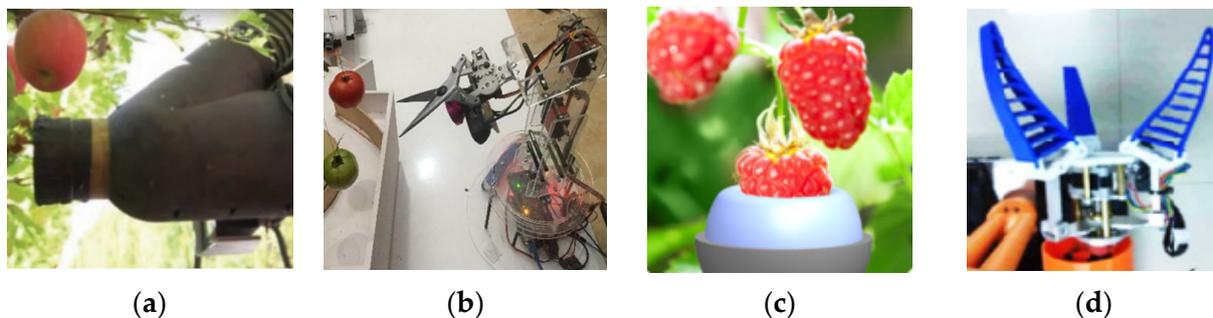
**Figure 7.** Four types of end-effectors for harvesting robots: (**a**) negative pressure adsorption end-effectors (ref. [99], 2018, Ramin Shamshiri et al.); (**b**) shearing-style end-effectors (ref. [63], 2020, Oktarina et al.); (**c**) cavity retrieval end-effectors (ref. [100], 2021, Navas et al.); (**d**) Flexible grasping end-effectors (ref. [59], 2021, Yu et al.).

#### 4.1.1. Negative-Pressure Adsorption End-Effectors

Negative-pressure adsorption end-effectors in agricultural harvesting robots represent a type of end-effector that utilizes negative-pressure adsorption forces during the picking of fruits and vegetables. This end-effector is typically comprised of a suction cup and a negative-pressure system. Such technology necessitates varying designs according to different crop shapes and sizes to ensure sufficient contact with the crop surface and generate ample adsorption force for secure harvesting. Relative to traditional mechanical claws and arms, this end-effector offers superior flexibility and precision, better accommodating crops of diverse shapes and sizes while simultaneously minimizing crop damage.

Over the past few years, significant advancements have been realized in the technology of negative-pressure adsorption end-effectors. Presently, this technology has found applicability in the harvesting of fruits and vegetables with relatively regular shapes, such as apples and tomatoes. A team led by Wang et al. [10] investigated a gripper composed of a flexible silicone funnel, as illustrated in Table 7 (a), which employs vacuum suction for apple harvesting. Through multiple prototype testing, the team designed an optimal funnel shape, considering parameters such as edge thickness, funnel angle, and size while striking a balance between flexibility and robustness. Experimental outcomes revealed that even prolonged exposure of the apple to relatively low vacuum levels did not inflict any damage. Pertaining to the pneumatic harvesting of apples, there are also techniques such as those exemplified in Table 7 (b), a vacuum mechanism robot for apple picking from Abundant Robotics (Hayward, CA, USA). It is a single-suction-cup end-effector capable of
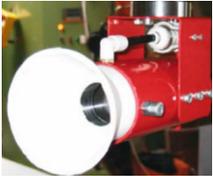
autonomous recognition and location of apples, accomplishing harvesting tasks through negative pressure adsorption. This robotic system demonstrated low damage rates and high harvesting precision in experimental trials [99]. As depicted in Table 7 (c), the harvesting gripper design consists of three components: adsorption, clamping, and twisting for the fruit. Upon the gripper's movement to the targeted location, it encapsulates the entire fruit within a sleeve. The rapid inflation of an airbag tightly clamps onto the surface of the fruit. Subsequently, the rotation of the sleeve enables the disengagement of the tomato fruit from its stem, thus culminating in the successful harvesting of the fruit. Nevertheless, challenges still exist for the negative-pressure adsorption end-effectors when it comes to harvesting crops with irregular shapes and soft textures. Variations in surface texture, size, and shape of such fruits and vegetables can affect the performance of the adsorption force. Relevant studies have indicated that negative-pressure adsorption end-effectors can be utilized not only for the harvesting of firm-textured and relatively regular crops, such as apples and tomatoes but also for delicate flowers, such as Hangzhou white chrysanthemums. As demonstrated in Table 7 (d), Yang et al. [44] specifically designed a unique end-effector with a special structure to avoid damage during the harvesting process of Hangzhou white chrysanthemums. This end-effector is equipped with an airbag device at its tip, allowing for the clamping of the chrysanthemum flowers through the inflation of the airbag. The utilization of a segmentation algorithm in conjunction with the end-effector featuring the airbag device effectively ensures the successful harvesting of Hangzhou white chrysanthemums.

In conclusion, considerable progress has been achieved both in research and practical application of the vacuum adsorption end-effector technology in agricultural harvesting robots. Beyond its application in agricultural harvesting, the vacuum adsorption end-effector has potential uses in other sectors, such as part handling and assembly in manufacturing industries. Ultimately, the vacuum adsorption end-effectors in agricultural harvesting robots will continue to evolve and improve in terms of automation, intelligence, multifunctionality, sustainability, and industrial promotion, thereby fostering significant transformation and progress in agricultural production. Therefore, further research and improvements can lead to broader applications and commercialization.

### 4.1.2. Shearing-Style End-Effectors

Shear-style end-effectors in agricultural harvesting robots are a prevalent type of end-effector, primarily utilized to sever the peduncles of fruits, thereby accomplishing the harvesting task. The following is a detailed overview and current development status of shear-style end-effectors in agricultural harvesting robots.

**Table 7.** Comparison of different styles of negative pressure adsorption end-effectors.

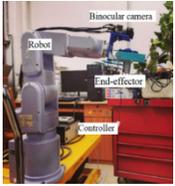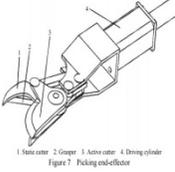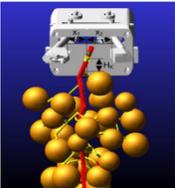| Figure | | Applied Crops | Advantages | Improvements | Gripper Size | Recognition Accuracy | Picking Success Rate | Picking Time | Ref. |
|---|---|---|---|---|---|---|---|---|---|
|  | a | Apple | Minimized damage rate | picking speed, and accuracy | Diameter: 10.5 cm | - | - | 10.3 s | [10] |
|  | b | Apple | Minimized damage rate, Less costly | Picking rate | - | 91% | 85% | 12 s | [99] |
|  | c | Tomato | Simple structure, Less costly | Positioning speed, picking speed | Diameter: 9 cm | - | - | 9.6 s | [10] |
|  | d | Hangzhou white chrysanthemum | High recognition rate | Picking efficiency and accuracy | Diameter: 3 cm | 90 | 80% | 12.5 s | [44] |

Xiong et al. [71] have dedicated their research to the development of a strawberry harvesting robot. After years of research and successive iterations, A novel strawberry harvesting robot has been developed in this study. This robot is comprised of a newly designed gripper mounted on an industrial arm, which, in turn, is mounted on a mobile base along with an RGB-D camera. The novel cable-driven gripper can open fingers to "swallow" a target. Since it is designed to target the fruit and not the stem, it only requires the fruit location for picking. Moreover, equipped with internal sensors, the gripper can sense and correct positional errors and is robust to the localization errors introduced by the vision module. Another important feature of the gripper is the internal container that is used to collect berries during picking. Since the manipulator does not need to go back and forth between each berry and a separate box, picking time is reduced significantly. The vision system uses color thresholding combined with a screening of the object area and the depth range to select ripe and reachable strawberries, which is fast for processing. These components are integrated into a complete system whose performance is analyzed, starting with the four main failure cases of the vision system: undetected, duplicate detections, inaccurate localization, and segmentation failure. The integration enables the robot to harvest continuously by moving the platform with a joystick. Field experiments show that the average cycle time of continuous single strawberry picking is 7.5 s and 10.6 s when including all procedures. This strawberry-harvesting robot can be considered the most advanced and intelligent strawberry-harvesting robot currently available in the agricultural machinery field.

In the context of harvesting cluster fruits, such as litchi, Ye et al. [62] from the South China Agricultural University developed a harvesting machine consisting primarily of an end-effector equipped with a terminal gripper and a rotating blade disc. The robot, during its harvesting operation, uses a collision-free motion planning algorithm, rendering the harvesting process safer and more convenient. Similarly, for the harvesting of cluster fruits, such as cherry tomatoes, Feng et al. [64] developed an end-effector akin to a pair of scissors. As illustrated in Table 8 (c), it is designed based on the mechanical characteristics of the stem, with dual cutting blades used for severing the stem. The handle, fixed to the cutting blades, can close or open to grasp or release the stem, allowing for reliable cutting and handling of the fruit and facilitating its separation from the plant. This design has enhanced the precision of the harvesting end-effector, providing superior stability during the harvesting process.

In the case of tomato harvesting, Oktarina et al. [63] from Indonesia have designed a tomato-harvesting robot, as shown in Table 8 (d). This robot features a simple yet vivid structure, with a scissor-style end-effector that is sharp and flexible. The harvesting process is facilitated through the drive of a servo motor. In a similar context to tomato harvesting, Jin et al. [61] developed an intelligent tomato harvesting robot system based on multimodal deep feature analysis. The end-effector of this system comprises two mechanical fingers and a three-degree-of-freedom mechanical cutter, utilizing digital servos for the rotation of the joints of the mechanical arm and cutter. This setup enhances the precision requirements of the mechanical arm and cutter, effectively addressing issues of labor shortages and high costs encountered in the tomato harvesting process.

For grape harvesting, Liu et al. [101] designed a harvesting hand, as shown in Table 8 (e), which represents a single-degree-of-freedom grasp-and-cut integrated end-effector. The opening and closing of the two fingers are driven by a helical and symmetrical oscillating linkage mechanism; one fingertip is composed of a floating clamp and blade. Upon contact with the pedicel, the compressive force closes the floating clamp around the pedicel while the blade continues to close, thus completing the cut. This type of end-effector enables low-vibration and rapid operation to minimize fruit drop.

**Table 8.** Comparison of different types of shear end-effectors.

| Figure | Number | Applied Crops | Advantages | Improvements | Gripper Size | Recognition Accuracy | Picking Success Rate | Picking Time | Ref. |
|---|---|---|---|---|---|---|---|---|---|
|  | a | Strawberry | Accurately separate obstacles | Targeting accuracy, picking rate | Maximum clamping diameter 60 mm, open diameter 45 mm | - | 96.8% | 10.6 s | [71] |
|  | b | Litchi | Non-destructive picking | Picking rate | - | - | - | - | [62] |
|  | c | Cherry Tomatoes | Stable clamping, low fruit falling rate | Picking success rate | - | - | 83% | 8 s | [64] |
|  | d | Tomatoes | Fast picking rate | Recognition accuracy | - | - | - | 9.676 s | [63] |
|  | e | Grape | Small size, Flexible | Robustness | Gripping shaft length of 30 mm | - | - | 9.6 s | [101] |

For sweet pepper harvesting, Ning et al. [47] developed a shear-style end-effector with a Robotic 2F-85 gripper as the terminal execution component of the robot harvesting system, providing a clamping force of 20–235 N and a payload capacity of 5 kg. The harvesting of sweet peppers is not confined to shear-style techniques; the following sections will also introduce flexible grabbing-style end-effectors, among others.
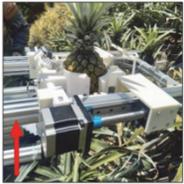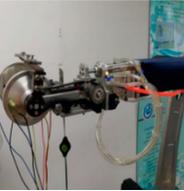
In summary, shear-style end-effectors have been extensively implemented in agricultural harvesting robots and have attracted growing attention from researchers. With continual technological development and innovation, shear-style end-effectors are anticipated to play a more significant role in future agricultural harvesting robots, offering more efficient and reliable solutions for automated agricultural production.

### 4.1.3. Cavity Retrieval End-Effectors

The cavity retrieval end-effector is another commonly used end-effector in agricultural harvesting robots. It accomplishes the harvesting task through mechanical clamping and encasing the crops. Its structure includes an external casing and an internal cavity, wherein the gas pressure within the cavity is controlled by the casing to either grip or release the crop. During the harvesting process, the cavity-insertion style end-effector needs to be adjusted according to the weight and size of the fruit to ensure that it is securely fixed within the cavity, thus preventing it from falling or being damaged during the harvesting process.

Cavity retrieval end-effectors are generally used for fruits with harder textures and regular shapes, such as pineapples and apples. To realize automated pineapple harvesting, Du et al. [45] designed a pineapple-harvesting gripper, as shown in Table 9 (a). It consists of a gripping mechanism and a cutting mechanism that can sequentially and cleanly sever the pineapple stem, thus minimizing damage to the stem. For the cavity-insertion style harvesting of apples, Wei et al. [102] developed a spherical double-finger structure gripper, as shown in Table 9 (c), which can effectively reduce fruit damage rates. Taking into account the shape characteristics of apples, Miao et al. [103] designed an end-effector, as shown in Table 9 (d), the advantage of which is that it does not damage the fruit during harvesting. The cavity-insertion style is not only used for harvesting pineapples and apples but is also suitable for harvesting softer fruits such as strawberries. The strawberry-harvesting robot developed by Xiong et al. [67] opens the cavity during harvesting, then "swallows" the fruit, and the blade severs the fruit stem, thus completing a cycle of strawberry harvesting.

**Table 9.** Comparison of different types of cavity extraction end-effectors.

| Figure | Number | Applied Crops | Advantages | Improvements | Gripper Size | Recognition Accuracy | Picking Success Rate | Picking Time | Ref. |
|---|---|---|---|---|---|---|---|---|---|
|  | a | pineapple | Minimized damage rate, High picking success rate | Picking rate | Cylindrical radius: 100 mm Blade diameter: 130 mm | 95% | 80% | 14.9 S | [45] |
|  | b | mulberry | Accurate separation of obstacles | Positioning accuracy | Maximum clamping diameter: 40 mm, open diameter: 25 mm | - | - | 10.6 s | [100] |
|  | c | Apple | Minimized damage rate | Picking rate and accuracy | - | - | - | 7.81 s | [102] |
|  | d | Apple | Minimized damage rate | Picking rate | Maximum diameter: 14 cm | 91% | 82% | 9.8 s | [103] |

#### 4.1.4. Flexible Gripping End-Effectors

Flexible harvesting end-effectors refer to terminal robotic arms that can mimic the actions of human fingers and palms, characterized by their flexibility, malleability, and ease of operation. The purpose of these effectors is to simulate human organs, such as fingers and palms, enabling precise picking and handling operations for objects of various shapes and sizes. They find wide-ranging applications in fields such as agriculture, manufacturing, and healthcare.

Apples represent one of the most commonly encountered fruit types within the broad spectrum of agricultural produce. Substantial scientific research and developmental efforts have been directed toward enhancing flexible apple-picking methodologies. For instance, Liu et al. [104] have devised a flexible gripper, as demonstrated in Figure 8a, which consists of two curved, flexible fingers. This apparatus has been extensively refined and optimized, enabling the harvesting not only of apples but also other fruits, such as pomegranates and grapefruits. In order to further minimize apple damage during the harvesting process, Pi et al. [105] were inspired by the physical properties of octopus tentacles to study and develop a biomimetic three-fingered flexible gripper, depicted in Figure 8b. Figure 8c illustrates the pneumatic pinch structure of an end-effector developed by Hohimer et al. [106]. This tool is capable of performing apple-picking tasks with significant flexibility and high precision.
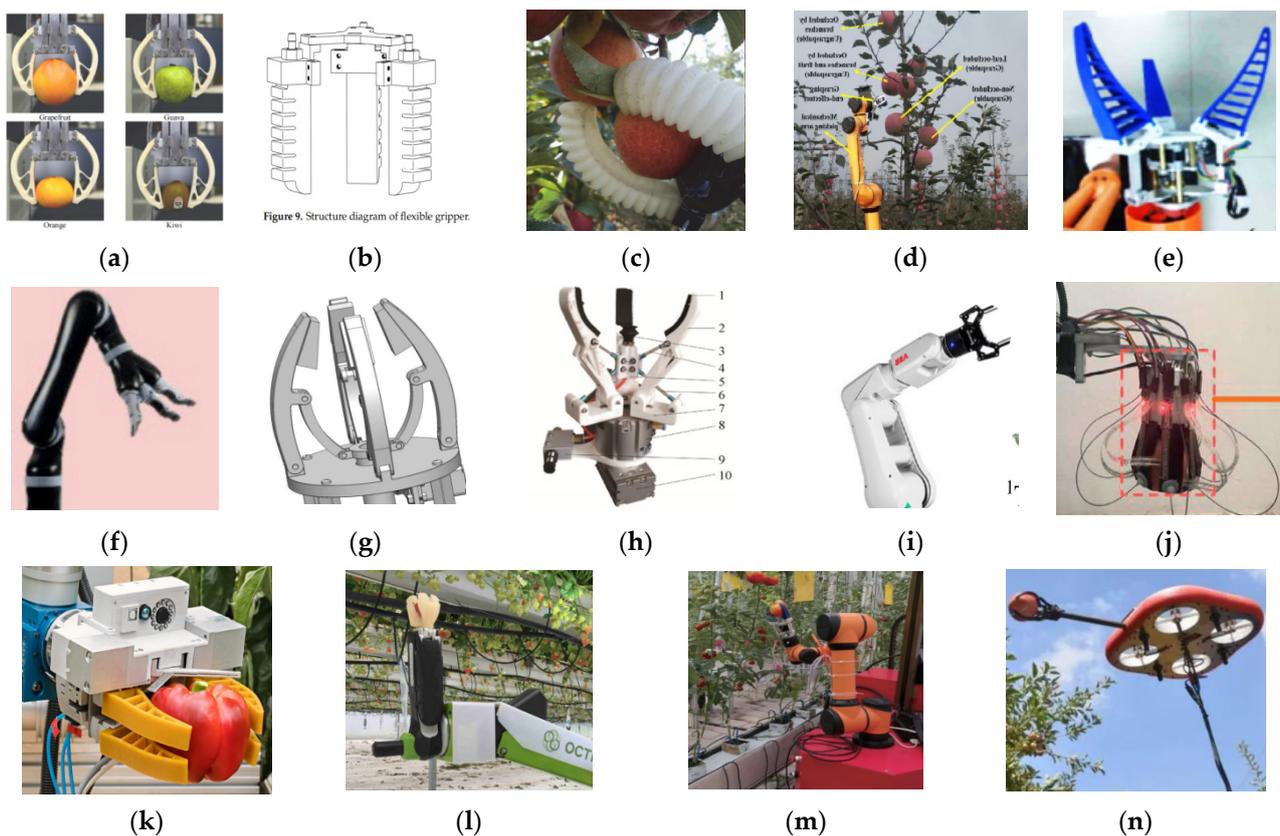


**Figure 8.** Flexible gripping end-effector: (**a**–**l**) (ref. [104], 2018, Liu et al.; ref. [105], 2021, Pi et al.; ref. [106], 2019, Hohimer et al.; ref. [86], 2021, Yan et al.; ref. [59], 2021, Yu et al.; ref. [65], 2020, Sepulveda et al.; ref. [107], 2020, Vu et al.; ref. [50], 2022, Yu et al.; ref. [108], 2021, Chen et al.; ref. [109], 2019, Yung et al.; ref. [110], 2021, Zhang et al.; ref. [111], 2021, Habegger et al.; ref. [112], 2018, Xiong et al.); (**m**,**n**) (https://www.tevel-tech.com/ (accessed on 27 June 2023)).

Furthering the field of flexible robotic technology, Yan et al. [86] developed a flexible gripper mounted on a six-axis robotic arm for apple harvesting, as shown in Figure 8d, which exhibits a high degree of flexibility. From an ergonomics perspective and based on

the structural characteristics of the human body, Yu et al. [59] have engineered a three-fingered gripper made from flexible materials, as illustrated in Figure 8e. This design takes advantage of the widely recognized fin effect, where a fin bends toward the direction of applied pressure and reverts to its original state once the pressure is relieved. Such a claw-like structure is conducive to protecting apples from damage, thus achieving the function of damage-free harvesting.

For the flexible harvesting of tomatoes, researchers around the globe have devoted considerable effort toward the cause. As depicted in Figure 8f, Sepulveda et al. [65] have developed an end-effector for harvesting that resembles the human hand, capable of swiftly and accurately picking tomatoes. Vu et al. [107]., on the other hand, have designed a four-fingered mechanical hand module equipped with an internal vacuum system, as shown in Figure 8g. This innovative vacuum claw module enhances the reliability and safety of fruit-picking operations. Yu et al. [50] have developed a flexible claw endowed with a thin film pressure sensor constructed of rubber via injection molding. This three-fingered device, driven by 42 stepping motors, delivers precise grip capabilities. As displayed in Figure 8h, Chen et al. [108] have designed an end-effector for a tomato-harvesting robot based on pneumatic damage-free clamping. This tool effectively reduces the damage rate during picking, thus achieving damage-free harvesting. Figure 8i shows an end-effector designed by Yung et al. [109] for harvesting tomato seedlings. This flexible harvesting tool boasts rapid collection speeds, further enhancing the efficiency of tomato harvesting.

Flexible grippers have also been employed in the harvesting of other fruits. For instance, as shown in Figure 8j, Zhang et al. [110] have investigated a robotic end-effector equipped with adaptive grasping and tactile sensors. This end-effector, using flexible fingers and integrated force and bend sensors, can measure the distribution of contact forces on the contact surface and the deformation of the fingers, enabling adaptive grasping of various spherical fruits. In Figure 8k, Habegger et al. [111] have designed a flexible end-effector specifically for sweet pepper harvesting composed of four fin-ray grippers. This mechanism ensures that no damage is inflicted upon the peppers during the harvesting process. As depicted in Figure 8l, a strawberry-harvesting robot developed by Preter et al. [112] features an end-effector consisting of two flexible two-fingered structures resembling a palm, greatly reducing damage to strawberries during the picking process. Figure 8m illustrates a tomato-harvesting robot whose end-effector is made of flexible materials, enabling damage-free harvesting. In Figure 8n, the Israel-based company Tevel Aerobotics has invented a fruit-picking drone that employs a simple and convenient end-effector structure suitable for picking a variety of fruits, including apples, nectarines, and plums.

In general, flexible end-effectors for harvesting exhibit high precision, strong flexibility, and easy operability, serving as vital tools for enhancing the efficiency of mechanized picking and handling tasks. The literature review above demonstrates that flexible gripper-style end-effectors have gained extensive application within the realm of agricultural harvesting robots, showing good adaptability and efficiency across diverse crop harvesting tasks.

End-effectors of agricultural harvesting robots are crucial components of agricultural robotic systems, with their performance and functionality directly influencing harvesting efficiency and quality. By leveraging ergonomic principles, these end-effectors can be optimized in terms of design and control to achieve more efficient, precise, and safe harvesting operations. The School of Mechanical and Electronic Engineering at Northwest A&F University has conducted extensive research and experimentation in this area, yielding a series of experimental conclusions [59,113–116].

Regarding flexible materials, they are essential components of flexible grippers. Common flexible materials currently used include elastomers, silicone, polyurethane, and airbags. These materials possess excellent flexibility and adaptability, enabling them to grasp objects of various shapes and sizes. When designing flexible grippers, it is necessary to consider the material's strength, durability, and elasticity to meet the requirements of harvesting operations.

Additionally, the grasping angle and gripper size are significant factors influencing the performance of end-effectors in agricultural harvesting robots. The grasping angle refers to the angle between the end-effector and the object during grasping. Different fruits and crops have varying horticultural characteristics, thus requiring different optimal grasping angles. Hence, it is crucial to determine the optimal grasping angle through research and analysis of crop characteristics. The gripper size should be designed based on the size and shape of different crops to ensure the end-effector can adapt to grasping fruits and crops of varying sizes and shapes. Harvesting modes also play a significant role in the performance of end-effectors in agricultural harvesting robots. Common harvesting modes include rotation, stretching, and combined rotation and stretching. Different harvesting modes are suitable for different fruits and crops. For example, rotational harvesting is suitable for smaller crops, while stretching is appropriate for larger fruits.

In summary, the design and control of end-effectors in agricultural harvesting robots necessitate considering multiple factors, including flexible materials, grasping angles, gripper sizes, and harvesting modes. In the future, advancements in materials, sensors, and control technologies can further enhance the performance and intelligence level of end-effectors in agricultural harvesting robots to meet the harvesting requirements of various fruits and crops, thus promoting the development and application of agricultural robotics technology.

*4.2. Overview of Harvesting Effect Evaluation Indicators*

Evaluating the harvesting performance of agricultural robots is of utmost importance as it directly reflects the efficiency and quality of harvesting, thereby influencing the profitability of agricultural production. This section provides an overview of common metrics used for evaluating harvesting performance (as shown in Table 10).

Recognition Rate: The recognition rate refers to the speed at which agricultural robots can identify fruits or vegetables during harvesting operations. Specifically, it measures the ratio between the number of images processed and recognized by the robot during visual recognition and the corresponding processing time. Generally, a higher recognition rate enables the robot to complete harvesting tasks more quickly, thereby improving harvesting efficiency. Indicators for the recognition rate include the number of images recognized per second, the number of items recognized per second, and the amount of data processed per second.

Harvesting Rate: The harvesting rate is a vital metric for assessing the efficiency of harvesting robots. It is closely related to the technical parameters of the harvesting robot, the complexity of the harvesting site, and the growth conditions of the crops.

Harvesting Quality: Harvesting quality is another important metric for evaluating the harvesting performance of robots. It encompasses indicators such as harvesting accuracy, damage rate, fruit drop rate, and average harvesting time. Harvesting accuracy refers to the consistency of the size, shape, color, and ripeness of the harvested fruits with predetermined standards, while the damage rate reflects the level of fruit damage during the harvesting process.

Harvesting Cost: Harvesting cost is one of the indicators used to measure the economic viability of harvesting robots. It includes factors such as equipment acquisition costs, maintenance and upkeep costs, and energy consumption costs.

Adaptability: Adaptability is a crucial metric for assessing the ability of harvesting robots to adapt to various crops and different harvesting environments. This includes aspects such as the flexibility, stability, and safety of the harvesting robot.

In conclusion, the evaluation metrics for harvesting performance serve as key indicators for assessing harvesting robots. In practical applications, it is necessary to select appropriate evaluation metrics based on specific harvesting tasks and requirements, thereby enabling a scientific and rational assessment and optimization process.

**Table 10.** Comparison of end-effector indicators for different fruits.

| Fruit | Gripper Category | Recognition Rate | Recognition Accuracy | Average Picking Time | Picking Success Rate | Ref. | Year |
|---|---|---|---|---|---|---|---|
| Apples | | - | 82.5% | 14.6 s | 72% | [59] | 2021 |
| | flexible grasping | 0.012 s | - | - | 100% | [105] | 2021 |
| | | - | - | 25.5 s | 96.67% | [93] | 2021 |
| | | - | - | 7.3 s | 67% | [106] | 2019 |
| | shearing-style | 0.015 s | - | - | - | [86] | 2021 |
| | - | 0.181 s | 89% | - | - | [55] | 2020 |
| | - | 0.235 s | 87.1% | 7 s | - | [52] | 2020 |
| Tomatoes | | - | 92.8% | - | 73.04% | [54] | 2021 |
| | shearing-style | - | - | 9.676 s | - | [63] | 2019 |
| | | 0.021 s | 94% | - | 100% | [117] | 2020 |
| | - | 0.096 s | - | - | 91.9% | [118] | 2018 |
| | - | - | 91.92% | - | - | [81] | 2020 |
| | flexible grasping | 0.016 s | - | 8 s | - | [113] | 2021 |
| | - | - | 98% | - | - | [110] | 2021 |
| | - | - | 89% | - | - | [119] | 2020 |
| Strawberries | cavity retrieval | 0.136 s | - | 6.1 s | 97.1% | [67] | 2019 |
| | | - | - | 10.6 s | 96.8% | [71] | 2019 |
| | | 0.086 s | 93.1% | 4 s | - | [112] | 2018 |
| | flexible grasping | - | - | 9.05 s | 96.8% | [73] | 2021 |
| | | 0.049 s | - | 10.62 s | 96.77% | [120] | 2018 |
| | - | - | 86.58% | - | - | [83] | 2022 |
| | | 0.062 s | 95.78% | - | - | [74]- | 2019 |
| | shearing-style | - | 94.43% | - | 84.35% | [79] | 2020 |
| | | - | - | 10.7 s | 84% | [66] | 2019 |
| Sweet papers | shearing-style | - | 91.84% | - | 90.04% | [47] | 2020 |
| | - | - | 96.91% | - | - | [84] | 2021 |
| | | 1.41 s | 86.4% | - | - | [46] | 2020 |
| Litchi fruits | shearing-style | 0.154 s | 93.5% | - | - | [121] | 2021 |
| | | 0.464 s | 83.33% | - | - | [89] | 2020 |
| | - | - | 96.78% | - | - | [77] | 2020 |
| Cherry tomatoes | flexible grasping | - | - | 6.4 s | 84% | [41] | 2022 |
| | shearing-style | - | - | 8 s | 83% | [64] | 2018 |
| | | - | - | 12.51 s | 99.81% | [49] | 2021 |

## 5. Challenges and Prospects

Agricultural fruit and vegetable-harvesting robots represent one of the rapidly evolving domains in recent years, offering vast potential and opportunities for agricultural production. However, with the continuous development of agricultural harvesting robots, a series of challenges and problems that need to be addressed have also emerged. Following are the six challenges and six prospects identified in this article concerning agricultural fruit and vegetable harvesting robots.

### 5.1. Challenges

5.1.1. Multi-Species, Multi-Form Fruit, and Vegetable Picking Is More Difficult

Due to the wide variety of crops, the size, shape, hardness, and other physical characteristics of different kinds of fruit are often very different, so it is difficult to design a robot that can adapt to the needs of multiple kinds of fruit picking at the same time. Generally, one kind of picking robot can only pick one specific kind of fruit, while other types of fruits and vegetables can only be significantly modified or redesigned, which will waste a lot of time and resources for researchers.

### 5.1.2. Difficulty in Picking in Complex Environment

Even when picking the same variety of fruit, the operating environment of the robot in different orchards is also complex and changeable. First of all, most picking robots are designed based on the structured or semi-structured picking environment, and in the actual operation process, different orchard structures are different, which may have a certain impact on the working effect of the picking robot. Secondly, weather conditions can also have an impact on the operation of robots, such as excessive wind power that may affect the stability, safety, and harvesting efficiency of robots. Therefore, fruit and vegetable-picking robots need to face the challenges of adapting and operating in different operating environments.

### 5.1.3. High Real-Time Requirements

The fruit and vegetable-picking robot operates in a real-time environment, requiring rapid recognition, positioning, and grasping of targets. At the same time, it needs to dynamically adjust control parameters during the movement process to maintain a stable motion trajectory, so its real-time requirements are very high. In different harvesting environments, robots also need to handle a variety of complex situations in real time, such as avoiding obstacles and adapting to various light conditions to ensure the efficient completion of harvesting tasks. Therefore, real-time performance represents a significant challenge for fruit and vegetable-harvesting robots.

### 5.1.4. Few Research on Walking Platforms and Navigation

At present, the research on fruit and vegetable-picking robots mainly focuses on visual systems, mechanical arms, terminal actuators, and picking-path planning, while the research on walking platforms and their navigation algorithms in the environment is less frequent. Of course, such studies have limited impact on picking robots operating in a structured environment, and some of the robots moving through slides rarely require navigation. However, under a semi-structural or unstructured working environment, the stability, driving speed, power, and other performances of the picking platform will have a great impact on the picking accuracy, harvesting efficiency, and even the fruit quality in the collecting device. The navigation and route planning of the walking platform in the orchard environment will have a great impact on the working efficiency of long-time picking. These are the questions we need to consider in the future.

### 5.1.5. The Working Height of Picking Robot Is Generally Limited

The existing harvesting robots are mainly designed based on ground mobile platforms, but the structural design of ground harvesting robots is generally fixed, and they usually have limitations in height and size. These limitations determine that the main target of ground-picking robots is some low-fruit trees, which are not suitable for picking higher-fruit trees. Even though some researchers have raised the picking height of the robot to some extent by using a liftable platform, the lifting is very limited, considering the power and balance problems. In this case, how to complete the picking of higher fruit trees has become an urgent problem.

### 5.1.6. High Costs

Currently, the research, development, and production costs of fruit and vegetable harvesting robots are high, which restricts their large-scale deployment in agricultural production. The costs of fruit and vegetable harvesting robots mainly comprise the costs of robot production research and development, maintenance, and labor, among others. At present, most fruit and vegetable harvesting robots necessitate substantial research and development and material costs, along with regular maintenance, which escalates the cost of robot usage. Hence, ensuring the efficiency and quality of robotic harvesting while reducing its costs is a significant issue that needs to be addressed in future research and development of fruit and vegetable harvesting robots.

*5.2. Prospects*

5.2.1. Modular Harvesting Robot

In the face of the difficulty of picking fruits and vegetables with multiple varieties and forms, in addition to using adjustable grippers or flexible end-effectors, modular design can also be used to solve the problem. By implementing modular design to achieve product versatility, this has been successfully applied and achieved good results in multiple types of military or civilian products. In the future field of harvesting robots, in addition to modular end-effectors, modular walking platforms, and even modular robotic arms and sensing systems should be widely developed. Through modular design, researchers can not only flexibly match various components based on the characteristics of the target fruit and the picking environment but also promote the standardization of picking robots while reducing costs. This will effectively promote the further promotion and application of intelligent picking robots.

5.2.2. Sensor Fusion and Algorithm Optimization

In the face of complex and ever-changing operating environments and high real-time requirements for harvesting, multiple strategies can be adopted to solve the problem. The sensing ability of picking robots can be improved by adopting the strategy of multi-mode sensor fusion and multi-algorithm fusion sensing so as to identify and locate the fruit and vegetable more accurately in a complex environment. In addition, the application of the neural network model in the picking field can effectively shorten the sensing time, and the real-time picking robot can be further improved by selecting a better hardware processor and continuously optimizing the control algorithm. In the future, with the continuous development of various new technologies, we have every reason to believe that fruit and vegetable harvesting robots will become more intelligent, efficient, and flexible, bringing further benefits to agricultural production.

5.2.3. Strengthening Research on Walking Platform and Navigation Algorithm

For the walking platform, first of all, it shall be ensured that the platform has good terrain adaptability so that it can maintain good stability and reliability in different terrain environments such as flat ground, slope, and grassland. Second, the walking platform shall be equipped with the necessary sensors, or the sensing system of the robot shall be invoked to serve the walking platform when the mechanical arm and the end actuator are idle so that the platform has basic environmental awareness and obstacle avoidance functions when moving between the picking areas. In addition, the research on the navigation algorithm of the walking platform can refer to the path-planning technology of the terminal actuator and the mechanical arm introduced in Section 3.3 and the obstacle avoidance strategy mentioned in Section 3.2. Using optimization algorithms to optimize traditional path-planning algorithms or using deep learning-based path-planning algorithms can shorten the movement time of robots between different picking areas, thereby effectively improving the efficiency of picking robots, especially for long-term operations.

5.2.4. The Development of Picking Drones

Due to the limited picking height of common ground picking robots, it is necessary to explore new picking modes in order to meet the picking needs of higher fruit trees. Thanks to the great flexibility of multi-rotor drones in three-dimensional space, harvesting operations based on multi-rotor drone platforms are an ideal way to solve such problems. At the same time, the tethered power supply mode can greatly extend the operation time of the picking drones, and theoretically, it can achieve 24 h of uninterrupted operation. In addition, if multiple harvesting drones work together, it can greatly improve harvesting efficiency. In terms of drone harvesting, Israel's Tevel company has achieved good results, but overall, there is still relatively little research and application of drone harvesting. In the future, it should improve its sensing ability and balance control ability in a complex environ-

ment and promote its further development and application in light-weight, miniaturization, clustering, and non-destructive picking.

### 5.2.5. Multi-Robot Collaborative Operation

With the expansion of agricultural scale and the increasing complexity of harvesting tasks, a single harvesting robot may not be able to efficiently complete all tasks. Therefore, multi-robot collaboration has become a trend that can improve the overall production efficiency and quality of harvesting. Multi-robot collaboration can be based on a distributed concept, where different tasks are assigned to multiple robots and executed. These robots can work collaboratively, share information, and allocate tasks via wireless communication or LAN. This cooperation is not only limited to the cooperation between different picking areas but also, each robot can perform different picking operations according to the task requirements and its own capabilities. For example, one robot is responsible for identifying and locating the target, and the other robot is responsible for precise grabbing and shearing, which can avoid the limitation of the picking robot's own vision. In addition, multi-robot cooperation can also better realize obstacle avoidance and provide a better picking environment for the picking robot so as to optimize the picking path and improve the picking efficiency. Of course, the advantages of the multi-robot cooperative picking mode are not limited to this; it has a huge development space and prospects, and its development and application in the future are worthy of expectation.

### 5.2.6. Reducing Costs

Researchers have adopted various methods to address the high costs of fruit and vegetable harvesting robots. One approach involves utilizing modular design and manufacturing to reduce production costs and increase efficiency. Another method employs new materials and manufacturing technologies to lower material costs and ease manufacturing difficulties. Moreover, with continuous technological advancements and market expansion, the production scale of fruit and vegetable harvesting robots will continue to grow, further reducing costs. Thus, although high costs constitute a significant issue, it will gradually be addressed with ongoing technological developments and market expansion.

## 6. Conclusions

This paper systematically reviewed the research progress of the "eye, brain, and hand" picking system in the past six years and discussed its potential impact and innovation value in the field of modern agriculture. Through the gradual analysis of each part of the article, we can understand the technical realization and application prospects of this intelligent agricultural picking robot, which brings unprecedented opportunities for the future of agricultural production.

In the detailed discussion of each section, this review provides insight into the core elements of the "eye–brain–hand" picking system. From the "eye" part using advanced sensors and image processing technology to achieve the accurate judgment of crop maturity, to the "brain" part through advanced algorithms to achieve real-time decision-making and guidance, and finally to the "hand" part to achieve accurate picking. This intelligent picking system not only improves agricultural production efficiency but also reduces resource waste and human investment, playing a positive role in the green and sustainable development of global agriculture.

The main contribution of this review is a comprehensive analysis of the "eye–brain–hand" picking system. From hardware modules to technical approaches, from potential challenges to future trends, this review provides valuable guidance for researchers. Section 2 provides a detailed introduction to the perception hardware system of the intelligent picking "eye" system. In terms of target sensing methods, we compare a variety of methods, which provide important guidance for achieving high-precision target detection. In Section 3, the intelligent picking "brain" system deeply studies the key issues such as regional division, task allocation, obstacle avoidance strategy, and path planning. In

this section, the importance of task allocation and obstacle avoidance strategies for robot agricultural operations is emphasized, providing key support for ensuring the efficient and safe operation of robots. Section 4 systematically reviews the performance indicators of four end-effectors, namely, negative pressure adsorption, shear, cavity trapping, and flexible grasping, for the intelligent harvesting "hand" system. Through analyzing the evaluation index of picking effect, we provide a valuable reference for the type selection and design of end-effectors for different kinds of fruit. In Section 5, "Challenges and Prospects", we identify the challenges faced by intelligent agricultural picking robots and also provide some prospects for their future development.

Through the comprehensive explanation of the above conclusions, this paper provides a deep insight into the agricultural picking robot field and provides important guidance and enlightenment for future research and application. In a word, the development of agricultural picking robots is not only the embodiment of technological progress but also the key step in the agricultural field toward sustainable development. We believe that the content of this review will make a beneficial contribution to the goals of agricultural modernization and sustainable development and promote the wide application and development of intelligent agriculture.

**Author Contributions:** Conceptualization, X.H. (Xiongkui He) and S.W.; methodology, W.J. and X.H. (Xianhao Huang); analysis, W.J. and X.H. (Xianhao Huang); investigation, S.W., W.J. and X.H. (Xianhao Huang); resources, X.H. (Xiongkui He), S.W., W.J. and X.H. (Xianhao Huang); data curation, W.J. and X.H. (Xianhao Huang); writing—original draft preparation, S.W., W.J. and X.H. (Xianhao Huang); writing—review and editing, X.H. (Xiongkui He), S.W., W.J. and X.H. (Xianhao Huang); visualization, W.J. and X.H. (Xianhao Huang); supervision, X.H. (Xiongkui He) and S.W.; project administration, X.H. (Xiongkui He) and S.W.; funding acquisition, X.H. (Xiongkui He) and S.W. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Gil, G.; Casagrande, D.E.; Cortés, L.P.; Verschae, R. Why the Low Adoption of Robotics in the Farms? Challenges for the Establishment of Commercial Agricultural Robots. *Smart Agric. Technol.* **2023**, *3*, 100069. [CrossRef]
2. Suresh Kumar, M.; Mohan, S. Selective Fruit Harvesting: Research, Trends and Developments towards Fruit Detection and Localization—A Review. *Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci.* **2023**, *237*, 1405–1444. [CrossRef]
3. Rose, D.C.; Bhattacharya, M. Adoption of Autonomous Robots in the Soft Fruit Sector: Grower Perspectives in the UK. *Smart Agric. Technol.* **2023**, *3*, 100118. [CrossRef]
4. Tang, Q.; Luo, Y.W.; Wu, X. Di Research on the Evaluation Method of Agricultural Intelligent Robot Design Solutions. *PLoS ONE* **2023**, *18*, e0281554. [CrossRef] [PubMed]
5. Kuta, Ł.; Li, Z.; Stopa, R.; Komarnicki, P.; Słupska, M. The Influence of Manual Harvesting on the Quality of Picked Apples and the Picker's Muscle Load. *Comput. Electron. Agric.* **2020**, *175*, 105511. [CrossRef]
6. Liu, J.; Peng, Y.; Faheem, M. Experimental and Theoretical Analysis of Fruit Plucking Patterns for Robotic Tomato Harvesting. *Comput. Electron. Agric.* **2020**, *173*, 105330. [CrossRef]
7. Xiong, Y.; Ge, Y.; From, P.J. Push and Drag: An Active Obstacle Separation Method for Fruit Harvesting Robots. In Proceedings of the IEEE International Conference on Robotics and Automation, Paris, France, 31 May–31 August 2020.

8.  Li, K.; Qi, Y. Motion Planning of Robot Manipulator for Cucumber Picking. In Proceedings of the 2018 3rd International Conference on Robotics and Automation Engineering, ICRAE 2018, Guangzhou, China, 17–19 November 2018.

9.  Wu, Z.; Du, H. Artificial Intelligence in Agricultural Picking Robot Displacement Trajectory Tracking Control Algorithm. *Wirel Commun. Mob. Comput.* **2022**, *2022*, 3105909. [CrossRef]

10. Wang, Z.; Xun, Y.; Wang, Y.; Yang, Q. Review of Smart Robots for Fruit and Vegetable Picking in Agriculture. *Int. J. Agric. Biol. Eng.* **2022**, *15*, 33–54. [CrossRef]

11. Li, Y.; Feng, Q.; Li, T.; Xie, F.; Liu, C.; Xiong, Z. Advance of Target Visual Information Acquisition Technology for Fresh Fruit Robotic Harvesting: A Review. *Agronomy* **2022**, *12*, 1336. [CrossRef]

12. Lin, G.; Tang, Y.; Zou, X.; Xiong, J.; Li, J. Guava Detection and Pose Estimation Using a Low-Cost RGB-D Sensor in the Field. *Sensors* **2019**, *19*, 428. [CrossRef]

13. Zheng, C.; Chen, P.; Pang, J.; Yang, X.; Chen, C.; Tu, S.; Xue, Y. A Mango Picking Vision Algorithm on Instance Segmentation and Key Point Detection from RGB Images in an Open Orchard. *Biosyst. Eng.* **2021**, *206*, 32–54. [CrossRef]

14. Garillos-Manliguez, C.A.; Chiang, J.Y. Multimodal Deep Learning and Visible-Light and Hyperspectral Imaging for Fruit Maturity Estimation. *Sensors* **2021**, *21*, 1288. [CrossRef]

15. Xu, N.; Song, Y.; Meng, Q. Application RFID and Wi-FI Technology in Design of IOT Sensor Terminal. In Proceedings of the Journal of Physics: Conference Series, Chongqing, China, 28–30 May 2021; Volume 1982.

16. Chen, M.; Tang, Y.; Zou, X.; Huang, Z.; Zhou, H.; Chen, S. 3D Global Mapping of Large-Scale Unstructured Orchard Integrating Eye-in-Hand Stereo Vision and SLAM. *Comput. Electron. Agric.* **2021**, *187*, 106237. [CrossRef]

17. Chen, M.; Tang, Y.; Zou, X.; Huang, K.; Huang, Z.; Zhou, H.; Wang, C.; Lian, G. Three-Dimensional Perception of Orchard Banana Central Stock Enhanced by Adaptive Multi-Vision Technology. *Comput. Electron. Agric.* **2020**, *174*, 105508. [CrossRef]

18. Mahanti, N.K.; Pandiselvam, R.; Kothakota, A.; Ishwarya, S.P.; Chakraborty, S.K.; Kumar, M.; Cozzolino, D. Emerging Non-Destructive Imaging Techniques for Fruit Damage Detection: Image Processing and Analysis. *Trends Food Sci. Technol.* **2022**, *120*, 418–438. [CrossRef]

19. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep Learning—Method Overview and Review of Use for Fruit Detection and Yield Estimation. *Comput. Electron. Agric.* **2019**, *162*, 219–234. [CrossRef]

20. Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and Localization Methods for Vision-Based Fruit Picking Robots: A Review. *Front. Plant Sci.* **2020**, *11*, 510. [CrossRef]

21. Mohd Ali, M.; Hashim, N.; Abd Aziz, S.; Lasekan, O. Utilisation of Deep Learning with Multimodal Data Fusion for Determination of Pineapple Quality Using Thermal Imaging. *Agronomy* **2023**, *13*, 401. [CrossRef]

22. Yang, F.; Ma, Z.; Xie, M. Image Classification with Superpixels and Feature Fusion Method. *J. Electron. Sci. Technol.* **2021**, *19*, 100096. [CrossRef]

23. Shivendra; Chiranjeevi, K.; Tripathi, M.K. Detection of Fruits Image Applying Decision Tree Classifier Techniques. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer Nature: Singapore, 2023; Volume 142.

24. Zhang, C.; Wang, H.; Fu, L.H.; Pei, Y.H.; Lan, C.Y.; Hou, H.Y.; Song, H. Three-Dimensional Continuous Picking Path Planning Based on Ant Colony Optimization Algorithm. *PLoS ONE* **2023**, *18*, e0282334. [CrossRef]

25. He, Z.; Ma, L.; Wang, Y.; Wei, Y.; Ding, X.; Li, K.; Cui, Y. Double-Arm Cooperation and Implementing for Harvesting Kiwifruit. *Agriculture* **2022**, *12*, 1763. [CrossRef]

26. Yang, C.; Liu, Y.; Wang, Y.; Xiong, L.; Xu, H.; Zhao, W. Research and Experiment on Recognition and Location System for Citrus Picking Robot in Natural Environment. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 14–22. [CrossRef]

27. Peng, H.; Shao, Y.; Chen, K.; Deng, Y.; Xue, C. Research on Multi-Class Fruits Recognition Based on Machine Vision and SVM. *IFAC-Pap.* **2018**, *51*, 817–821. [CrossRef]

28. Udhaya, K.; Miruthula, R.; Pavithra, G.; Revathi, R.; Suganya, M. FPGA-Based Hardware Acceleration for Fruit Recognition Using SVM. *Ir. Interdiscip. J. Sci. Res.* **2022**, *06*, 22–29. [CrossRef]

29. Xu, L.; Cao, M.; Song, B. A New Approach to Smooth Path Planning of Mobile Robot Based on Quartic Bezier Transition Curve and Improved PSO Algorithm. *Neurocomputing* **2022**, *473*, 98–106. [CrossRef]

30. Guo, Y.; Wang, W.; Wu, S. Modeling Method of Mobile Robot Workspace. *IEEE* **2017**, 2146–2150.

31. Chen, W.; Xu, T.; Liu, J.; Wang, M.; Zhao, D. Picking Robot Visual Servo Control Based on Modified Fuzzy Neural Network Sliding Mode Algorithms. *Electronics* **2019**, *8*, 605. [CrossRef]

32. Dai, Y.; Zhang, R.; Ma, L. Path Planning and Tracking Control of Picking Robot Based on Improved A* Algorithm. *J. Chin. Agric. Mech.* **2022**, *43*, 138. [CrossRef]

33. Ma, Y.; Zhang, W.; Qureshi, W.S.; Gao, C.; Zhang, C.; Li, W. Autonomous Navigation for a Wolfberry Picking Robot Using Visual Cues and Fuzzy Control. *Inf. Process. Agric.* **2021**, *8*, 15–26. [CrossRef]

34. Zhang, F.; Chen, Z.; Wang, Y.; Bao, R.; Chen, X.; Fu, S.; Tian, M.; Zhang, Y. Research on Flexible End-Effectors with Humanoid Grasp Function for Small Spherical Fruit Picking. *Agriculture* **2023**, *13*, 123. [CrossRef]

35. Xu, L.; Liu, X.; Zhang, K.; Xing, J.; Yuan, Q.; Chen, J.; Duan, Z.; Ma, S.; Yu, C. Design and Test of End-Effector for Navel Orange Picking Robot. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 53–61. [CrossRef]

36. Guo, T.; Zheng, Y.; Bo, W.; Liu, J.; Pi, J.; Chen, W.; Deng, J. Research on the Bionic Flexible End-Effector Based on Tomato Harvesting. *J. Sens.* **2022**, *2022*, 1–14. [CrossRef]

37. Gharakhani, H.; Thomasson, J.A.; Lu, Y. An End-Effector for Robotic Cotton Harvesting. *Smart Agric. Technol.* **2022**, *2*, 100043. [CrossRef]

38. Xiao, X.; Wang, Y.; Jiang, Y. End-Effectors Developed for Citrus and Other Spherical Crops. *Appl. Sci.* **2022**, *12*, 7945. [CrossRef]

39. Hu, G.; Chen, C.; Chen, J.; Sun, L.; Sugirbay, A.; Chen, Y.; Jin, H.; Zhang, S.; Bu, L. Simplified 4-DOF Manipulator for Rapid Robotic Apple Harvesting. *Comput. Electron. Agric.* **2022**, *199*, 107177. [CrossRef]

40. Chen, M.; Chen, F.; Zhou, W.; Zuo, R. Design of Flexible Spherical Fruit and Vegetable Picking End-Effector Based on Vision Recognition. In *Proceedings of the Journal of Physics: Conference Series*; IOP Publishing: Bristol, UK, 2022; Volume 2246.

41. Gao, J.; Zhang, F.; Zhang, J.; Yuan, T.; Yin, J.; Guo, H.; Yang, C. Development and Evaluation of a Pneumatic Finger-like End-Effector for Cherry Tomato Harvesting Robot in Greenhouse. *Comput. Electron. Agric.* **2022**, *197*, 106879. [CrossRef]

42. Lu, W.; Wang, P.; Du, X.; Ma, Z. *Design and Experiment of a Multi-Knuckle End-Effector for Tomato Picking Robot*; American Society of Agricultural and Biological Engineers: St. Joseph Charter Township, MI, USA, 2022.

43. Oliveira, F.; Tinoco, V.; Magalhaes, S.; Santos, F.N.; Silva, M.F. End-Effectors for Harvesting Manipulators-State Of The Art Review. In Proceedings of the 2022 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2022, Santa Maria da Feira, Portugal, 29–30 April 2022.

44. Yang, Q.; Luo, S.; Chang, C.; Xun, Y.; Bao, G. Segmentation Algorithm for Hangzhou White Chrysanthemums Based on Least Squares Support Vector Machine. *Int. J. Agric. Biol. Eng.* **2019**, *12*, 127–134. [CrossRef]

45. Du, X.; Yang, X.; Ji, J.; Jin, X.; Chen, L. Design and Test of a Pineapple Picking End-Effector. *Appl. Eng. Agric.* **2019**, *35*, 1045–1055. [CrossRef]

46. Lin, G.; Tang, Y.; Zou, X.; Xiong, J.; Fang, Y. Color-, Depth-, and Shape-Based 3D Fruit Detection. *Precis. Agric.* **2020**, *21*, 1–17. [CrossRef]

47. Ning, Z.; Luo, L.; Ding, X.M.; Dong, Z.; Yang, B.; Cai, J.; Chen, W.; Lu, Q. Recognition of Sweet Peppers and Planning the Robotic Picking Sequence in High-Density Orchards. *Comput. Electron. Agric.* **2022**, *196*, 106878. [CrossRef]

48. 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]

49. Zhang, Q.; Liu, F.; Jiang, X.; Xiong, Z.; Xu, C. Motion Planning Method and Experiments of Tomato Bunch Harvesting Manipulator. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 149–156.

50. Yu, F.; Zhou, C.; Yang, X.; Guo, Z.; Chen, C. Design and Experiment of Tomato Picking Robot in Solar Greenhouse. *Nongye Jixie Xuebao/Trans. Chin. Soc. Agric. Mach.* **2022**, *53*, 41–49. [CrossRef]

51. Li, T.; Qiu, Q.; Zhao, C.; Xie, F. Task Planning of Multi-Arm Harvesting Robots for High-Density Dwarf Orchards. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 1–10. [CrossRef]

52. Kang, H.; Zhou, H.; Chen, C. Visual Perception and Modeling for Autonomous Apple Harvesting. *IEEE Access* **2020**, *8*, 62151–62163. [CrossRef]

53. Sarabu, H.; Ahlin, K.; Hu, A.P. Graph-Based Cooperative Robot Path Planning in Agricultural Environments. In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM, Hong Kong, China, 8–12 July 2019; Volume 2019.

54. Gong, L.; Wang, W.; Wang, T.; Liu, C. Robotic Harvesting of the Occluded Fruits with a Precise Shape and Position Reconstruction Approach. *J. Field Robot.* **2022**, *39*, 69–84. [CrossRef]

55. Fu, L.; Majeed, Y.; Zhang, X.; Karkee, M.; Zhang, Q. Faster R–CNN–Based Apple Detection in Dense-Foliage Fruiting-Wall Trees Using RGB and Depth Features for Robotic Harvesting. *Biosyst. Eng.* **2020**, *197*, 245–256. [CrossRef]

56. Xiong, Y.; Ge, Y.; From, P.J. An Improved Obstacle Separation Method Using Deep Learning for Object Detection and Tracking in a Hybrid Visual Control Loop for Fruit Picking in Clusters. *Comput. Electron. Agric.* **2021**, *191*, 106508. [CrossRef]

57. Lv, J.; Wang, Y.; Xu, L.; Gu, Y.; Zou, L.; Yang, B.; Ma, Z. A Method to Obtain the Near-Large Fruit from Apple Image in Orchard for Single-Arm Apple Harvesting Robot. *Sci. Hortic.* **2019**, *257*, 108758. [CrossRef]

58. Wang, L.; Li, H.R.; Zhou, K.; Mu, B. Design of Binocular Vision System for Fruit and Vegetable Picking Based on Embedded Arm. *Guangdianzi Jiguang/J. Optoelectron. Laser* **2020**, *31*, 71–80. [CrossRef]

59. Yu, X.; Fan, Z.; Wang, X.; Wan, H.; Wang, P.; Zeng, X.; Jia, F. A Lab-Customized Autonomous Humanoid Apple Harvesting Robot. *Comput. Electr. Eng.* **2021**, *96*, 107459. [CrossRef]

60. Zhou, T.; Zhang, D.; Zhou, M.; Xi, H.; Chen, X. System Design of Tomatoes Harvesting Robot Based on Binocular Vision. In Proceedings of the 2018 Chinese Automation Congress, CAC 2018, Xi'an, China, 30 November 2019.

61. Jin, Z.; Sun, W.; Zhang, J.; Shen, C.; Zhang, H.; Han, S. Intelligent Tomato Picking Robot System Based on Multimodal Depth Feature Analysis Method. In *Proceedings of the IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2020; Volume 440.

62. Ye, L.; Duan, J.; Yang, Z.; Zou, X.; Chen, M.; Zhang, S. Collision-Free Motion Planning for the Litchi-Picking Robot. *Comput. Electron. Agric.* **2021**, *185*, 106151. [CrossRef]

63. Oktarina, Y.; Dewi, T.; Risma, P.; Nawawi, M. Tomato Harvesting Arm Robot Manipulator; A Pilot Project. In Proceedings of the Journal of Physics: Conference Series, South Sumatera, Indonesia, 9–10 October 2020; Volume 1500.

64. Feng, Q.; Zou, W.; Fan, P.; Zhang, C.; Wang, X. Design and Test of Robotic Harvesting System for Cherry Tomato. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 96–100. [CrossRef]

65. Sepulveda, D.; Fernandez, R.; Navas, E.; Armada, M.; Gonzalez-De-Santos, P. Robotic Aubergine Harvesting Using Dual-Arm Manipulation. *IEEE Access* **2020**, *8*, 121889–121904. [CrossRef]

66. Feng, Q.; Chen, J.; Zhang, M.; Wang, X. Design and Test of Harvesting Robot for Table-Top Cultivated Strawberry. In Proceedings of the WRC SARA 2019—World Robot Conference Symposium on Advanced Robotics and Automation 2019, Beijing, China, 21–22 August 2019.

67. Xiong, Y.; Ge, Y.; Grimstad, L.; From, P.J. An Autonomous Strawberry-harvesting Robot: Design, Development, Integration, and Field Evaluation. *J. Field Robot.* **2020**, *37*, 202–224. [CrossRef]

68. Zhuang, J.; Hou, C.; Tang, Y.; He, Y.; Guo, Q.; Zhong, Z.; Luo, S. Computer Vision-Based Localisation of Picking Points for Automatic Litchi Harvesting Applications towards Natural Scenarios. *Biosyst. Eng.* **2019**, *187*, 1–20. [CrossRef]

69. Septiarini, A.; Hamdani, H.; Hatta, H.R.; Anwar, K. Automatic Image Segmentation of Oil Palm Fruits by Applying the Contour-Based Approach. *Sci. Hortic.* **2020**, *261*, 108939. [CrossRef]

70. Mao, S.; Li, Y.; Ma, Y.; Zhang, B.; Zhou, J.; Kai, W. Automatic Cucumber Recognition Algorithm for Harvesting Robots in the Natural Environment Using Deep Learning and Multi-Feature Fusion. *Comput. Electron. Agric.* **2020**, *170*, 105254. [CrossRef]

71. Xiong, Y.; Peng, C.; Grimstad, L.; From, P.J.; Isler, V. Development and Field Evaluation of a Strawberry Harvesting Robot with a Cable-Driven Gripper. *Comput. Electron. Agric.* **2019**, *157*, 392–402. [CrossRef]

72. Liu, D.; Shen, J.; Yang, H.; Niu, Q.; Guo, Q. Recognition and Localization of Actinidia Arguta Based on Image Recognition. *EURASIP J. Image Video Process.* **2019**, *2019*, 21. [CrossRef]

73. Kurpaska, S.; Bielecki, A.; Sobol, Z.; Bielecka, M.; Habrat, M.; Śmigielski, P. The Concept of the Constructional Solution of the Working Section of a Robot for Harvesting Strawberries. *Sensors* **2021**, *21*, 3933. [CrossRef]

74. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment Based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]

75. Yang, C.H.; Xiong, L.Y.; Wang, Z.; Wang, Y.; Shi, G.; Kuremot, T.; Zhao, W.H.; Yang, Y. Integrated Detection of Citrus Fruits and Branches Using a Convolutional Neural Network. *Comput. Electron. Agric.* **2020**, *174*, 105469. [CrossRef]

76. Fu, L.; Duan, J.; Zou, X.; Lin, J.; Zhao, L.; Li, J.; Yang, Z. Fast and Accurate Detection of Banana Fruits in Complex Background Orchards. *IEEE Access* **2020**, *8*, 196835–196846. [CrossRef]

77. Liang, C.; Xiong, J.; Zheng, Z.; Zhong, Z.; Li, Z.; Chen, S.; Yang, Z. A Visual Detection Method for Nighttime Litchi Fruits and Fruiting Stems. *Comput. Electron. Agric.* **2020**, *169*, 105192. [CrossRef]

78. 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]

79. Yu, Y.; Zhang, K.; Liu, H.; Yang, L.; Zhang, D. Real-Time Visual Localization of the Picking Points for a Ridge-Planting Strawberry Harvesting Robot. *IEEE Access* **2020**, *8*, 116556–116568. [CrossRef]

80. Xu, Z.F.; Jia, R.S.; Sun, H.M.; Liu, Q.M.; Cui, Z. Light-YOLOv3: Fast Method for Detecting Green Mangoes in Complex Scenes Using Picking Robots. *Appl. Intell.* **2020**, *50*, 4670–4687. [CrossRef]

81. Xu, Z.F.; Jia, R.S.; Liu, Y.B.; Zhao, C.Y.; Sun, H.M. Fast Method of Detecting Tomatoes in a Complex Scene for Picking Robots. *IEEE Access* **2020**, *8*, 55289–55299. [CrossRef]

82. Chen, J.; Wang, Z.; Wu, J.; Hu, Q.; Zhao, C.; Tan, C.; Teng, L.; Luo, T. An Improved Yolov3 Based on Dual Path Network for Cherry Tomatoes Detection. *J. Food Process. Eng.* **2021**, *44*, e13803. [CrossRef]

83. Wang, Y.; Yan, G.; Meng, Q.; Yao, T.; Han, J.; Zhang, B. DSE-YOLO: Detail Semantics Enhancement YOLO for Multi-Stage Strawberry Detection. *Comput. Electron. Agric.* **2022**, *198*, 107057. [CrossRef]

84. Li, X.; Pan, J.; Xie, F.; Zeng, J.; Li, Q.; Huang, X.; Liu, D.; Wang, X. Fast and Accurate Green Pepper Detection in Complex Backgrounds via an Improved Yolov4-Tiny Model. *Comput. Electron. Agric.* **2021**, *191*, 106503. [CrossRef]

85. Wang, L.; Zhao, Y.; Liu, S.; Li, Y.; Chen, S.; Lan, Y. Precision Detection of Dense Plums in Orchards Using the Improved YOLOv4 Model. *Front. Plant Sci.* **2022**, *13*, 839269. [CrossRef] [PubMed]

86. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [CrossRef]

87. Qian, Y.; Jiacheng, R.; Pengbo, W.; Zhan, Y.; Changxing, G. Real-Time Detection and Localization Using SSD Method for Oyster Mushroom Picking Robot. In Proceedings of the 2020 IEEE International Conference on Real-Time Computing and Robotics, RCAR 2020, Asahikawa, Japan, 28–29 September 2020.

88. Liu, J.; Zhao, M.; Guo, X. A Fruit Detection Algorithm Based on R-FCN in Natural Scene. In Proceedings of the 32nd Chinese Control and Decision Conference, CCDC 2020, Hefei, China, 22–24 August 2020.

89. Li, J.; Tang, Y.; Zou, X.; Lin, G.; Wang, H. Detection of Fruit-Bearing Branches and Localization of Litchi Clusters for Vision-Based Harvesting Robots. *IEEE Access* **2020**, *8*, 117746–117758. [CrossRef]

90. Peng, H.; Xue, C.; Shao, Y.; Chen, K.; Xiong, J.; Xie, Z.; Zhang, L. Semantic Segmentation of Litchi Branches Using Deeplabv3+ Model. *IEEE Access* **2020**, *8*, 164546–164555. [CrossRef]

91. Xiong, Y.; Ge, Y.; From, P.J. An Obstacle Separation Method for Robotic Picking of Fruits in Clusters. *Comput. Electron. Agric.* **2020**, *175*, 105397. [CrossRef]

92. Mghames, S.; Hanheide, M.; Ghalamzan, E.A. Interactive Movement Primitives: Planning to Push Occluding Pieces for Fruit Picking. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Las Vegas, NV, USA, 24 October 2020–24 January 2021.

93. Cao, X.; Yan, H.; Huang, Z.; Ai, S.; Xu, Y.; Fu, R.; Zou, X. A Multi-Objective Particle Swarm Optimization for Trajectory Planning of Fruit Picking Manipulator. *Agronomy* **2021**, *11*, 2286. [CrossRef]

94. Chen, J.; Qiang, H.; Wu, J.; Xu, G.; Wang, Z. Navigation Path Extraction for Greenhouse Cucumber-Picking Robots Using the Prediction-Point Hough Transform. *Comput. Electron. Agric.* **2021**, *180*, 105911. [CrossRef]

95. Colucci, G.; Botta, A.; Tagliavini, L.; Cavallone, P.; Baglieri, L.; Quaglia, G. Kinematic Modeling and Motion Planning of the Mobile Manipulator Agri.Q for Precision Agriculture. *Machines* **2022**, *10*, 321. [CrossRef]

96. Le Flécher, E.; Durand-Petiteville, A.; Cadenat, V.; Sentenac, T. Visual Predictive Control of Robotic Arms with Overlapping Workspace. In Proceedings of the ICINCO 2019—Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, Prague, Czech Republic, 29–31 July 2019; 2019; Volume 1, pp. 130–137.

97. Ji, W.; Zhang, J.; Xu, B.; Tang, C.; Zhao, D. Grasping Mode Analysis and Adaptive Impedance Control for Apple Harvesting Robotic Grippers. *Comput. Electron. Agric.* **2021**, *186*, 106210. [CrossRef]

98. Wang, Y.; Zhang, H.; Wang, L.; Li, G.; Zhang, Y.; Liu, X. Development of Control System for Cotton Picking Test Bench Based on Fuzzy PID Control. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2018**, *34*, 23–32. [CrossRef]

99. Ramin Shamshiri, R.; Weltzien, C.; Hameed, I.A.; Yule, I.J.; Grift, T.; Balasundram, S.; Pitonakova, L.; Ahmad, D.; Chowdhary, G. Research and Development in Agricultural Robotics: A Perspective of Digital Farming. *Int. J. Agric. Biol. Eng.* **2018**, *11*, 1–11. [CrossRef]

100. Navas, E.; Fernández, R.; Sepúlveda, D.; Armada, M.; Gonzalez-de-Santos, P. Soft Grippers for Automatic Crop Harvesting: A Review. *Sensors* **2021**, *21*, 2689. [CrossRef]

101. Liu, J.; Yuan, Y.; Gao, Y.; Tang, S.; Li, Z. Virtual Model of Grip-and-Cut Picking for Simulation of Vibration and Falling of Grape Clusters. *Trans ASABE* **2019**, *62*, 603–614. [CrossRef]

102. Wei, J.; Yi, D.; Bo, X.; Guangyu, C.; Dean, Z. Adaptive Variable Parameter Impedance Control for Apple Harvesting Robot Compliant Picking. *Complexity* **2020**, *2020*, 1–15. [CrossRef]

103. Miao, Y.; Zheng, J. Optimization Design of Compliant Constant-Force Mechanism for Apple Picking Actuator. *Comput. Electron. Agric.* **2020**, *170*, 105232. [CrossRef]

104. Liu, C.H.; Chiu, C.H.; Chen, T.L.; Pai, T.Y.; Chen, Y.; Hsu, M.C. A Soft Robotic Gripper Module with 3d Printed Compliant Fingers for Grasping Fruits. In Proceedings of the IEEE/ASME International Conference on Advanced Intelligent Mechatronics, AIM, Auckland, New Zealand, 9–12 July 2018; Volume 2018.

105. Pi, J.; Liu, J.; Zhou, K.; Qian, M. An Octopus-Inspired Bionic Flexible Gripper for Apple Grasping. *Agriculture* **2021**, *11*, 1014. [CrossRef]

106. Hohimer, C.J.; Wang, H.; Bhusal, S.; Miller, J.; Mo, C.; Karkee, M. Design and Field Evaluation of a Robotic Apple Harvesting System with a 3D-Printed Soft-Robotic End-Effector. *Trans ASABE* **2019**, *62*, 405–414. [CrossRef]

107. Vu, Q.; Ronzhin, A. Models and algorithms for design robotic gripper for agricultural products. *Comptes Rendus De L'Academie Bulg. Des Sci.* **2020**, *73*, 103–110.

108. Chen, Z.; Yang, M.; Li, Y.; Yang, L. Design and Experiment of Tomato Picking End-Effector Based on Non-Destructive Pneumatic Clamping Control. *Nongye Gongcheng Xuebao/Trans. Chin. Soc. Agric. Eng.* **2021**, *37*, 27–35. [CrossRef]

109. Yung, I.; Maccarana, Y.; Maroni, G.; Previdi, F. Partially Structured Robotic Picking for Automation of Tomato Transplantation. In Proceedings of the 2019 IEEE International Conference on Mechatronics, ICM 2019, Ilmenau, Germany, 18–20 March 2019.

110. Zhang, J.; Lai, S.; Yu, H.; Wang, E.; Wang, X.; Zhu, Z. Fruit Classification Utilizing a Robotic Gripper with Integrated Sensors and Adaptive Grasping. *Math. Probl. Eng.* **2021**, *2021*, 1–15. [CrossRef]

111. Habegger, R.; Bergamo, E.; Schwab, W.; Berninger, T.; Rixen, D. Impact of Intensive Modification of Sweet Pepper Plants on Performance of End Effectors for Autonomous Harvesting. *Eur. J. Hortic. Sci.* **2021**, *86*, 354–359. [CrossRef]

112. De Preter, A.; Anthonis, J.; De Baerdemaeker, J. Development of a Robot for Harvesting Strawberries. *FAC-PapersOnLine* **2018**, *51*, 14–19. [CrossRef]

113. Li, Z.; Miao, F.; Yang, Z.; Wang, H. An Anthropometric Study for the Anthropomorphic Design of Tomato-Harvesting Robots. *Comput. Electron. Agric.* **2019**, *163*, 104881. [CrossRef]

114. Li, Z.; Miao, F.; Yang, Z.; Chai, P.; Yang, S. Factors Affecting Human Hand Grasp Type in Tomato Fruit-Picking: A Statistical Investigation for Ergonomic Development of Harvesting Robot. *Comput. Electron. Agric.* **2019**, *157*, 90–97. [CrossRef]

115. Hou, Z.; Li, Z.; Fadiji, T.; Fu, J. Soft Grasping Mechanism of Human Fingers for Tomato-Picking Bionic Robots. *Comput. Electron. Agric.* **2021**, *182*, 106010. [CrossRef]

116. Öz, E.; Jakob, M. Ergonomic Evaluation of Simulated Apple Hand Harvesting by Using 3D Motion Analysis. *Ege Üniversitesi Ziraat Fakültesi Derg.* **2020**, *57*, 249–256. [CrossRef]

117. Liu, X.; Xu, H.; Chen, F. Research on Vision and Trajectory Planning System for Tomato Picking Robots. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020, Harbin, China, 25–27 December 2020.

118. Zhang, L.; Jia, J.; Gui, G.; Hao, X.; Gao, W.; Wang, M. Deep Learning Based Improved Classification System for Designing Tomato Harvesting Robot. *IEEE Access* **2018**, *6*, 67940–67950. [CrossRef]

119. Horng, G.J.; Liu, M.X.; Chen, C.C. The Smart Image Recognition Mechanism for Crop Harvesting System in Intelligent Agriculture. *IEEE Sens. J.* **2020**, *20*, 2766–2781. [CrossRef]

120. Xiong, Y.; From, P.J.; Isler, V. Design and Evaluation of a Novel Cable-Driven Gripper with Perception Capabilities for Strawberry Picking Robots. In Proceedings of the IEEE International Conference on Robotics and Automation, Brisbane, QLD, Australia, 21–25 May 2018.
121. Zhong, Z.; Xiong, J.; Zheng, Z.; Liu, B.; Liao, S.; Huo, Z.; Yang, Z. A Method for Litchi Picking Points Calculation in Natural Environment Based on Main Fruit Bearing Branch Detection. *Comput. Electron. Agric.* **2021**, *189*, 106398. [CrossRef]