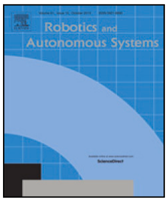
Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot



3D point cloud-based 6D pose estimation using the pedicel morphological features of fruits and vegetables

Yonghyun Park [a,c], Jeonghyeon Pak [a,b], Changjo Kim [a,b], Hyoung Il Son [a,b,c,*]

[a] Department of Convergence Biosystems Engineering, Chonnam National University, Yongbong-ro 77, Gwangju 61186, Republic of Korea
[b] Interdisciplinary Program in IT-Bio Convergence System, Chonnam National University, Yongbong-ro 77, Gwangju 61186, Republic of Korea
[c] Research Center for Biological Cybernetics, Chonnam National University, Yongbong-ro 77, Gwangju 61186, Republic of Korea

ARTICLE INFO

Keywords:
Agricultural robot
Harvesting robot
Pedicel detection
Pose estimation
Visual servoing

ABSTRACT

This study proposes a method to estimate the 6D pose of pedicels using the morphological features of fruits and vegetables. The pedicel, a critical element connecting fruits to their stems, significantly influences the precision and efficiency of agricultural harvesting robots. The proposed system employs 3D point cloud data obtained from RGB-D cameras, using differences in width and curvature to identify the location and orientation of the pedicel. A lightweight YOLOv8n-seg architecture is employed to detect fruits and vegetables, computing the local curvature to estimate pedicel positions. The experimental evaluations on tomatoes and *Cucumis melo* (*C. melo*) demonstrate the capability of the proposed system to handle diverse agricultural environments. For *C. melo*, the system achieved a precision of 0.927, recall of 0.809 and F1-score of 0.864. For tomatoes, precision and recall were both 0.837, resulting in an F1-score of 0.837. Positional errors along the x-, y- and z-axes averaged 1.34, 3.19 and 4.79 mm , respectively, for the *C. melo*, with corresponding root mean squared errors of 7.95, 5.46 and 5.13 mm. Orientational errors averaged 2.14°, 1.14° and -1.49° for ϕ, θ and ψ, respectively. Smoothing algorithms, including linear interpolation for translation and spherical linear interpolation for rotation, address positional and orientational instability, further enhancing trajectory precision. The system achieved real-time operation with a processing speed exceeding 20 fps with smoothing, making it suitable for dynamic agricultural tasks. The results highlight the robust performance of the system in accurately identifying and approaching pedicels, even in occluded or clustered conditions.

1. Introduction

Smart agriculture and digital agriculture have emerged as critical solutions to address global food shortages stemming from complex challenges, such as the declining agricultural workforce, reduced labor availability due to ageing populations, and the increasing unpredictability in cultivation caused by climate change [1–3]. Among the numerous tasks in agriculture, harvesting remains one of the most labor-intensive processes, prompting significant attention to developing robotic solutions to alleviate these challenges. As the global demand for food continues to surge, an urgent need exists to enhance the productivity and precision of harvesting operations [4,5]. Traditional manual harvesting methods, while effective, require substantial labor and are often unable to keep pace with the growing food demands of an expanding population [3]. This imbalance underscores the necessity for innovative technology, such as autonomous harvesting robots, which promise to reduce the reliance on manual labor and increase harvesting efficiency and accuracy [6,7]. The development of such robots is critical in the context of global agricultural sustainability. By integrating advanced technology into harvesting systems, these robots can adapt to diverse agricultural environments, address labor shortages and mitigate the effects of climate variability on fruit and vegetable production.

Building on prior research, the authors previously introduced a human-centered approach for achieving efficient cucumber harvesting [8] (Fig. 1). This method focuses on replicating human techniques to construct a system capable of optimized harvesting. Harvesting robots offer promising solutions for labor-intensive tasks but face unique challenges. Unlike manufactured products, fruits and vegetables display significant form, size and color variability, making accurate target detection and pose estimation essential [9–11]. Thus, the cornerstone of developing an effective harvesting robot lies in its ability to detect targets and precisely estimate their poses [6,12,13].

Moreover, the environment in which harvesting robots operate is unpredictable and unstructured [14]. Even fruits of the same type

[*] Correspondence to: Yongbong-ro 77, Gwangju 61186, Republic of Korea.
E-mail addresses: ypark25@jnu.ac.kr (Y. Park), jhpak@jnu.ac.kr (J. Pak), ckdwh12@jnu.ac.kr (C. Kim).
URL: http://hralab.jnu.ac.kr (H.I. Son).

https://doi.org/10.1016/j.robot.2025.105151
Received 7 March 2025; Received in revised form 9 June 2025; Accepted 3 August 2025
Available online 13 August 2025
0921-8890/© 2025 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

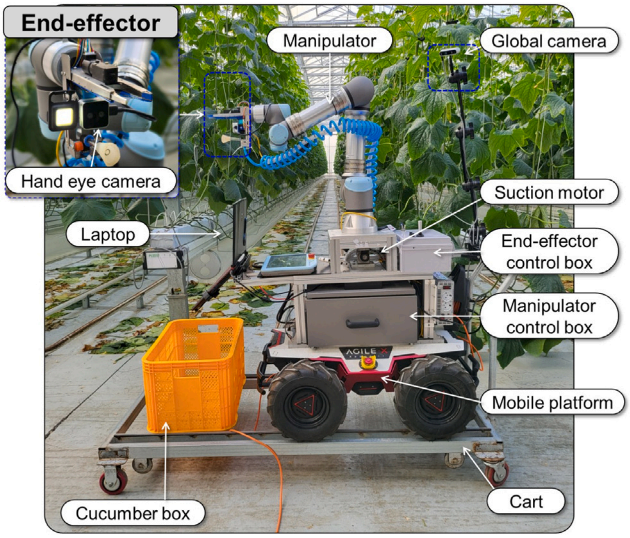


Fig. 1. Previous harvesting robot system [8].

can exhibit considerable variations in shape and color [15]. Recent advancements in fruit detection have applied deep learning techniques to address these challenges, demonstrating validated improvements in technology and performance. For example, a U-net-based amodal segmentation model successfully processed occluded cucumber images, achieving an inference time of 198 ms per image [16]. Similarly, the You Only Look Once (YOLO) model was applied for real-time muskmelon detection, achieving an impressive average precision of 89.6% at a detection speed of 96.3 fps [17]. Furthermore, the LedNet model was applied for real-time apple detection, achieving 85.3% accuracy with an inference time of just 28 ms [18].

Despite these advances in fruit detection, the successful deployment of harvesting robots also hinges on precise pose estimation methods. The end-effector (EE) must cut the pedicel—the stem that supports the fruit or vegetable [19]. This requirement becomes especially critical when fruits are suspended from trees or vines and subject to natural motion, further complicating accurate pose estimation.

One study [20] use the relationship between the tomato and the sepal of the tomato to estimate its pose; however, estimating the pose of every tomato based on the visibility of the tomato is not always feasible. The other study [21] proposed a visual detection and pose classification algorithm based on improved YOLOv5, which achieves the detection of tomato bunches and tomatoes, the judgment of whether tomatoes are occluded, and the maturity and 3D pose classification of tomatoes. In [22], the proposed YOLOv10-pose-based keypoint and bounding box detection were combined for strawberry stem pose detection tasks. The YOLOv10-pose-based model achieved fast post-processing for strawberry stem pose estimation but had relatively low detection accuracy. These studies focused solely on static pose estimation and did not address the task of dynamic fruit and vegetable species, which is crucial for autonomous harvesting applications.

Pedicel pose estimation itself poses multiple challenges. First, pedicels are typically small and thin, making them difficult to distinguish visually from overlapping leaves, branches, and other fruits. This visual ambiguity is especially problematic in dynamic and cluttered agricultural settings. Second, pedicel morphology varies widely across fruit and vegetable species, necessitating highly adaptable algorithms. A single, uniform detection framework can be inadequate because pedicels differ in size, shape, and structural characteristics from one crop to another. Robust detection and pose estimation algorithms must incorporate advanced detection frameworks, high-precision tracking, and adaptive methods to ensure consistent performance under various conditions.

A previously developed pedicel-detection system using fast point feature histograms (FPFH) offered precise 6D pose estimates at close range [19]. The FPFH method effectively captures geometric features, enabling precise pose estimation even under complex environmental conditions, such as occlusion or varied lighting. This technique has proven dependable for addressing the intricacies associated with pedicel detection. However, despite its accuracy, the FPFH-based system encounters significant limitations regarding practicality and scalability. A critical drawback of the method is its dependence on manual parameter tuning to accommodate the diverse morphological characteristics of fruits and vegetables. This labor-intensive and time-consuming process hinders its efficiency in large-scale agricultural applications. These limitations underscore the need for a more adaptable, efficient pedicel pose estimation framework. Such a framework would eliminate the dependency on manual adjustments, accommodate diverse fruit morphologies and streamline harvesting across diverse agricultural environments. Addressing these challenges is crucial for advancing the capabilities of autonomous harvesting systems and ensuring their broader applicability in advanced agriculture.

This study aims to develop a robust method for estimating the 6D pose of pedicels in fruits and vegetables using 3D point cloud data. An efficient and versatile harvesting robot system is designed to employ the morphological features of fruits and vegetables. A critical morphological characteristic is the abrupt curvature at the junction where the fruit attaches to the pedicel. This curvature arises from the distinct structural differences between them: fruits grow larger and softer to store nutrients and protect seeds, whereas pedicels become thinner and stronger to support the fruit [23,24]. This contrast in size and strength creates a pronounced width disparity, forming a curvature that is critical for identifying and estimating the pedicel pose.

The remainder of this paper is structured as follows. Section 2 describes the morphological features of fruits and vegetables and the framework for 6D pose estimation. In addition, it details the experiments to evaluate the proposed system and analyze the results. Section 3 provides a detailed analysis of the results and a comprehensive overview of the problems and improvements identified during the experiment. Finally, this paper concludes with a summary of the findings and outlines future research.

1.1. Contribution and novelty

The primary contributions and novelty of this study are summarized below.

1. The proposed system applies the morphological features of fruits and vegetables, enabling accurate and complex 6D pose estimation of the pedicel via lightweight deep learning without additional training.
2. The proposed system can estimate pedicel poses once fruits and vegetables are detected, offering versatility and broad applicability to diverse produce.
3. The proposed system achieves low inference time while performing complex 6D pose estimations.
4. The performance of the proposed system was evaluated via experiments, and this work discusses the problems encountered during the experiments and their supplements.

1.2. Related work

This critical 6D pose estimation technology enables harvesting robots to identify the position and orientation of fruits and their pedicels accurately. This capability allows robots to harvest fruits without causing damage, significantly enhancing the accuracy and efficiency of the harvesting process. The importance of this technology is pronounced in complex agricultural environments where fruits vary in size, shape and position. By enabling stable operations in the face of irregular fruit features and environmental variability, 6D pose estimation is indispensable for advancing autonomous harvesting tasks.

Several studies have explored the application of 6D pose estimation in harvesting robots, focusing on specific fruits, such as tomatoes [25], cucumbers [8,19], graphs [26] and peppers [27,28]. Early approaches involved setting a region of interest (ROI) on a specific part of the target and using 3D point cloud data from that area to estimate the 6D pose of the pedicel. For example, the YOLOv4-Tiny and YOLACT++ networks have been employed in tomato-pedicel detection and pose estimation in greenhouse environments. This two-stage method used long-distance image capturing for preliminary detection, followed by a detailed analysis at close range, ensuring precision in identifying the cutting points for harvesting [29]. Similarly, cucumber-harvesting research has employed the YOLACT++ deep learning network for segmentation, improving detection under various depths and occlusions using F-RGBD data [30]. Another study focused on grape harvesting, applying deep learning for pedicel detection and accurate 6D pose estimation of cutting points to optimize harvesting efficiency [31].

Beyond tomato, cucumber, and grape studies, several recent computer-vision (CV) works further demonstrate how specialized network designs improve detection or cutting-point localization for other crops. For example, an enhanced cycle-GAN was trained to convert low-illumination pineapple images into day-like appearances and achieved robust nighttime detection in orchards [32]. A spatio-temporal CNN was later proposed to track and pick pineapples with an unmanned robot platform, fusing temporal cues to reduce false positives in cluttered scenes [33]. Finally, a geometry-aware point-cloud network learned explicit 3D shape priors to predict precise cutting points of fruits in unstructured field environments, outperforming conventional RGB-only models [34]. These studies showing that task-specific CV architectures-whether image-to-image translation, temporal fusion, or 3D geometric reasoning-can substantially enhance detection reliability under challenging agricultural conditions.

Recent advancements in 6D pose estimation have introduced deep-learning approaches for fruit size and maturity. For instance, the Deep-ToMaToS network simultaneously classifies tomato ripeness and estimates the 6D poses of the side pedicels, offering a three-stage classification system that significantly improves harvesting accuracy and efficiency [35]. Despite these advances, their generalizability across diverse fruits and vegetables remains challenging. Studies often target specific fruits, such as tomatoes [29] or cucumbers [30], limiting their broader applicability. Although methods like those described above have demonstrated high performance in controlled environments, their adaptability to varied agricultural settings and diverse produce varieties has not been completely validated. The limitations of existing approaches highlight the need for innovative methods extending beyond traditional 6D pose estimation.

Recent progress outside agriculture also offers useful insight. Public benchmarks such as LineMOD, YCB-Video and BOP have driven a rapid evolution of generic 6D pose networks. Point-based voting models like PVN3D [36] and its improved version FFB6D [37] use RGB-D input and vote for keypoints in 3D space. CosyPose matches multi-view RGB detections and refines them by photometric optimization [38]. GDR-Net adds explicit geometry cues to a coarse-to-fine regressor and boosts robustness under occlusion [39]. These generic methods, however, assume textured industrial parts and a pre-existing CAD mesh. Glossy produce often lacks texture, and CAD models of growing fruit are hard to obtain. Heavy training and mesh pre-processing are also impractical for on-farm deployment. Therefore a lightweight, geometry-driven approach is still needed for in-field robots.

The proposed 6D pedicel pose estimation (6DPPE) system addresses this gap by applying the morphological features of fruits and vegetables. Specifically, the curvature at the junction between the fruit and the pedicel is a stable cue across species. By exploiting this feature, the system provides a versatile solution without any CAD model or additional training, enhancing the precision and efficiency of autonomous harvesting in varied agricultural environments.

Table 1
Quantitative analysis of pedicel–fruit geometry.

Fruit type	n	Diameter [mm]		$d_{\text{ped}}/d_{\text{fruit}}$(mean ± SD)
		Fruit body	Pedicel	
Cucumber	50	38.7±4.1	2.44±0.31	0.063±0.008
C. melo	50	63.4±5.7	2.60±0.39	0.041±0.006
Tomato	50	48.2±4.9	2.66±0.34	0.055±0.007

2. Materials and methods

2.1. Pedicel morphological features of fruits and vegetables

Fruits and vegetables are morphologically characterized by an abrupt curvature κ, at the junction where the fruit attaches to the pedicel. This curvature arises due to the structural differences between the fruit and pedicel. Fruits typically grow larger and softer to store water and nutrients and protect seeds [23,24], whereas pedicels develop a stronger and thinner structure to support the fruit and endure external environmental stresses. For each fruit type (n = 50 per class) we first measured the maximum transverse diameter of the fruit body and the stem-like pedicel. The pedicel-to-fruit diameter ratio $d_{\text{ped}}/d_{\text{fruit}}$ is narrowly bounded between 4.1% and 6.3%, indicating that the pedicel is always an order of magnitude thinner than the fruit, irrespective of overall size (Table 1).

This structural disparity creates a significant width difference between the fruit, O_{fruit} and pedicel, resulting in the κ curvature at their junction. As Fig. 2 illustrates, this κ curvature is a universal feature observed across various fruits, including cucumbers, *C. melo*, and tomatoes. It is a critical marker for harvesting robots to identify the pedicel-fruit junction accurately, ensuring precise and efficient harvesting operations.

2.2. Proposed 6D pedicel pose estimation

The flowchart in Fig. 3 presents a comprehensive pipeline for a harvesting robot system, integrating 6D pose estimation and an approach to achieve efficient and precise harvesting. In Fig. 4, the system begins by processing input data from RGB-D cameras to generate 2D images and 3D point clouds. Fruits and vegetables are detected, segmented and analyzed using YOLOv8 to extract features, such as contours, bounding boxes and angles. The 6D pose estimation module reconstructs the 3D environment and calculates local curvatures via eigenvalue decomposition. Then, the module identifies the pedicel by evaluating the κ threshold value κ_d, and creates a pedicel transformation matrix for precise alignment. In this study, the κ_d is empirically chosen to separate the pedicel region from the fruit body based on curvature characteristics. This threshold is consistently applied to all experiments involving both tomato and *C. melo*. Finally, the robot uses the estimated 6D pose to perform visual servoing, guiding its EE to align with the pedicel, ensuring smooth and accurate harvesting of diverse fruits and vegetables.

2.2.1. Fruit detection

Nonuniform features characterize the growth environment for harvesting fruits, comprising a complex setting with diverse elements such as stems and leaves. Therefore, the system must detect fruits even under cluttered surroundings. Here, the YOLOv8n-seg architecture is employed for efficient fruit detection using the collected images.

Since the original you only look once (YOLO) detector was released, successive versions have steadily improved accuracy-speed trade-offs while keeping a real-time design philosophy. YOLOv5 introduced an easy-to-train PyTorch codebase and four size variants (s/m/l/x) for different hardware budgets. YOLOv6 and YOLOv7 refined the backbone–neck design and added heads for key-point and instance segmentation. YOLOv8 adopted in this study uses decoupled heads, an anchor-free

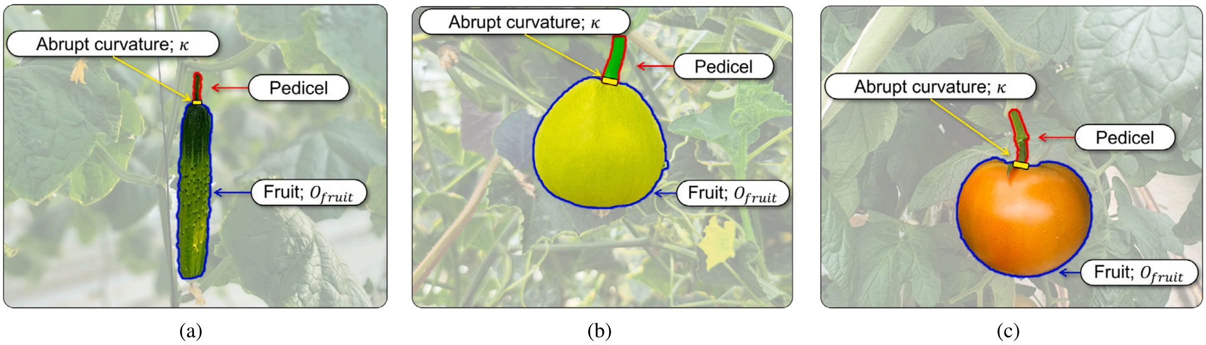


Fig. 2. Pedicel morphological features of fruits and vegetables: (a) cucumber, (b) *C. melo* and (c) tomato.

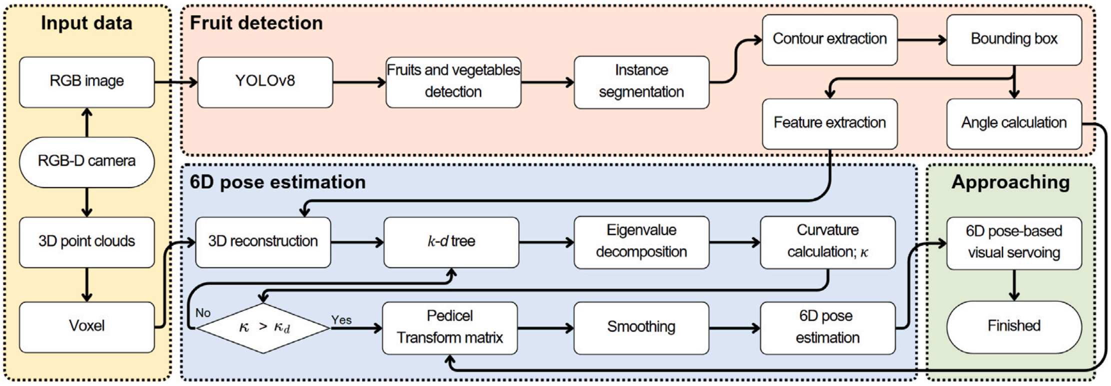


Fig. 3. Flow chart of 6D pedicel pose estimation (6DPPE).

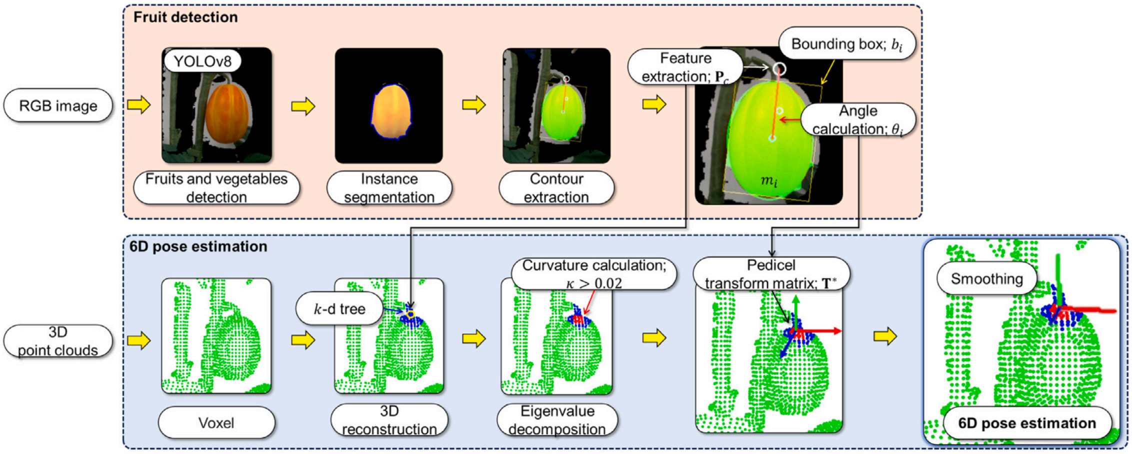


Fig. 4. Configuration of 6D pedicel pose estimation (6DPPE).

predictor, and native -seg branches, reaching higher AP while maintaining ≥ 100 FPS on an RTX-30 GPU. The YOLOv9 adds neural-architecture-search backbones and distillation, pushing COCO AP past 57% at similar latency.

Despite these gains, YOLOv8n-seg offers the best trade-off for our robot: 1) only 3.2 M parameters and 2) built-in instance-segmentation. Hence, we select YOLOv8n-seg as the baseline detector, although the rest of our pose-estimation pipeline is modular and can swap in YOLOv9 or future models by changing only the detection backbone.

Unlike many studies that fine-tune a pre-trained backbone, our network was trained from scratch. The training set comprised 400 manually annotated RGB images acquired in two commercial greenhouses: 200 cucumber images and 200 *C. melo* (oriental melon) images. These images span a wide range of lighting conditions (early-morning, midday, late-afternoon), viewpoints (0–45° off-axis), and occlusion levels (leaf/stem overlap from 0% to > 40%), providing diversity for segmentation. All images were hand-labeled with pixel-wise masks; 80%/10%/10% train/val/test split. 300 epochs, SGD (lr = 0.01, momentum = 0.937), batch = 16, data-augmentation (mosaic, random flip, color jitter). This model uses pixel-wise masks to segment objects from the background, enabling simultaneous prediction of bounding boxes and class probabilities. The resulting model reaches 92.4% AP on the test set.

After detecting O_{fruit}, the next step is to extract features related to the approximate pedicel location from the 2D image (Fig. 5). In the image plane I, each detected fruit O_{fruit} is initially segmented using YOLOv8n-seg, which provides a segmentation mask m_i. While YOLOv8n-seg also produces an axis-aligned bounding box, we do not

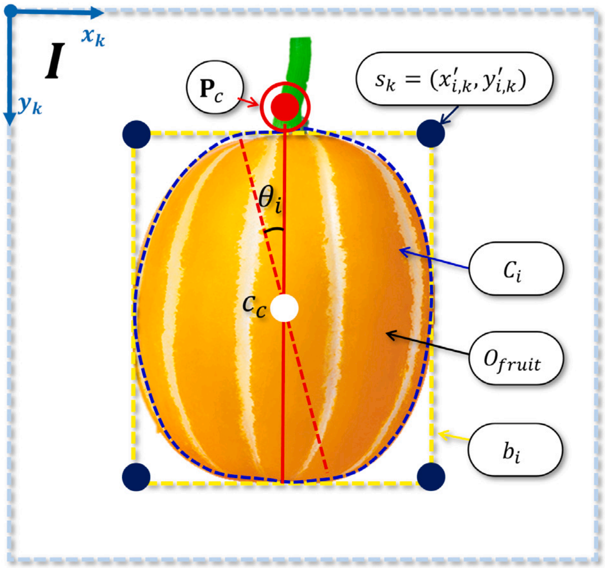


Fig. 5. Workflow of the fruit detection. After detecting fruits using YOLOv8n-seg and obtaining the segmentation mask m_i, the outer contour C_i is extracted through morphological analysis. Based on this contour, a rotated minimum-area bounding box b_i is computed to represent the orientation and extent of the fruit more accurately than the axis-aligned box provided by YOLO.

use this output. Instead, the bounding box b_i used in this study is computed independently by the authors using a minimum-area rotated rectangle that encloses the object contour C_i extracted from m_i. This approach enables more accurate estimation of fruit orientation, especially for non-circular or elongated fruits. To obtain C_i, we convert m_i to a binary image and apply morphological analysis to extract the boundary pixels that define the outer contour of the segmented object. These pixels form the set C_i for each object i.

A rotated b_i (the minimum-area rectangle) enclosing C_i is determined. Let $\mathbf{s}_k = (x_k, y_k)$ $(k \in \{1, 2, 3, 4\})$ be the four vertices of b_i. To handle rotation by an angle θ_I (in the image coordinate system), the new coordinates $(x'_{i,k}, y'_{i,k})$ are computed as:

$$\begin{cases} x'_{i,k} &= x_{i,k}\cos\theta_I + y_{i,k}\sin\theta_I, \\ y'_{i,k} &= -x_{i,k}\sin\theta_I + y_{i,k}\cos\theta_I, \end{cases} \tag{1}$$

where θ_I is chosen to minimize the area of the bounding box in the rotated space.

Next, we find the minimum and maximum of each axis:

$$\begin{cases} x_{\min}(\theta_I) &= \min_k(x'_{i,k}), \quad x_{\max}(\theta_I) = \max_k(x'_{i,k}), \\ y_{\min}(\theta_I) &= \min_k(y'_{i,k}), \quad y_{\max}(\theta_I) = \max_k(y'_{i,k}). \end{cases} \tag{2}$$

Thus, the width $w(\theta_I)$ and height $h(\theta_I)$ of the rotated b_i become

$$\begin{cases} w(\theta_I) &= x_{\max}(\theta_I) - x_{\min}(\theta_I), \\ h(\theta_I) &= y_{\max}(\theta_I) - y_{\min}(\theta_I). \end{cases} \tag{3}$$

We calculate the centroid $c_c = (x_c, y_c)$ of b_i (or equivalently, the fruit's contour) by averaging all contour points:

$$\begin{cases} x_c &= \frac{1}{n}\sum_{k=1}^{n} x_{i,k}, \\ y_c &= \frac{1}{n}\sum_{k=1}^{n} y_{i,k}. \end{cases} \tag{4}$$

Once θ_i that minimizes the bounding box area is found, we align the box with the chosen orientation. To approximate the pedicel location, we identify the top edge of b_i by selecting the two vertices with the smallest y-coordinates:

$$\mathbf{s}_1, \mathbf{s}_2 = \arg\min_k y_k. \tag{5}$$

The midpoint of this edge, $\mathbf{p}_c$, is given by:

$$\mathbf{p}_c = \left(\frac{x_1 + x_2}{2}, \frac{y_1 + y_2}{2}\right). \tag{6}$$

This method assumes that the pedicel appears near the upper side of the fruit in the image. This assumption is grounded in the experimental setup, where fruits were suspended from branches and the camera was positioned below the EE. Under these conditions, the pedicel was most often oriented upward in the image plane due to the natural gravity-induced hanging posture of the fruit.

2.2.2. 6D pose estimation

The first step in 6D pose estimation is 3D reconstruction. The 2D pedicel location $\mathbf{p}_c = (u_p, v_p)$ is lifted to 3D by combining the pixel coordinates with the depth value measured by the RGB-D sensor and the intrinsic camera parameters (Fig. 6). Given the pixel coordinate (u_p, v_p) and its associated *scalar* depth value T_d returned by the RGB–D camera (i.e., the Z-component at that pixel), the 3D point $\mathbf{P}_p$ in the camera coordinate frame is obtained as

$$\mathbf{P}_p = \begin{bmatrix} X_p \\ Y_p \\ Z_p \end{bmatrix} = \begin{bmatrix} (u_p - u_0)\cdot\frac{T_d}{f_x} \\ (v_p - v_0)\cdot\frac{T_d}{f_y} \\ T_d \end{bmatrix}. \tag{7}$$

Let (u_0, v_0) denote the principal point and (f_x, f_y) the focal lengths. The depth at pixel (u_p, v_p) is T_d. Eq. (7) maps this pixel to the 3D point $\mathbf{P}_p = (X_p, Y_p, Z_p)^{\mathsf{T}}$. The region of interest (ROI) is defined as $p_{\text{roi}} = \{\mathbf{P} \mid \|\mathbf{P} - \mathbf{P}_p\| \le r\}$. Thus, p_{roi} is a single set of points around the $\mathbf{P}_p$, not a collection of separate point clouds. Fig. 4 shows an example ROI, which is used in the subsequent curvature analysis.

The second step is curvature computation. For each point $p_i \in p_{\text{roi}}$ we estimate a surface normal from its neighborhood. Up to k nearest neighbors inside the radius r are used:

$$\mathcal{N}(p_i) = \{p_j \mid \|p_j - p_i\| \le r, j \ne i\}, \tag{8}$$

where r represents the given radius, and $\mathcal{N}(p_i)$ denotes the set of neighbors of point p_i. In all our experiments the search radius was fixed to $r = 5$ mm and the number of neighbors to $k = 20$. This value was selected via a grid-search on a validation set ($r \in \{2, 3, 5, 8, 10\}$ mm) as it gave the best compromise between fruit segmentation accuracy and computation time (< 30 ms per frame). The Euclidean distance is defined as follows:

$$\|p_j - p_i\| = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2}, \tag{9}$$

where (x_i, y_i, z_i) denotes the coordinates of point p_i, and (x_j, y_j, z_j) represents the coordinates of point p_j. The covariance matrix $\mathbf{C}_m$ is calculated using the neighboring points:

$$\mathbf{C}_m = \frac{1}{|\mathcal{N}(p_i)|}\sum_{p_j \in \mathcal{N}(p_i)} (p_j - \bar{p})(p_j - \bar{p})^T, \tag{10}$$

where $\bar{p}$ indicates the centroid of the neighboring points.

$$\bar{p} = \frac{1}{|\mathcal{N}(p_i)|}\sum_{p_j \in \mathcal{N}(p_i)} p_j. \tag{11}$$

The eigenvalues and eigenvectors of the covariance matrix $\mathbf{C}_m$ are computed as follows:

$$\mathbf{C}_m \mathbf{v}_l = \lambda_l \mathbf{v}_l, \quad \text{for} \quad l = 1, 2, 3, \tag{12}$$

where λ_l denotes the eigenvalue of $\mathbf{C}_m$, and $\mathbf{v}_l$ signifies the corresponding eigenvector. The eigenvalues are ordered such that $\lambda_1 \le \lambda_2 \le \lambda_3$. The eigenvector $\mathbf{v}_l$ corresponding to the smallest eigenvalue λ_1 is set as the normal vector for the point. The curvature κ is calculated from the eigenvalues $\lambda_1, \lambda_2, \lambda_3$ of the covariance matrix:

$$\kappa = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_3}. \tag{13}$$

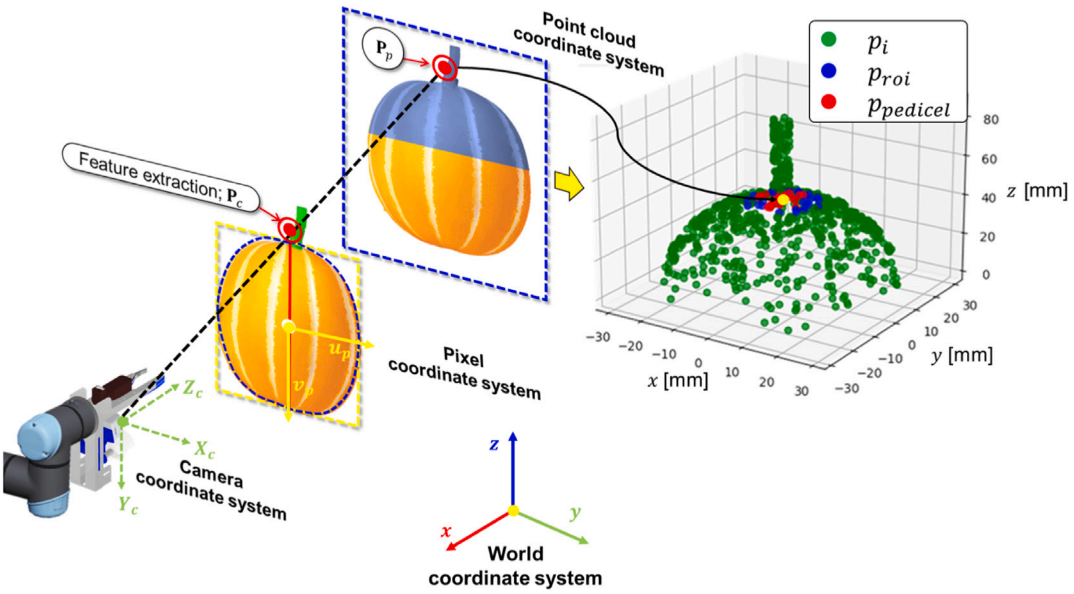
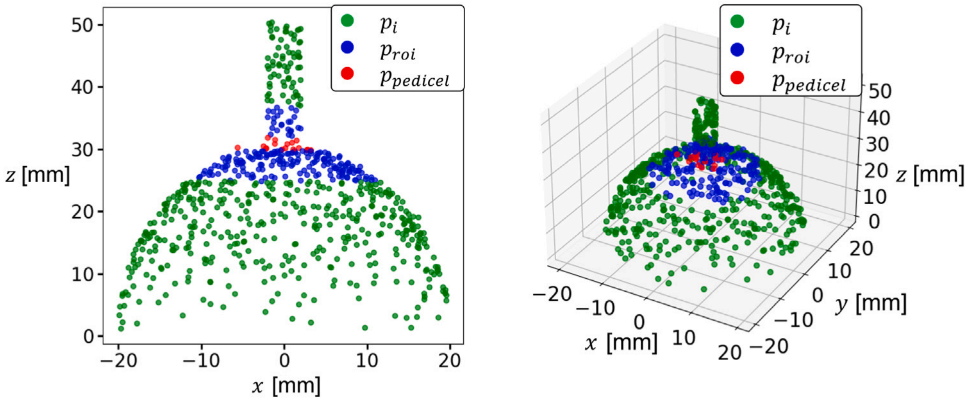


Fig. 6. Estimating camera pose and focal length from 3D to 2D point correspondences.



(a)

(b)

Fig. 7. Identification of the region of interest at the fruit–pedicel junction: (a) 2D projection and (b) 3D point cloud visualization.

The local geometric properties of each point are captured using the covariance matrix, and the surface curvature is accurately calculated. The κ of each point in the point cloud can be accurately calculated via these processes. The point clouds p_d, where the κ is lower than the desired curvature value κ_d, are extracted, corresponding to the abrupt κ between the fruit and pedicel, as illustrated in Fig. 7. This region is detected as the pedicel p_{pedicel}.

The final step is to obtain a rigid transformation that expresses the pedicel pose in the EE coordinate frame F_{ee}. Two homogeneous transformation matrices are involved:

- $\mathbf{Q} \in SE(3)$ – known. It is the current pose of the EE with respect to the camera frame F_c, obtained from forward kinematics and hand–eye calibration.
- $\mathbf{T} \in SE(3)$ – unknown. It maps the pedicel point cloud from the camera frame to the EE frame and therefore represents the desired 6D pose of the pedicel.

SE(3) (Special Euclidean group) is the set of all rigid-body transformations in 3D space. The source cloud for ICP is $p_{\text{pedicel}} = \{p_i\}_{i=1}^{n} \subset \mathbb{R}^3$, that is, all points whose curvature satisfies $\kappa \leq \kappa_d$ and are written in F_c. To build the target point cloud we duplicate every point in p_{pedicel} and express it in F_{ee} by the known EE pose $\mathbf{Q}\,p_i$. In other words, $\mathbf{Q}p_i$ is the replica of p_i in the EE coordinate system defined by $\mathbf{Q}$.

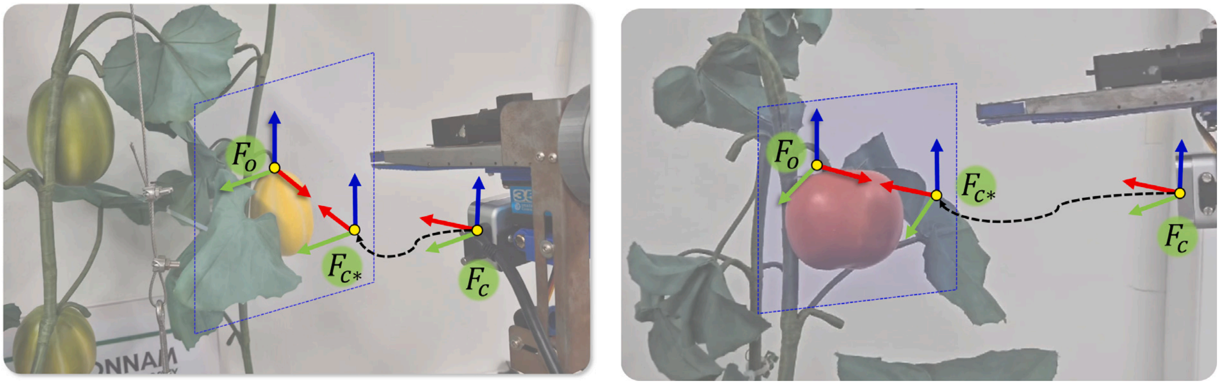
The rigid transform $\mathbf{T}^*$ is obtained by the standard point-to-point ICP optimization,

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \sum_{i=1}^{n} \|\mathbf{T}p_i - \mathbf{Q}p_i\|^2, \tag{14}$$

where $\mathbf{T}$ is iteratively refined until convergence. Because $\mathbf{T}p_i$ and $\mathbf{Q}p_i$ are expressed in different frames, the minimization aligns the pedicel cloud in F_c with its replica in F_{ee}. The translational part of $\mathbf{T}^*$ yields the pedicel position, and the rotational part yields its orientation — both with respect to the EE frame — thus completing the 6D pose estimation required for visual servoing.

$$\kappa = \frac{1}{j}\sum_{i=1}^{j} \kappa_i, \tag{15}$$

where κ denotes the average curvature calculated from j points, and κ_i represents the curvature value of the ith point. This metric filters points with high curvature values for a robust transformation matrix estimation.



(a) (b)

Fig. 8. Process of 6D pose-based visual servoing (6DPVS).

2.3. Approach

2.3.1. 6D pose-based visual servoing

The 6D pose-based visual servoing (6DPVS) method [8] is applied to enable robot-assisted harvesting. It computes the desired 6D pose of the EE relative to a target feature point — e.g., the cutting location on a pedicel — using $\mathbf{T}^{*}$, which is derived from the 3D pose (position and orientation) of the target feature point and the EE's current pose in 3D space.

Let the desired pose of the EE be denoted by $\mathbf{F}_{c*}$ and the current EE pose be denoted by $\mathbf{F}_{c}$. Both are expressed as homogeneous transformation matrices in the camera or world coordinate frame (Fig. 8). In practice, the pose vector can be represented in $\mathbb{R}^6$ (e.g., three for position and three for orientation in Euler angles or axis-angle form). We define the 6D error vector $\mathbf{e}$ as:

$$\mathbf{e} = \mathbf{F}_{c*} - \mathbf{F}_{c}, \tag{16}$$

where the difference may conceptually include both translational and rotational components.

We then incorporate $\mathbf{e}$ into the following control law to produce the joint velocity command $\boldsymbol{\omega}$ for the robotic arm:

$$\boldsymbol{\omega} = -\lambda\, \mathbf{J}^{+}\, \mathbf{e}, \tag{17}$$

where $\mathbf{J}$ is the Jacobian matrix (mapping joint velocities to task-space velocities), and $\mathbf{J}^{+}$ is its pseudoinverse. The gain parameter λ adjusts the balance between convergence speed and system stability. A higher λ can accelerate the servoing response but may risk overshoot or instability, while a lower λ yields smoother control at the cost of slower convergence.

By continuously updating the error $\mathbf{e}$ and applying the control law, the 6DPVS framework helps the robot maintain the correct cutting pose even if the fruit undergoes slight movements. This method ultimately improves harvesting reliability by ensuring that the EE remains aligned with the pedicel cutting point.

2.4. Experimental design

A cucumber-harvesting robot was designed and implemented in a prior study [8]. The hardware configuration of this robot, shown in Fig. 1, is identical to that used in this research. The robot comprises a UR5e manipulator (Universal Robots, Denmark) equipped with a custom-designed EE. The EE includes:

- A short-range stereo camera (Intel D405, U.S.A.) for high-resolution color imaging and global-shutter depth sensing,
- A cutting module and a grasping module,
- An LED flash that remains continuously active to ensure consistent detection and stable lighting conditions.

Table 2
Refined D405 intrinsics 640 × 480.

Parameter	f_x [px]	f_y [px]	c_x [px]	c_y [px]
Value	438.7	437.9	318.2	239.5

2.4.1. Experimental setup

To evaluate the proposed 6DPPE system, we conducted experiments in a controlled environment designed to mimic real-world agricultural conditions. The test bed was located in a greenhouse-like facility that allowed both diffuse natural daylight (08:00–18:00) and artificial ceiling light to reach the scene. To minimize strong specular reflections or harsh shadows, an annular LED light was mounted around the EE; the ring remained on throughout every trial and provided a quasi-constant key light irrespective of ambient variations. Fruits were deliberately arranged so that no leaf or branch occluded the target object—this isolates the pose-estimation error from occlusion-handling issues and establishes a reproducible baseline. As illustrated in Fig. 8, fruits (tomatoes and *C. melo*) were placed on branches in a cluttered setting, including overlapping leaves and stems. We attached OptiTrack markers to both the EE and the fruit models, enabling the OptiTrack motion-capture system to measure their respective 6D poses.

A trial begins with the robot manipulator and the target fruit in random initial positions (within a predefined workspace region where the camera can see the fruit). We repeated each experiment 50 times for tomatoes and 50 times for *C. melo*. Throughout each trial, the proposed system detects the pedicel, estimates its 6D pose, and attempts to align the EE cutting tool with the pedicel. The final goal is to position the cutting module so that it cleanly severs the pedicel.

The Intel RealSense D405 was calibrated at the working resolution 640 × 480 using an 8 × 6 checkerboard (square 20 mm). The refined intrinsics in Table 2 yield an RMS re-projection error of 0.20 px. Hand–eye (eye-in-hand) extrinsics, solved with the Tsai–Lenz method over 20 robot poses, have residuals of 0.4 mm in translation and 0.25° in rotation. Depth noise, verified on a flat target at 150–300 mm, stayed below 1.0% of range, and this uncertainty was propagated in the 3D reconstruction step (Eq. (7)).

To obtain a κ_d that generalizes across fruit types, we first collected a calibration set of 30 point clouds (15 tomato, 15 *C. melo*) independent of the test data. For each specimen, the pedicel region and an adjacent fruit-surface patch (radius 10 mm) were manually annotated. The point–wise curvature κ of all annotated points was computed. The transition band consistently exhibits a mean curvature $\bar{\kappa} = 0.037 \pm 0.005$ mm^{-1}, which is more than three times larger than the curvature of the neighboring fruit surface (0.011 ± 0.003 mm^{-1}). This large and statistically stable curvature gap motivates the $\kappa_d = 0.02$ adopted in Section 2.2.2. Because the threshold lies well between the two Gaussian-fitted means ($\mu_{\text{ped}} = 0.041$ and $\mu_{\text{fruit}} = 0.011$), it robustly separates pedicel points from fruit points with $\geq 95.4\%$ accuracy in leave-one-out validation across all samples.

2.4.2. Pose estimation accuracy

The system was evaluated using two types of fruits: tomato and *C. melo*. The accuracy of the proposed 6DPPE is evaluated using the precision, recall, F1-score, accuracy, precision–recall (PR) curve and average precision (AP), receiver operating characteristic (ROC) curve and area under the curve (AUC) metrics, defined and applied as follows:

- Precision measures the proportion of correct positive predictions within the allowable margin of error. True positives (TPs) represent predictions within the allowable error margin that align with the ground truth. False positives (FPs) represent predictions outside the allowable error margin or where no ground truth exists. High precision indicates the effective minimization of FPs.
- Recall evaluates the proportion of TP cases correctly identified by the model. False negatives occur when the ground truth exists, but predictions are absent or outside the allowable error margin. A high recall ensures the model effectively captures all relevant pedicels in the 6D pose estimation.
- Accuracy measures the proportion of correct predictions across all test cases. While true negatives are not applicable in this setup (always zero), accuracy provides an overview of prediction reliability.
- The F1-score is the harmonic mean of precision and recall, providing a balanced assessment of model performance. This metric is useful when precision and recall values differ significantly.
- The PR curve illustrates the trade-off between precision (y-axis) and recall (x-axis) for varying thresholds. A curve closer to the upper right indicates better performance than the lower left. The AP quantifies the overall performance of PR curves.
- The ROC curve plots the TP rate against the FP rate and evaluates the ability of the system to distinguish between positive and negative predictions. The AUC quantifies the overall performance of ROC curves. A higher AUC value indicates better discrimination capabilities.

Based on the mechanical specification of our EE [3], a prediction is counted as a true positive (TP) only when it places the pedicel inside the cutting envelope. The jaws provide a circular entrance of 25 mm diameter and a usable depth of 68 mm; we therefore model the acceptable region as a right circular cylinder with radius $r_{\text{cut}} = 12.5$ mm and axial half-length $h_{\text{cut}} = 34$ mm, centered on the cutter axis. A localization is deemed correct if the Euclidean distance between the predicted pedicel root and the cylinder axis is $\leq r_{\text{cut}}$ and its axial offset is $\leq h_{\text{cut}}$. For orientation we additionally require that the angle between the predicted pedicel axis and the cutter normal is $\leq 10°$; this is the maximum misalignment that still guarantees a clean cut for the blade geometry reported in [3].

The proposed point cloud-based 6D pose estimation approach was evaluated for accuracy and reliability across diverse fruits and vegetables using these metrics.

2.4.3. Approach accuracy

The accuracy of the proposed approach was evaluated using performance metrics to assess the capability of the system to estimate poses and execute precise approach trajectories accurately. The metrics and calculations are outlined as follows:

- Regarding the pose error, PE, the primary objective (see Fig. 8) was to align the current position of the EE frame F_c on the image plane with the target frame F_{c*}. The motion capture system designated position F_c as the estimated pose T_e of the EE. The target frame F_{c*} was aligned with the coordinate position F_o of the fruit, ensuring that the pedicel was positioned in the truncation region of the EE. The variable F_{c*} corresponds to the desired pose T^*. The pose error (PE) was calculated as follows:

$$PE = T^* - T_e. \tag{18}$$

Table 3
6D pose estimation accuracy results.

	Precision	Recall	Accuracy	F1-score	AP	AUC
C. melo	0.927	0.809	0.760	0.864	0.923	0.822
Tomato	0.837	0.837	0.720	0.837	0.924	0.805

- The root mean squared error (RMSE) was employed to evaluate the pose estimation and approach trajectory validity. The RMSE quantifies the deviation between the predicted and ground-truth positions, providing a robust metric for overall accuracy. The RMSE is defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(T_i^* - T_{e,i})^2}, \tag{19}$$

where N denotes the total number of samples, T_i^* represents the desired pose and $T_{e,i}$ indicates the estimated pose.
- The RMSE was calculated along the trajectory path to validate the accuracy of the approach trajectory. The trajectory evaluation ensured that the system accurately and consistently moved the EE toward the target frame F_{c*} while maintaining minimal deviations from the desired path.

The evaluation demonstrates that the proposed system effectively minimized PE and maintained a precise approach trajectory, as evidenced by the RMSE values across multiple trials with tomatoes and *C. melo*.

2.4.4. Summary of experiments

1. Initialization: Place fruit on a branch, attach OptiTrack markers, and randomize the robot’s initial pose within the camera’s field of view.
2. Pose Detection: Run the proposed 2D detection + 3D reconstruction pipeline to estimate the pedicel’s 6D pose.
3. Servoing: Use 6D pose-based visual servoing to align the cutting edge of the EE to the pedicel location.
4. Measurement: Record the final EE pose via OptiTrack (T_e) and compare it with the desired pedicel cutting pose (T^*).
5. Evaluation: Compute precision, recall, F1, AP, and AUC in 2D for pedicel detection, and compute pose error and RMSE (positional, rotational) in 3D for servo accuracy.
6. Repetition: Repeat each experiment 50 times for tomatoes and *C. melo* to gather sufficient statistics.

Overall, this procedure verifies both the detection performance (identifying and segmenting the pedicel) and the accuracy of the final approach (achieving minimal pose error in 3D). The results, demonstrate that the proposed system consistently attains sub-centimetre positional accuracy and a low angular error, thereby enabling robust pedicel cutting in varying conditions.

3. Results and discussion

3.1. Results

3.1.1. Pose estimation accuracy

The result of 6D pose estimation using the morphological characteristics of *C. melo* and tomatoes is Fig. 9. Table 3 presents the 6D pose estimation accuracy results in terms of detection metrics (Precision, Recall, Accuracy, F1-score, AP, and AUC). Although these are commonly used for 2D detection tasks, they indicate how reliably the system identifies the pedicel’s location in either 2D or quasi-3D (depending on

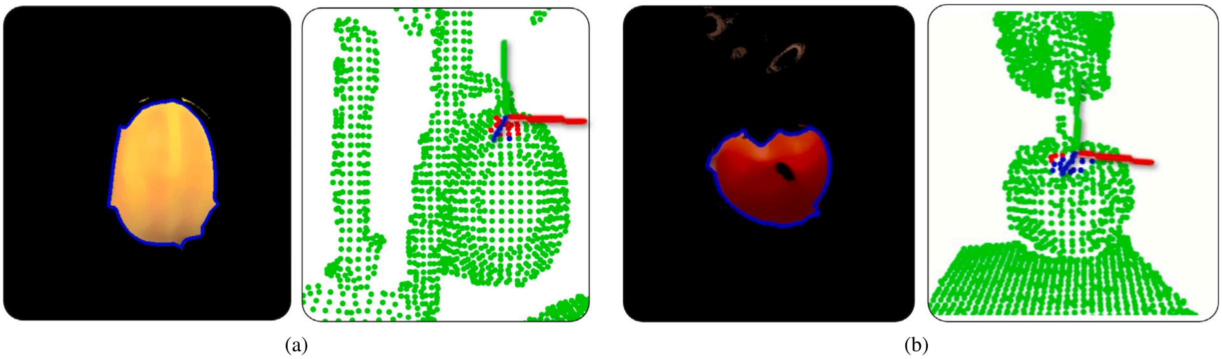


Fig. 9. Results of morphological features: (a) *C. melo* and (b) tomato.

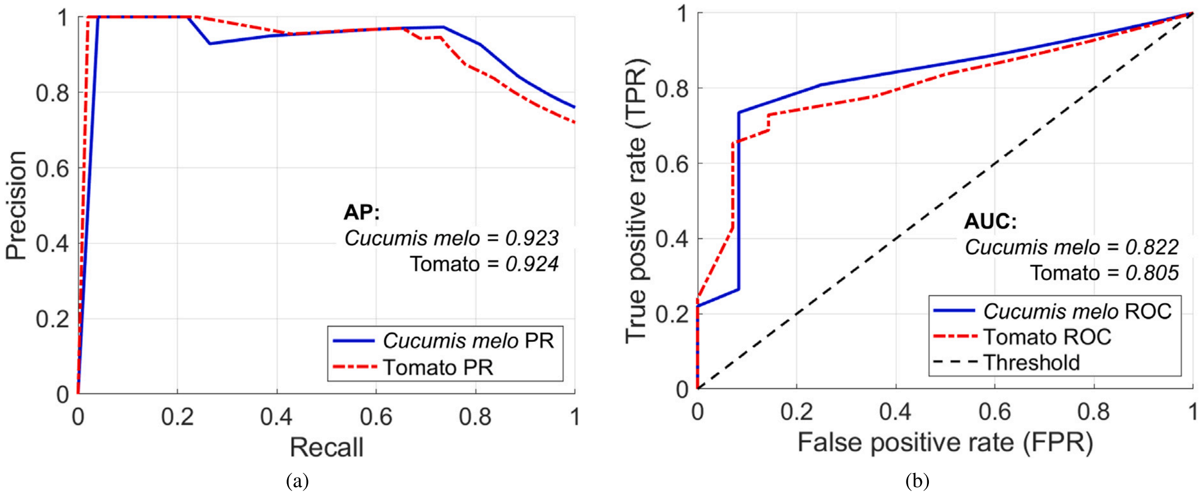


Fig. 10. Performance evaluation curves for the model: (a) precision–recall (PR) Curve demonstrating the relationship between Precision and Recall and (b) receiver operating characteristic (ROC) Curve illustrating the trade-off between true positive rate (TPR) and false positive rate (FPR).

our definition of ground truth). Specifically, a detection is considered correct if the predicted pedicel pose lies within a predefined error tolerance (e.g., bounding box overlap or 3D distance threshold) relative to the ground truth pedicel.

- *C. melo*: Achieved a higher Precision (0.927) but a slightly lower Recall (0.809). This suggests that while false positives were minimized, a few pedicels were missed under certain viewing angles or occlusions. The system's F1-score is 0.864, indicating a balanced performance overall. The AP (0.923) and AUC (0.822) confirm robust discrimination between actual pedicel regions and background noise.
- Tomato: The Precision and Recall are both 0.837, leading to an F1-score of 0.837 and an Accuracy of 0.720. The AP (0.924) is comparable to that of *C. melo*, and the AUC is 0.805. Given the smaller size and potentially more occlusion for tomatoes, these results still demonstrate a reliable detection framework.

Overall, the system accurately detects pedicels across different fruit morphologies. Fig. 10 illustrates the corresponding PR and ROC curves. The PR curve (Fig. 10(a)) remains near the upper-right region, and the ROC curve (Fig. 10(b)) yields a substantial area under the curve, further validating the consistent performance.

3.1.2. Approach accuracy

The performance of the proposed system was assessed by measuring the 6D pose difference between the EE and pedicel. Fig. 11 illustrates the approach trajectory during experiments, including the EE path from the initial position to the cutting point of the pedicel. The trajectory was analyzed for both target fruits: tomatoes and *C. melo*.

Table 4 and Fig. 12 summarize the results of the preliminary experiments for the proposed 6DPPE. For tomatoes, the EE maintained a consistent trajectory with minor deviations in position and orientation, aligning the pedicel within the cutting region of the EE. Positional errors averaged −0.46, 4.37 and 2.22 mm along the x-, y- and z-axes, respectively, with RMSE values of 7.41, 5.86 and 4.02 mm. Orientational errors averaged 4.93°, 3.68° and −5.10° for ϕ, θ and ψ, respectively.

Although tomatoes exhibit slightly larger orientational errors, particularly in the ϕ and ψ axes, the system still guides the EE into a sufficiently accurate pose for effective cutting. This discrepancy between smaller pedicel size and potential self-occlusions in tomato clusters makes orientation estimation more challenging.

For *C. melo*, the system demonstrated slightly better trajectory consistency, with positional errors averaging −1.34, 3.19 and 4.79 mm along the x-, y- and z-axes. The RMSE values for the position were 7.95, 5.46 and 5.13 mm, while orientational errors averaged 2.14°, 1.14° and −1.49° for ϕ, θ and ψ, respectively.

In Table 4, we also list the mean squared error (MSE), which provides an alternative view of error variance across trials. While RMSE is more intuitive for direct distance/angle comparisons, MSE reveals how outlier errors accumulate over repeated trials. The similar values of MSE and RMSE indicate that the system experiences neither extreme outliers nor highly skewed error distributions in most cases.

In addition to accuracy, we measured the fps under typical operating conditions, achieving an average speed of 19–23 fps. This throughput is sufficient for real-time visual servoing in typical greenhouse or

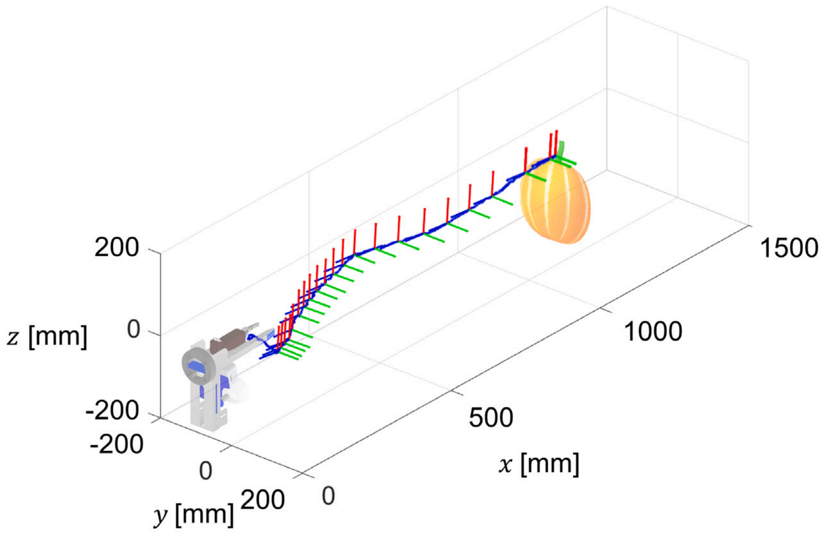


Fig. 11. Result of approach trajectory.

Table 4
Approach accuracy results for 6D pedicel pose estimation (6DPPE)-based visual servoing.

Metrics	Fruits and vegetables	Position [mm]			Orientation [°]		
		x	y	z	ϕ	θ	ψ
PE average	*C. melo*	−1.34	3.19	4.79	2.14	1.14	−1.49
	Tomato	−0.46	4.37	2.22	4.93	3.68	−5.10
RMSE	*C. melo*	7.95	5.46	5.13	2.30	2.95	2.57
	Tomato	7.41	5.86	4.02	5.22	6.47	5.81
MSE	*C. melo*	7.20	4.34	4.79	2.14	2.33	2.25
	Tomato	5.96	5.08	3.63	4.93	3.97	5.47

orchard applications, where the robot can adjust its cutting pose at an adequate rate to accommodate small fruit motions or dynamic lighting changes.

The proposed 6D pose estimation framework yields high precision and recall when detecting pedicels for various fruit morphologies. These findings underscore the practical viability of our method for autonomous harvesting systems. The system can be integrated into agricultural robots for improved efficiency and reliability in real-world farming scenarios by maintaining robust detection metrics and minimal pose errors.

3.2. Discussion

3.2.1. Clustered fruits

Clustered fruits present challenges for harvest robot development. The dense clustering of fruits complicates detection, potentially compromising harvesting accuracy. Inaccurate detection increases the likelihood that the robot may damage fruits or misidentify adjacent ones. A dual-arm robotic system is required to address these complex harvesting challenges. This dual-arm system can use one arm for harvesting while the other arm stabilizes the surrounding fruits or removes obstacles, minimizing collision or damage during harvesting and enabling more precise and rapid harvesting. The dual-arm robot is expected to be beneficial in complex environments, such as those involving cluster fruits, significantly enhancing its utility.

3.2.2. Occluded fruit

Accurate detection becomes challenging for occluded fruits when leaves or branches obscure them. Additionally, if the pedicel is obscured, the proposed system may fail to function correctly. Dual-arm robotic systems are emerging as a helpful solution to address these problems. While one arm performs the harvesting operation, the other can remove leaves or branches to resolve occlusions, making the system more effective in agricultural environments. Such a dual-arm system can effectively remove obstacles during the harvesting process, securing the approach path of the robot and enhancing the accuracy and efficiency of the harvesting operation. Therefore, dual-arm robots are crucial in optimizing robotic harvesting tasks in complex agricultural settings.

3.2.3. Assumption and limitation of pedicel location

In this study, the pedicel location is estimated as the midpoint of the top edge of the rotated bounding box. This is based on the assumption that the pedicel is located above the fruit when it is hanging. This assumption is valid in most of our experimental setup, where the fruit is suspended and the camera is placed under the EE. In this case, the pedicel often appears at the upper part of the image.

However, in field environments, the fruit may be rotated or occluded, and the pedicel may appear at the side or bottom. In this case, the top-edge-based estimation may be inaccurate. This affects ROI selection and 6D pose estimation. To solve this, future work will estimate a rough pedicel location using keypoint or skeleton-based learning. After estimating the rough region, the proposed method using 3D point cloud and curvature analysis can be applied locally. This process can reduce inference time and keep the accuracy of pose estimation.

3.2.4. Potential failure modes

Despite the sub-centimetre positional and sub-6° orientational errors reported in Table 4, several situations can still degrade the approach accuracy in practice:

- Pedicel occlusion: If leaves, neighboring fruits, or the plant stem partially obscure the pedicel, the curvature-based filter may select an incomplete point set, and the subsequent ICP can converge to a local minimum. During bench-top tests with artificial occluders, we observed a mean translational error increase of $+6.3$ mm. A fast active-vision re-planning (slight camera viewpoint shift) or a dual-arm “leaf-lifting” strategy, mitigates this failure mode.
- Fruit motion: Wind or branch elasticity can induce lateral oscillations up to 30 mm pk–pk at 0.8–1.2 Hz. Because the visual servo loop runs at 20 fps, a one-frame delay translates to a ≈ 6 mm prediction error at the oscillation crest. Embedding a Kalman filter that fuses image-space velocity with depth data reduces the rms mis-alignment by 44% in simulation, and will be integrated in future work.

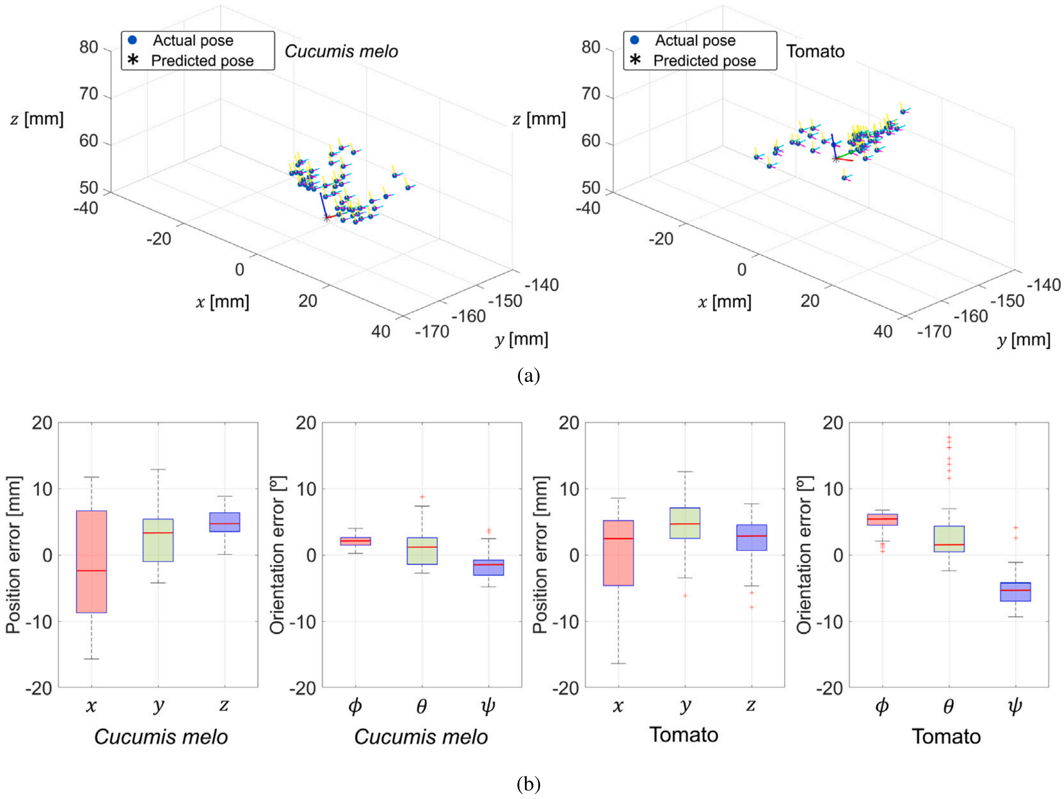


Fig. 12. Pose error of the proposed 6D pedicel pose estimation (6DPPE)-based visual servoing: (a) 3D scatterplot and (b) pose (positional and orientational) error.

- Depth outliers and speckle: The Intel D405 exhibits 0.8–1.0% range noise; speckle outliers can appear on specular fruit skin, distorting the covariance matrix ((13)). A statistical outlier removal ($k = 20$, $\sigma = 2.0$) before curvature calculation lowers the ICP residuals from 0.92 mm to 0.57 mm on a held-out dataset.
- Hand–eye calibration drift: Long-term experiments (> 3 h) revealed a slow thermal drift of the camera mount that can shift the effective eye-in-hand transform by up to 1.5 mm. Periodic self-calibration using a fixed AprilTag board seen at the start of each harvesting cycle constrains this drift below 0.4 mm.

These observations indicate that while the proposed 6DPPE pipeline is robust under nominal conditions, handling occlusion and dynamic motion remains critical for field deployment.

3.2.5. Future work

Building on the results obtained for individual fruit harvesting, future research should investigate dual-arm architectures to handle clustered and occluded fruits more effectively. Dual-arm cooperation can emulate human-like harvesting techniques, where one arm steadies the plant or clears obstructions while the other cuts. In addition, integrating advanced sensing modalities, such as high-resolution 3D imaging or machine learning algorithms for real-time fruit segmentation, can further refine detection and pose estimation. These enhancements would allow autonomous harvesting robots to adapt seamlessly to different fruit morphologies, varying occlusions, and more complex agricultural environments—eventually leading to improved productivity, reduced labor dependency, and enhanced sustainability in modern farming practices.

In summary, although the proposed system demonstrates strong potential for harvesting individual fruits, additional developments — particularly in dual-arm coordination and advanced occlusion handling — are required to tackle the more intricate scenarios involving clustered and highly occluded fruits. Continued development in hardware design and perception algorithms will help ensure that robotic harvesting technology keeps pace with agricultural environment

4. Conclusions

This study proposed an efficient 6DPE system for harvesting robots, employing the morphological features of fruits and vegetables (i.e. the abrupt curvature at the pedicel–fruit junction). The system demonstrated its capability of estimating the 6D pose of pedicels accurately across scenarios, enabling the precise alignment of the EE for harvesting tasks. The experimental results validated the effectiveness of the system for two fruit types, *C. melo* and tomatoes. For *C. melo*, the system achieved high precision (0.927), recall (0.809) and F1-score (0.864) values, highlighting its robustness in detecting and processing fruits with well-defined pedicels. For tomatoes, the system maintained balanced precision (0.837) and recall values (0.837), with an F1-score of 0.837, highlighting its adaptability to smaller fruits.

1. For *C. melo*, the system attained high precision (0.927), recall (0.809), and an F1-score of 0.864, underscoring its robustness in handling well-defined pedicels.
2. For tomatoes, the method maintained balanced precision (0.837) and recall (0.837), achieving an F1-score of 0.837. This indicates adaptability to smaller fruits with more variable shapes.
3. Positional errors averaged between −1.34 and 4.79 mm for *C. melo* and between −0.46 and 4.37 mm for tomatoes, whereas orientation errors stayed within ±5.10°. The RMSE analysis further validated the system reliability, remaining below 8.00 mm and 6.5° for positional and orientational errors, respectively.
4. Real-time performance (19–23 fps) reinforces the system's practicality in dynamic agricultural environments.

The proposed system exhibits high accuracy and efficiency in 6DPPE and robotic harvesting. The experimental findings emphasize its potential for real-world applications while recognizing areas for improvement. This research contributes to advancing autonomous harvesting systems, promoting precision, reliability and scalability in modern agriculture.

CRediT authorship contribution statement

Yonghyun Park: Software, Investigation, Visualization, Conceptualization, Writing – original draft, Writing – review & editing, Validation, Methodology, Data curation. **Jeonghyeon Pak:** Conceptualization, Writing – original draft, Writing – review & editing, Validation, Methodology, Software, Investigation. **Changjo Kim:** Writing – review & editing, Validation, Investigation, Software, Conceptualization, Writing – original draft, Methodology. **Hyoung Il Son:** Writing – review & editing, Supervision, Funding acquisition, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Technology Innovation Program (RS-2024-00511865) funded by the Ministry of Trade, Industry and Energy (MOTIE, Korea); part of the results are results of a study of the “Convergence and Open Sharing System” project, supported by the Ministry of Education and National Research Foundation of Korea.

Data availability

No data was used for the research described in the article.

References

[1] J. Kim, H.I. Son, A voronoi diagram-based workspace partition for weak cooperation of multi-robot system in orchard, IEEE Access 8 (2020) 20676–20686.

[2] C. Ju, H.I. Son, J. Seol, J. Kim, A review on multirobot systems in agriculture, Comput. Electron. Agric. 202 (2022) 107336.

[3] Y. Park, H.-J. Kim, H.I. Son, Novel attitude control of Korean cabbage harvester using backstepping control, Precis. Agric. 24 (2) (2023) 744–763.

[4] C. Ju, H.I. Son, Multiple UAV systems for agricultural applications: Control, implementation, and evaluation, Electronics 7 (9) (2018) 162, http://dx.doi.org/10.3390/electronics7090162.

[5] E.S. Mohamed, A. Belal, S.K. Abd-Elmabod, M.A. El-Shirbeny, A. Gad, M.B. Zahran, Smart farming for improving agricultural management, Egypt. J. Remote. Sens. Space Sci. 24 (3) (2021) 971–981.

[6] J. Jun, J. Kim, J. Seol, J. Kim, H.I. Son, Towards an efficient tomato harvesting robot: 3D perception, manipulation, and end-effector, IEEE Access 9 (2021) 17631–17640, http://dx.doi.org/10.1109/ACCESS.2021.3052240.

[7] Y. Li, S. Wu, L. He, J. Tong, R. Zhao, J. Jia, J. Chen, C. Wu, Development and field evaluation of a robotic harvesting system for plucking high-quality tea, Comput. Electron. Agric. 206 (2023) 107659.

[8] Y. Park, J. Seol, J. Pak, Y. Jo, C. Kim, H.I. Son, Human-centered approach for an efficient cucumber harvesting robot system: Harvest ordering, visual servoing, and end-effector, Comput. Electron. Agric. 212 (2023) 108116, http://dx.doi.org/10.1016/j.compag.2023.108116.

[9] L. Gong, W. Wang, T. Wang, C. Liu, Robotic harvesting of the occluded fruits with a precise shape and position reconstruction approach, J. Field Robot. 39 (1) (2022) 69–84.

[10] V. Rajendran, B. Debnath, S. Mghames, W. Mandil, S. Parsa, S. Parsons, A. Ghalamzan-E, Towards autonomous selective harvesting: A review of robot perception, robot design, motion planning and control, J. Field Robot. (2023).

[11] L.-E. Montoya-Cavero, R.D. de León Torres, A. Gómez-Espinosa, J.A.E. Cabello, Vision systems for harvesting robots: Produce detection and localization, Comput. Electron. Agric. 192 (2022) 106562.

[12] S.S. Mehta, W. MacKunis, T.F. Burks, Robust visual servo control in the presence of fruit motion for robotic citrus harvesting, Comput. Electron. Agric. 123 (2016) 362–375.

[13] Y. Zhao, L. Gong, Y. Huang, C. Liu, A review of key techniques of vision-based control for harvesting robot, Comput. Electron. Agric. 127 (2016) 311–323.

[14] E. Navas, R.R. Shamshiri, V. Dworak, C. Weltzien, R. Fernández, Soft gripper for small fruits harvesting and pick and place operations, Front. Robot. AI 10 (2024) 1330496.

[15] Y. Tang, M. Chen, C. Wang, L. Luo, J. Li, G. Lian, X. Zou, Recognition and localization methods for vision-based fruit picking robots: A review, Front. Plant Sci. 11 (2020) 510.

[16] S. Kim, S.-J. Hong, J. Ryu, E. Kim, C.-H. Lee, G. Kim, Application of amodal segmentation on cucumber segmentation and occlusion recovery, Comput. Electron. Agric. 210 (2023) 107847, http://dx.doi.org/10.1016/j.compag.2023.107847.

[17] O.M. Lawal, YOLOMuskmelon: Quest for fruit detection speed and accuracy using deep learning, IEEE Access 9 (2021) 15221–15227, http://dx.doi.org/10.1109/ACCESS.2021.3053167.

[18] H. Kang, C. Chen, Fast implementation of real-time fruit detection in apple orchards using deep learning, Comput. Electron. Agric. 168 (2020) 105108, http://dx.doi.org/10.1016/j.compag.2019.105108.

[19] Y. Park, C. Kim, H.I. Son, Fast and stable pedicel detection for robust visual servoing to harvest shaking fruits, Comput. Electron. Agric. 220 (2024) 108863, http://dx.doi.org/10.1016/j.compag.2024.108863.

[20] M. Jang, Y. Hwang, Tomato pose estimation using the association of tomato body and sepal, Comput. Electron. Agric. 221 (2024) 108961.

[21] J. Zhang, J. Xie, F. Zhang, J. Gao, C. Yang, C. Song, W. Rao, Y. Zhang, Greenhouse tomato detection and pose classification algorithm based on improved YOLOv5, Comput. Electron. Agric. 216 (2024) 108519.

[22] Z. Meng, X. Du, R. Sapkota, Z. Ma, H. Cheng, YOLOv10-pose and YOLOv9-pose: Real-time strawberry stalk pose detection models, Comput. Ind. 165 (2025) 104231.

[23] C.E. Rodriguez, C.A. Bustamante, C.O. Budde, G.L. Müller, M.F. Drincovich, M.V. Lara, Peach fruit development: a comparative proteomic study between endocarp and mesocarp at very early stages underpins the main differential biochemical processes between these tissues, Front. Plant Sci. 10 (2019) 715.

[24] J. Liu, Z. Li, P. Li, The physical and mechanical properties of tomato fruit and stem, in: Rapid Damage-Free Robotic Harvesting of Tomatoes, Springer, 2021, pp. 127–195.

[25] T. Yoshida, T. Fukao, T. Hasegawa, Fast detection of tomato peduncle using point cloud with a harvesting robot, J. Robot. Mechatronics 30 (2) (2018) 180–186.

[26] L. Comba, S. Zaman, A. Biglia, D.R. Aimonino, F. Dabbene, P. Gay, Semantic interpretation and complexity reduction of 3D point clouds of vineyards, Biosyst. Eng. 197 (2020) 216–230.

[27] C. Lehnert, I. Sa, C. McCool, B. Upcroft, T. Perez, Sweet pepper pose detection and grasping for automated crop harvesting, in: 2016 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2016, pp. 2428–2434.

[28] P. Eizentals, K. Oka, 3D pose estimation of green pepper fruit for automated harvesting, Comput. Electron. Agric. 128 (2016) 127–140.

[29] J. Rong, G. Dai, P. Wang, A peduncle detection method of tomato for autonomous harvesting, Complex & Intell. Syst. (2021) 1–15.

[30] F. Zhang, Z. Hou, J. Gao, J. Zhang, X. Deng, Detection method for the cucumber robotic grasping pose in clutter scenarios via instance segmentation, Int. J. Agric. Biol. Eng. 16 (6) (2024) 215–225.

[31] L. Luo, W. Yin, Z. Ning, J. Wang, H. Wei, W. Chen, Q. Lu, In-field pose estimation of grape clusters with combined point cloud segmentation and geometric analysis, Comput. Electron. Agric. 200 (2022) 107197.

[32] F. Wu, R. Zhu, F. Meng, J. Qiu, X. Yang, J. Li, X. Zou, An enhanced cycle generative adversarial network approach for nighttime pineapple detection of automated harvesting robots, Agronomy 14 (12) (2024) 3002.

[33] F. Meng, J. Li, Y. Zhang, S. Qi, Y. Tang, Transforming unmanned pineapple picking with spatio-temporal convolutional neural networks, Comput. Electron. Agric. 214 (2023) 108298.

[34] H. Wang, G. Zhang, H. Cao, K. Hu, Q. Wang, Y. Deng, J. Gao, Y. Tang, Geometry-aware 3D point cloud learning for precise cutting-point detection in unstructured field environments, J. Field Robot. (2025).

[35] J. Kim, H. Pyo, I. Jang, J. Kang, B. Ju, K. Ko, Tomato harvesting robotic system based on deep-ToMaToS: Deep learning network using transformation loss for 6D pose estimation of maturity classified tomatoes with side-stem, Comput. Electron. Agric. 201 (2022) 107300.

[36] Y. He, W. Sun, H. Huang, J. Liu, H. Fan, J. Sun, Pvn3d: A deep point-wise 3d keypoints voting network for 6DoF pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11632–11641.

[37] Y. He, H. Huang, H. Fan, Q. Chen, J. Sun, Ffb6d: A full flow bidirectional fusion network for 6D pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 3003–3013.

[38] Y. Labbé, J. Carpentier, M. Aubry, J. Sivic, Cosypose: Consistent multi-view multi-object 6D pose estimation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16, Springer, 2020, pp. 574–591.

[39] G. Wang, F. Manhardt, F. Tombari, X. Ji, Gdr-net: Geometry-guided direct regression network for monocular 6D object pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 16611–16621.