Estimation of asparagus stem height and diameter in complex environments by integrating improved YOLOv5 with point cloud

Weiwei Hong[1,2], Zenghong Ma[1,3,4], Bingliang Ye[1,3,4*], Gaohong Yu[1,3,4], Tao Tang[1], Mingfeng Zheng[1,2]

(1. *Faculty of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China*;
2. *Special Equipment Institute, Hangzhou Vocational & Technical College, Hangzhou 310018, China*;
3. *Zhejiang Province Key Laboratory of Transplanting Equipment and Technology, Hangzhou 310018, China*;
4. *Key Laboratory of Agricultural Equipment for Hilly and Mountainous Areas in Southeastern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs, Hangzhou 310018, China*)

Abstract: Identifying the maturity of asparagus is a crucial step for machine-assisted harvesting of asparagus in complex environments. This study proposes an innovative method to evaluate the height and diameter of asparagus stems, combining an enhanced YOLOv5 detection algorithm with point cloud data. In this method, first, the YOLOv5 model was improved, enabling efficient recognition and detection of asparagus in complex environments. Subsequently, a RealSense L515 radar camera was deployed to capture both the original RGB images and the point cloud information. The improved YOLOv5 algorithm was then employed to detect asparagus instances within the RGB images, with the pixel positions of the detection frames mapped onto the point cloud dataset to extract comprehensive 3D point cloud details of the asparagus. Finally, noise was reduced through statistical filtering and Euclidean clustering, and asparagus height was determined using the oriented bounding box methodology. Slices, each with a thickness of 10 mm, were extracted at designated measurement points, and the asparagus diameter was calculated by assessing the disparity between the maximum and minimum coordinates perpendicular to the growth direction of the asparagus. Experimental results showed that the mean average precision, precision, and recall of the improved YOLOv5 model increased by 4.85%, 5.09%, and 3.4%, reaching 98.21%, 97.11%, and 95.33%, respectively, which are higher than those of the YOLOv5 prototype network. Therefore, the proposed method could effectively detect asparagus. The algorithm exhibited a mean absolute error of 1.08 cm, a mean absolute percentage error of 4.06%, and a root mean square error of 1.60 cm in its estimation of asparagus height. For asparagus diameter estimation, the algorithm achieved a mean absolute error of 0.86 mm, a mean absolute percentage error of 7.98%, and a root mean square error of 1.23 mm. These results confirm that the proposed method can estimate the height and diameter of asparagus stems accurately, thereby providing invaluable technical support for machine harvesting of asparagus.

Keywords: asparagus, height, diameter, improved YOLOv5, point cloud
DOI: 10.25165/j.ijabe.20251805.9262

Citation: Hong W W, Ma Z H, Ye B L, Yu G H, Tang T, Zheng M F. Estimation of asparagus stem height and diameter in complex environments by integrating improved YOLOv5 with point cloud. Int J Agric & Biol Eng, 2025; 18(5): 268–277.

1 Introduction

Asparagus, hailed as the "king of vegetables" and considered among the top 10 most popular vegetables worldwide[1,2], is a nutritious and healthful vegetable. In 2022, China's asparagus cultivation spanned approximately 1.45 million hectares, yielding an output of 7.35 million tons, thus being the global leader in asparagus cultivation[3,4]. However, the maturity of edible asparagus shoots is inconsistent, requiring manual selective harvesting, which relies on the measurement of asparagus stem height and diameter. Unfortunately, this approach has a number of shortcomings, such as inefficiency, labor-intensive demands, and high operational costs, which present challenges to the sustainable development of the asparagus industry[5]. Hence, conducting research on machine-assisted selective harvesting of green asparagus is urgently needed.

A crucial factor in machine-assisted selective harvesting of asparagus is the identification of asparagus maturity. Among various phenotypic parameters, the height and diameter of asparagus stems are important metrics for estimating maturity[6-10]. Sakai et al. developed an innovative green asparagus selective harvesting apparatus, using lidar technology for asparagus recognition and precise measurement of stem diameter and height. The system screened and identified harvestable targets on the basis of predefined harvesting criteria, achieving a 75% success rate in asparagus identification[11]. Leu et al. of the University of Bremen in Germany pioneered an autonomous ambulating green asparagus selective harvesting robot. This advanced system harnessed RGB-D camera technology to capture 3D data of asparagus and ridge surfaces, subsequently performing filtering, clustering, and point cloud segmentation to extract size information and assess asparagus

Received date: 2024-07-31 **Accepted date:** 2025-07-15
Biographies: Weiwei Hong, PhD candidate, research interest: intelligent agricultural equipment, Email: will_hong@yeah.net; **Zenghong Ma**, PhD, Associate Professor, research interest: intelligent agricultural equipment, Email: mzh2018@zstu.edu.cn; **Gaohong Yu**, PhD, Professor, research interest: intelligent agricultural equipment, Email: yugh@zstu.edu.cn; **Tao Tang**, PhD candidate, research interest: intelligent agricultural equipment, Email: ttzstu@163.com; **Mingfeng Zheng**, PhD candidate, research interest: intelligent agricultural equipment, Email: 37209351@qq.com.
***Corresponding author: Bingliang Ye**, PhD, Professor, research interest: intelligent agricultural equipment, Faculty of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou 310018, China. Tel: +86-13336060776, Email: zist_ybl@zstu.edu.cn.

maturity[12]. The Geiger-Lund company of the United States developed a field-treaded green asparagus selective harvesting machine, which uses optical detectors to gauge asparagus height. This measurement determines asparagus maturity, thus allowing one to decide whether a given specimen can be harvested[13]. Peebles et al. compared the sensor technologies used for asparagus harvesting and investigated the methods used to determine the position of the ridge surface in an asparagus harvesting scenario[14-16]. Kennedy et al. proposed the concept of perceptual channels based on multiple cameras to localize green asparagus[17]. Notably, although China has conducted asparagus detection, the focus on asparagus maturity identification technology remains relatively limited. Liu et al. pioneered the application of an enhanced mask R-CNN for target segmentation of asparagus in straightforward spring settings[18]. Hong et al. proposed the use of the refined YOLOv5 algorithm for asparagus target detection in complex scenarios[19].

In recent years, the integration of point cloud data[20-22] and deep learning detection algorithms[23-26] for use in the field of agriculture has produced valuable insights into the estimation of asparagus stem height and diameter in complex environments. Sun et al. employed SOLOv2 to delineate apple trees in RGB images by utilizing threshold parameters in the *X*-, *Y*-, and *Z*-axes to intercept the point cloud model[27]. This approach facilitated the identification of grafting positions on apple trees and enabled long-distance estimation of trunk diameters for grafted apple trees. Qiu et al. introduced an RGB-D camera-based method to compute maize plant height and diameter by extracting pertinent point cloud feature information[28]. Yang et al. placed an RGB-D camera over cucumber seedlings and segmented clustered spots within the collected color three-dimensional point cloud data[29]. This procedure enables localization of individual seedlings and facilitates plant height measurements. McGlade et al. used Kinect V2 technology to measure outdoor tree diameters and heights, achieving a root mean square error (RMSE) of 35.3 mm after excluding non-circular diameters[30]. Montoya et al. devised a tool for extracting point cloud data to estimate tree stem heights and diameters[31]. Pires et al. employed lasers to scan forests to intercept cross-sections of point clusters, which were fitted with circles, and extracted tree diameters[32]. Li et al. used YOLOv3 in conjunction with RGB-D cameras for precise 3D positioning of outdoor tea picking points, achieving an accuracy rate of 93.1%[33].

In summary, although both domestic and international studies have proposed methods for identifying green asparagus and assessing its maturity primarily in indoor or uncomplicated scenarios, few studies have been conducted on the identification of asparagus maturity in complex environments. Therefore, the use of asparagus maturity recognition technology in complex growth environments needs to be investigated. This approach requires the full potential of RGB-D data to be harnessed and thus enhance the generality and resilience of feature design while reducing computational demands to align with practical applications. This study introduces a methodology for estimating asparagus stem height and diameter in complex environments, thereby providing crucial technical support for intelligent machine-assisted asparagus harvesting. This approach involves the real-time collection of RGB images and point cloud data via an RGB-D camera, followed by the utilization of the improved YOLOv5 algorithm to detect the green asparagus region. 3D asparagus information was extracted from the point cloud dataset by using the pixel positions within the detection frames in the RGB image. Filtering, clustering, fitting, and related procedures were then performed, thereby determining asparagus stem height and diameter accurately, which served as maturity indicators.

This study is organized as follows: Section 2 delineates the complete process of asparagus height and diameter estimation. Section 3 describes the evaluation index of the proposed estimation method, the detection effect of the improved YOLOv5 algorithm, the appropriate slice thickness threshold setting, and the evaluation of the height and diameter estimation effect. The last section summarizes the study.

2 Materials and methods

2.1 Analysis overview

During summer and autumn, the stems and leaves of asparagus plants are clustered densely, creating a complex cultivation environment. The similarity in both shape and color between asparagus shoots and stems further complicates the matter, which is aggravated by the presence of various stacking conditions. The color of asparagus sprouts is similar to that of surrounding leaves and weeds. As a result, many leaves and weeds may obscure the sprouts or serve as the background. These factors collectively affect the success rate of asparagus detection. During point cloud processing, the 3D information of asparagus may be mixed with these extraneous factors, which can affect the final calculation results, thereby making the identification of asparagus maturity difficult.

To address these challenges, this study introduced a method for maturity identification of asparagus stems by integrating YOLOv5[34-37] and point cloud technologies, respectively. This process is illustrated in Figure 1. First, an RGB color image of asparagus, along with the corresponding point cloud data, was acquired using an RGB-D camera (Intel, CA, USA). The original RGB image and depth map were captured directly using an RGB camera and depth sensor, respectively. Subsequently, Intel RealSense SDK2.0 was utilized to align the depth map with the original RGB image, thereby generating point cloud data on the basis of this synchronized depth map. The *Z*-coordinate value for each point within the point cloud was derived directly from the aligned depth map. The *X*- and *Y*-coordinates for each point in the point cloud were computed using the *Z*-coordinate value, pixel coordinates of the point, and internal parameters of the Intel RealSense L515 camera (Intel, CA, USA).

Subsequently, the YOLOv5 algorithm was enhanced. The foundational backbone network was substituted with FasterNet[38], enhancing the capability and efficiency of feature extraction. Moreover, a dynamic head based on the attention mechanism was incorporated to improve the detection head's ability to identify asparagus[39]. Wise IoU was implemented as the bounding box regression loss function, speeding up convergence and increasing accuracy[40]. The improved YOLOv5 model was then used to perform detection on the RGB image, producing detection boxes and acquiring pixel coordinates. These pixel coordinates were subsequently mapped to the point cloud dataset to facilitate the extraction of the 3D points that correspond to the detection frame. Finally, noise was meticulously removed through a combination of filtering and Euclidean clustering. An enveloping box was constructed to obtain the height information of asparagus. Additionally, point cloud data corresponding to the measured location of the asparagus diameter were obtained through a slicing operation and then analyzed, thus determining the asparagus diameter.

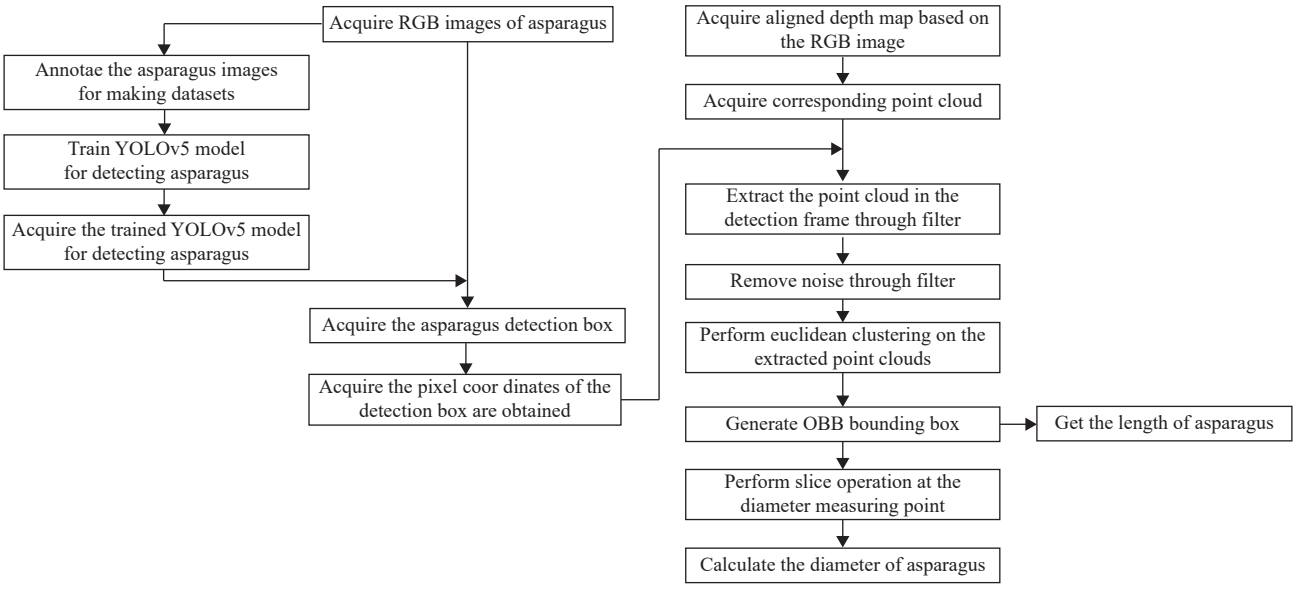


Figure 1 Workflow of asparagus height and diameter estimation

2.2 Data acquisition

2.2.1 3D data acquisition

The data acquisition system utilized for capturing asparagus data comprised a RealSense L515 depth camera, a computer, and a tripod. According to the asparagus production process, the asparagus undergoes stem treatment after being harvested in spring. Therefore, asparagus stems and leaves are already present during summer and autumn, and they become lush. As a result of the bottom–up growth pattern of asparagus, the leaves on the top provide sufficient shade for the asparagus. The thick foliage of asparagus during the summer and autumn seasons was avoided by mounting a RealSense L515 depth camera strategically on a tripod. This camera collected data approximately 750 mm away and 475 mm above the ground alongside the asparagus. The collected dataset encompasses RGB images, depth maps, and the corresponding point cloud data, as shown in Figure 2. The depth values, that is, the *Z*-coordinate values within the point cloud, were acquired using the L515 depth camera. The *X*- and *Y*-coordinates of the points within the point cloud can be computed by using the *Z*-coordinates, the internal parameters of the RealSense L515, and the pixel coordinates of the points, as shown in Equations (1) and (2).

The original depth map has a resolution of 640×480 pixels, whereas the RGB image has a resolution of 1280×720 pixels. Therefore, the depth map needs to be registered with the RGB map, thereby expanding the generated point cloud from 640×480 to 1280×720 pixels accordingly.

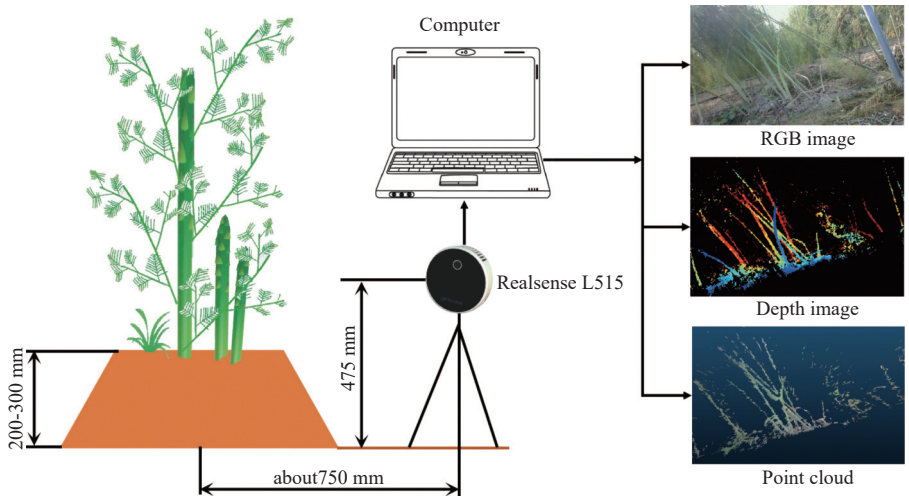


Figure 2 Data acquisition system

Data were collected within the asparagus planting base of Hangzhou Jiahui Agricultural Development Co., Ltd. using the Jialu No. 1 asparagus variety. The asparagus greenhouse is about 6×60 m with four rows. The dimensions of each row are shown in Figure 1. Sixty sets of summer asparagus RGB maps, depth maps, and point cloud datasets were collected. For these 60 asparagus specimens, the diameter and height measurements were manually acquired and documented, serving as reference values for comparison with the results generated by the algorithm.

In accordance with the typical asparagus production process, a measuring point located 5 cm from the ground was designated as the reference point for asparagus diameter assessment, as illustrated in Figure 3.

$$x = \frac{(u - c_x)z}{f_x} \tag{1}$$

$$y = \frac{(v - c_y)z}{f_y} \tag{2}$$

where, u and v are the pixel coordinates of the point. z is the depth value of the point collected by the depth sensor of the RealSense L515. c_x, c_y, f_x, and f_y are the internal parameters of the RealSense L515.

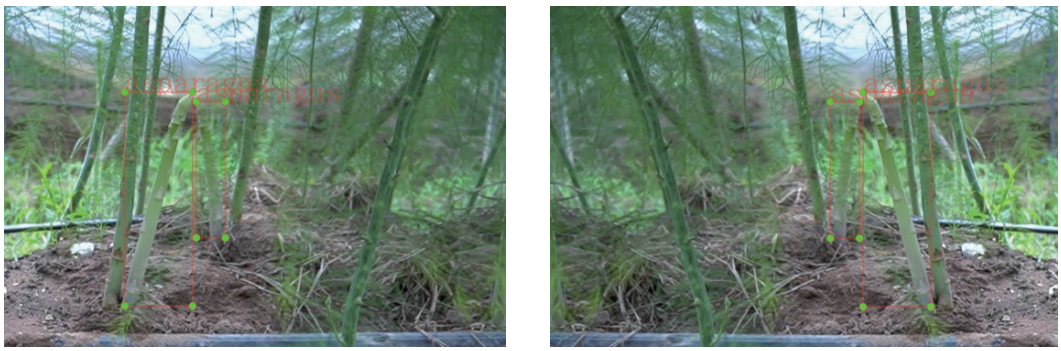
a. Growing environments of asparagus in summer and autumn b. Measuring point location

Figure 3 Diameter measuring point location

2.2.2 RGB dataset acquisition

The summer asparagus harvest period, which is primarily in May and June, corresponds to the time when asparagus reaches its peak in terms of growth. Consequently, a comprehensive dataset comprising 1302 photographs of asparagus was compiled during this period, capturing the plants from various angles under various lighting conditions. A total of 701 photos were collected under sunny conditions on May 31, 2022, and 601 photos were collected under cloudy conditions on June 3, 2022. The data collection location was the asparagus planting base of Hangzhou Jiahui Agricultural Development Co., Ltd. LabelImg software was used for annotation (Intel, CA, USA). The harvesting process involved the machine exclusively targeting the current row, which is why the annotations were restricted to asparagus specimens within the immediate row, with those situated in the background considered as part of the backdrop.

Subsequently, dataset augmentation measures were implemented to enhance the adaptability and robustness of the model. A straightforward horizontal flip operation was applied to the images to recognize the vertical alignment of asparagus during harvesting. This augmentation process resulted in a dataset of 2604 samples, as shown in Figure 4. Following best practices, the dataset was divided into training, validation, and test sets at an 8:1:1 ratio. Consequently, 2084 photos were allocated for training, 260 photos were used for validation, and an additional 260 photos were used for testing.



a. original photo b. photo after being flipped horizontally

Figure 4 Data enhancement

2.3 Improved YOLOv5

The YOLO model was employed to enhance the accuracy and real-time performance of asparagus detection in complex environments. Replacing FasterNet with a backbone feature extraction network can improve the feature extraction capability and efficiency. The detection head was adapted to the dynamic head to focus on the asparagus characteristics and enhance the detection capability. Wise IoU was employed as the bounding box regression loss function to accelerate the convergence rate of the network, thus allowing it to more effectively detect asparagus in complex environments.

2.3.1 FasterNet backbone

The growth environment of asparagus in summer and autumn is complex. FasterNet was applied to the YOLOv5 network as the backbone to extract feature information and thus achieve fast and effective extraction of asparagus features in complex scenarios. FasterNet utilizes a novel partial convolution (PConv), capitalizing on the redundancy in the feature graph and applying regular convolution (Conv) on a fraction of the input channels while leaving the remaining channels unaffected. As a result, PConv has lower FLOPs than regular Conv does and has higher FLOPs than DWConv/GConv does. The computing power of the device can be used more efficiently to identify the spatial features, as illustrated in Figure 5. A new neural network family, FasterNet, based on PConv is constructed, which has a fast running speed, strong feature extraction ability, and high efficiency.

2.3.2 Dynamic head

The dynamic head is an innovative detection head that integrates scale, spatial, and task awareness. The attention mechanism is deployed separately on each specific dimension of the

feature, namely, level, spatial, and channel. Scale-aware attention modules are deployed only on the level dimension, learning the relative importance of various semantic levels to enhance features at an appropriate level based on the scale. The spatial-aware attention module is deployed on the spatial dimension (i.e., height×width) and learns coherent discriminative representations in spatial locations. The task-aware attention module is deployed on channels and guides different feature channels to favor different tasks according to different convolution kernel responses from the objects, as shown in Figure 6. The brown rectangular box π_L in front of the green square represents the attention weight coefficient obtained at the layer level. This coefficient is then multiplied with multiple slices in front of the green square to obtain the attention feature map at the layer level; this is known as scale-aware attention. The black rectangular box π_S located directly above the green square represents the weight coefficient of spatial attention. It is multiplied with several spatial slices of the green square positioned directly above to obtain the attention feature map at the spatial level; this is referred to as spatial-aware attention. The yellow rectangular box π_C on the right side of the green square represents the attention weight coefficient obtained at the channel level. It is multiplied with several channel slices of the green square on the right side to obtain the attention feature map at the channel level; this is known as task-aware attention.

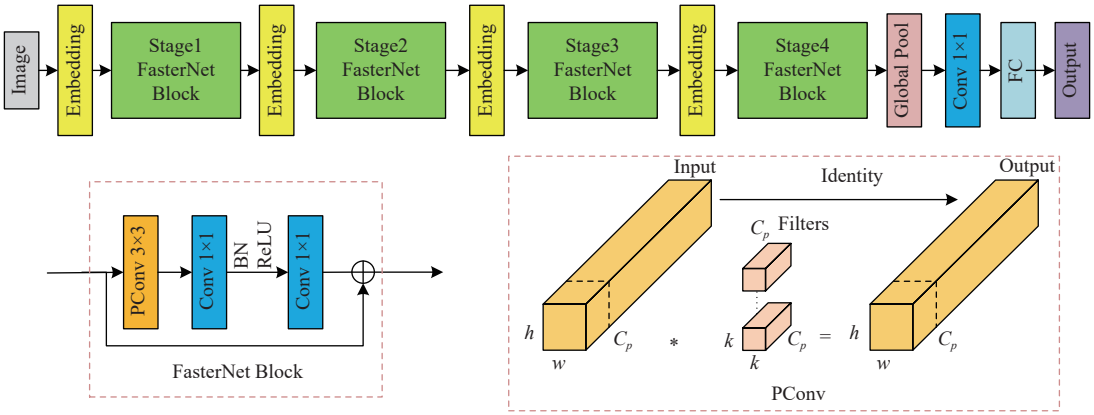


Figure 5 FasterNet backbone structure

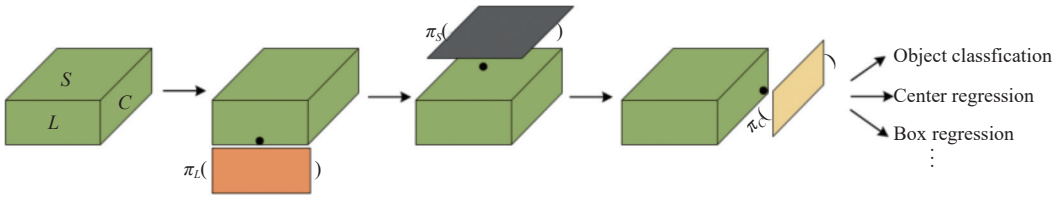


Figure 6 Illustration of dynamic head

The feature map is more sensitive to the proportion difference of foreground objects after the scale-awareness attention module. It becomes sparser and focused on discriminating the spatial location of foreground objects after they pass through the spatial awareness attention module. Finally, after passing through the task-awareness attention module, these feature maps reform different activations according to the needs of different downstream tasks.

2.3.3 Wise IoU

The original YOLOv5s adopted CIoU loss as the loss function of the prediction frame. Based on DIoU, the CIoU loss function introduces the aspect ratio between the predicted boundary frame and the real boundary frame, which can accelerate the regression speed of the prediction frame to a certain extent. Network optimization is hindered by the lack of clarity in defining the aspect ratio of the two boundary frames. In this study, the Wise-IoU v3 loss function is used. Unlike the former, Wise-IoU v3 does not introduce the aspect ratio calculation. Instead, it adopts a dynamic non-monotone focusing mechanism on the attention-based bounding frame loss Wise-IoU v1, using the outlier to describe the quality of the anchor frame.

First, the attention-based bounding frame loss Wise-IoU v1 is calculated as follows:

$$L_{\mathrm{WIoUv1}} = R_{\mathrm{WIoU}} L_{\mathrm{IoU}} \tag{3}$$

$$R_{\mathrm{WIoU}} = \exp\left(\frac{(x-x_{\mathrm{gt}})^2+(y-y_{\mathrm{gt}})^2}{(W_g^2+H_g^2)^*}\right) \tag{4}$$

where, Wise-IoU v1 is the boundary frame loss; L_{IoU} is the boundary frame loss IoU; R_{IoU} is the distance attention; x and y are the horizontal and vertical coordinates of the center point of the prediction frame, respectively; x_{gt} and y_{gt} are the horizontal and vertical coordinates of the center point of the real frame, respectively; and W_g and H_g are the width and height of the minimum external rectangle of the prediction frame and the real frame, respectively. * indicates the separation of the operation from the computed graph to make it constant without a gradient.

The outlier is then used to describe the quality of the anchor frame, defined as follows:

$$\beta = \frac{L_{\mathrm{IoU}}^*}{\overline{L_{\mathrm{IoU}}}} \in [0,+\infty) \tag{5}$$

where, $\overline{L_{\mathrm{IoU}}}$ is the sliding average.

Finally, the non-monotone focusing frame loss Wise-IoU v3 is constructed using the outlier degree β, and the calculation formula is as follows:

$$L_{\mathrm{WIoUv3}} = kL_{\mathrm{WIoUv1}} \tag{6}$$

where, L_{WIoUv3} is the boundary frame loss Wise-IoU v3, $k = \frac{\beta}{\delta\alpha^{\beta-\delta}}$, and the hyperparameters α and δ in this study are set to 1.9 and 3.0,

respectively.

2.4 Target point cloud extraction

The point cloud data were transformed into three 1280×720 coordinate matrices to precisely determine the pixel coordinates that index the position of asparagus. These matrices comprise the *X*, *Y*, and *Z* coordinates of the points within the point cloud. The point cloud data encapsulates comprehensive three-dimensional information about the object, encompassing the *X*-, *Y*-, and *Z*-coordinates, all expressed in meters. In this context, the *x*-axis corresponds to the height direction, the *y*-axis represents the diameter direction, and the *z*-axis represents the depth direction.

During RealSense L515 data acquisition, the coordinate values of the points are stored in the PLY format point cloud and arranged in columns to align with the depth map. Subsequently, pixel coordinate transformation is facilitated by employing the center point of the prediction frame. Once the pixel coordinates of the center point within the prediction box are obtained, the corresponding points within the point cloud are matched using coordinate transformation formulas, as shown in Equations (7) and (8).

$$x_p = 1279 - x_i \tag{7}$$

$$y_p = 719 - y_i \tag{8}$$

where, (x_i, y_i) is the pixel coordinate of one point in the original RGB image, and (x_p, y_p) is the coordinate of the corresponding point in the three coordinate matrices.

A point cluster near the center point was isolated by setting the thresholds in the *X*- and *Y*-directions as half of the height and width of the prediction box, respectively. All points that meet the criteria outlined in formulas (9) and (10) were used in calculating asparagus height and diameter.

$$x_{\text{height}} - b_h/2 \le x \le x_{\text{height}} + b_h/2 \tag{9}$$

$$y_{\text{width}} - b_w/2 \le y \le y_{\text{width}} + b_w/2 \tag{10}$$

where, (x, y) are the coordinates of the points extracted to calculate the asparagus information, $(x_{\text{height}}, y_{\text{width}})$ are the coordinates of the center point of the prediction box, and b_h, b_w are the height and width of the prediction box, respectively. This process is illustrated in Figure 7.

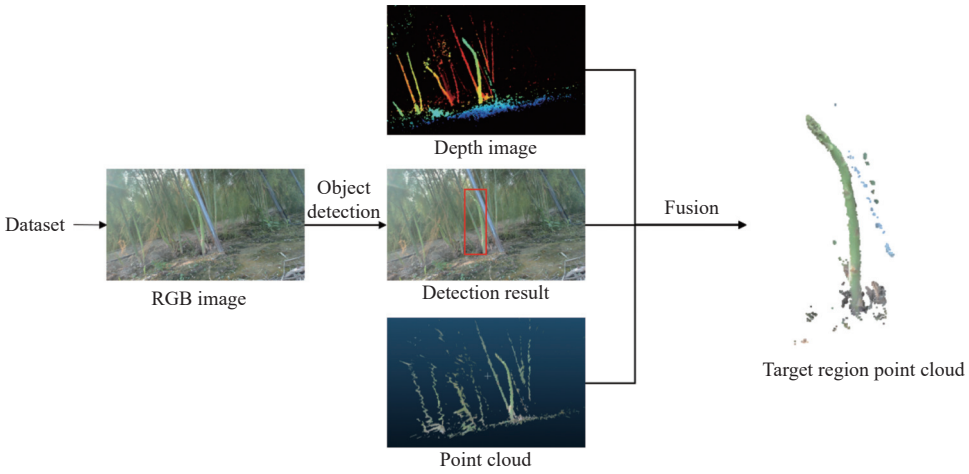


Figure 7 Target point cloud extraction

2.5 Height and diameter calculation

A series of essential steps was performed to ascertain the height and diameter of the asparagus. Initially, statistical filtering was applied to the point cloud, followed by Euclidean clustering to eliminate noise, thereby yielding a clean asparagus point cloud. In statistical filtering, the nb_neighbors (number of neighbors around the target point) was set to 5 and the std_ratio (standard deviation ratio) was set to 1. This balances computational efficiency with filtering intensity, enabling rapid removal of significant outliers. In Euclidean clustering, the tolerance (maximum allowable Euclidean distance between two points for them to be considered part of the same cluster) was set to 0.012, the minimum cluster size (minimum number of points required to form a valid cluster) to 1200, and the maximum cluster size (maximum number of points allowed in a single cluster) to 10 000. This specific configuration achieved precise object segmentation and efficient noise filtering based on the physical constraints and point cloud density characteristics of the typical asparagus greenhouse environment. Subsequently, an oriented bounding box (OBB) was constructed to obtain height information regarding asparagus. The primary direction was determined through principal component analysis transformation, with the *X*-direction corresponding to the height and the *Y*-direction to the diameter. A suitable threshold was established along the *X*-direction to precisely measure the diameter of the asparagus. Subsequently, a section of the point cloud was sliced at a distance of 50 mm from the bottom of the OBB, which represented the point for measuring the asparagus diameter. With the use of the OBB, the asparagus diameter was determined by calculating the difference between the maximum and minimum coordinate values along the *Y*-axis direction, as illustrated in Figure 8.

3 Results

3.1 Training environment

Model training was conducted on a platform equipped with an Intel i5-6500 (3.2 GHz) 4-core CPU, an Nvidia GeForce RTX 3060 GPU (with 12 GB video memory), and a Windows 10 64-bit operating system with 24 GB of RAM. The deep learning framework employed was PyTorch, with PyCharm serving as the compilation tool. GPU acceleration was facilitated through CUDA 11.1 and CUDNN 8.0.5.

All images were adjusted to a standardized size of 640×640 pixels to adhere to the input requirements of the model algorithm. Given the constraints imposed by computer hardware, the batch size was set to 16. Network optimization was executed using the SGD optimizer, with a momentum parameter of 0.937, an initial learning rate of 0.01, a decay index of 0.0005, and 300 training rounds.

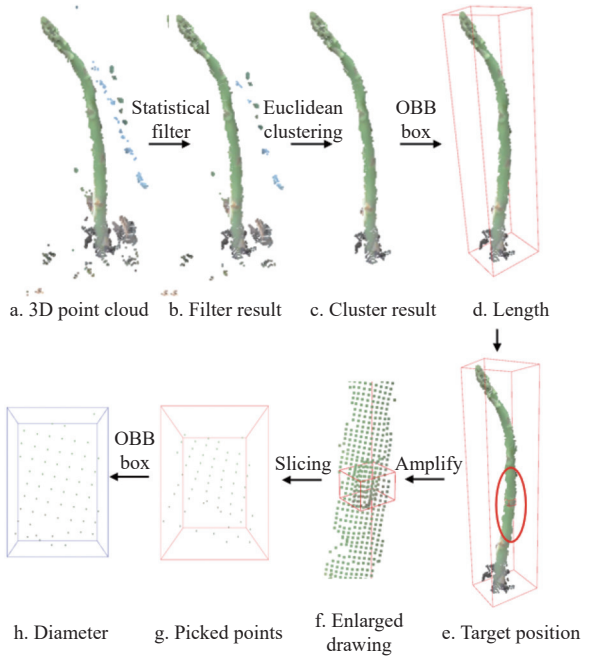


Figure 8 Obtaining the height and diameter of asparagus

3.2 Performance evaluation

Detection accuracy and speed are important metrics for assessing the YOLO detection model. Detection accuracy encompasses precision (P), recall (R), average precision (AP), and mean average precision (mAP), which are defined in Equations (11) to (14).

$$P = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FP}} \times 100\% \tag{11}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \times 100\% \tag{12}$$

$$\mathrm{AP} = \int_0^1 P(R)\mathrm{d}R \tag{13}$$

$$\mathrm{mAP} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{AP}_i \tag{14}$$

where, TP represents the count of correctly predicted positive samples, FP denotes the number of erroneously predicted positive samples, FN signifies the count of incorrectly predicted negative samples, n corresponds to the number of target categories detected, and AP_i represents the average precision (AP) for the i th target class.

The performance assessment of the proposed method for height and diameter estimation in this study was conducted using the evaluation metrics of mean absolute error (MAE), mean absolute percentage error (MAPE), and RMSE when compared with manual measurements as the ground truth. MAE, MAPE, and RMSE were calculated using Equations (15)-(17), respectively.

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|(f_i - g_i)| \tag{15}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{f_i - g_i}{g_i}\right| \tag{16}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i - g_i)^2} \tag{17}$$

where, n is the number of samples used to estimate the height and diameter, f_i is the estimated value of the i^{th} sample, and g_i is the ground truth of the i^{th} sample.

3.3 Performance of YOLOv5 model

3.3.1 Training assessment of YOLOv5 model

Throughout the model training process, the efficacy of the model was assessed on the basis of two crucial metrics: mAP value and loss value. The training trajectory for mAP and loss is depicted in Figure 9. The mAP indicates asparagus detection accuracy, whereas the loss value reflects the model's fitting performance during training. After 273 training epochs, the loss curve converged gradually and stabilized. The minimum recorded box loss and object loss were 0.0172 and 0.01, respectively, with mAP@0.5 reaching a high value of 0.97. This finding indicates that the model effectively captures the distinctive characteristics of asparagus, enabling efficient asparagus detection. This result forms a robust foundation for subsequent asparagus height and diameter estimations.

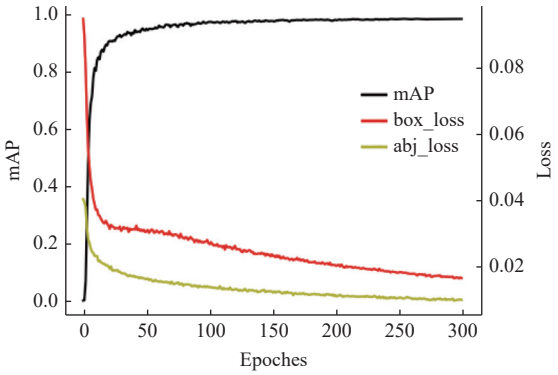


Figure 9 Loss value and mAP curve

3.3.2 Detection results of improved YOLOv5 model

The improved YOLOv5 model demonstrated commendable performance in asparagus detection, as illustrated in Figure 10. The detection results of the model served as the foundation for subsequent asparagus height and diameter measurements. The test results for the test set are summarized in Table 1, achieving a precision value of 97.11%, recall value of 95.33%, and mAP@0.5 value of 98.21%. Compared with those of the YOLOv5 prototype network, the mAP@0.5, precision value, and recall value of the improved YOLOv5 model increased by 4.85%, 5.09%, and 3.4%, respectively. Compared to advanced detection algorithms, the detection effect is slightly better than YOLOv8. These results show that the model is able to detect asparagus rapidly and accurately in complex environments, effectively meeting the requirements for asparagus height and diameter measurements in later stages.

Table 1 Detection results of improved YOLOv5 model

Model	P/%	R/%	mAP@0.5
Improved YOLOv5s	97.11	95.33	98.21
YOLOv8s	96.98	95.17	98.14
YOLOv5s	92.02	91.93	93.36
YOLOv4	93.50	79.74	92.28
YOLOv3	90.48	76.24	88.84
Faster-RCNN	76.98	93.36	92.92

3.4 Determining the threshold in the *x*-direction

After the asparagus point cloud has been extracted, slicing needs to be performed at the designated asparagus diameter measurement point to compute the asparagus diameter accurately.

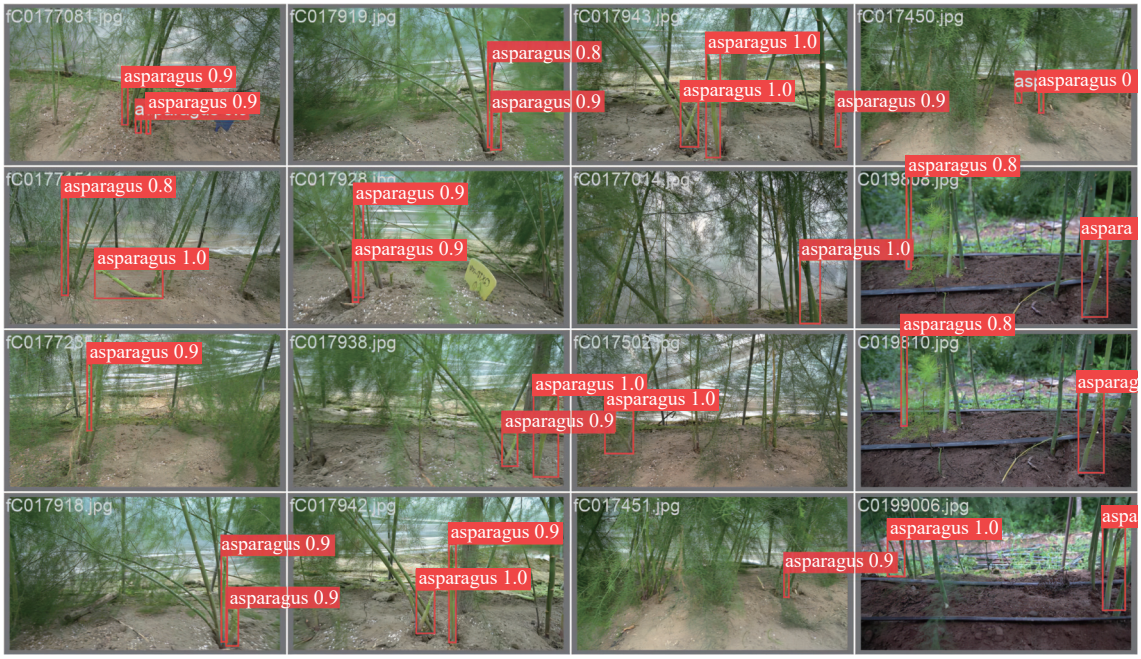


Figure 10 Asparagus detection effect

The thickness of this slice, which denotes the distance along the *x*-axis after principal component analysis, affects the diameter calculation results. Hence, determining the appropriate slice thickness is crucial to ensure that the asparagus diameter is estimated accurately.

Five distance thresholds—5.0, 7.5, 10.0, 12.5, and 15.0 mm—were established to investigate the effects of various section thicknesses on the diameter estimation results. The proposed method was subsequently applied to estimate asparagus diameter under each of these conditions. Figure 11 depicts the ratio between the diameter estimated by this method and the manually measured values across the five slice thicknesses.

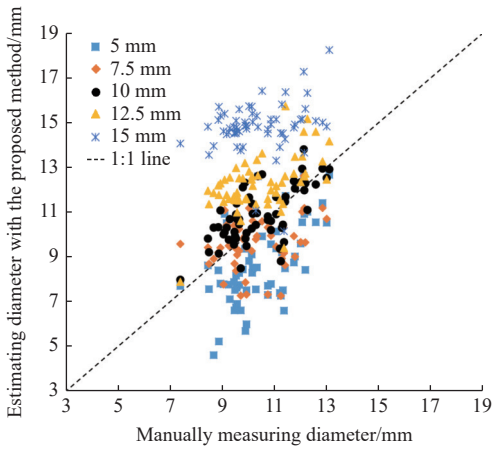


Figure 11 Results of estimating diameter with different distance thresholds

The appropriate slice thickness was determined by configuring five distinct distance thresholds along the *x*-axis, and the proposed method was applied for estimation. Table 2 lists the MAE and MAPE values for the five thresholds. Notably, when the slice thickness was less than 10 mm, both the MAE and MAPE values increased, which was primarily attributed to the sparseness of the point cluster captured by the RealSense L515 sensor. Conversely, when the slice thickness exceeded 10 mm, an excess of nontarget points within the slice contributed to elevated MAE and MAPE values. Remarkably, the most favorable results were achieved with a thickness of 10 mm, yielding the smallest MAE and MAPE values of 0.86 mm and 7.98%, respectively. Thus, establishing a threshold of 10 mm was optimal for obtaining accurate results.

Table 2 MAEs and MAPEs with different thresholds

Distance threshold in *X*-direction/mm	MAE/mm	MAPE
5.0	1.99	26.19%
7.5	1.10	12.02%
10.0	0.86	7.98%
12.5	1.86	15.07%
15.0	4.31	28.86%

3.5 Evaluation of estimating height and diameter

In this study, a method for estimating the height and diameter of asparagus that integrates deep learning and point cloud data acquired from a depth camera was introduced to facilitate selective asparagus harvesting. The measurement results obtained using this method are presented in Figures 12 and 13. With regard to the asparagus height measurement, the MAE, MAPE, and RMSE were 1.08 mm, 4.06%, and 1.60 cm, respectively. For asparagus diameter

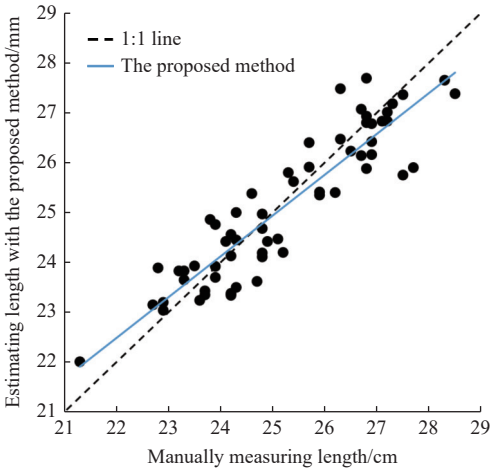


Figure 12 Results of estimating height compared with manual measurement

measurement, the corresponding metrics were 0.86 mm, 7.98%, and 1.23 mm, respectively. A substantial correlation between the proposed approach and the manual measurements can be observed.

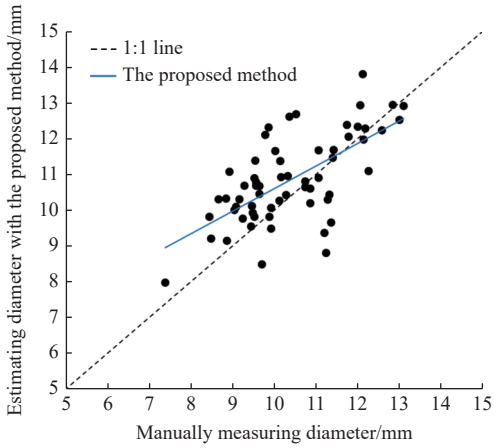


Figure 13 Results of estimating diameter compared with manual measurement

4 Discussion

To enhance the detection of asparagus targets in complex environments, this study made several modifications. First, the backbone network of YOLOv5s was replaced with FasterNet. Second, the detection head was replaced with dynamic head. Lastly, Wise IoU was implemented as the boundary box regression loss function. These changes considerably improved the success rate of detecting green asparagus in complex environments. Compared with Peebles's use of Faster-RCNN for spring asparagus detection, this method achieved a 12.21% increase in detection success rate, reaching 98.21%. Furthermore, when compared with Hong et al.'s utilization of YOLOv5m for green asparagus detection in complex environments, the improved YOLOv5s achieved a similar detection success rate. However, 27 cases of false detection and missing detection still exist. Analysis revealed that these cases primarily occur in two situations: 1) when asparagus is heavily obstructed by asparagus stems, leaves, and grass, resulting in errors and missed detections; 2) when asparagus is positioned at the image's edge, especially if the asparagus bud head is located outside the image, leading to a higher occurrence of incorrect and missed detections. These conditions occur alone or together. The specific statistics are listed in Table 3. Moving forward, the future research will focus on investigating the characteristics of green asparagus stems (excluding the asparagus bud heads), allowing accurate identification of green asparagus even when the bud heads are heavily obscured or out of sight.

Table 3 Statistics of the main reasons for detection failure

Main cause	Asparagus is heavily obstructed	Asparagus is positioned at the image's edge
Quantity	18	15

This study integrates YOLOv5s and point cloud technology to design an algorithm for estimating the length and width of green asparagus stems. The algorithm enables the extraction of 3D information, noise removal, and accurate estimation of stem length and width. Currently, no existing studies have estimated the length and width of green asparagus specifically, but previous research examined the length and width estimation of other plants such as corn and apple trees. For instance, Sun et al. employed SOLOv2 to segment apple trees in RGB maps, mapped and intercepted the point cloud model, and successfully estimated the trunk diameter of grafted apple trees from a distance. In comparison, estimating the dimensions of green asparagus is more challenging due to its smaller size and unique shape. However, the present study achieves an MAPE of only 7.98% and an MAE of merely 0.86 mm, considerably outperforming Sun et al.'s measurement method. Qiu et al. utilized an RGB-D depth camera to collect point cloud data for calculating the height and stem diameter of maize. Corn stems have a cylindrical and straightforward shape, while green asparagus stems exhibit more curvature. Despite the complexity in measuring the length and width of green asparagus, this study successfully reduces the MAE and MAPE. Qiu et al. employed traditional methods for image segmentation processing, while this study combines deep learning with point cloud analysis to achieve a more robust and real-time estimation of asparagus height and diameter. However, this study's algorithm may introduce large errors when estimating the dimensions of green asparagus with significant curvature. For future work, a crucial task is to further consider the influence of asparagus curvature on length and width estimation and refine the algorithm accordingly to enhance accuracy.

5 Conclusions

To address the challenge of machine-based asparagus maturity identification in complex environments, this study introduces a method for estimating asparagus stem height and diameter by leveraging depth camera technology and integrating the improved YOLOv5 with a point cloud. The improved YOLOv5 model was employed to detect asparagus by pinpointing its precise location, thus enabling robust detection even in complex surroundings. Subsequently, 3D point cloud information for asparagus was derived from the pixel positions. The height of asparagus was estimated by constructing an OBB via point cloud data processing. Finally, asparagus diameter was estimated by slicing the asparagus at the designated measurement point.

The results showed that mAP@0.5, precision, and recall of the improved YOLOv5 model increased by 4.85%, 5.09%, and 3.4%, reaching 98.21%, 97.11%, and 95.33%, respectively. These outcomes indicate a highly successful asparagus detection. Furthermore, this study elucidates the influence of slice thickness on asparagus diameter, identifying 10 mm as the optimal thickness. For asparagus height estimation, the proposed method yielded MAE, MAPE, and RMSE values of 1.08 cm, 4.06%, and 1.60 cm, respectively. For asparagus diameter estimation, the corresponding values were 0.86 mm, 7.98%, and 1.23 mm, respectively. These findings highlight the effectiveness of the method in accurately estimating asparagus height and diameter under complex environmental conditions, thereby offering valuable technical support for selective machine harvesting of asparagus.

Acknowledgements

The authors gratefully acknowledge the support provided by the National Natural Science Foundation of China (Grant No. 32171899). We also thank Hangzhou Jiahui Agricultural Development Co., Ltd., for their support in this study.

[References]

[1] Lu B. Development status and trend of asparagus industry in China. Shanghai Vegetables, 2018; 4: 3, 4, 12. (in Chinese) doi: 10.3969/j.issn.1002-1469.2018.04.002.

[2] Peng L L, Yu Y F, Zhou K H. Analysis of the industrial layout

optimization of asparagus in China. Chinese Journal of Agricultural Resources and Regional Planning, 2015; 36(1): 123–127. (in Chinese)
[3] Li D, Lin Z. Study on international trade and industrial competitiveness of asparagus in China since 2000. Chinese Agricultural Science Bulletin, 2022; 38(26): 158–164. (in Chinese)
[4] Li D, Lin Z, Li L. Analysis of global asparagus production situation in recent fifty years. Hunan Agricultural Sciences, 2019; 9: 96–99.
[5] He C X. Development status and prospect of asparagus industry in China. Vegatables, 2022; 5: 33–39. (in Chinese)
[6] Gao F F, Fang W T, Sun X M, Wu Z C, Zhao G N, Li G, et al. A novel apple fruit detection and counting methodology based on deep learning and trunk tracking in modern orchard. Computers and Electronics in Agriculture, 2022; 197: 107000.
[7] Kotaro I, Taichiro Y, Masayuki I, Yoshiko K. Estimating tree height and diameter at breast height (DBH) from digital surface models and orthophotos obtained with an unmanned aerial system for a Japanese Cypress (Chamaecyparis obtusa) forest. Remote Sensing, 2018; 10(2): 13.
[8] Swayze N C, Tinkham W T, Vogeler J C, Hudak A T. Influence of flight parameters on UAS-based monitoring of tree height, diameter, and density. Remote Sensing of Environment, 2021; 263: 112540.
[9] Wu Z C, Li G, Yang R Z, Fu L S, Li R, Wang S J. Coefficient of restitution of kiwifruit without external interference. Journal of Food Engineering, 2022; 327: 111060.
[10] Zhang C, Serra S, Quiros-Vargas J, Musacchi W, Sankaran S. Non-invasive sensing techniques to phenotype multiple apple tree architectures. Information Processing in Agriculture, 2023; 10(1): 136–147.
[11] Sakai H, Shiigi T, Kondo N, Ogawa Y, Taguchi N. Accurate position detecting during asparagus spear harvesting using a laser sensor. Engineering in Agriculture, 2013; 6(3): 105–110.
[12] Leu A, Razavi M, Langstadtler L, Ristic-Durrant D, Raffel H, Schenck C, et al. Robotic green asparagus selective harvesting. IEEE/ASME Transactions on Mechatronics, 2017; 22(6): 2401–2410.
[13] Clary C D, Ball T, Ward E, Fuchs S, Durfey J E, Cavalieri R P, et al. Performance and economic analysis of a selective asparagus harvester. Applied Engineering in Agriculture, 2007; 23(5): 571–577.
[14] Peebles M, Lim S H, Duke M, McGuinness B. Investigation of optimal network architecture for asparagus spear detection in robotic harvesting. IFAC-PapersOnLine, 2019; 52(30): 283–287.
[15] Peebles M, Shen H L, Duke M, Chi K A. Overview of sensor technologies used for 3d localization of asparagus spears for robotic harvesting. Applied Mechanics and Materials, 2018; 884: 77–85.
[16] Peebles M, Shen H L, Streeter L, Duke M, Chi K A. Ground plane segmentation of time-of-flight images for asparagus harvesting. In: 2018 International Conference on Image and Vision Computing New Zealand (IVCNZ), Auckland, New Zealand: IEEE, 2018; pp.1–6.
[17] Kennedy G, Ila V, Mahony R. A Perception pipeline for robotic harvesting of green asparagus. IFAC-PapersOnLine, 2019; 52(30): 288–293.
[18] Liu X P, Wang D N, Li Y N, Guan X Q, Qin C J. Detection of green asparagus using Improved Mask R-CNN for automatic harvesting. Sensors, 2022; 22(23): 9270.
[19] Hong W W, Ma Z H, Ye B L, Yu G H, Tang T, Zheng M F. Detection of green asparagus in complex environments based on the improved YOLOv5 algorithm. Sensors, 2023; 23(3): 1562.
[20] Qian Y, Xu Q J, Yang Y Y, Lu H, Li H, Feng X B, et al. Classification of rice seed variety using point cloud data combined with deep learning. Int J Agric & Biol Eng, 2021; 14(5): 206–212.
[21] Kang H W, Chen C. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. Computers and Electronics in Agriculture, 2020; 171: 105302.
[22] Ji J T, Han Z H, Zhao K X, Li Q W, Du S C. Detection of the farmland plow areas using RGB-D images with an improved YOLOv5 model. Int J Agric & Biol Eng, 2024; 17(4): 156–165.
[23] Hua C J, Zou X T, Jiang Y, Yu J F, Chen Y. RGB-D visual saliency detection of stacked fruits under poor lighting. Int J Agric & Biol Eng, 2025; 18(1): 230–237.
[24] Wang Z, Wang Q Y, Lou M Z, Wu F, Zhu Y N, Hu D, et al. Geometric based apple suction strategy for robotic packaging. Int J Agric & Biol Eng, 2024; 17(3): 12–20.
[25] Hu H M, Kaizu Y, Zhang H D, Xu Y W, Imou K, Li M, et al. Recognition and localization of strawberries from 3D binocular cameras for a strawberry picking robot using coupled YOLO/Mask R-CNN. Int J Agric & Biol Eng, 2022; 15(6): 175–179.
[26] Wang J P, Xu L, Mei S, Hu H R, Zhou J L, Chen Q. Fast and accurate detection of kiwifruits in the natural environment using improved YOLOv4. Int J Agric & Biol Eng, 2024; 17(5): 222–230.
[27] Sun X M, Fang W T, Gao C Q, Fu L S, Majeed Y Q, Liu X J, et al. Remote estimation of grafted apple tree trunk diameter in modern orchard with RGB and point cloud based on SOLOv2. Computers and Electronics in Agriculture, 2022; 199: 107209.
[28] Qiu R C, Miao Y L, Ji Y H, Zhang M, Li H, Liu G. Method for measurement of maize stem diameters based on RGB-D camera. Transactions of the Chinese Society for Agricultural Machinery, 2017; 33(s1): 211–219. (in Chinese)
[29] Yang S, Gao W L, Mi J Q, Wu M L, Wang M J, Deng L H. Method for measurement of vegetable seedlings height based on RGB-D camera. Transactions of the Chinese Society for Agricultural Machinery, 2019; 50(S1): 128–135. (in Chinese)
[30] McGlade J, Wallace L, Hally B, White A, Reinke K, Jones S. An early exploration of the use of the Microsoft Azure Kinect for estimation of urban tree Diameter at Breast Height. Remote Sensing Letters, 2020; 11(11): 963–972.
[31] Montoya O, Icasio-Hernández O, Salas J. TreeTool : A tool for detecting trees and estimating their DBH using forest point clouds. SoftwareX, 2021; 16: 100889. doi: 10.1016/j.softx.2021.100889.
[32] Pires R d P, Olofsson K, Persson H J, Lindberg E, Holmgren J. Individual tree detection and estimation of stem attributes with mobile laser scanning along boreal forest roads. ISPRS J. Photogramm. Remote Sensing, 2022; 187: 211–224.
[33] Li Y T, He L Y, Jia J M, Lv J, Chen J N, Qiao X, et al. In-field tea shoot detection and 3D localization using an RGB-D camera. Computers and Electronics in Agriculture, 2021; 185: 106149.
[34] Bochkovskiy A, Wang C-Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection. arXiv: 2004.10934, 2020; In press. doi: 10.48550/arXiv.2004.10934.
[35] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA: IEEE, 2016; pp.779–788.
[36] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA: IEEE, 2017; pp.6517–6525.
[37] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv: 1804.02767, 2018; In press. doi: arxiv-1804.02767.
[38] Chen J R, Kao S H, He H, Zhuo W P, Wen S, Lee C H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada: IEEE. 2023; pp.12021–12031.
[39] Dai X Y, Chen Y P, Xiao B, Chen D D, Liu M C, Yuan L, et al. Dynamic head: Unifying object detection heads with attentions. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA: IEEE, 2021; pp.7369–7378.
[40] Tong Z J, Chen Y H, Xu Z W, Yu R. Wise-IoU: Bounding box regression loss with dynamic focusing mechanism. arXiv: 2301.10051, 2023; In press. doi: 10.48550/arXiv.2301.10051.