

Article

Not peer-reviewed version

---

# 3D Spatial Perception of Blueberry Fruits Based on Improved YOLOv11 Network

---

[Kun Zhao](#)\*, [Yuhuan Li](#), [Zunmin Liu](#)

Posted Date: 4 February 2025

doi: 10.20944/preprints202502.0232.v1

Keywords: improved YOLOv11; object detection; 3D localization and pose calculation; blueberry picking



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

*Article*

# 3D Spatial Perception of Blueberry Fruits Based on Improved YOLOv11 Network

Kun Zhao \*, Yuhuan Li and Zunmin Liu

School of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao 266520, China

\* Correspondence: zhao\_kun@qut.edu.cn

**Abstract:** Automated harvesting of blueberries using a picking robot places a greater demand on the 3D spatial perception performance of blueberry fruits, as the robot's grasping mechanism needs to pick blueberry fruits accurately at specific positions and in particular poses. To achieve this goal, this paper presents a method for blueberry detection, 3D spatial localization, and pose estimation using visual perception, which can be deployed on an OAK depth camera. Firstly, a blueberry and calyx scar detection dataset is constructed to train the detection network and evaluate its performance. Secondly, the blueberry and calyx scar detection model based on a lightweight yolov11 (the eleventh version of the You Only Look Once) network with an improved depth-wise separable convolution (DSC) module is designed, and a 3D coordinate system relative to the camera is established to calculate the 3D pose of the blueberry fruits. Finally, the above detection model is deployed onto the OAK depth camera, leveraging its depth estimation module and three-axis gyroscope to obtain the 3D coordinates of the blueberry fruits. Experimental results demonstrate that the method proposed in this paper can accurately perceive blueberry fruits at various maturity levels, achieving a detection accuracy of 95.8% mAP50-95, a positioning error of less than 1 centimeter within 0.5 meters, and an average 3D pose error of 19.2 degrees while maintaining the detection frame rate of 13.4 FPS (Frames Per Second) on the OAK depth camera, which can provide effective picking guidance for the mechanical arm of picking robot.

**Keywords:** improved YOLOv11; object detection; 3D localization and pose calculation; blueberry picking

## 1. Introduction

The blueberry harvesting cycle is short, and fruit ripening is inconsistent. Manual picking is labor-intensive, inefficient, and costly [1]. Mechanical harvesting allows large-scale harvesting using methods such as shaking, or tapping. Although the harvesting efficiency is high, the fruit loss rate is significant, and cleaning and sorting of the fruits are required in the later stage [2]. Utilizing robots for automated and precise picking has become an effective solution to address these challenges. For picking robots, the priority is to perceive blueberry fruits in the surrounding environment, accurately determine their position in space, and provide data support for the robotic arm to grasp the fruits. Consequently, accurately perceiving blueberries is essential for robots to execute precise picking tasks.

With the help of some mature object detection frameworks [3,4], research on blueberry perception has been extensive. These researches primarily enhance the blueberry detection performance in the following ways:

(1) By introducing the attention mechanism, the model could learn more detailed and comprehensive representations of blueberry features.

The Convolutional Block Attention Module (CBAM) [5] was introduced to enhance the feature fusion capability of the blueberry detection network [6]. Feng et al. [7] introduced the SE attention module [8] into the YOLOv9 (the ninth version of the You Only Look Once) [9] network, whereas Zhai et al. [10] combined the SE module with BiFPN [11] in the YOLOv5 network to enhance the

model's attention to relevant blueberry features while reducing the influence of unnecessary information. Gai et al. [12] introduced the Multiplexed Coordinated Attention (MPCA) module in the last layer of the backbone to improve the feature extraction capability during network training. YOLO-BLBE [13] proposed the GhostNet [14] module embedded with the Coordinate Attention (CA) mechanism [15] to replace the backbone of YOLOv5s. Yang et al. [16] designed an NCBAM module in the YOLOv5 network, which improved the blueberry feature extraction capability of the backbone network. Liu et al. [17] replaced the C3 module with the Multi-Head Self-Attention (MHSA) module [18] before the SPPF module, enabling the network to learn more comprehensive feature representations and enhancing its understanding of complex spatial relationships and contextual information in blueberry images.

(2) The real-time performance of the detection network is improved by introducing a lightweight convolution module into the network.

Feng et al. [6] introduced the ShuffleNet [19] module, whereas Yang et al. [16] proposed the C3Ghost module, and Wang et al. [13] introduced the GhostNet [14] module in the YOLOv5 networks. Feng et al. [7] introduced the SCConv [20] module in the YOLOv9c network. These lightweight convolution modules significantly reduced the computational complexity of the detection network, which was very advantageous for deploying detection models on terminal devices with limited computing resources.

(3) Improving the detection capability of small-scale blueberry fruits.

A small object detection layer was introduced to improve the multi-scale recognition capability of blueberries [16]. Liu et al. [17] eliminated the maximum object detection layer from the backbone and head of the network to enhance the model's detection capacity for small-scale blueberry fruits.

(4) Addressing occlusion issues.

TL-YOLOv8 [12] proposed the Multi-scale Separation and Occlusion-Aware Module (MultiSEAM) in the network. The multi-scale separation module captured the multi-scale features of blueberries by segmenting the input image at different scales, while the occlusion-aware module could infer the possible shape and position of the occluded part by analyzing the features of the surrounding area of the blueberry fruits. Feng et al. [7] introduced the MDPIoU loss function [21] to enhance the accuracy of the blueberry detection network in occluded environments.

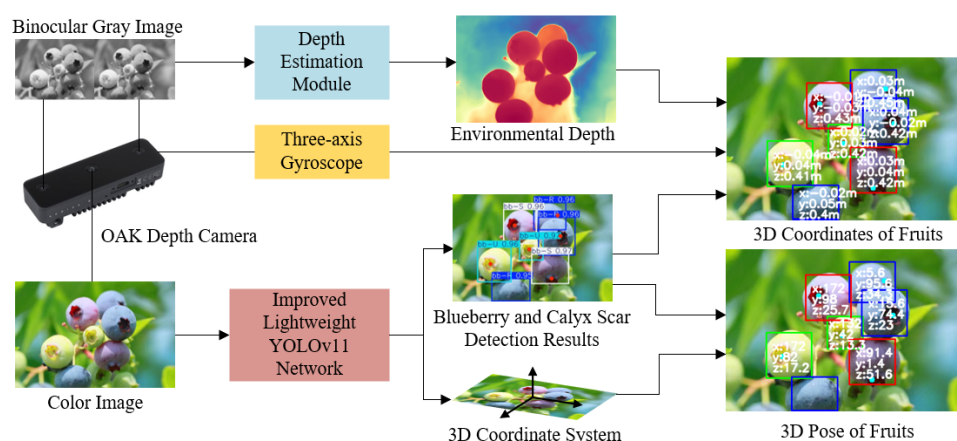
Although the detection performance of the methods above is evaluated on private datasets, which cannot be compared horizontally, the results indicate that their detection accuracy can exceed 80%, and some can even exceed 95%, which can guide robotic arms in picking blueberry fruits. However, compared to other fruits, blueberry fruits have a unique characteristic: their skin is thin, which makes them prone to damage during picking. Particularly in the early stages of blueberry fruit ripening, the connection between the fruit stem and the fruit is quite strong, and grasping fruits directly can easily tear the skin, as shown in Figure 1. To avoid this problem, experienced pickers commonly hold the plane perpendicular to the fruit axis and rotate it to detach from the fruit stem. Therefore, when the robotic arm's picking mechanism picks fruits according to the above pose and action, relying solely on the 3D spatial position of fruits without considering spatial pose is no longer sufficient to satisfy the needs of precise picking by the robotic arm.



**Figure 1.** Peel tearing caused by incorrect picking techniques.

Currently, research on the 3D pose perception of fruits is still limited. Li et al. [22] calculated the symmetry axis using the symmetry plane in the point cloud information of fruits and then evaluated the pose of the fruit in 3D space through the symmetry axis. Christopher Lehnert et al. [23] fitted a hyperelliptic model onto segmented sweet peppers and used nonlinear least squares optimization to estimate the 6DOF pose of the sweet peppers. Lin et al. [24] estimated the 3D pose of the pomegranate fruit by determining its center position and the nearest branch information. Zhang et al. [25] extracted the contours of clustered fruit stems through morphological segmentation, and then employed multidimensional vectors along with Gaussian mixture models to extract the poses of fruit stems. Yun et al. [26] proposed a method to determine the 3D pose of a pomelo by analyzing its boundary shape and applying projection geometry principles using binocular stereo-vision technology.

There is still no research on the 3D spatial pose perception of blueberry fruits. To this end, this paper proposes a blueberry detection, 3D spatial localization, and pose estimation method using visual perception, which can be deployed on an OAK depth camera. As shown in Figure 2, this method employs an improved lightweight YOLOv11 network to detect blueberry fruits and their calyx scars from the color image. Next, a 3D coordinate system is established based on the image plane, and the 3D pose of every blueberry fruit is calculated using the pixel coordinates of the fruit center and calyx scar center. Then, the depth estimation module in the OAK camera is used to gather the environmental depth information, which is combined with the sensing data from the three-axis gyroscope to calculate the 3D coordinates of each blueberry fruit relative to the camera. Finally, the 3D spatial position and pose information of blueberry fruits obtained through this method can be transmitted to the subsequent picking decision and control system.



**Figure 2.** Schematic diagram of blueberry 3D spatial localization and pose calculation.

The remainder of the work is organized as follows. In Section II, the blueberry and calyx scar detection method is described, including the blueberry and calyx scar dataset, and the improved lightweight YOLOv11 detection network. Section III introduces the 3D spatial localization and pose calculation method for blueberry fruits, based on the detection results of blueberries and calyx scars. The experimental results are given in Section VI, and the work of this paper is summarized in Section V.

## 2. Materials and Methods

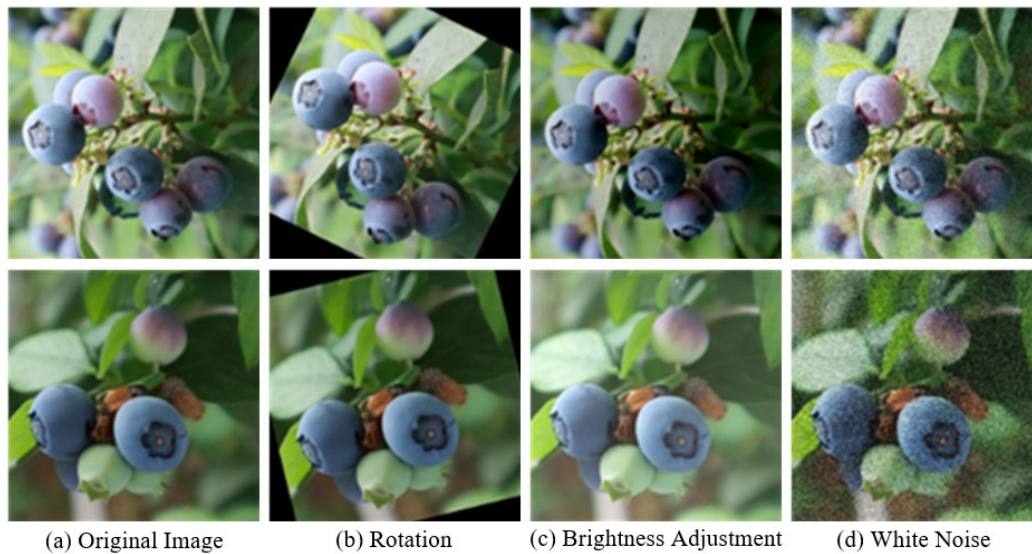
### 2.1. Blueberry and Calyx Scar Detection Dataset

#### 2.1.1. Collection and Partition of Blueberry Images

This paper primarily employs three ways to collect blueberry images. The first way is to take live-action shots at the Wolin Blueberry Base in Qingdao, Shandong, China, obtaining 642 original images, including different blueberry images under sunny, cloudy, backlit, and occluded conditions. The second way is to capture online images using web crawling technology, obtaining 1358 images.



The third way involves data augmentation of 1500 images collected through the first two ways, generating 3000 images, including translation, flipping, rotation, brightness adjustment, and adding white noise. The partial augmentation is shown in Figure 3.



**Figure 3.** Schematic diagram of partial data augmentation.

The dataset, which contains 5000 blueberry images collected through the above three ways, is divided into the training set, validation set, and testing set in an 8:1:1 ratio for detection network training and performance testing. It should be noted that data augmentation can lead to correlations between data. To avoid overfitting in the trained detection network, the 500 samples in the testing set are only the original images that have not undergone data augmentation in the first two collection ways.

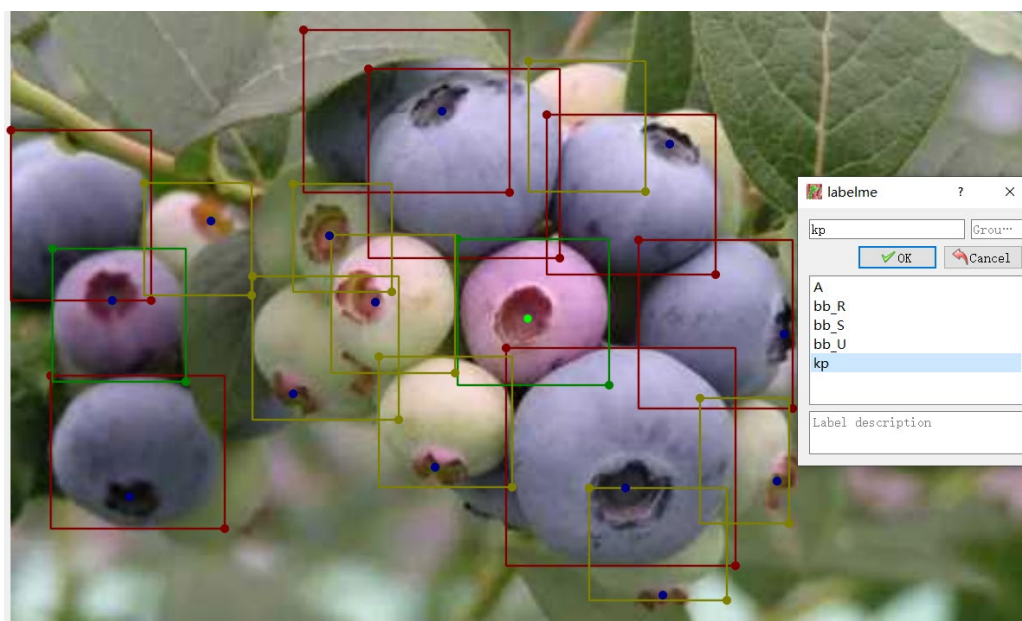
### 2.1.2. Data Annotations

This paper uses a manual annotation software called LabelMe to complete the data labeling task. This software can generate a JSON file containing the image name, size, position, and specific annotation information for each image. During the annotation process, the blueberry fruits are roughly divided into three categories based on maturity: ripe blueberries, semi-ripe blueberries, and unripe blueberries, with colors corresponding to purple red, light red, and cyan, respectively.

The axis information of the blueberry fruit in the image can reflect its 3D pose, which can be determined by connecting the stem or calyx scar center to the fruit center. For clustered blueberries, it is more convenient to detect the calyx scar as their visibility is stronger than that of the fruit stem. Therefore, to obtain the fruit pose, the calyx scar centers of the blueberry fruits are also marked with dots, as shown in Figure 4. Finally, the annotation information of blueberry and calyx scar converted to YOLO format can be represented as:

$$Lable = [cls, c_x, c_y, w, h, kp_x, kp_y, vis], \quad (1)$$

where,  $cls$  represents the blueberry categories, and  $bb\_R$ ,  $bb\_S$ , and  $bb\_U$  represent the three states of blueberry in ripe, semi-ripe, and unripe, respectively;  $c_x$  and  $c_y$  represent the center coordinates of the bounding box, and  $w$  and  $h$  represent the length and width of the bounding box.  $kp_x$  and  $kp_y$  represent the coordinates of the calyx scar center;  $vis$  represents the visibility of the calyx scar center, 'zero' is visible, and 'two' is invisible. When the calyx scar center is invisible, the values of  $kp_x$  and  $kp_y$  are both 'zero'.



**Figure 4.** Schematic diagram of blueberries and calyx scar annotation.

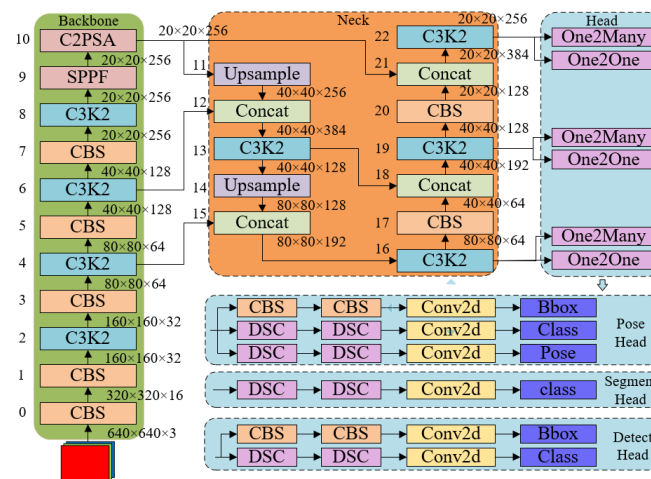
Blueberries grow in clusters and have many individual fruits, making annotation time-consuming and laborious. To reduce workload, this paper divides the annotation into three stages. In the first stage, 500 images are labeled, and a preliminary model for blueberry and calyx scar detection is trained based on the YOLO network. In the second stage, the trained model is used to detect other unlabeled blueberry images, and the detection results are converted into JSON format as prelabeled information. In the last stage, the inaccurate prelabeled results are refined to generate accurate labeling information in LabelMe and finally converted into annotation files that are readable by YOLO.

## 2.2. Blueberry and Calyx Scar Detection Network

When utilizing a robot for blueberry picking, it is crucial to improve the picking efficiency. The YOLO series networks represent a seminal real-time detection framework in the object detection field, renowned for its exceptional detection accuracy coupled with minimal computational overhead, making it a ubiquitous choice for a wide array of embedded devices. The YOLO series networks currently have 11 versions and have spawned multiple variants based on them. YOLOv10 [27] and YOLOv11 [28] adopt the consistent dual assignments strategy, which significantly reduces computational overhead and meets the requirements for blueberry detection.

### 2.2.1. Basic Structure of the YOLOv11n Network

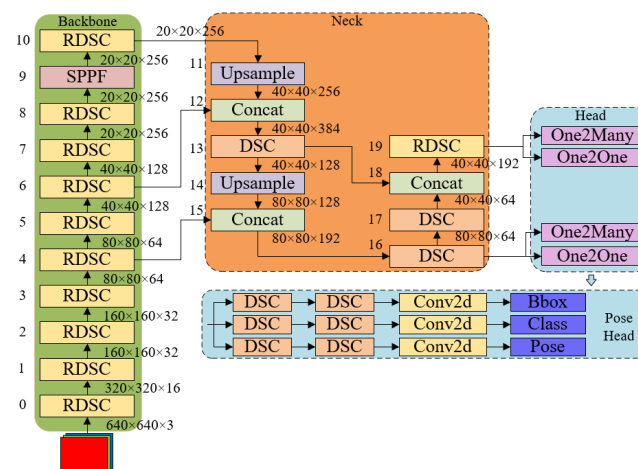
As shown in Figure 5, the YOLOv11n network consists of three parts: Backbone, Neck, and Head. Backbone gradually extracts high-order features of the target by stacking CBS, C3k2, SPPF, and C2PSA modules. Neck has a pyramid structure, including modules such as Concat, Upsample, C3k2, and CBS, which enhance the resolution of high-order features extracted by Backbone and output three feature layers with different scales. Head consists of two branches: one2many and one2one, used for NMS-Free training and inference. The structure of the Head varies depending on different visual tasks such as object detection, pose detection, and instance segmentation.



**Figure 5.** Schematic diagram of YOLOv11n network framework.

### 2.2.2. Improved Lightweight YOLOv11 Network for Blueberry and Calyx Scar Detection

Although the YOLOV11 is designed with different parameter scales to adapt to perception tasks of varying complexity, the parameters of the YOLOV11 network are still redundant for blueberry detection, due to the low diversity of fruit shape (i.e., circular or elliptical) and color (i.e., cyan, red, or purple). Therefore, considering the unique characteristics and requirements for blueberry detection, this paper designs a lightweight detection network based on the consistent dual assignment strategy and YOLOv11 framework, to maximize the accuracy and efficiency of blueberry detection. The specific network structure is shown in Figure 6.

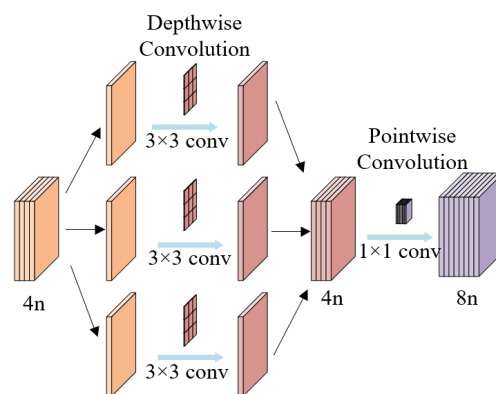


**Figure 6.** Schematic diagram of blueberry and calyx scar detection network framework.

Comparing Figure 5 and Figure 6, this paper has made improvements in the following aspects of blueberry and calyx scar detection:

- Proposing an improved depth-wise separable convolution (DSC) module to replace the CBS and C3K2 modules.

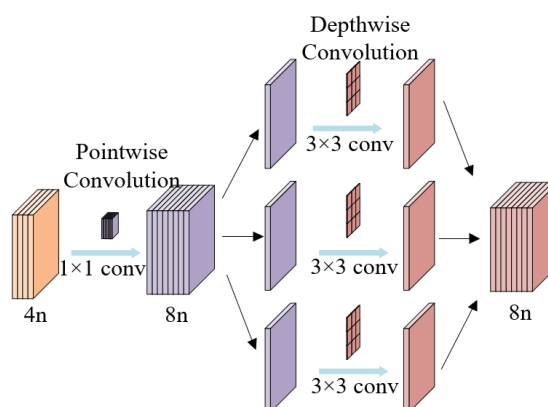
As shown in Figure 7, the DSC module decomposes the traditional convolution into two sequential processes: depth-wise convolution and point-wise convolution. Depth-wise convolution extracts spatial information from each feature channel independently, and point-wise convolution mixes features between channels, significantly reducing network computation while sacrificing some model performance.



**Figure 7.** Schematic diagram of the DSC module.

Subsequent experimental results show that directly using the DSC module instead of the convolution module in YOLOv11 will significantly decrease the accuracy of blueberry and calyx scar detection. We have analyzed this phenomenon and found that the convolution process generally involves an increase or decrease in feature dimensionality. The order of depth-wise convolution and point-wise convolution will affect the ability of each layer to extract blueberry spatial features and the computational complexity of the network. According to the order of depth-wise convolution and point-wise convolution in Figure 7, the feature dimension increases from  $4n$  to  $8n$ , while the extracted spatial features are only  $4n$ , which is insufficient to learn the spatial features of blueberry fruits well.

Therefore, this paper proposes an improved DSC that can adaptively change the order of depth-wise convolution and point-wise convolution based on whether the feature dimension is increased or decreased. When the feature dimension increases, the reverse DSC depicted in Figure 8 is employed to extract features, conversely, the DSC illustrated in Figure 7 is utilized, ensuring that the network is capable of learning more spatial features of blueberries.



**Figure 8.** Schematic diagram of reverse DSC module.

- Removing the attention mechanism module C2PSA at the end of the Backbone network.  
Our experiments find that the attention mechanism module C2PSA did not significantly improve the performance of blueberry detection, but instead produced inefficient computations. Therefore, it is removed from our detection network.
- Retaining only the relatively large scale outputs of features with  $80 \times 80$  pixels and  $40 \times 40$  pixels.  
Given the comparatively small scale of the blueberry fruits in the image, the detection head only outputs detection results at the larger scales of  $80 \times 80$  pixels and  $40 \times 40$  pixels. This diminishes computational complexity and prompts the model to be more apt at learning the intricate features of blueberries, such as the details around the calyx scar.

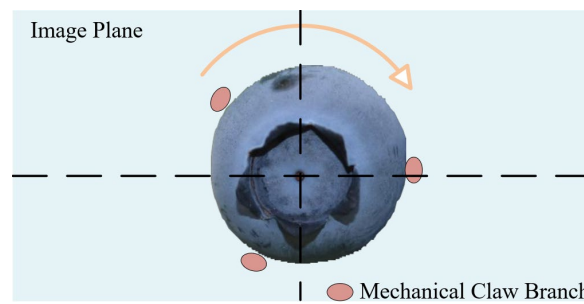


### 2.3. 3D Spatial Localization and Pose Calculation of Blueberry Fruits

#### 2.3.1. 3D Pose Calculation

During blueberry picking, the gripper mechanism must grasp the blueberry fruit in a plane roughly perpendicular to its axis and then rotate around this axis to detach the fruit from the fruit stem, preventing damage to the fruit's skin. Clarifying the pose of blueberry fruits in 3D space is a prerequisite for achieving the above operation.

Assuming the camera is installed at the end of the robotic arm, the ideal picking pose of the blueberry fruit in the image should be as shown in Figure 9. That is, the blueberry fruit is located at the center of the image, and the fruit axis is approximately perpendicular to the image plane. Therefore, we establish a 3D coordinate system relative to the camera using the image plane as the reference plane. This coordinate system takes the image center as the origin, the image plane as the XY-plane, and the direction perpendicular to the image plane as the Z-axis.

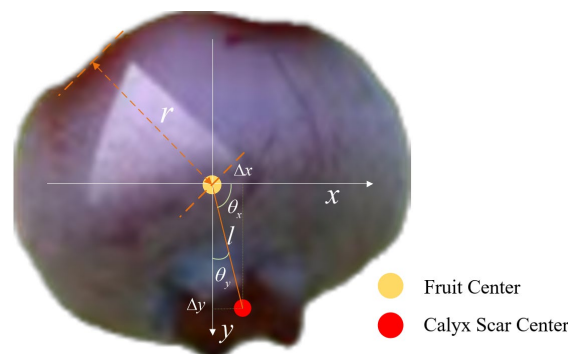


**Figure 9.** Ideal picking pose of blueberry fruit in the image.

As shown in Figure 10, according to the detection results of blueberries and calyx scars, the angles between the fruit axis and the X-axis and Y-axis can be calculated by obtaining the pixel distance  $l$  between the fruit center and the calyx scar center, as well as their components  $\Delta x$  and  $\Delta y$  on the X-axis and Y-axis, respectively. Similarly, the angle between the fruit axis and the Z-axis can be calculated by taking the arcsine value of the pixel distance  $l$  and the fruit radius  $r$ . The calculation formula for blueberry fruits' 3D pose is as follows:

$$Pose_{3D} = \begin{cases} \theta_x = \arccos(\frac{\Delta x}{l}) \\ \theta_y = \arccos(\frac{\Delta y}{l}) \\ \theta_z = \arcsin(\frac{l}{r}) \end{cases} \quad (2)$$

where,  $r$  is the radius of the fruit, due to the shape of blueberry fruits being mostly circular or elliptical in the image plane, half of the average length and width of the bounding box is taken as the approximate pixel radius of the blueberry fruit.



**Figure 10.** Schematic diagram of 3D pose calculation of blueberry fruits.

From Figure 10 and Equation (2), it can be seen that the smaller the pixel distance  $l$ , the smaller the angle between the fruit axis and the Z-axis, which is closer to the ideal picking pose of the blueberry fruit.

2.3.2. 3D Spatial Localization

To ensure the precise picking of blueberry fruits by the robotic arm, it is crucial to obtain accurate relative positional information between the gripper mechanism and the fruits. The OAK-D-Pro depth camera used in this paper contains a depth estimation module that can directly obtain environmental depth information. The depth estimation module has two ranging methods, namely structured light ranging and binocular depth ranging. Through algorithm fusion, the ranging error of the OAK camera within 4 meters is less than 2%, meeting the requirements for blueberry picking. By combining the depth information, the blueberry detection results, and the sensor data from BMI270 three-axis gyroscopes, the 3D coordinate of each fruit relative to the OAK camera can be ascertained.

In this way, by obtaining the 3D spatial position and pose of blueberry fruits, the robotic arm electronic control system can guide the gripper mechanism to reach the designated position and perform precise picking operations according to specific poses and actions.

3. Results and Discussion

The experiment in this paper is divided into three stages. The first stage is to construct and train a blueberry and calyx scar detection model based on the deep learning framework PyTorch, which can obtain the detection results of blueberry fruit and calyx scar center on the image plane in the form of rectangular boxes and key points, respectively. We compared the detection performance of models with different parameter scales and modules, and visualized some of the detection results. The second stage is to calculate the 3D pose information of blueberry fruits on the constructed 3D coordinate system based on the detection results of the first stage. We evaluated the errors of pose estimation in three directions and analyzed them. In the third stage, the trained detection model is converted into blob format and deployed to the OAK depth camera. It can detect blueberry fruits on terminal devices and obtain the 3D coordinates of blueberry fruits relative to the depth camera with the help of its internal depth estimation module. We analyzed whether the positioning error meets the requirements under the blueberry picking task.

3.1. Blueberry and Calyx Scar Detection Results

3.1.1. Experimental Result Statistics

As shown in Table 1, on the blueberry and calyx scar detection dataset, our method can achieve an accuracy of 99.2% mAP50 and 95.8% mAP50-90 for blueberry detection, as well as an accuracy of 87.0% mAP50 and 86.6% mAP50-90 for calyx scar detection. The time consumption of our model to detect a 640 × 640 pixels image on NVIDIA GeForce GTX 4090 and 13900KF CPU is 0.4 ms, making it suitable for picking robots that require fast and precise blueberry detection. Furthermore, Table 1 also displays the specific performance of blueberry detection at three maturity levels, which shows that the detection accuracy of blueberries at all maturity levels is very high.

Table 1. Detection results on blueberries and calyx scar testing set.

Category	Blueberry Detection		Calyx Scar Detection		Time Consumption (ms)
	mAP50 (%)	mAP50-95 (%)	mAP50 (%)	mAP50-95 (%)	
All	99.2	95.8	87.0	86.6	0.4
bb_R	99.5	96.2	87.2	86.8	
bb_S	99.1	94.9	85.4	84.9	
bb_U	99.0	96.2	88.3	88.2	

In addition, we present some ablation experiments, which evaluate the detection performance of DSC, improved DSC (ours), and CBS modules within the YOLO framework across three parameter scales: N, S, and M. For detailed parameters, please refer to the description in YOLOv11.

Table 2 shows that all models can achieve over 93.5% mAP50-90 for blueberry detection, which we believe meets the demand for blueberry picking. For calyx scar detection, the mAP50-95s of all models are relatively low, only around 84.1% to 88.2%. Specifically, the models with the M scale don't improve detection performance compared to models with the S scale, but increase the models' time consumption significantly. Therefore, these models with the M scale are unsuitable for blueberry and calyx scar detection. At the N and S scales, models with DSC modules have the fastest detection speed, but their detection accuracies are only 84.1% and 85.2% mAP50-95. The detection accuracies of models with our improved DSC modules are slightly lower than models with the CBS modules, at 0.7 % and 0.3 % mAP50-95, respectively, but their detection speeds are faster. It can be seen that the improved DSC module proposed in this paper significantly improves detection speed while slightly sacrificing detection accuracy.

**Table 2.** Performance comparison of detection models with different parameter scales and modules.

Para. Scale	Module	Blueberry Detection		Calyx Scar Detection		Time Consumption (ms)
		mAP50-95 (%)	Improvement (%)	mAP50-95 (%)	Improvement (%)	
N	DSC	93.5	base	84.1	base	0.3
	Ours	95.8	+ 2.3	86.6	+ 2.5	0.4
	CBS	96.2	+ 2.7	87.3	+ 3.2	0.6
S	DSC	93.8	+ 0.3	85.2	+ 1.1	0.4
	Ours	96.0	+ 2.5	87.7	+ 3.6	0.5
	CBS	96.3	+ 2.8	88.0	+ 3.9	0.8
M	DSC	93.9	+ 0.4	84.9	+ 0.8	1.6
	Ours	96.1	+ 2.6	87.5	+ 3.4	1.8
	CBS	96.4	+ 2.9	88.2	+ 4.1	2.1

For the blueberry picking task, detection speed is more important than detection accuracy, because missed or obstructed blueberries will become detected blueberries as the viewing angle changes during the continuous picking process. Therefore, this paper selects the model with our improved DSC module at the N scale as the blueberry and calyx scar detection model.

### 3.1.2. Visualization of Detection Instances

To verify the practicality of our detection network, some blueberry images are selected from the testing set for the visual representation of detection results, as shown in Figure 11. As can be seen, our detection network can accurately identify blueberry fruits of different maturity levels, which is beneficial for picking only ripe blueberries and reducing fruit damage rates. At the same time, the point detection of the calyx scar is also basically accurate, which is beneficial for the 3D pose calculation of blueberry fruits.

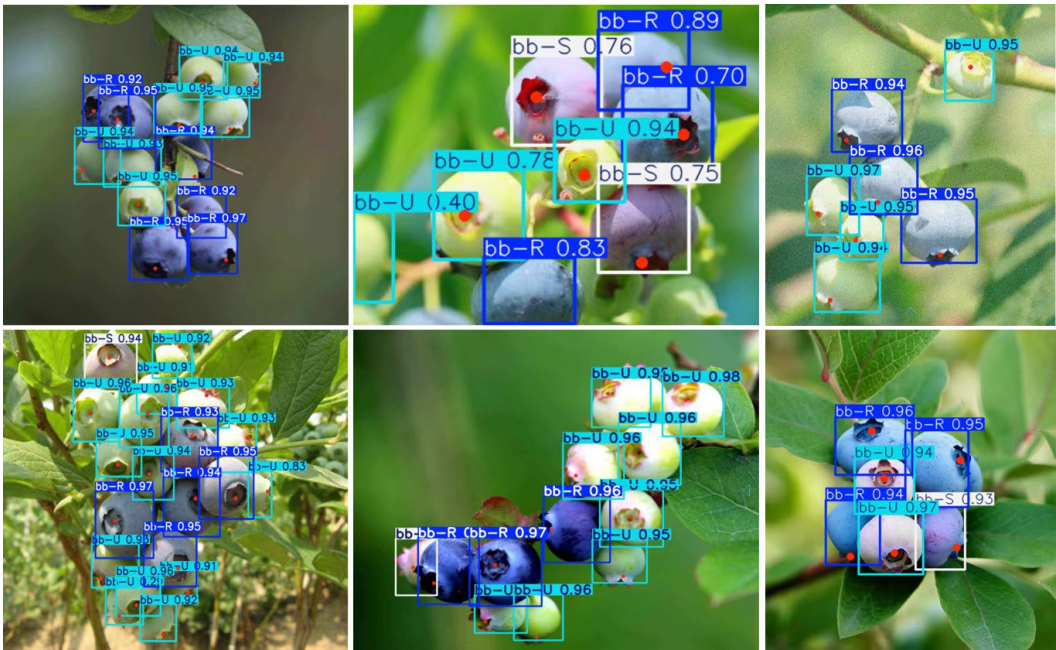


Figure 11. Visualization of blueberry and calyx scar detection results.

3.2. 3D Pose Calculation Results of Blueberry Fruits

According to the calculation in section 2.3.1, the 3D poses of blueberry fruits are shown in Figure 12. It should be noted that blueberry fruits with invisible calyx scars will not display their 3D poses.

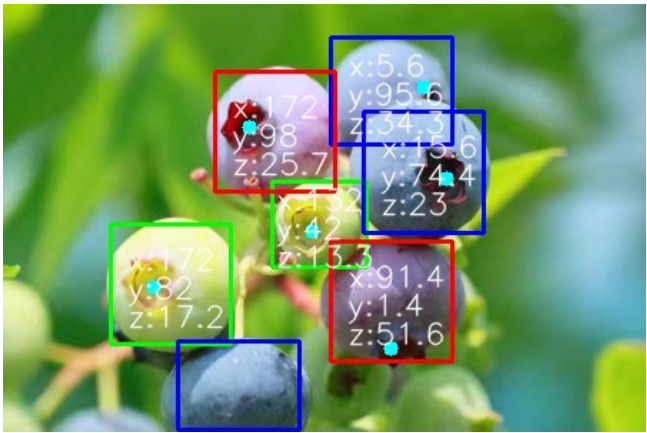


Figure 12. 3D poses of blueberry fruits.

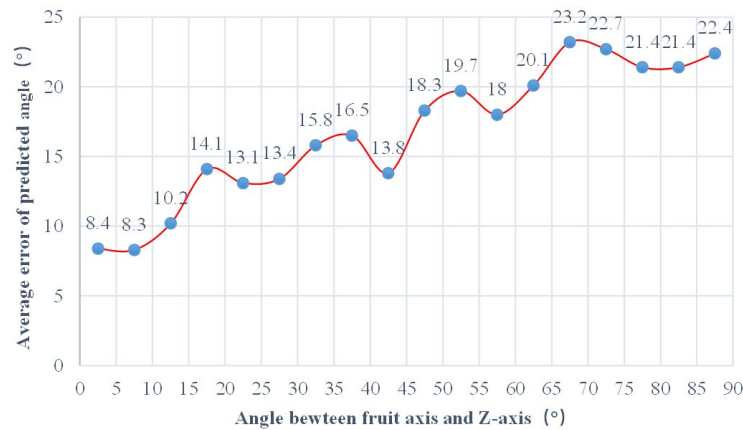
Based on the recorded detection results of blueberries and calyx scars in the testing set, compared with the ground truth, this paper calculated the average error of the predicted 3D pose estimation of blueberry fruits in three directions. The statistical results are shown in Table 3.

Table 3. Average error of 3D pose estimation of blueberry fruits in three directions.

Category	Average Errors (°)		
	X-axis	Y-axis	Z-axis
All	13.7	14.2	19.2
bb_R	14.5	15.4	17.8
bb_S	12.4	14.9	19.4
bb_U	14.1	12.4	20.4



We find that the average error in the Z-axis direction is slightly higher than that in the X-axis and Y-axis directions. In fact, the picking robot is more concerned about the prediction error of the fruit axis in the Z-axis direction (i.e. perpendicular to the image plane), as it determines whether the fruit is in the ideal picking pose. Therefore, this paper conducted a detailed analysis of the errors on the Z-axis and calculated the errors at different angles of the fruit axis (every 5 degrees). The statistical results are shown in Figure 13.

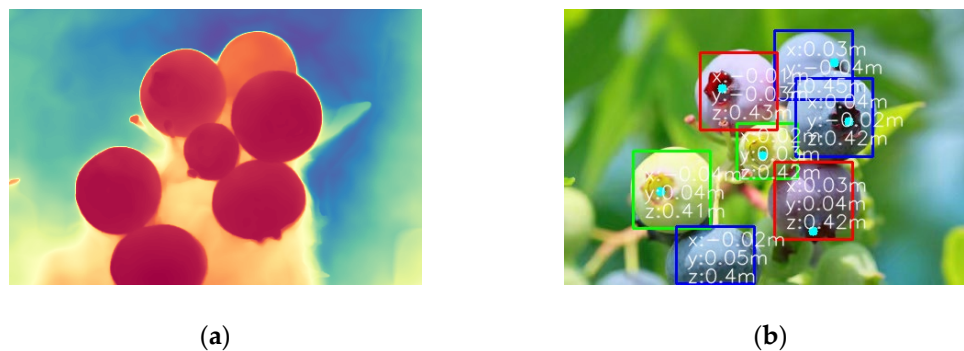


**Figure 13.** Error statistics of fruit axis within different angle ranges in the Z-axis direction.

It can be seen that the angle between the fruit axis and the Z-axis is smaller, the error of the prediction angle is also smaller, and conversely, the error is larger. This is primarily due to the fact that, during the calculation of the angle between the fruit axis and the Z-axis, an approximation of the blueberry's radius is employed by utilizing half of the average length and width of the bounding box. This approximation error will amplify the calculation error of the inverse trigonometric function in Equation (2). However, as the robotic arm moves, the angle between the fruit axis and the Z-axis will become smaller, and the prediction error will also become smaller, eventually converging to a smaller angle range (approximately 10 degrees in Figure 13), which is sufficient to meet the automation and precision picking requirements of robotic arms.

### 3.3. 3D Localization Results of Blueberry Fruits

The blueberry and calyx scar detection model trained on the server is a pt file. Through the official tutorial provided by OAKChina, this paper converts the pt file into the blob format supported by the OAK depth camera. The OAK camera can directly locate the 3D coordinates of the detected blueberry fruits by using the internal depth estimation module. The schematic diagram of environmental depth estimation and blueberry fruit localization is shown in Figure 14.



**Figure 14.** Environmental depth estimation (a) and 3D coordinates of blueberry fruits (b).

We did not conduct experiments on the positioning accuracy of blueberry fruits but evaluated it based on the technical parameters of the OAK camera. According to the parameters provided by

OAKchina, the ranging accuracy of the OAK depth camera is less than 2% within 4 meters. Therefore, for the blueberry picking, the theoretical positioning error within 0.5 meters can be controlled within 1 centimeter, and the actual error should be smaller, which can fully satisfy the positioning accuracy requirements for robotic arms. In addition, the model with our improved DSC module deployed on the OAK depth camera can achieve a detection speed of 13.4 FPS (Frames Per Second), while the model with the CBS module has a detection speed of only 9.3 FPS. Therefore, our detection model can better satisfy the requirements of terminal devices for detection efficiency.

## 4. Conclusions

To satisfy the demand for automated and precise blueberry picking operations, this paper proposes a blueberry fruit detection, 3D spatial localization, and pose calculation method using visual perception. An improved lightweight Yolov11 model proposed by this method can perceive blueberry fruits and calyx scars of different maturity levels in images, and calculate the 3D pose information of blueberry fruits relative to the camera. The detection model deployed to the OAK camera can use its internal depth estimation module and the three-axis gyroscope to obtain the 3D coordinates of blueberry fruits. The experimental results show that the blueberry perception method can achieve a detection accuracy of 95.2% mAP50-95 for blueberry fruits. The 3D positioning error of blueberry fruits can be controlled within 1 centimeter within a range of 0.5 meters, and the 3D pose estimation error can be controlled within 19.2%. This performance can fulfill the requirement for 3D spatial perception of blueberries during automated and precise picking operations by robotic arms. Therefore, the research in this paper is of great significance for improving the picking efficiency of blueberries, reducing picking costs and fruit damage rates, and has an extensive application prospect.

**Author Contributions:** Conceptualization, K.Z. and Y.L.; methodology, K.Z.; software, K.Z.; validation, K.Z., Y.L. and Z.L.; investigation, K.Z.; resources, Y.L. and Z.L.; data curation, K.Z. and Y.L.; writing—original draft preparation, K.Z.; writing—review and editing, K.Z., Y.L. and Z.L.; visualization, K.Z.; supervision, K.Z.; project administration, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** Please add: This research was funded by Shandong Provincial Natural Science Foundation Youth Project, grant number ZR2022QE231.

**Data Availability Statement:** The datasets presented in this article are not readily available because the data are part of an ongoing study. Requests to access the datasets should be directed to the corresponding author.

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

## References

1. Lobos, G. A.; Bravo, C.; Valdés, M.; Graell, J.; Ayala, I. L.; Beaudry, R. M.; Moggia, C. Within-plant variability in blueberry (*Vaccinium corymbosum* L.): maturity at harvest and position within the canopy influence fruit firmness at harvest and postharvest. *Postharvest Biol. Tec.* **2018**, *146*, 26-35.
2. Brondino, L.; Briano, R.; Massaglia, S.; Giuggioli, N. R. Influence of harvest method on the quality and storage of highbush blueberry. *J. Agric. Food Res.* **2022**, *10*, 100415-100423.
3. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: unified, real-time object detection. In Proceedings of the 2016 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2016), Las Vegas, NV, USA, 27-30 June **2016**.
4. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*(6), 1137-1149.
5. Woo, S.; Park, J.; Lee, J. Y.; Kweon, I. S. CBAM: Convolutional block attention module. In Proceedings of the 15th European Conference on Computer Vision (ECCV 2018), Munich, Germany, 8-14 September **2018**.

6. Xiao, F.; Wang, H.; Xu, Y.; Shi, Z. A Lightweight Detection Method for Blueberry Fruit Maturity Based on an Improved YOLOv5 Algorithm. *Agriculture* **2024**, *14*, 36.
7. Feng, W.; Liu, M.; Sun, Y.; Wang, S.; Wang, J. The use of a blueberry ripeness detection model in dense occlusion scenarios based on the improved YOLOv9. *Agronomy* **2024**, *14*, 1860.
8. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, USA, 18-22 June **2018**.
9. Wang, CY.; Yeh, IH.; Mark Liao, HY. YOLOv9: learning what you want to learn using programmable gradient information. In Proceedings of the 18th European Conference on Computer Vision (ECCV 2024), Milan, Italy, 29 September - 4 October **2024**.
10. Zhai, X.; Zong, Z.; Xuan, K. et al. Detection of maturity and counting of blueberry fruits based on attention mechanism and bi-directional feature pyramid network. *Food Measure* **2024**, *18*, 6193–6208.
11. Tan, M.; Pang, R.; Le, QV. EfficientDet: scalable and efficient object detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, USA, 14-19 June **2020**.
12. Gai, R.; Liu, Y.; Xu, G. TL-YOLOv8: a blueberry fruit detection algorithm based on improved YOLOv8 and Transfer Learning. *IEEE Access* **2024**, *12*, 86378-86390.
13. Wang, C.; Han, Q.; Li, J.; Li, C.; Zou, X. YOLO-BLBE: A Novel Model for Identifying Blueberry Fruits with Different Maturities Using the I-MSRCR Method. *Agronomy* **2024**, *14*, 658.
14. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: more features from cheap operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2020), Seattle, USA, 14-19 June **2020**.
15. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2021), Nashville, TN, USA, 20-25 June **2021**.
16. Yang, W.; Ma, X.; Hu, W.; Tang, P. Lightweight Blueberry Fruit Recognition Based on Multi-Scale and Attention Fusion NCBAM. *Agronomy* **2022**, *12*, 2354.
17. Liu, Y.; Zang, W.; Ma, H.; Liu, Y.; Zhang, Y. Lightweight YOLO v5s blueberry detection algorithm based on attention mechanism. *J. Henan Agric. Sci.* **2024**, *53*(3), 151-157.
18. Vaswani A. et al. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, California, USA, 4-9 December **2017**.
19. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. ShuffleNet: an extremely efficient convolutional neural network for mobile devices. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2018), Salt Lake City, USA, 18-22 June **2018**.
20. Li, J.; Wen, Y.; He, L. SCConv: spatial and channel reconstruction convolution for feature redundancy. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2023), Vancouver, Canada, 18-22 June **2023**.
21. Li, F.; Na, X.; Shi, J.; Sun, Q. FM-YOLOv8: Lightweight gesture recognition algorithm. *IET Image Processing* **2024**, *18*(13), 4023-4031.
22. Li, H.; Zhu, Q.; Huang, M.; Guo, Y.; Qin, J. Pose estimation of sweet pepper through symmetry axis detection. *Sensor* **2018**, *18*(9), 3083-3096.
23. Lehnert, C.; Sa, I.; McCool, C.; Upcroft, B.; Perez, T. Sweet pepper pose detection and grasping for automated crop harvesting. In Proceedings of the 2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, May **2016**.
24. 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*(2), 428-442.

25. Zhang, Q.; Gao, G. Grasping model and pose calculation of parallel robot for fruit cluster. *Transactions of the CSAE* **2019**, *35*, 37-47. (in Chinese with English abstract)
26. Yun, S.; Zhang, Q. A method of recognizing the grapefruit axial orientation by binocular vision. *Journal of Jiangnan University (Natural Science Edition)* **2012**, *11*, 33-37.
27. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z., Han, J.; Ding, G. Yolov10: real-Time end-to-End object detection. *arXiv* **2024**, arXiv:2405.14458.
28. Khanam, R.; Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.