

Immature Green Apple Detection and Sizing in Commercial Orchards using YOLOv8 and Shape Fitting Techniques

Ranjan Sapkota¹, Dawood Ahmed¹, Martin Churuviya¹ and Manoj Karkee¹

¹ Center for Precision and Automated Agricultural Systems, Washington State University, 24106 N Bunn Rd, Prosser, 99350, Washington, USA

ABSTRACT Detecting and estimating size of apples during the early stages of growth is crucial for predicting yield, pest management, and making informed decisions related to crop-load management, harvest and post-harvest logistics, and marketing. Traditional fruit size measurement methods are laborious and time-consuming. This study employs the state-of-the-art YOLOv8 object detection and instance segmentation algorithm in conjunction with geometric shape fitting techniques on 3D point cloud data to accurately determine the size of immature green apples (or fruitlet) in a commercial orchard environment. The methodology utilized two RGB-D sensors: the Intel RealSense D435i and the Microsoft Azure Kinect DK. Notably, the YOLOv8 instance segmentation models exhibited proficiency in immature green apple detection, with the YOLOv8m-seg model clinching the highest AP@0.5 and AP@0.75 scores of 0.94 and 0.91, respectively. Leveraging the ellipsoid fitting technique on images from the Azure Kinect, we observed remarkable metrics, including an RMSE of 2.35, MAE of 1.66, MAPE of 6.15, and an R-squared value of 0.9. Challenges such as partial occlusion, where YOLOv8 sometimes misinterpreted immature green apple clusters, were recognized. In a comparison of 102 outdoor samples, the Microsoft Azure Kinect showed better performance than the Intel Realsense D435i, as supported by the MAE data. This study emphasizes the combined effectiveness of shape-fitting methods and 3D sensors in improving fruitlet sizing for precision agriculture.

Keywords: YOLOv8, deep learning, machine vision, automation, agriculture, fruit thinning

I. INTRODUCTION

Detecting and estimating size of apples during the early stages of growth is crucial for predicting yield and market, pest management, harvest and post-harvest logistics, and making informed decisions related to crop-load management [1]. Accurate size and number information during this stage enables farmers to strategize and prepare for harvest and post-harvest logistics including the workforce, equipment, and storage requirements [2]. Additionally, green fruit size serves as an indicator of tree health and vigor [3]. If the fruit does not attain its expected size during the growth stage, it could signify insufficient nutrients or pest infestation [4], [5]. By monitoring apple size and growth pattern, farmers can address these issues and improve crop production and quality to the desired level. Furthermore, accurate green fruit size estimates can help farmers predict future market reception of their crop. For instance, optimal size apples may go for higher prices than small or large apples. By knowing the size of green fruits, farmers can assess the tree's capacity to support the number of fruits and identify any potential

problems that may be affecting the tree's ability to produce the desired yield of high-quality fruit [6] [7][8][9].

In the commercial production of tree fruit crops such as apples, farmers generally use hand tools such as measuring tape and specifically designed fruit sizer to measure fruit size [7]. However, these techniques are labor-intensive and time-consuming, necessitating a considerable workforce in the orchard. Concurrently, commercial tree fruit growers have been grappling with labor shortages for the past two decades [8]. The COVID-19 pandemic has further exacerbated the labor shortage situation, posing a significant threat to global food security [9], [10]. Furthermore, hiring and training a seasonal or temporary workforce can be costly. Consequently, the tree fruit industry urgently requires alternatives to human labor [15]. The advancement in precision agriculture technologies, such as remote sensing, machine vision, and machine learning, opens the potential for development of non-destructive, rapid methods to detect fruitlet, estimate their size and predict crop yields [16].

Adoption of such a system can facilitate informed decisions on nutrient and water management, disease and pest control, and harvesting schedules, thereby reducing production costs and enhancing crop productivity and quality [17].

In the rapidly advancing realm of precision agriculture, the accuracy and efficacy of sensor systems stand as paramount determinants of successful crop management and yield prediction. As technology evolves, multiple state-of-the-art sensors emerge, each boasting unique capabilities and specifications. However, without a comprehensive comparison of these sensors, it remains challenging for researchers, agriculturalists, and technologists to discern their true potential and limitations in real-world applications. Such comparisons not only elucidate the strengths and weaknesses of each sensor but also provide invaluable insights into their adaptability, resilience, and performance under varied environmental conditions. A thorough assessment is, therefore, indispensable and following this, the study was guided by two primary objectives:

1. **Automated Green Fruit Detection and Size Estimation:** The primary objective was set to use YOLOv8, an advanced object detection algorithm, whereby green apples in a natural environment were identified. Following this identification, their sizes were estimated using various geometric shape fitting techniques, namely: Sphere Fitting (through Least Squares and RANSAC) and Ellipse Fitting.
2. **Sensor Comparative Analysis:** A comprehensive comparison between two widely-used and state-of-the-art machine vision sensors, the Intel RealSense 435i and the Microsoft Azure Kinect DK, was conducted as the secondary objective. Their performance in estimating green apple size under diverse orchard conditions was evaluated.

II. RELATED STUDY

In the past, researchers explored traditional image processing methods for fruit detection and size estimation. These conventional approaches typically involved image segmentation to separate fruits from the tree canopy, followed by the application of morphological operations. For instance, Behera et al. [11] employed color thresholding and the Randomized Hough Transform (RHT) technique to detect mangoes in canopy images. Following segmentation, an ellipse fitting technique was applied to delineate fruit estimate mango size in images (# of pixels). Similarly, Wang et al. [12] estimated mango size in trees using RGB-D images and a cascade detection method, combining 'histogram of gradients' features with Otsu's thresholding. However, this study was conducted under artificial nighttime lighting, limiting its practicality in variable natural lighting conditions.

Lin et al. [13] fused geometrical properties of a "kite" to estimate strawberry fruit size, as the kite shape resembles that of a strawberry. By segmenting calyx from strawberry fruit portions in 2D RGB images, the author identified the boundary pixels of the fruit. However, this approach is only applicable when all fruits are equidistant from the camera, and the study was conducted in a controlled environment for post-harvest grading and sorting purposes, which does not address the needs of crop load management during the growing season.

Gongal et al. [14] developed a 3D machine vision system for apple fruit size estimation in tree canopies by integrating a 2D color camera with a 3D time-of-flight (TOF) camera. They utilized histogram equalization in HSI color space to enhance color differences, followed by Otsu's thresholding and Circular Hough Transform for apple segmentation. While the study reported modest accuracy (69.1% using 3D coordinates and 84.8% using 2D pixel size), it focused on mature apples during harvest season, which cannot be used for automated crop-load management during the growing season.

Tsoulias et al. [15] recently estimated apple diameter using a LiDAR Laser scanner. They extracted radiometric and geometric features and applied the density-based scan algorithm (DBSCAN) to group segmented LiDAR points on the apple surface. The accuracy of this method was marginal ($R^2 = 0.78$) particularly because of challenges posed by occlusion due to branches and leaves.

Likewise, Apolo-Apolo et al. [16] developed an unmanned aerial vehicle (UAV) and deep learning-based approach to identify, count, and estimate citrus fruit sizes on tree canopies. Using a Faster R-CNN-based method, they detected citrus fruits in aerial images captured by a DJI phantom 3 drone. Although the study reported R^2 value of 0.80, the study was performed only over the parts of tree fruits that were visible in the top-view images.

Recently advance machine learning techniques have been extensively applied to fruit size estimation in agriculture [17]. For example, Omeed et al. [18] used a Convolutional Neural Network (CNN) model for on-tree kiwifruit detection and size estimation. Li et al. [19] employed a Random Forest algorithm to estimate apple size using 3D images captured with a structured light-based imaging system. Fu et al. [20] utilized a Faster R-CNN-based model to detect and segment apples in RGB images and applied a regression model for size estimation. Tobias et al. [21] recently introduced a viewpoint planning approach for identifying fruit position and size in plants, overcoming challenges posed by occlusions from leaves. By constructing an octree with labeled fruit regions of interest, the method evaluated viewpoint candidates using a utility function that considered expected information gain,

resulting in improved fruit detection and size estimation in both simulated and real-world scenarios. However, this study was limited to glasshouse environment.

III. MOTIVATION

The current state of research on fruit size estimation primarily focused on matured fruits near harvest that exhibit distinct colors compared to the background in tree canopy images. This color difference simplified the segmentation process, making it relatively easier to differentiate the fruit from the canopy. However, studies on accurately estimating fruit size during the early growth stages when fruit are small and have similar green color to other canopy parts is limited. Robotic crop load management applications require information about the size of fruits during their immature, green stages, as this data are crucial for various operations. For example, during fruitlet thinning, growers remove the smallest apples and retain the larger ones for future harvests. Furthermore, having accurate green fruit size information during the early growing season is essential for predicting crop yield and conducting automated fruit thinning in orchards. Addressing this research gap in accurately estimating fruit size during early growth stages can significantly improve the efficiency of crop load management practices, both manually and using robotic machines.

IV. MATERIALS AND METHODS

This study consisted of four major steps, as depicted in Figure 1, beginning with the acquisition of RGB-D data in orchard environment, followed by early stage green fruit detection, shape fitting, and finally fruit size estimation. The study was conducted in a commercial apple orchard with unstructured, natural environment by employing deep learning methodologies. Fruit size estimation was achieved through the reconstruction of their corresponding geometrical shapes in a 3D, utilizing sphere and ellipse fitting techniques. Additionally, the proposed approaches were rigorously validated using two state-of-the-art camera sensors, the Microsoft Azure Kinect (Microsoft Corporation, Redmond, USA) and the Intel RealSense 435i (Intel Corporation, California, USA), to ascertain the accuracy and robustness of the methodologies in diverse conditions.

A. Study site and data acquisition

This study on early fruit size estimation was conducted in a commercial orchard located in Prosser, Washington State, USA, where the predominant apple variety was Jazz (Scifresh). The orchard was characterized by trees reaching a height of 9 feet, with plant spacing set at 3 feet and a consistent row spacing of 10 feet.

In 2022, data collection was undertaken manually with machine vision sensors, gathering fruitlet images at varying

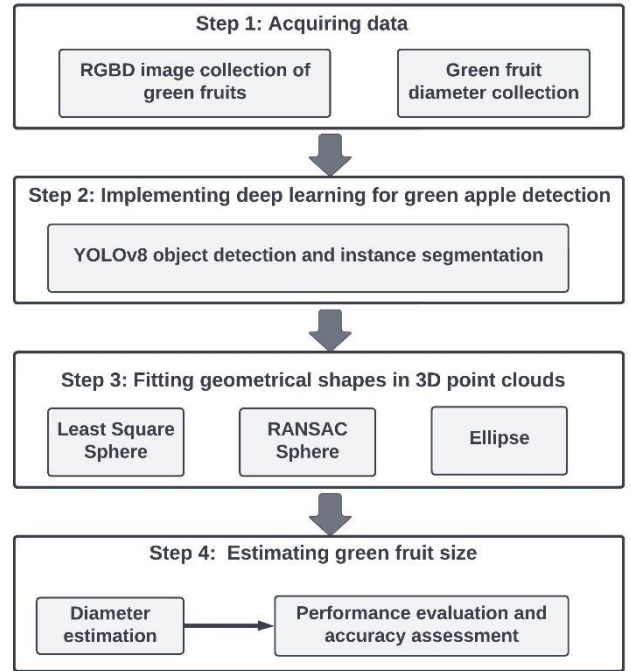


Figure 1: Block diagram illustration workflow for green apple detection and size estimation

distances ranging from 1 to 4 feet from the tree canopy. From this exercise, a dataset comprising 1079 images of fruitlets was gathered. The distribution of these images was almost equal between the two cameras, with 534 images captured through the Intel RealSense and 545 via the Microsoft Azure. These collected images were subsequently utilized to train a deep learning model tailored for immature green apple detection.

By June 2023, in anticipation of future real-time agricultural automation where sensors would be integral to robotic platforms, a strategic methodological shift was initiated. To simulate this future setup and ensure consistent environmental conditions for data acquisition, the two sensors were mounted on top of each other on a UR5e robotic arm (Universal Robots, Boston, USA) as shown in Figure 2. This arm was seamlessly integrated into a Warthog unmanned ground vehicle robot (Clearpath Robotics, Inc., Ontario, Canada) that had been retrofitted with an UR5e industrial manipulator as illustrated in Figure 2 a and 2b.. This design ensured that both cameras concurrently captured the tree canopy under identical conditions. Aligning them with precision, they were set perpendicular to the canopy and were stationed approximately 1 meter away from the target tree trunks. Precision in maintaining this specified distance was ensured using a measuring tape.

B. Machine-Vision Sensors Used

1) INTEL REALSENSE D435I

The Intel RealSense D435i is a depth-sensing camera featuring active infrared (IR) stereo vision system and an inertial measurement unit (IMU). The camera includes a 2-megapixel RGB sensor and a depth sensor with a resolution of 1280 x 720 pixels, and depth range up to 10 meters. The depth sensor operates using structured light technology, utilizing a pattern projector to create disparities between the stereo images, which are captured by two IR cameras. With a speed up to 90 frames per second (fps), and a 69.4° horizontal field-of-view (HFOV) and 42.5° vertical field-of-view (VFOV), the D435i offers flexibility for various application scenarios. The camera also includes a 6-axis IMU that provides accurate orientation data, enabling improved depth data alignment and scene understanding. The Intel RealSense D435i is compact, lightweight, and offers robust performance for depth and RGB data acquisition in a wide range of environments.

2) MICROSOFT AZURE

The Azure Kinect DK sensor is equipped with a 12-megapixel color camera and a 1-megapixel depth sensor. The depth sensor operates on the Time of Flight of Light (ToF) principle and features a global shutter with analog binning, resulting in pixel-synchronized capture and reduced noise [25]. With a

modulation frequency ranging from 200 to 320 MHz, the sensor offers various modes for resolution, range, and frame rate. The depth sensor has two operational depth modes: Narrow Field-of-View (NFOV) and Wide Field-of-View (WFOV). The sensor functions within a temperature range of 0°C to 27°C for the Azure Percept Vision SoM assembly with housing, and -10°C to 70°C for the Vision SoM chip. The data presented in this paper were collected using the NFOV depth mode.

C. Robot Manipulation for image collection

The UR5e robotic arm underwent a precise programming sequence, initiating its movement vertically upwards from the tree's base as depicted in Figure 2b while orienting the sensors to face the tree. This ensured comprehensive coverage, capturing images from the trunk base all the way to the fifth branch layer. Throughout this maneuver, the arm consistently progressed at a speed of 0.1 m/s ensuring that both sensors remained as close to perpendicular to the canopy as possible.

D. Training the Deep Learning Model YOLOv8 for Green Fruitlet Detection

This research utilized the YOLOv8 object detection algorithm, a recent addition from Ultralytics (Ultralytics, Maryland, USA) as of January 2023. Building upon the legacy of YOLOv5, YOLOv8 introduces enhancements that amplify its



Figure 2: Image acquisition of fruitlets in a commercial jazz apple orchard in Washington State, USA. (a) Illustrating two machine vision sensors in-action, (b) Robot and sensors interfacing canopy (fruitlets), and (c) Integrated system for image collection

accuracy and efficiency in object detection tasks [22]. The YOLOv8 model represents a significant advancement in object detection, being anchor-free in its approach. It integrates pioneering techniques such as Pseudo Ensemble or Pseudo Supervision (PS), the adept Darknet-53 architecture, and an enhanced Feature Pyramid Network, termed YOLOv8PAFPN[23]. The principle behind PS is the concurrent training of various models, each with unique configurations, on an identical dataset. The Darknet-53 architecture, a deep 53-layer convolutional neural network, augments feature extraction, thereby enhancing object detection capabilities [24]. Anchor-free detection method emphasizes direct estimation of an object's center instead of its relative distance from a predetermined anchor box. This approach reduces the box predictions, streamlining the Non-Maximum Suppression (NMS) process, a computationally intensive post-processing step employed to refine potential detections post-inference [25]–[27].

The YOLOv8 model offers a suite of five distinct configurations for tasks such as identification, segmentation, and classification: YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. All these configurations were tested in this study for segmentation of immature green apples from the collected imagery. The images were annotated using Labelbox (Labelbox, California, USA). This meticulous process yielded 5,921 labels of immature green apples after roughly 44 hours of manual annotation.

1. Hardware and Software

The YOLOv8 model was trained on a workstation with an Intel Xeon® W-2155 CPU @ 3.30 GHz x20 processor, NVIDIA TITAN Xp Collector's edition/PCIe/SSE2 graphics card, 31.1 gigabyte memory, and Ubuntu 16.04 LTS 64-bit operating system. The training process involved setting several parameters including the learning rate, batch size, and the number of iterations to optimize the performance of the model. In this study, the learning rate was set to 0.001, while the batch size was set to 32. To prevent overfitting, a dropout rate of 0.5 was used. The model was trained for 1000 iterations while monitoring the loss function to assess the progress on model training. The images were resized to 640x640 pixels to make them compatible for YOLOv8 format.

2. Model Training and Testing

The YOLOv8 model was trained with early stopping if the validation loss did not improve for 200 iterations. A batch size of 16 and an image size of 640 x 640 pixels were utilized for training. Training was executed on devices labeled as '0' and '1', which are likely references to specific GPU devices. To expedite data loading and preprocessing, eight worker threads were employed. These worker threads parallelize tasks such as reading and transforming data, ensuring a consistent and efficient supply of data to the training process and preventing

unnecessary idle times in the GPU or CPU. Using multiple worker threads is beneficial in reducing data loading times, especially when dealing with large datasets.

The Stochastic Gradient Descent (SGD) optimization algorithm was employed with a learning rate of 0.01. The momentum and weight decay parameters were set to 0.937 and 0.0005, respectively, based on empirical evidence and prior research that suggested these settings offer a balance between fast convergence and model stability, reducing the chances of overfitting. Likewise, a warm-up phase was implemented during the first three epochs, with a warm-up momentum of 0.8 and a warm-up bias learning rate of 0.1. More data augmentation and regularization parameters used during the training process is presented in Table 1.

Table 1: Data augmentation and regularization parameters used in YOLOv8 training

Methods Applied	Value
Hue augmentation (fraction)	0.015
Saturation augmentation (fraction)	0.7
Value augmentation (fraction)	0.4
Rotation	0.0
Translation	0.1
Scale	0.5
Flip left-right (probability)	0.5
Mosaic (probability)	1.0
Weight decay	0.0005

E. Performance evaluation of YOLOv8 for Immature green fruit segmentation

The detection and segmentation performance of the YOLOv8 algorithm for green apple detection was evaluated using Mean Intersection over Union (MIoU), average precision (AP), and mean average precision (mAP), mean average recall (mAR), and F1-score. MIoU, also known as the Jaccard index, assesses the accuracy of the segmented mask with respect to the target object, calculated as follows:

$$MIoU = \frac{Area\ Overlap}{Area\ Union} = \frac{TP}{FP + TP + FN}$$

Equation 1

Where,

- TP is true True Positives, which counts correctly identified apples
- FP is False Positives, indicating non-apples mistakenly identified as apples.
- FN is False Negatives, denoting apples that were missed.

Precision evaluates the accuracy of the predicted positive detections, calculated as

$$Precision = \frac{TP}{TP + FP}$$

Equation 2

Recall, on the other hand, indicates how many of the actual positives our model can identify, and it is computed as:

$$Recall = \frac{TP}{TP + FN}$$

Equation 3

The F1-score, which considers both precision and recall, is calculated as:

$$F1 - Score = \frac{2(Precision * Recall)}{Precision + Recall}$$

Equation 4

AP provides a measure of the model's performance across different threshold levels, quantifying the area enclosed by the recall rate, the precision rate, and the horizontal axis. Meanwhile, mAP is a single consolidated metric to represent the model's overall detection performance. It averages the AP for all classes, providing a holistic view of the model's

capability in both target detection and instance segmentation tasks.

F. Size Estimation using shape fitting in 3D point clouds for segmented immature green apples

Upon successful segmentation of immature green apples with the YOLOv8 model, the next vital step was the size estimation, which would be essential for various applications including growth monitoring, harvesting predictions and possible automation. Point clouds, which are sets of data points in space, representing the segmented green apples, were meticulously visualized and extracted using Open3D library (Open3D Engine, California, USA). To estimate the size of immature green apples, shape fitting technique in 3D space was employed. The rationale behind shape fitting is to create an accurate 3D representation of the apples. By fitting shapes to the segmented apples in the point clouds, the physical dimensions of the apples, such as their diameter or major and minor axes, can be deduced. Earlier, there are several studies that have explored sphere fitting approach to estimate fruit size of peaches [28], guava[29], apples[30], pomegranate[31], tomato[32] and citrus [33]. Additionally, have explored ellipsoid fitting technique to estimate the size of non-spherical fruits such as banana [34] , watermelon [35] and mushroom [36]. Based on these studies, this study also explored the utilization of sphere and ellipsoid to reconstruct the immature green apples in 3D space using the image collected by Intel RealSense and Microsoft Azure cameras.

The ground truth measurement of immature green apples was collected using a digital caliper as shown in Figure 5b. The following shape fitting approaches were explored and compared against the ground truth:

- Sphere Fitting Using Least Squares Method: This approach was particularly useful for immature apples that maintained a somewhat spherical shape.

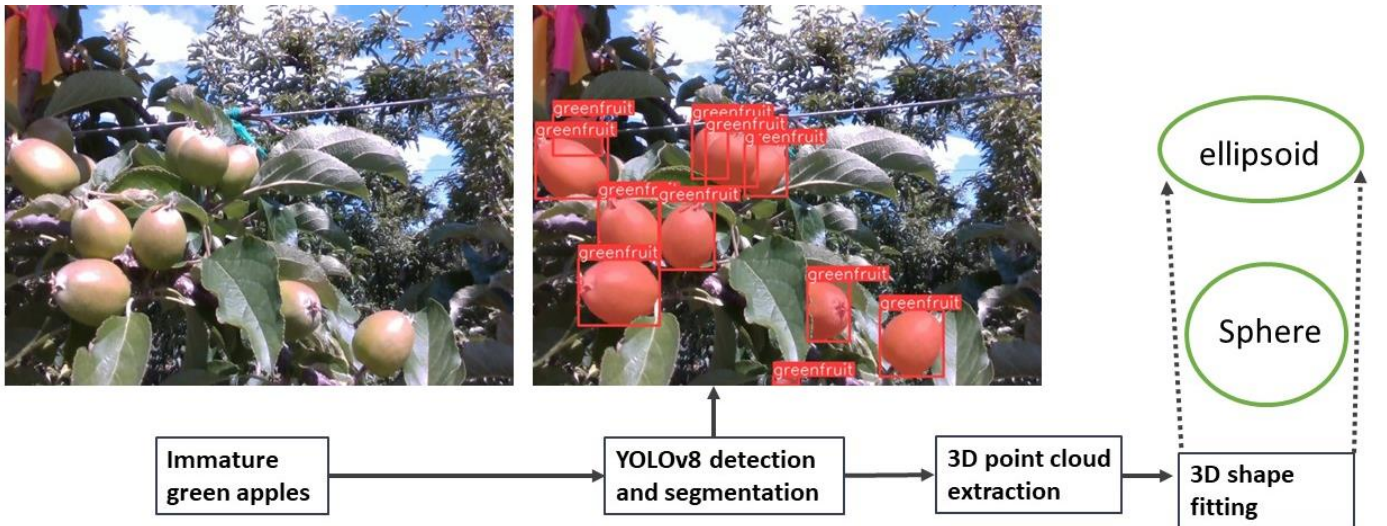


Figure 3: YOLOv8 segmentation and size estimation workflow chart. The model was trained to segment the immature green apples and their segmented mask was used to further image processing. 3d point clouds from the segmented mask was utilized to fit two kind of 3D geometry : sphere and ellipsoid

- b. **Sphere Fitting Using Random Sample Consensus (RANSAC):** While similar in aim to the previous method, this approach is robust against outliers, ensuring more accurate fittings even with occasional erroneous data points.
- c. **Ellipse Fitting:** Given that not all immature green apples are perfectly spherical, the ellipse fitting method catered to those apples with more elongated shapes, allowing for a better fit and thus more accurate size estimations.

1) SPHERE FITTING USING LEAST SQUARES

The Least Squares Fitting method focuses on the sum of squared distances from the data points to the sphere to determine the most fitting sphere that could represent the point clouds. From a set of 3D coordinates, represented as $\{p_1, p_2, \dots, p_n\}$ where each point $p_i = (x_i, y_i, z_i)$, the center of the sphere, denoted as $C = (x_0, y_0, z_0)$, and its radius, R , were identified to best encapsulate the immature apple. This relationship is encapsulated in equation:

$$(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2 = R^2$$

Equation 5

In this equation, $X = (x_0, y_0, z_0, R)$ details the specifics of the determined sphere: its center and radius. To ensure this sphere closely represented the immature green apple data as shown in Figure 4a, the sum of squared differences between each data point and the sphere's surface was minimized, as shown in equation 6:

$$S = \sum(f(x)^2)$$

Equation 6

In this context, $f(x)$ denotes the distance between a specific data point and the modeled sphere. Through the optimization of this function, the representation of the sphere ensured to align closely with the real distribution of 3D point clouds extracted from segmented immature apple.

2) SPHERE FITTING USING RANSAC

The **RANSAC Sphere Fitting algorithm**, inherently adept at handling outliers, was implemented with a process that begins with the RANSAC algorithm choosing three random points from the dataset to hypothesize a potential sphere. These points serve to delineate the sphere's center and radius. Subsequently, all dataset points are cross-referenced with this hypothetical sphere. Points lying within a predefined proximity to the sphere's surface are termed 'inliers', reflecting their appropriateness to the shape, while the rest are labeled 'outliers', indicating potential errors or deviations. This process iterates, with each cycle generating a new hypothetical sphere based on a fresh trio of random points. The iteration yielding

the highest inlier count identifies the most fitting sphere, whose dimensions, in turn, indicate the apple's size. Such shape fitting methodologies, like RANSAC, are paramount because they offer a rigorous, mathematically sound approach to estimate size, especially in datasets susceptible to errors or noise. Given that the immature apples can vary in shape, exploring different shape fitting options ensures comprehensive and accurate size estimations for all samples. The mathematical underpinning for the RANSAC sphere fitting remains consistent with the distance equation employed in the least squares method, as denoted in equation 5 and 6. Figure 4b represents the RANSAC sphere fitting over the segmented 3D point clouds of immature apples.

3) ELLIPSE FITTING

For the optimal size estimation of immature green apples, especially those possessing an elliptical shape, an ellipsoid fitting approach was employed to the point cloud data representation of each immature apple, aiming to ascertain both the major and minor axis lengths for a detailed size profile. The ellipsoid fitting as depicted in Figure 4c, at its core, tries to determine the most fitting ellipsoid's center, orientation, and semi-axes lengths for a collection of data points in a three-dimensional realm. Mathematically, an ellipsoid is characterized by a set of points such that the sum of the squares of the distances from these points to two distinct foci remains constant. This relationship is mathematically captured by equation 3.

$$(x - a)^2/a^2 + (x - b)^2/b^2 = 1$$

Equation 7

Here, (a,b) represents the ellipsoid's center, while 'a' and 'b' indicate the semi-major and semi-minor axes, respectively.

In our endeavor, the Löwner-John ellipse fitting algorithm was employed. Initially, a foundational ellipse is derived using the least-squares method, serving as a launching pad for the subsequent steps governed by the Löwner differential equation. This equation refines the ellipse iteratively, adapting it until the best-fitting ellipsoid for the data points emerges. This refined ellipsoid is the key to deducing the apple's size. As the algorithm progresses, it fine-tunes the weights of the points and the encompassing ellipsoid's parameters. Continuation ensues until the variation between consecutive weight sets is marginal, falling below a predetermined threshold. The culmination of this process yields a matrix 'A' and vector 'c', which encapsulate the ellipsoid's defining equation. This matrix 'A' dictates the ellipsoid's shape, while the vector 'c' pinpoints its spatial positioning. To extract the ellipsoid's semi-axes lengths, Singular Value Decomposition (SVD) is performed on matrix 'A'.

This breakdown reveals the longest (major) and shortest (minor) semi-axis lengths, crucial for the apple's size estimation. This meticulous process, underpinned by robust mathematical logic, is a testament to the comprehensive nature of our size estimation approach. It also highlights the importance of discerning between different shape-fitting techniques to cater to the diverse apple shapes encountered. The entire procedure and its associated code can be accessed via github at <https://github.com/rmsandu/Ellipsoid-Fit>.

$$\frac{\partial E}{\partial t} = \left(\frac{1}{2}\right) \left(1 - \frac{(x-a)^2}{a^2} - \frac{(y-b)^2}{b^2}\right) E$$

Equation 8

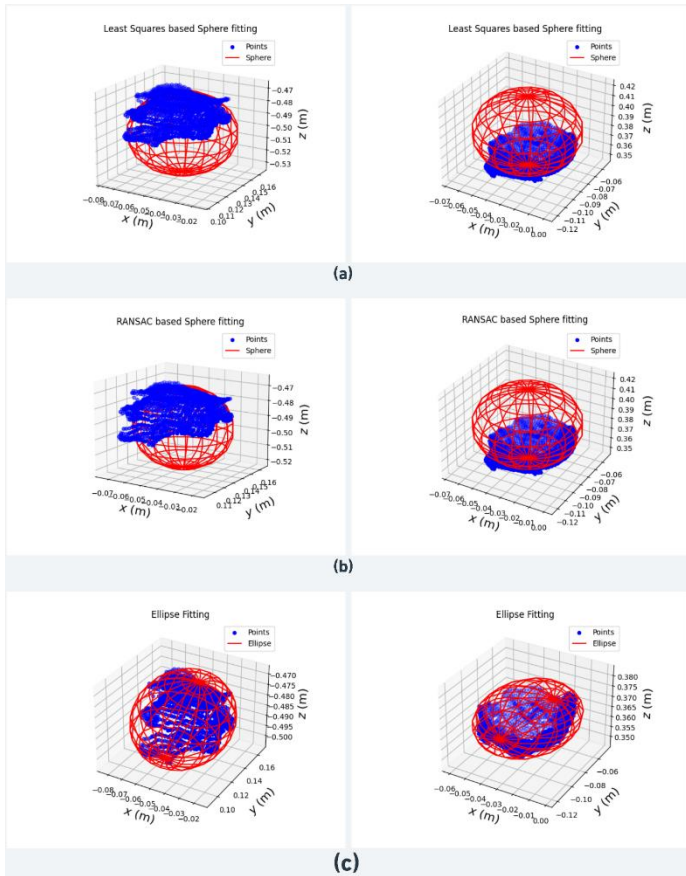


Figure 4: 3D geometry fitting in the point clouds extracted from the segmented green apple from YOLOv8 algorithm (a) Sphere Fitting using Least Squares (b) Sphere Fitting using RANSAC (c) Ellipse fitting. At each image in this figure, the geometry fitting illustration shown in left side is for RealSense 435i camera and on the right is Microsoft Azure camera.

G. Evaluation of estimated fruit size

To evaluate the performance of the shape-fitting techniques for size estimation using the two cameras, two distinct experimental setups across different years were employed. In 2022, a sample of 31 green apples with varying diameters (24, 27, 30, and 70 mm) was utilized in controlled indoor conditions, as illustrated in Figure 5a. The diverse range of synthetic green apple sizes was specifically selected to account for the variability in the field including different growth stages of green apples. Building on this, in 2023, the shape fitting techniques were further challenged by the inclusion of real-world scenarios. A comprehensive set of 102 samples was collected from a commercial jazz apple orchard environment, as had been previously detailed, with the assistance of the imaging setup controlled by the Warthog and UR5e arm.

To assess the performance in fruit sizing, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R2) were calculated.

Root Mean Squared Error (RMSE): RMSE is given by Equation 9, where n denotes the total number of observations, predicted_i corresponds to the predicted value for the i th observation, and actual_i represents the actual value for the i th observation.

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum (\text{predicted}_i - \text{actual}_i)^2}$$

Equation 9

Mean Absolute Error (MAE): MAE calculates the average absolute difference between the estimated fruit sizes and the actual fruit sizes as given by Equation 10. It is less sensitive to outliers than RMSE, as it does not square the differences.

$$MAE = \left(\frac{1}{n}\right) * \sum |y_i - \hat{y}_i|$$

Equation 10

where:

n is the number of samples

y_i is the actual size of immature apple

\hat{y}_i is the estimated size of immature apple

In your case, MAE can be used to quantify the average difference in size estimation for green apples. A lower MAE indicates a better model performance.



4 different sizes of synthetic green apples in a controlled environment (Indoor test)



Immature green apples in a commercial orchard environment

Figure 5: (a) Indoor environment to estimate the size of green apples of different sizes, (b) Commercial orchard environment, measuring the size of immature green apples using a digital caliper

Mean Absolute Percentage Error (MAPE): MAPE presents the average relative error between the estimated and the actual fruit sizes, expressed as a percentage. This allows for easier comparison across different datasets with varying scales.

$$MAPE = \left(\frac{100}{n} \right) * \sum |y_i - \hat{y}_i| / y_i$$

Equation 11

where:

n is the number of samples

y_i is the actual size of immature apple

\hat{y}_i is the estimated size of immature apple

In your case, MAPE can be used to assess the relative performance of the size estimation workflow for green fruit. A lower MAPE indicates a better model performance.

R-squared (R^2): R-squared is a statistical measure that represents the proportion of the variance in the dependent variable (fruit size) that can be explained by the independent

variables (features used in the estimation model), which is given by.

$$R - Squared = 1 - \left(\frac{\left(\sum (y_i - \hat{y}_i)^2 \right)}{\left(\sum (y_i - y_{mean})^2 \right)} \right)$$

Equation 12

where:

y_i is the actual fruit size

\hat{y}_i is the estimated fruit size

y_{mean} is the mean of the actual fruit sizes

V. RESULTS AND DISCUSSION

A. Immature green apple segmentation using YOLOv8

A total of 108 images containing immature green apples were utilized to evaluate the performance of the YOLOv8 model in segmenting the green apples. These images, captured in a commercial orchard, featured immature apples of varying sizes and orientations, providing a diverse dataset for the assessment. Figure 6 provides a visual representation of the immature green apple detection results using the YOLOv8 model. Among the various configurations of the YOLOv8 algorithm tested in this study, The YOLOv8m-seg model achieved the highest precision at 0.9. The YOLOv8x-seg model marked the highest recall of 0.9. In terms of the F1 score, the YOLOv8m-seg configuration stood out with a score of 0.89. Moreover, when evaluated against the AP@0.5 metric, the YOLOv8m-seg model registered a score of 0.94. A comprehensive breakdown of the performance metrics for all tested YOLOv8 configurations is presented in Table 2.

The Precision-Recall curve, Recall-Threshold curve, F1-Score Confidence curve, and the area under the Precision-Recall curve (PR-AUC) for the YOLOv8 model are presented in Figure 7 parts a, b, c, and d, respectively. From the Precision-Recall curve, a peak precision of 0.91 was observed for green fruit detection across all classes. A maximum recall of 0.98 was revealed by the Recall-Threshold curve. A value of 0.94 was recorded for the PR-AUC, suggesting that a robust precision-recall performance across varied thresholds was exhibited. Similarly, a value of 0.94 was recorded for the mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 for all was shown that F1 score of 0.89 was achieved for immature green apple segmentation.

Although the YOLOv8 object detection and segmentation algorithm demonstrated promising capabilities with high accuracy in detecting and segmenting green apples in complex orchard environments with similar green color of background objects such as leaves, certain challenging situations still resulted in detection failures. These challenges included occlusions caused by leaves and low light or shadow conditions. Figure 6a and 6b presented examples of failed detections (or false negatives) due to leaf occlusions on green apples, while Figure 6c illustrated the failure in detecting and segmenting green apples under low

improvements could be made by training the algorithm on a larger dataset encompassing more diverse lighting conditions and varied fruit orientation to enhance its performance and robustness in addressing these challenges.

In recent years, there has been a surge in research exploring the use of YOLO-based algorithms for agricultural applications. Noteworthy among them is the study by [Sun et al.], which focused on recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm [37]. Another important contribution was the study that delved into green fruit segmentation and orientation estimation for robotic green fruit thinning of apples [38]. There has also been an increasing trend towards channel-pruned and optimized versions of YOLO for fruit detection, as highlighted in studies like "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet de tectioin before fruit thinning"[39], and "Green citrus detection and counting in orchards based on YOLOv5-CS and AI edge system"[40]. Furthermore, the potential of the YOLOv5 model was showcased in research topics such as the detection of litchi fruits for yield estimation [41], spikelet detection in grapes [42] and green pepper detection using NSGA-II-based pruned YOLOv5l in the field environment [43]

While these studies made significant strides in their respective domains, they predominantly utilized standard RGB images or those captured using non-specialized machine vision sensors. Furthermore, most of the existing studies were performed under indoor/greenhouse agriculture. In contrast, the research presented here was based on data collected from a commercial orchard using two of the most renowned sensors widely recognized in agricultural automation studies. The findings obtained from the YOLOv8 model, as employed in this research, surpassed those reported in the aforementioned studies, particularly in terms of processing speed and accuracy. Despite the advancements and capabilities of the YOLOv8 model in the context of immature apple detection, it remains susceptible to challenges arising from occlusion.

Table 2: Report on Precision, Recall, F1-score, AP@0.5 and AP@0.75 of all five YOLOv8 configurations

Model	Precision	Recall	F1 Score	AP@0.5	AP@0.75
YOLOv8n-seg	0.88	0.88	0.88	0.93	0.89
YOLOv8s-seg	0.89	0.85	0.87	0.93	0.89
YOLOv8m-seg	0.9	0.88	0.89	0.94	0.91
YOLOv8l-seg	0.87	0.87	0.87	0.93	0.89
YOLOv8x-seg	0.86	0.9	0.88	0.93	0.90

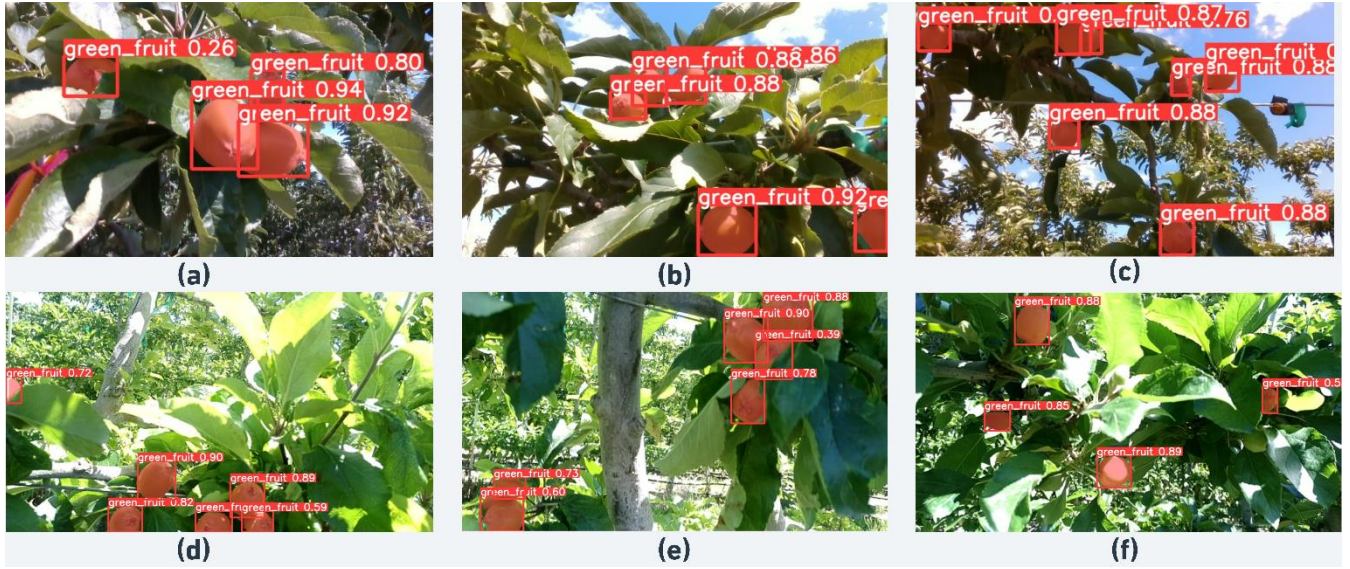


Figure 6: YOLOv8 detection and segmentation results illustration (a, b and c) segmentation examples of green apples in the images captured by IntelRealsense D435i in variable lighting condition and occlusions, (d, e and f) showing segmentation results illustration in complex orchard environment in the images captured using Microsoft Azure Camera

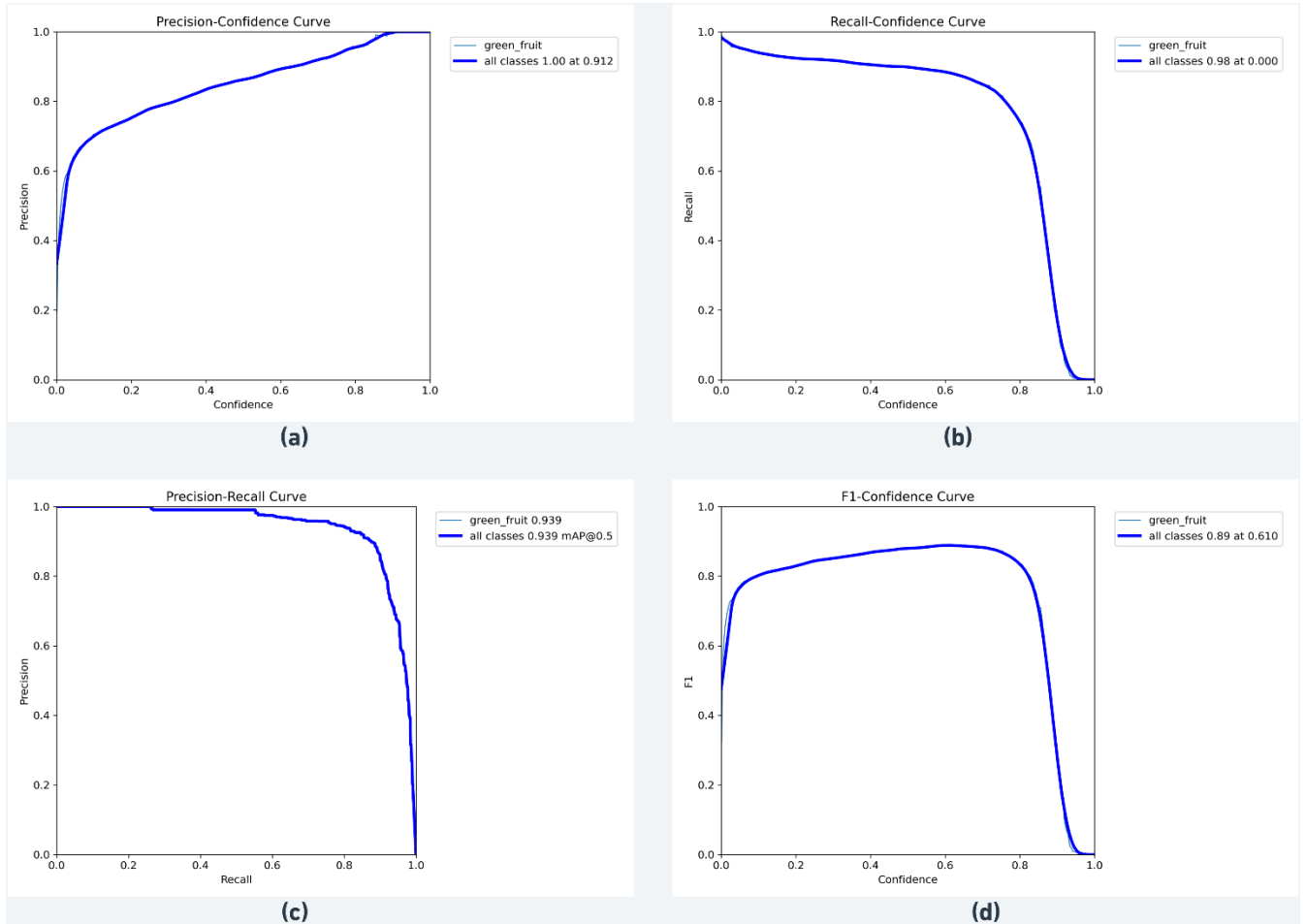


Figure 7: Immature green apples segmentation results achieved with YOLOv8m-seg; (a) Precision-confidence curve; (b) Recall-confidence curve; (c) Precision-Recall Curve; and (d) F1-Confidence curve

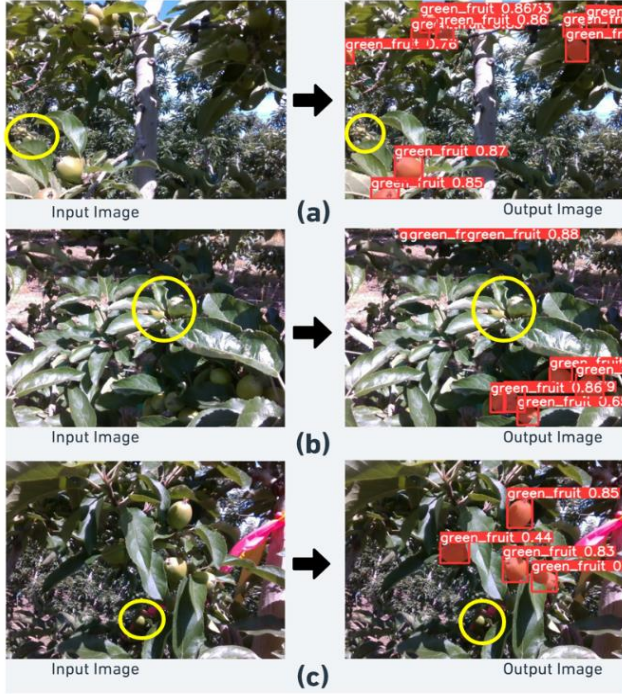


Figure 8: Occlusion examples. Yellow region indicates the affected region; (a) showing occlusion due to leaves; (b) Occlusion due to leaves and stem; and (c) Foliage and shadow effect resulting in failure in detection.

As illustrated in Figure 8a, an immature apple partially obscured by overlapping leaves and branches goes undetected by the model, marking a false negative. Similarly, Figure 8b and 8c depicts an interference from neighboring fruits and foliage, which also results in a false negative. Such occlusions can not only introduce inaccuracies in the segmentation mask but also subsequently affect the precise estimation of the green apple sizes.

Training a higher number of data samples with more occlusion examples could possibly increase the detection and segmentation of such immature apples.

B. Results for immature green apple size estimation

Utilizing the ellipsoid fitting technique, the measurement distribution for sizes over images captured by the Microsoft Azure Kinect camera yielded optimal results in terms of RMSE, MAE, MAPE, and R-squared values. Similarly, the ellipsoid fitting method registered commendable results with an RMSE of 2.35, MAE of 1.66, MAPE of 6.15, and an R-squared value of 0.9 for images taken with the Microsoft Azure Kinect camera. Figure 9a and 9b provide visualization of the diameter estimates for 102 immature green apple samples, as deduced from the three employed techniques on images from Microsoft and IntelRealsense camera, respectively. The comparative metrics—RMSE, MAE, MAPE, and R-squared—obtained from the three techniques on both sets of camera images are also encapsulated in Figure 10. The indoor validation study (Figure 5a) conducted on 2022 over synthetic green apples with varying sizes, including diameters of 24mm, 27mm, 30mm, and 70mm also showed that the combination of Microsoft Azure with ellipsoid fitting approach achieved maximum accuracy. The RMSE values for the size estimation using Least Square-based sphere fit, RANSAC-based sphere fit, and Ellipse fit for the images captured with the Intel RealSense camera were 32.11, 50.22, and 15.36, respectively. In contrast, when utilizing the Microsoft Azure Kinect camera, the corresponding RMSE values were found to be 6.21, 8.74, and 5.74, respectively. Overall, Microsoft Azure camera achieved better performance in terms of predicting size using the all methods applied in this study. Figure 11 shows the whisker box plot comparing RMSE, MAPE, and MAE for all three shape fitting technique that were investigated.

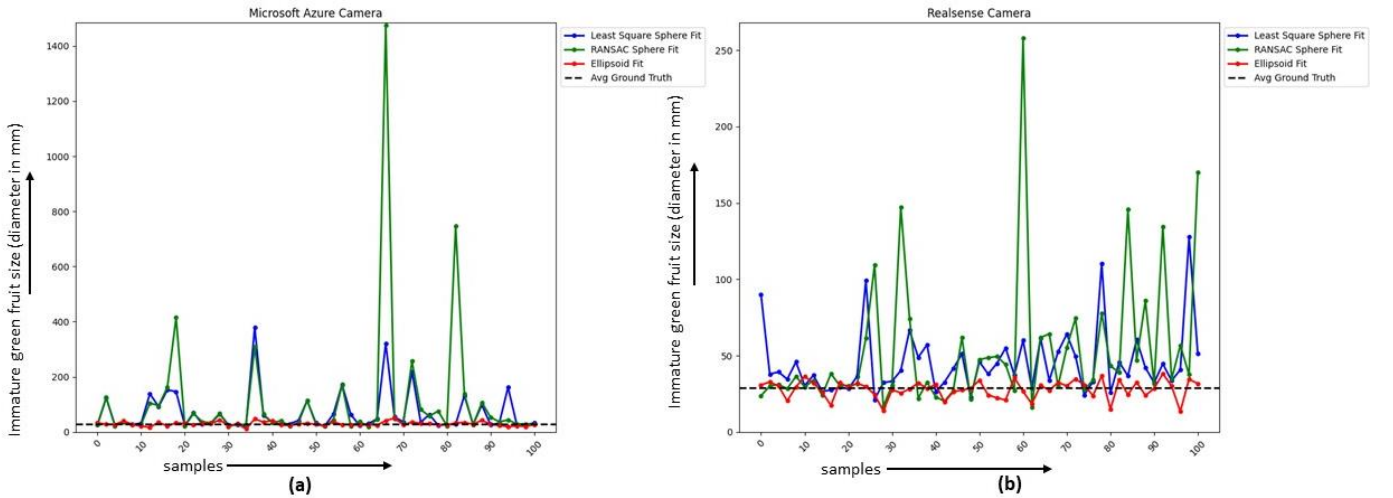


Figure 9: Line diagram showing the trend of results of size estimation using Least Square sphere Fitting, RANSAC sphere fitting and Ellipsoid fitting on the image collected using (a) Microsoft Azure Kinect DK camera and (b) IntelRealsense d435i camera

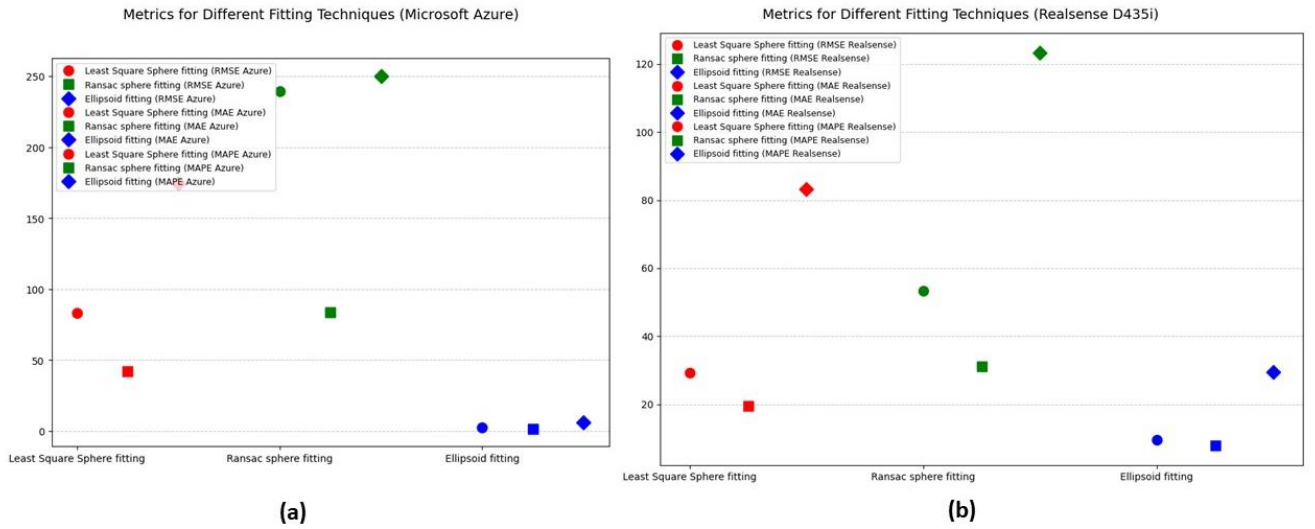


Figure 10: RMSE, MAE, MAPE, and R-squared distribution achieved using the 3 shape fitting techniques on the images collected using; (a) Microsoft Azure camera and (b) Intel RealSense D435i camera

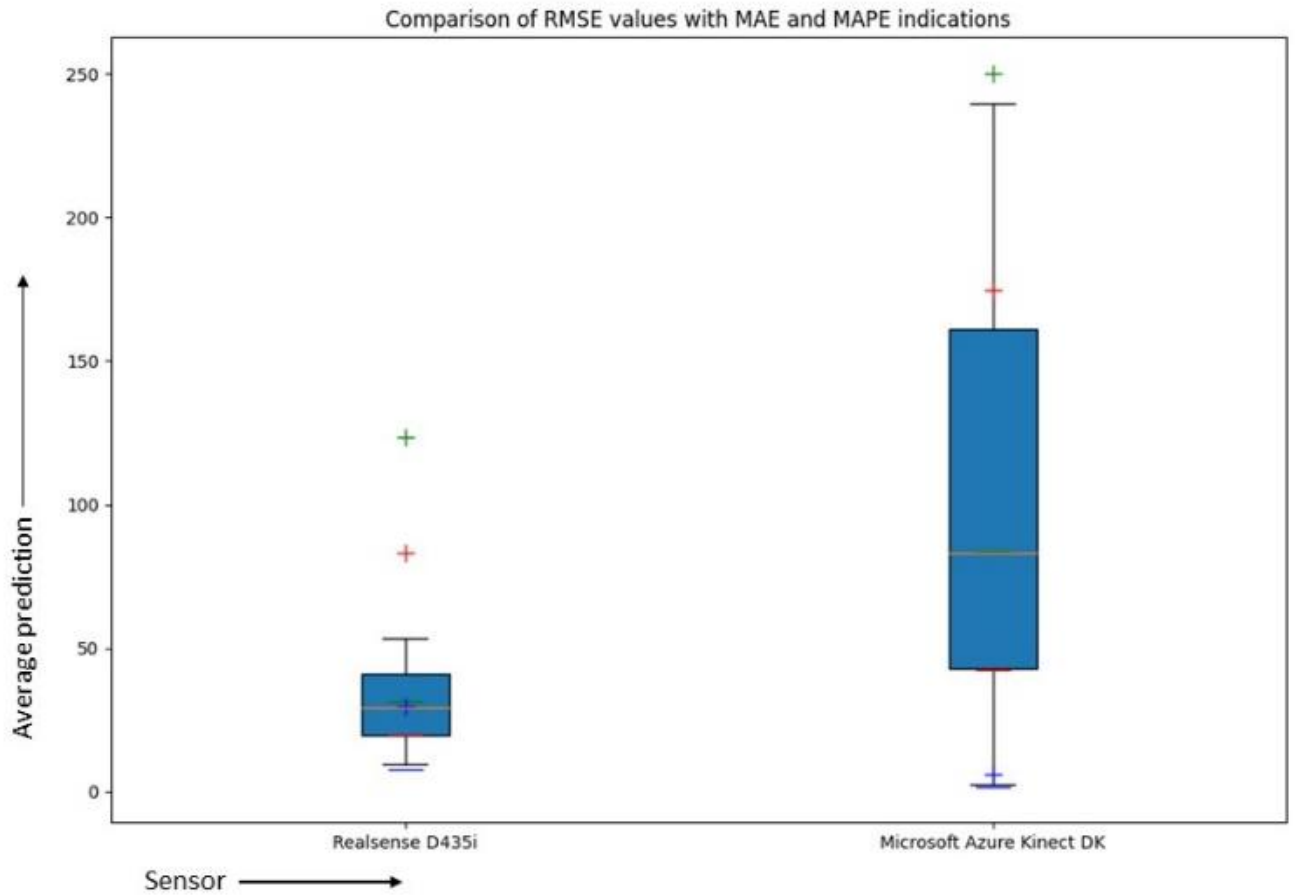


Figure 11: Boxplot comparing RMSE values of Realsense and Azure cameras for green apple dimension estimation. Inside each box, underscored markers represent MAE and plus markers signify MAPE across three techniques, highlighting the precision differences between the cameras.

Figure 12 showcases a scenario where the YOLOv8 algorithm resulted in overmasking and undermasking in the segmentation of immature apples. Figure 12a illustrates a condition where two immature apples are segmented as a single immature apple. This resulted in a bigger mask segmentation than the actual mask in 3D image processing, ultimately affecting the size estimation of the immature apple. Partial occlusion is a common issue in computer vision, where objects in the image overlap with each other, causing challenges for the deep learning models to accurately segment them out. Therefore, it is crucial to consider such limitations while developing and evaluating deep learning-based algorithms for object detection and segmentation.

Likewise, Figure 12b demonstrates a condition where the immature apple is occluded by the leaves and foliage. Since only half (approximate) part of the immature apple has been segmented by the YOLOv8 deep learning model, the extracted point clouds could not be enough to fit the shape to estimate the size of the apple.

Building on our observations with Microsoft Azure Kinect and Intel Realsense D435i data, the literature reveals a rich tapestry of efforts to size fruit in agricultural settings. Prior studies, such as [44], deployed calibration spheres on trees, using them as reference scales to gauge segmented apple sizes. Likewise, a shift towards 3D solutions saw the advent of reconstruction-centric techniques by the likes of [45], who turned to 3D models synthesized from multi-sensor inputs. Despite their innovative approach, these methods grapple with computational demands and falter in scenarios with occlusions, leading to incomplete reconstructions. Subsequent solutions by [46] and [47] ventured into automated shape completion, fitting ellipsoids to gathered point clouds. Yet, they faced hurdles, either due to algorithms like the Iterative Closest Point failing amidst the plant's dynamic structure or the sheer computational heft of raycasting operations.

Not to be deterred, [48] embarked on non-reconstruction-based 3D sizing, aligning an apple's major-axis with 3D points from a solitary camera shot combined with a time-of-flight sensor.

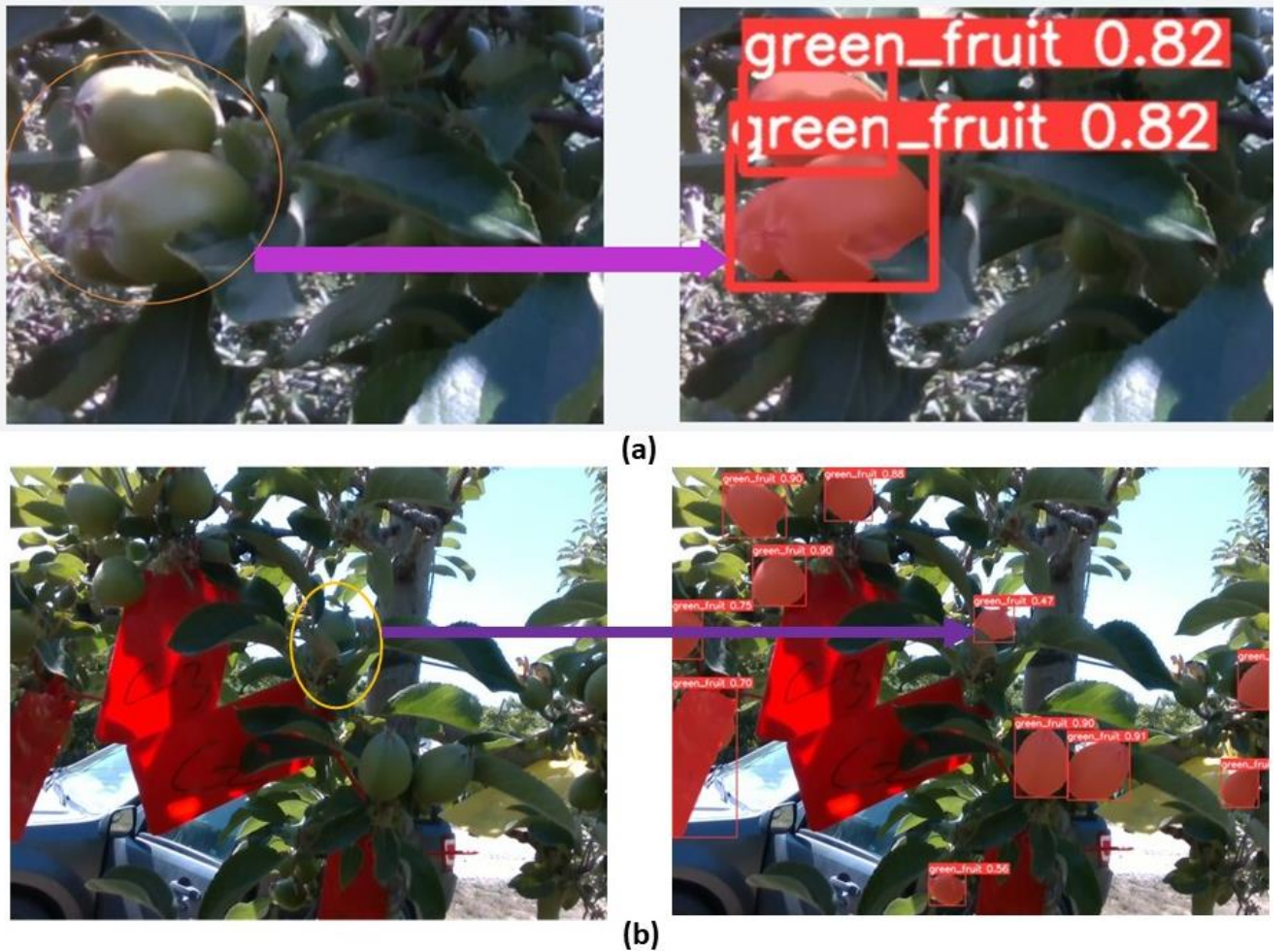


Figure 12: Overmasking and under masking during the YOLOv8 segmentation of immature green apples ;(a) showing two immature apples segmented as one immature apple, and (b) only the top part of the immature apple is segmented

Their results, though pioneering, left room for improvement with an accuracy mark of 69.1%. Adding to the 3D arsenal, [49] sized apples from singular images by fitting them in 3D spheres, however, this technique not quite suited for immature apples given their diminutive and irregular shape.

VI. CONCLUSION AND FUTURE SUGGESTIONS

The importance of detecting and estimating the size of immature apples during their early growth stages is paramount for numerous agricultural processes, from predicting yield and market reception to managing pests and making informed decisions about crop-load. Traditional methods, while effective, are labor-intensive and the recent labor shortages, accentuated by the COVID-19 pandemic, underscore the pressing need for automated solutions. In this context, the present study endeavored to harness advanced precision agriculture technologies like YOLOv8 object detection and geometric shape fitting techniques, with a specific focus on comparing the efficiency of two state-of-the-art machine vision sensors: the Intel RealSense 435i and the Microsoft Azure.

The key findings of this study are:

- The YOLOv8 object detection model demonstrated superior performance in the detection and segmentation of immature green apples.

VII. ACKNOWLEDGEMENT

This research is funded by the National Science Foundation and United States Department of Agriculture, National Institute of Food and Agriculture through the "AI Institute for Agriculture" Program (Award No. AWD003473). The authors gratefully acknowledge Dave Allan (Allan Bros., Inc.) for providing access to the orchard during the data collection and field evaluation.

REFERENCES

- [1] D. Stajko, M. Lakota, and M. Hočevár, "Estimation of number and diameter of apple fruits in an orchard during the growing season by thermal imaging," *Comput Electron Agric*, vol. 42, no. 1, pp. 31–42, 2004.
- [2] M. C. González-Araya, W. E. Soto-Silva, and L. G. A. Espejo, "Harvest planning in apple orchards using an optimization model," *Handbook of operations research in agriculture and the agri-food industry*, pp. 79–105, 2015.
- [3] S. R. Drake, J. T. Raese, and T. J. Smith, "TIME OF NITROGEN APPLICATION AND ITS INFLUENCE ON 'GOLDEN DELICIOUS' APPLE YIELD AND FRUIT QUALITY," *J Plant Nutr*, vol. 25, no. 1, pp. 143–157, 2002.
- [4] C. Carranca, G. Brunetto, and M. Tagliavini, "Nitrogen nutrition of fruit trees to reconcile productivity and environmental concerns," *Plants*, vol. 7, no. 1, p. 4, 2018.
- [5] D. B. Walsh *et al.*, "Drosophila suzukii (Diptera: Drosophilidae): invasive pest of ripening soft fruit expanding its geographic range and damage potential," *J Integr Pest Manag*, vol. 2, no. 1, pp. G1–G7, 2011.
- [6] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, "Deep learning—Method overview and review of use for fruit detection and yield estimation," *Comput Electron Agric*, vol. 162, pp. 219–234, 2019.
- [7] I. Sa, Z. Ge, F. Dayoub, B. Upcroft, T. Perez, and C. McCool, "Deepfruits: A fruit detection system using deep neural networks," *sensors*, vol. 16, no. 8, p. 1222, 2016.
- [8] D. Laborde, W. Martin, J. Swinnen, and R. Vos, "COVID-19 risks to global food security," *Science (1979)*, vol. 369, no. 6503, pp. 500–502, 2020.
- [9] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE international conference on robotics and automation (ICRA)*, IEEE, 2017, pp. 3626–3633.
- [10] C. C. Ukwoma, Q. Zhiguang, M. B. Bin Heyat, L. Ali, Z. Almaspoor, and H. N. Monday, "Recent advancements in fruit detection and classification using deep learning techniques," *Math Probl Eng*, vol. 2022, pp. 1–29, 2022.
- [11] S. K. Behera, S. Sangita, P. K. Sethy, and A. K. Rath, "Image processing based detection & size estimation of fruit on mango tree canopies," *International Journal of Applied Engineering Research*, vol. 13, no. 4, pp. 6–13, 2018.
- [12] Z. Wang, K. B. Walsh, and B. Verma, "On-tree mango fruit size estimation using RGB-D images," *Sensors*, vol. 17, no. 12, p. 2738, 2017.
- [13] L. M. Oo and N. Z. Aung, "A simple and efficient method for automatic strawberry shape and size estimation and classification," *Biosyst Eng*, vol. 170, pp. 96–107, 2018.
- [14] A. Gongal, M. Karkee, and S. Amatya, "Apple fruit size estimation using a 3D machine vision system," *Information Processing in Agriculture*, vol. 5, no. 4, pp. 498–503, 2018.
- [15] N. Tsoulas, K. K. Saha, and M. Zude-Sasse, "In-situ fruit analysis by means of LiDAR 3D point cloud of normalized difference vegetation index (NDVI)," *Comput Electron Agric*, vol. 205, p. 107611, 2023.
- [16] O. E. Apolo-Apolo, M. Pérez-Ruiz, J. Martínez-Guanter, and J. Valente, "A cloud-based environment for generating yield estimation maps from apple orchards using UAV imagery and a

- The Ellipse fitting method consistently showcased higher accuracy and efficiency in estimating fruit sizes over other techniques.
- The Microsoft Azure Kinect DK sensor outperformed the Intel RealSense 435i in estimating immature apple size.

Reflecting on this study's contributions, it's clear that integrating advanced machine vision and object detection technologies can greatly streamline and enhance traditional agricultural practices. The automation of green fruit detection and sizing not only addresses current labor challenges but also holds the potential for substantial cost savings and improved crop quality. For future research endeavors, it would be beneficial to explore the integration of even more advanced machine learning algorithms or deep learning techniques. This could further refine the fruit segmentation and sizing processes. Embracing a multi-sensor fusion approach and real-time processing capabilities might also adapt the system to different environmental conditions and a broader spectrum of crop types. The continued evolution of these technologies promises a future where agriculture is more efficient, precise, and sustainable.

- deep learning technique,” *Front Plant Sci*, vol. 11, p. 1086, 2020.
- [17] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine learning in agriculture: A review,” *Sensors*, vol. 18, no. 8, p. 2674, 2018.
- [18] O. Mirbod, D. Choi, P. H. Heinemann, R. P. Marini, and L. He, “On-tree apple fruit size estimation using stereo vision with deep learning-based occlusion handling,” *Biosyst Eng*, vol. 226, pp. 27–42, 2023.
- [19] L. Fu, F. Gao, J. Wu, R. Li, M. Karkee, and Q. Zhang, “Application of consumer RGB-D cameras for fruit detection and localization in field: A critical review,” *Comput Electron Agric*, vol. 177, p. 105687, 2020.
- [20] L. Fu, Y. Majeed, X. Zhang, M. Karkee, and Q. Zhang, “Faster R-CNN-based apple detection in dense-foliage fruiting-wall trees using RGB and depth features for robotic harvesting,” *Biosyst Eng*, vol. 197, pp. 245–256, 2020.
- [21] T. Zaenker, C. Smitt, C. McCool, and M. Bennewitz, “Viewpoint planning for fruit size and position estimation,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2021, pp. 3271–3277.
- [22] F. M. Talaat and H. ZainEldin, “An improved fire detection approach based on YOLO-v8 for smart cities,” *Neural Comput Appl*, pp. 1–16, 2023.
- [23] A. Aboah, B. Wang, U. Bagci, and Y. Adu-Gyamfi, “Real-time multi-class helmet violation detection using few-shot data sampling technique and yolov8,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5349–5357.
- [24] Jacob Solawetz and Francesco, “<https://blog.roboflow.com/whats-new-in-yolov8/#what-is-yolov8>.”
- [25] Sovit Rath, “<https://learnopencv.com/ultralytics-yolov8/#YOLOv8-vs-YOLOv5>.”
- [26] RangeKing, “<https://github.com/ultralytics/ultralytics/issues/189>.”
- [27] D. Ahmed, R. Sapkota, M. Churuvija, and M. Karkee, “Machine Vision-Based Crop-Load Estimation Using YOLOv8,” *arXiv preprint arXiv:2304.13282*, 2023.
- [28] G. Wu, Q. Zhu, M. Huang, Y. Guo, and J. Qin, “Automatic recognition of juicy peaches on trees based on 3D contour features and colour data,” *Biosyst Eng*, vol. 188, pp. 1–13, 2019.
- [29] G. Lin, Y. Tang, X. Zou, J. Xiong, and J. Li, “Guava detection and pose estimation using a low-cost RGB-D sensor in the field,” *Sensors*, vol. 19, no. 2, p. 428, 2019.
- [30] T. Li, Q. Feng, Q. Qiu, F. Xie, and C. Zhao, “Occluded apple fruit detection and localization with a frustum-based point-cloud-processing approach for robotic harvesting,” *Remote Sens (Basel)*, vol. 14, no. 3, p. 482, 2022.
- [31] T. Yu, C. Hu, Y. Xie, J. Liu, and P. Li, “Mature pomegranate fruit detection and location combining improved F-PointNet with 3D point cloud clustering in orchard,” *Comput Electron Agric*, vol. 200, p. 107233, 2022.
- [32] X. Du, Z. Meng, Z. Ma, W. Lu, and H. Cheng, “Tomato 3D pose detection algorithm based on keypoint detection and point cloud processing,” *Comput Electron Agric*, vol. 212, p. 108056, 2023.
- [33] J. Xiong *et al.*, “Real-time localization and 3D semantic map reconstruction for unstructured citrus orchards,” *Comput Electron Agric*, vol. 213, p. 108217, 2023.
- [34] Y. Jiang, J. Duan, X. Xu, Y. Ding, Y. Li, and Z. Yang, “Measurement of the banana pseudo-stem phenotypic parameters based on ellipse model,” *International Journal of Agricultural and Biological Engineering*, vol. 15, no. 3, pp. 195–202, 2022.
- [35] J. Rong *et al.*, “Development and Evaluation of a Watermelon-Harvesting Robot Prototype: Vision System and End-Effector,” *Agronomy*, vol. 12, no. 11, p. 2836, 2022.
- [36] G. Retsinas, N. Efthymiou, D. Anagnostopoulou, and P. Maragos, “Mushroom Detection and Three Dimensional Pose Estimation from Multi-View Point Clouds,” *Sensors*, vol. 23, no. 7, p. 3576, 2023.
- [37] S. Sun, M. Jiang, D. He, Y. Long, and H. Song, “Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm,” *Biosyst Eng*, vol. 187, pp. 201–213, 2019.
- [38] M. Hussain, L. He, J. Schupp, D. Lyons, and P. Heinemann, “Green fruit segmentation and orientation estimation for robotic green fruit thinning of apples,” *Comput Electron Agric*, vol. 207, p. 107734, 2023.
- [39] D. Wang and D. He, “Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning,” *Biosyst Eng*, vol. 210, pp. 271–281, 2021.
- [40] S. Lyu, R. Li, Y. Zhao, Z. Li, R. Fan, and S. Liu, “Green citrus detection and counting in orchards based on YOLOv5-CS and AI edge system,” *Sensors*, vol. 22, no. 2, p. 576, 2022.
- [41] L. Wang, Y. Zhao, Z. Xiong, S. Wang, Y. Li, and Y. Lan, “Fast and precise detection of litchi fruits for yield estimation based on the improved YOLOv5 model,” *Front Plant Sci*, vol. 13, p. 965425, 2022.
- [42] W. Du, Y. Zhu, S. Li, and P. Liu, “Spikelets detection of table grape before thinning based on improved YOLOV5s and Kmeans under the complex environment,” *Comput Electron Agric*, vol. 203, p. 107432, 2022.
- [43] Y. Nan, H. Zhang, Y. Zeng, J. Zheng, and Y. Ge, “Faster and accurate green pepper detection using NSGA-II-based pruned YOLOv5l in the field environment,” *Comput Electron Agric*, vol. 205, p. 107563, 2023.
- [44] H. Cheng, L. Damerow, Y. Sun, and M. Blanke, “Early yield prediction using image analysis of apple fruit and tree canopy features with neural networks,” *J Imaging*, vol. 3, no. 1, p. 6, 2017.
- [45] T. Jadhav, K. Singh, and A. Abhyankar, “Volumetric estimation using 3D reconstruction method for grading of fruits,” *Multimed Tools Appl*, vol. 78, pp. 1613–1634, 2019.
- [46] S. Marangoz, T. Zaenker, R. Menon, and M. Bennewitz, “Fruit mapping with shape completion for autonomous crop monitoring,” in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, IEEE, 2022, pp. 471–476.
- [47] C. Lehnert, I. Sa, C. McCool, B. Upcroft, and 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.
- [48] A. Gongal, M. Karkee, and S. Amatya, “Apple fruit size estimation using a 3D machine vision system,” *Information Processing in Agriculture*, vol. 5, no. 4, pp. 498–503, 2018.
- [49] E. Grilli, R. Battisti, and F. Remondino, “An advanced photogrammetric solution to measure apples,” *Remote Sens (Basel)*, vol. 13, no. 19, p. 3960, 2021.