Ingénierie des Systèmes d'Information
Vol. 31, No. 2, February, 2026, pp. 579-589
Journal homepage: http://iieta.org/journals/isi

Stereo Vision-Based Detection and 3D Localization of Loose Oil Palm Fruits Using YOLOv8 and Hungarian Matching

Elly Warni[1], Indrabayu[1*], Andani Achmad[2], Syafruddin Syarif[2], Rudi[3], Nadya Petroya Sadi[1]

[1] Informatics Department, Universitas Hasanuddin, Gowa 92171, Indonesia
[2] Electrical Engineering Department, Universitas Hasanuddin, Gowa 92171, Indonesia
[3] Mechanical Engineering Department, Universitas Hasanuddin, Gowa 92171, Indonesia

Corresponding Author Email: indrabayu@unhas.ac.id

https://doi.org/10.18280/isi.310225

Received: 19 October 2025
Revised: 5 January 2026
Accepted: 18 February 2026
Available online: 28 February 2026

Keywords:
loose oil palm fruits, stereo vision, YOLOv8, Hungarian matching, distance estimation, 3D localization

ABSTRACT

Accurate detection and localization of loose oil palm fruits are essential for improving harvesting efficiency, reducing post-harvest losses, and supporting automation in oil palm plantations. This study develops and evaluates a stereo vision-based perception system that integrates YOLOv8 object detection with Hungarian matching to detect loose oil palm fruits and establish reliable correspondence between stereo image pairs. Three YOLOv8 variants, namely YOLOv8n, YOLOv8s, and YOLOv8m, were systematically evaluated under different camera heights ranging from 20 to 50 cm and object distances from 20 to 120 cm. The results show that YOLOv8s achieved the best overall detection and matching performance, reaching an accuracy of 94.79% and an F1-score of 0.972, with the most favorable detection performance observed at a camera height of 30 cm. For distance estimation, stereo triangulation combined with ground-plane projection produced the lowest error at a camera height of 50 cm, achieving a Mean Absolute Percentage Error (MAPE) of 0.86% and a Mean Absolute Error (MAE) of 0.62 cm. These findings demonstrate that the proposed stereo vision system can localize loose oil palm fruits with centimeter-level accuracy under varying operational settings. The developed framework provides a practical and reliable perception module for autonomous harvesting robots and offers technical support for precision agriculture applications in oil palm plantations.

1. INTRODUCTION

The agricultural sector stands as a foundational pillar of Indonesia's economy, with the palm oil subsector playing a particularly critical role in its global standing and domestic economic health. As the world's largest producer of crude palm oil (CPO), Indonesia's vast plantations, spanning approximately 15.34 million hectares in 2022, contribute significantly to both national revenue and the global supply of vegetable oils [1]. Palm oil's importance extends beyond its dietary applications, serving as a crucial raw material for a myriad of industries, including cosmetics, pharmaceuticals, chemicals, and increasingly, as a feedstock for biodiesel production [2, 3]. The persistent growth in global population and the escalating demand for vegetable oils underscore the imperative for enhanced efficiency in palm oil harvesting practices to maintain economic competitiveness and ensure sustainable production [4]. This escalating demand necessitates optimizing every stage of the palm oil production lifecycle, making efficiency improvements in harvesting paramount to avoiding potential yield stagnation or decline and ensuring the industry's long-term viability.

Within the intricate process of palm oil cultivation, the collection of loose fruits, locally known as loose fruits palm oil, holds a particularly pivotal position, directly impacting both the volume and quality of the final CPO yield [5-7]. Loose Fruits Palm Oil are fruits that detach naturally from bunches upon reaching optimal ripeness and are highly valued for their superior oil content and quality. However, the current methods for collecting these fallen fruits are often inefficient, leading to significant production losses estimated to be several percentage points of potential yield due to spoilage, incomplete gathering, or physical loss during manual operations [8-11]. Furthermore, suboptimal collection practices can contribute to a decline in plantation hygiene and inadvertently foster weed growth, introducing further operational inefficiencies and costs [12, 13]. In an increasingly competitive global market where efficiency is a key differentiator, minimizing these losses is not merely desirable but an essential strategic objective for enhancing profitability and maintaining market leadership. The economic implications of these losses, when aggregated across millions of hectares, represent a substantial drain on the industry's potential, highlighting a critical area ripe for technological intervention and process optimization to bolster both economic output and operational sustainability [14].

Despite the profound economic and operational significance of Loose Fruits Palm Oil collection, the process in most palm oil plantations remains predominantly manual, necessitating laborers to repeatedly bend and gather the scattered fruits [15,

16]. This labor-intensive approach presents a cascade of systemic challenges that hinder productivity and scalability. The task is inherently time-consuming, physically arduous, and consequently, financially costly due to the substantial human resource input required. Workers frequently experience fatigue, which directly correlates with reduced productivity and introduces variability in the efficiency and consistency of the harvest. Moreover, the reliance on manual labor inherently limits the capacity for scaling operations, particularly as plantation areas continue to expand or as labor availability fluctuates. Research has consistently pointed to manual harvesting as a significant bottleneck in overall productivity, diminishing the industry's competitiveness in a sector where operational efficiency is paramount [17, 18]. These persistent issues emphasize the urgent and unmet need for technological advancements that can modernize and optimize the loose fruits palm oil harvesting process, moving beyond traditional methods to embrace innovative solutions that enhance efficiency, reduce costs, and improve working conditions.

The advent and rapid advancement of computer vision and robotics are offering transformative opportunities to automate various agricultural processes. Computer vision, situated at the confluence of artificial intelligence, digital imaging, and machine learning, equips machines with the capacity to interpret and act upon visual data in ways analogous to human perception [19-22]. Its applications are broad, spanning healthcare, transportation, and critically, agriculture, which has emerged as one of the most promising domains for its deployment [23]. Within the specific context of oil palm harvesting, computer vision holds the potential to revolutionize operations by enabling the development of intelligent robotic systems capable of autonomously detecting, localizing, and collecting loose fruits palm oil. Object detection, a core capability within computer vision, allows algorithms to distinguish palm loose fruits from their complex and often cluttered environment, even under challenging conditions such as variable lighting, background heterogeneity, and partial occlusions. This technology promises to transform a laborious manual task into an efficient automated operation, thereby addressing many of the long-standing challenges faced by the industry [24, 25].

Among the diverse array of object detection methodologies, single-stage detectors such as YOLO (You Only Look Once) have garnered significant attention due to their effective balance between accuracy and speed, making them particularly suitable for real-time applications [26, 27]. The YOLO architecture processes entire images in a single forward pass, enabling high frames-per-second (FPS) performance which is crucial for dynamic environments like agricultural fields, where rapid and continuous detections are required. Compared to traditional two-stage detectors like Faster R-CNN, YOLO offers substantially higher processing speeds while maintaining competitive accuracy metrics [28, 29]. This efficiency makes YOLO an ideal candidate for integration into robotic systems intended for agricultural tasks. However, merely detecting the loose fruits palm oil is insufficient for autonomous robotic harvesting. Precisely estimating the spatial distance between the detected object and the robotic end-effector is equally critical for enabling accurate interaction with the environment. In the context of Loose Fruits Palm Oil collection, accurate distance estimation ensures that robotic manipulators can approach and grasp the fruits without causing damage to the fruit itself or the surrounding plantation infrastructure, a key requirement for a fully functional automated system [30].

Traditional approaches to distance estimation in agriculture have frequently relied on monocular vision, which uses a single camera to infer depth information. This method typically employs geometric models, such as the pinhole camera model, to estimate distances based on object size and image features. However, monocular vision inherently possesses limitations in accurately estimating distances, particularly for small objects or at greater ranges where scale ambiguity becomes a significant problem. Studies have demonstrated that while monocular approaches can achieve acceptable accuracy under specific controlled conditions, such as particular camera heights (e.g., 40 cm), their performance tends to deteriorate significantly at shorter or longer distances [31]. These limitations hinder the practical deployment of monocular vision in dynamic plantation environments where loose fruits appear in a wide variety of positions, scales, and contexts, making it an unreliable sole solution for precise robotic action.

To surmount the inherent limitations of monocular vision, stereo vision has emerged as a more robust and accurate alternative. By employing two cameras positioned at a fixed, known distance apart, stereo vision mimics human binocular perception. This setup allows the system to calculate depth information through the disparity (difference) between corresponding features detected in the left and right images. This approach provides richer and more reliable depth information, enabling more accurate estimation of object distances across a wider range of scenarios. Furthermore, stereo vision systems are generally better suited for scenes containing multiple objects, as depth derivation is based on geometric triangulation rather than solely on assumptions about object size, which can be unreliable. Therefore, integrating stereo vision with state-of-the-art object detection algorithms represents a promising pathway to significantly enhance the accuracy and robustness of loose fruits palm oil detection and distance estimation systems, addressing a critical need for precision in automated harvesting.

The integration of the Hungarian Matching algorithm further strengthens this combined framework. This combinatorial optimization algorithm is widely recognized for its efficacy in solving assignment problems, such as matching detected objects across different frames or viewpoints (in this case, between left and right stereo images) while minimizing an overall cost function [32]. In the context of stereo vision, applying Hungarian Matching ensures that detected bounding boxes in the left image are accurately associated with their corresponding detections in the right image. This process significantly reduces instances of false matches and avoids duplicate counting of the same object, thereby enhancing the reliability of the stereo perception pipeline. Once objects are reliably detected and matched across the stereo pair, the system can then utilize established geometric principles, such as the Pythagorean theorem, to compute precise three-dimensional coordinates and subsequently derive the horizontal distance between the camera system and each detected loose fruits palm oil. This calculation completes the pipeline necessary for enabling a robotic manipulator to approach and collect the fruit effectively.

Previous research has explored various components of this integrated framework independently. For instance, studies have successfully implemented YOLO-based detection of palm loose fruits and investigated monocular vision for

distance estimation [33]. While these works demonstrated the feasibility of automated loose fruits palm oil detection and provided initial insights into distance estimation challenges, they also highlighted significant limitations. Specifically, monocular approaches struggled to consistently maintain low error margins, particularly at longer ranges, and single-stage detection algorithms, when used without robust stereo matching, risked misidentification or inconsistent counting of fruits. Consequently, there exists a clear research gap in developing a system that synergistically combines stereo vision, advanced object detection (like YOLOv8), and a reliable matching algorithm to provide both accurate detection and precise spatial localization for loose fruits palm oil. Addressing this gap is crucial for the practical implementation of autonomous harvesting robots. More critically, no existing study has attempted to synergistically combine stereo vision, state-of-the-art single-stage detection (YOLOv8) and a principled assignment algorithm (Hungarian Matching) into a unified pipeline specifically designed for the 3D localization of loose palm oil fruits. Furthermore, the systematic influence of camera height on the dual tasks of detection accuracy and distance estimation accuracy remains unexplored in the palm oil harvesting context.

This study is therefore motivated by the need to develop and rigorously evaluate an integrated stereo vision-based system for detecting palm loose fruits and estimating their ground-plane distance. The primary objectives are to: (1) compare the performance of different YOLOv8 variants (nano, small, medium) when integrated into a stereo vision pipeline for loose fruits palm oil detection; (2) investigate the effectiveness of Hungarian Matching in establishing reliable correspondences between stereo image pairs; (3) determine the optimal camera height for both accurate detection-matching and precise distance estimation; and (4) quantify the system's performance using established metrics such as accuracy, F1-score, Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE).

The principal novelty of this research lies in three specific contributions that distinguish it from prior studies. First, unlike monocular approaches such as those in the study [31], which suffer from scale ambiguity and significant accuracy degradation at varying distances, our system employs a calibrated stereo vision framework that provides geometrically grounded depth estimation through triangulation, thereby enabling highly precise distance measurements. Second, in contrast to detection-only studies [33] that do not address the stereo correspondence problem, we introduce the Hungarian Matching algorithm as a structured assignment mechanism to associate detections across stereo pairs, thereby eliminating duplicate counting and reducing false matches. Third, we present the first systematic empirical investigation of the task-dependent relationship between camera height and system performance in loose palm fruit applications. The results indicate that the optimal height for detection differs from the optimal height for distance estimation, a finding with direct implications for robotic platform design. This synergistic integration enables centimeter-level distance estimation and high detection accuracy, providing a robust perception backbone for the autonomous deployment of loose palm fruit harvesting robots. Ultimately, the system contributes to reducing fruit loss, enhancing harvesting productivity, and improving sustainability in the palm oil industry. The scope of this study encompasses the development, integration, and comprehensive evaluation of the perception system under varied operational conditions within a controlled plantation environment.

2. METHODOLOGY

The methodology encompasses field stereo data acquisition, YOLOv8-based detection of loose fruits, inter-lens correspondence, and 3D geometric reconstruction to determine direct camera to object and ground-plane horizontal distances. This design is consistent with established computer vision practices for stereo systems and single-stage detection, proving efficient in unstructured environments such as oil palm plantations.

2.1 Stages of research

The research commenced with a comprehensive literature review and requirements identification phase. This was followed by the acquisition of stereo video data, the subsequent design and implementation of a detection stereo pipeline, rigorous testing, analysis of the obtained results, and finally, the compilation of the research report. The workflow diagram clearly delineates the logical sequence of these stages, commencing with the establishment of the theoretical foundation, progressing through system design, and concluding with empirical evaluation. This sequence adheres to established methodological guidelines for computer vision based engineering research. Figure 1 illustrates the procedural flow of the research. The diagram outlines the sequential stages, beginning with the initial literature review and system design, followed by implementation and empirical testing, and concluding with the analysis of results.

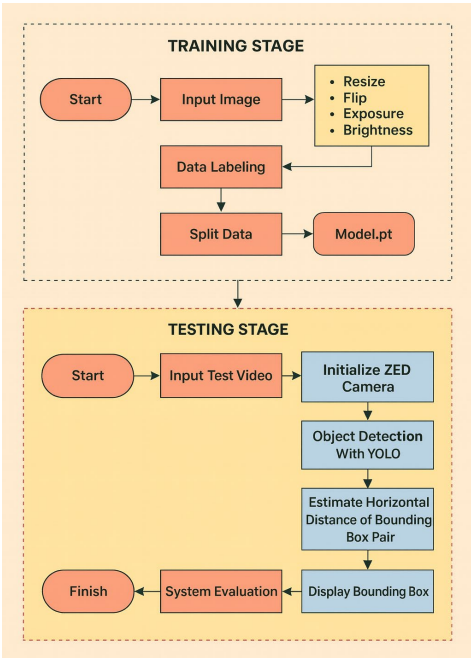

Figure 1. Research stages

2.2 Training stages

2.2.1 Dataset

The dataset utilized in this study was designed to train and evaluate the performance of object detection and distance

estimation models for loose palm fruits, comprising a total of 1,520 annotated images in YOLO format. This dataset is a combination of primary and secondary data. The primary data, consisting of 759 images, were collected directly using a ZED 2 stereo camera under diverse oil palm plantation conditions, with variations in camera height (20-50 cm) and object distance (20-120 cm) to encompass real world scenarios, including variations in lighting and soil texture. To broaden the variation and enhance model generalization, an additional 761 secondary images were obtained from public repositories such as Roboflow. Bounding box annotations were performed manually using Roboflow tools. Furthermore, 41 recorded videos using the ZED 2 were prepared for system testing under continuous real world conditions. To further enhance diversity and mitigate overfitting, image data augmentation strategies were also applied, including horizontal flipping, adjustments to brightness, saturation, exposure, and random cropping. Figure 2 shows an example of a dataset.

Figure 2. Example of a dataset

2.2.2 Preprocessing and data augmentation

Preprocessing was essential to standardize input images and enhance the training process. All images were resized to 640×640 pixels, the default resolution for YOLOv8, balancing computational efficiency with detection precision. Data augmentation techniques included horizontal flipping, variations in brightness, saturation, and exposure, as well as random cropping. These operations simulated real world variability, reducing overfitting and increasing model robustness in diverse plantation conditions.

2.2.3 YOLOv8 object detection model

The object detection process using YOLOv8 can be seen in Figure 3, follows a sequential pipeline that transforms raw image inputs into final detection outputs. Initially, the input image is processed by the backbone, consisting of convolutional layers and C2f modules designed to extract representative visual features. These features are then forwarded to the PAN-FPN, which aggregates multi scale information to enhance the detection of objects with varying sizes.

Subsequently, the decoupled head performs three specialized tasks: (1) objectness prediction to determine the presence of the target object, (2) bounding box regression to localize objects with high precision, and (3) confidence scoring to evaluate the reliability of classification. The outputs of these heads are optimized through the YOLO loss function, which minimizes localization and classification errors.

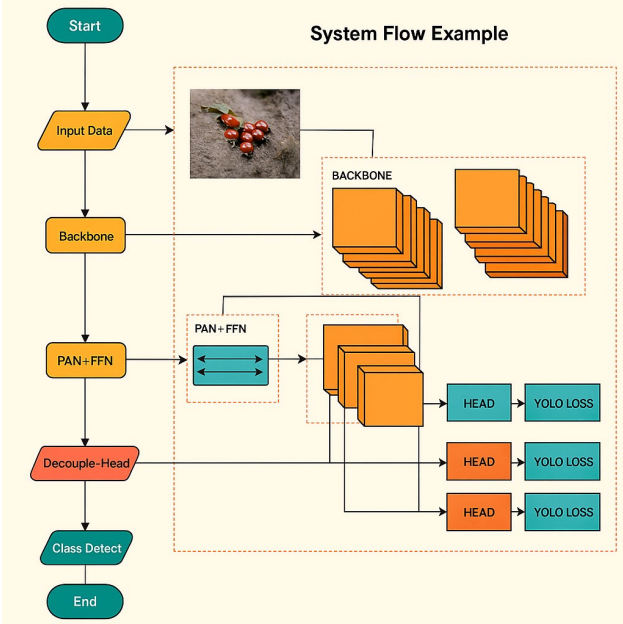

Figure 3. YOLOv8 architecture

Finally, the model produces bounding boxes and confidence scores corresponding to the single target class, thereby ensuring efficient and accurate object detection.

2.2.4 Model training scenarios

Three YOLOv8 model variants YOLOv8n, YOLOv8s, and YOLOv8m were trained to evaluate the accuracy-speed trade-off on edge computing devices. The training dataset comprised 3,998 annotated images (a composite of augmented primary and secondary data). The models were trained using specific hyperparameters: 500 epochs, a batch size of 32, the SGD optimizer, and an initial learning rate of 0.01. This configuration aligns with standard practices for training single-stage detectors in scenarios involving small to medium-sized objects. Comparative analysis of the model variants, as detailed in Table 1, highlights the incremental increases in Mean Average Precision (mAP) and Floating Point Operations (FLOPs) from the nano to the medium variants, alongside their respective inference speeds on CPU and GPU. These findings are consistent with the observed scaling laws within the modern YOLO family.

2.3 Testing stage

In the testing phase, stereo video is acquired, followed by object detection on both lenses. Subsequently, assignment based stereo correspondence is performed, culminating in 3D reconstruction via triangulation to determine both the straight line and horizontal distances.

Table 1. Technical specifications of YOLOv8 models for nano, small, and medium variants

Model	Size (pixels)	mAP	Speed CPU ONNX	Speed A100 TensorRT	Params	FLOPs
YOLOv8n	640	37.3	80.4	0.99	3.2	08.07
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9

2.3.1 Test video input

Video acquisition was captured using the ZED 2 Stereo Camera. This camera has two parallel lenses forming a stereo camera setup. The camera was placed on a previously constructed pushing device. The test video acquisition is shown in Figure 4.

Figure 4. Data acquisition testing

The stereo camera simultaneously produces two synchronized *frames*, the right and left frames. The two *frames* taken have a bolt in front of them as a detection. The video has a resolution from the camera input of 1920 × 1080. The results of the ZED 2 camera are shown in Figure 5.

Figure 5. The output obtained from the ZED 2 camera

2.3.2 ZED 2 camera initialization

The ZED 2 camera system is initialized and configured using its factory-provided parameters, which encompass intrinsic and extrinsic calibration data, prior to processing SVO (Stereolabs Video) data. The test video, utilizing the .svo2 file format, provides essential depth and unit information necessary for estimating the horizontal distance to the target object (here, loose fruits). It is important to note that all experiments in this study were conducted on offline, pre-recorded video data rather than real-time streaming; the SVO format preserves the full stereo and depth information captured during field acquisition, enabling reproducible offline analysis. The instantiation of the ZED camera object is performed via zed = sl.Camera(), enabling subsequent operations. Default initialization parameters are then established by constructing init_params = sl.InitParameters(). The init_params.depth_mode is set to sl.DEPTH_MODE.PERFORMANCE, selecting a high-performance depth sensing algorithm that prioritizes processing speed over maximum accuracy, which is suitable for offline batch processing of small objects like loose fruits and yielding improved point cloud processing speed. Furthermore, init_params.coordinate_units is set to sl.UNIT.METER to ensure all spatial measurements (depth, X/Y/Z coordinates) are consistently reported in meters, aligning with the applied Euclidean transformations. This initialization enables the extraction of individual frames, depth maps, and point clouds from the pre-recorded video captured by the ZED 2 camera.

2.3.3 Object detection with YOLOv8

The research implemented a comparative analysis of three YOLOv8 variants (YOLOv8n, YOLOv8s, YOLOv8m) for the detection of palm loose fruits. Processing data from both the left and right frames of stereo imagery, the YOLOv8 architecture was employed to predict bounding boxes, assign unique identifiers to matched object pairs, and estimate distances. To refine these predictions, bounding box values were filtered using a confidence threshold (0.3) and an IoU threshold (0.3). The confidence threshold served to discard detections with low probability scores, while the IoU threshold eliminated redundant overlapping bounding boxes by evaluating prediction score similarity. This comparative assessment, using consistent confidence and IoU parameters across all YOLOv8 variants, aimed to identify the most suitable model for further processing. Post-detection, the Hungarian Matching algorithm was employed to establish correspondences between detected loose fruits across the stereo frames, assigning final pair IDs. Representative examples of the YOLOv8 detection results are visually documented in Figure 6.

Figure 6. YOLOv8 detection results

2.3.4 Objects with stereo matching (Hungarian)

The process of matching detected objects across the right and left frames involves quantifying the proximity of bounding box centers. Subsequently, an optimal pairing of loose fruits is determined using the Hungarian Matching algorithm. Initially, the minimum and maximum coordinates $(xmin, ymin, xmax, ymax)$ of each detected bounding box are extracted to compute valid pixel center coordinates, resulting in two lists representing detections in the left and right frames, respectively. Following this, an $N\ x\ M$ cost matrix is constructed, where $N\ and\ M$ denote the number of detections in the left and right frames, correspondingly. Each matrix element quantifies the Euclidean distance between the centers of detected objects from the left and right frames to measure their spatial closeness. The Hungarian Matching algorithm is then applied to find the optimal assignment across rows and columns that minimizes the total cumulative cost. While the Hungarian algorithm guarantees a globally minimal cost assignment, it may occasionally produce pairings that are spatially disparate. Therefore, a post-assignment filtering step is implemented using a predefined threshold to ensure the validity of the selected pairs, stating that only pairs within this maximum distance threshold are considered correct by the system. Therefore, a post-assignment filtering step is implemented using a maximum distance threshold of 50 pixels to ensure the validity of the selected pairs. This threshold was empirically determined based on the maximum expected horizontal disparity for loose fruits within the operational distance range (20–120 cm) at the ZED 2 camera's stereo baseline of 120 mm. Pairs exceeding this threshold are discarded as invalid correspondences. In our experiments, this

filtering step eliminated approximately 3–5% of initially matched pairs, predominantly at extreme distances where disparity estimation becomes less reliable, thereby improving overall matching precision.

2.3.5 Estimation of distance in bounding box pairs

Subsequent to the matching procedure, bounding box coordinates for each matched pair (i, j) are extracted from both stereo frames. Three-dimensional (3D) spatial information, comprising X, Y, and Z coordinates, is subsequently computed via triangulation. The X-coordinate quantifies the lateral spatial displacement, with positive and negative values indicating positions to the right and left, respectively, relative to the left camera's optical axis. The Y-coordinate represents the vertical distance from the object to the camera plane, accounting for camera pitch and height offsets. The Z-coordinate signifies the depth, or forward distance, from the camera to the object. While the X and Y coordinates are derived using Eq. (1), the Z-coordinate computation relies on triangulation principles outlined in Eq. (2). The coordinate system is then adjusted for camera pitch rotation, resulting in transformed Y and Z values as presented in Eq. (3). Figure 7 offers a graphical representation of these spatial coordinates.

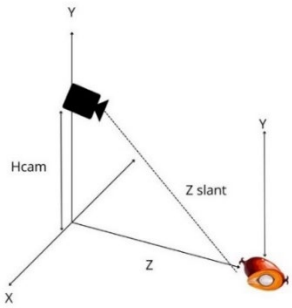

Figure 7. Illustration of the calculation of the straight distance of the camera lens

$$\begin{aligned} Yworld &= \cos(\theta) \times Y - \sin(\theta) \times Z \\ Zworld &= \sin(\theta) \times Y + \cos(\theta) \times Z \end{aligned} \quad (1)$$

The X, Y, and Z spatial coordinates are instrumental in determining the direct, slant-line distance from the camera lens to the loose fruit, as formulated in Eq. (2). Following this calculation, the horizontal distance from the camera to the fruit is derived using the Pythagorean theorem, detailed in Eq. (3). Figure 8 provides an illustrative depiction of the methodology employed for this distance estimation.

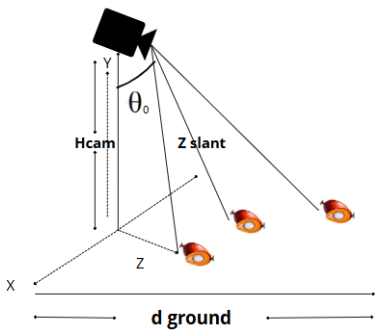

Figure 8. Illustration of distance estimation from the camera to loose fruits palm oil

$$Z_{slant} = \sqrt{(X)^2 + (Yworld)^2 + (Zworld)^2} \quad (2)$$

$$\begin{aligned} d_{ground} &= \sqrt{(Z_{slant})^2 - (H_{cam})^2} \\ d_{cm} &= d_{ground} \times 100 \end{aligned} \quad (3)$$

Description:

- Z_{slant} is Direct distance from the camera to the object (loose fruits palm oil).
- d_{ground} is Horizontal distance on the ground plane.
- d_{cm} is Distance measured in centimeters.
- H_{cam} is Height of the camera from the ground.
- X is Horizontal lateral distance (left or right). If the object is on the right side of the left lens, the value of X is positive; conversely, if the object is on the left side of the left lens, the value of X is negative.
- Y_World is Vertical distance (object height relative to the camera) after applying pitch rotation.
- Z_World is Horizontal forward distance (depth) from the camera to the object after applying pitch rotation.

2.4 YOLOv8 testing

During testing, stereo video sequences captured by the ZED 2 camera were processed frame by frame. Object detection was performed on both left and right images independently. Predictions were filtered based on confidence thresholds (0.3) and intersection over union (IoU) thresholds (0.3) to ensure reliable bounding box generation. The outputs of the detection stage were then used as inputs for the Hungarian Matching algorithm to establish correspondences between stereo pairs.

3. RESULTS AND DISCUSSION

This chapter presents the empirical performance of the proposed stereo vision pipeline for the detection and distance estimation of loose palm oil fruits. The results are systematically analyzed across four key areas: (1) the detection and matching performance of different YOLOv8 variants, (2) the impact of detection height on system accuracy, (3) the accuracy of distance estimation at various heights, and (4) an overall comparative synthesis of the findings.

For all experimental conditions, unless otherwise specified, stereo video frames (left and right) were processed independently for object detection. Subsequent matching of detected objects between frames was performed using the Hungarian Matching algorithm. The performance of the detection and matching module was quantitatively evaluated using accuracy and F1-score, metrics commonly employed to assess classification and localization tasks. For distance estimation, the MAE and MAPE were utilized to quantify the discrepancy between actual and estimated distances. It is important to note that, by design of instance-level evaluation, the count of True Negatives (TN) is zero, as frames lacking any detected objects (and consequently, bounding boxes) are not included in the box-level performance metrics. This design choice may inflate the reported Accuracy metric, since Accuracy = (TP + TN) / (TP + TN + FP + FN), and TN = 0 removes a component that would otherwise contribute to the denominator. However, given the high target-presence rate in our test videos (loose fruits appear in the vast majority of frames), the practical impact of this inflation is minimal. The F1-score, which is independent of TN, serves as a more

reliable and complementary performance indicator in this context. Future work could incorporate mean Average Precision (mAP) as a supplementary detection metric.

3.1 Detection and matching performance

The detection stage compared three YOLOv8 variants nano (YOLOv8n), small (YOLOv8s), and medium (YOLOv8m) trained with identical settings and evaluated on 41 stereo videos covering camera heights of 20, 30, 40, and 50 cm and object distances from 20 to 120 cm. Hungarian Matching associated left/right detections using a cost matrix on box centers with a filtering threshold to reject implausible correspondences.

The performance of the detection and matching system was systematically evaluated using accuracy and F1-score, with results detailed in the provided confusion matrices. Analysis began with the YOLOv8n model paired with Hungarian Matching, as shown in Table 2. This configuration yielded an overall accuracy of 93.56% and an F1-score of 0.966. The highest performance was recorded at a camera height of 30 cm, achieving 97.96% accuracy and an F1-score of 0.99. However, the YOLOv8n model's accuracy showed greater variability with distance, performing less reliably at closer ranges. The total counts of false positives (FP) and false negatives (FN) were 2574 and 4706, respectively, indicating a notable number of misclassifications.

Moving to the YOLOv8s model with Hungarian Matching, presented in Table 3, the results indicated superior overall performance. This variant achieved an accuracy of 94.79% and an F1-score of 0.972. Similar to YOLOv8n, the peak performance was observed at a camera height of 30 cm, with an accuracy of 97.63% and an F1-score of 0.99. Importantly, the YOLOv8s model demonstrated more consistent performance across varying distances and exhibited a reduction in both false positives (1982) and false negatives (3787) compared to YOLOv8n. This suggests an improved ability in accurately detecting and matching objects between stereo frames.

The YOLOv8m model, evaluated in Table 4, achieved an overall accuracy of 93.00% and an F1-score of 0.964. While it performed well at specific heights, notably 50 cm (97.12% accuracy, 0.99 F1-score), its aggregate performance was surpassed by YOLOv8s. The higher number of false positives (2023) and false negatives (5391) observed with YOLOv8m compared to YOLOv8s points to potential difficulties in precisely detecting smaller objects like palm kernels. This could be attributed to YOLOv8m's larger receptive field, which might lead to less focused detection on fine details.

Table 2. Confusion matrix across all heights for YOLOv8n + Hungarian matching

Camera Height (cm)	Frame	Ture Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	Accuracy (%)	F1 Score
20	1562	27062	1332	0	2322	88.10	0.94
30	1570	29913	389	0	233	97.96	0.99
40	1415	25572	383	0	1741	92.33	0.96
50	1260	23271	470	0	410	96.36	0.98
Total	5807	105818	2574	0	4706	93.56	0.966

Table 3. Confusion matrix across all heights for YOLOv8s + Hungarian matching

Camera Height (cm)	Frame	Ture Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	Accuracy (%)	F1 Score
20	1562	26891	521	0	2040	91.30	0.95
30	1570	29974	369	0	360	97.63	0.99
40	1415	25846	528	0	999	94.42	0.97
50	1260	22349	564	0	388	95.91	0.98
Total	5807	105060	1982	0	3787	94.79	0.972

Table 4. Confusion matrix across all heights for YOLOv8m + Hungarian matching

Camera Height (cm)	Frame	Ture Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	Accuracy (%)	F1 Score
20	1562	20959	672	0	4241	81.01	0.90
30	1570	28995	607	0	388	96.68	0.98
40	1415	25323	379	0	441	96.86	0.98
50	1260	23170	365	0	321	97.12	0.99
Total	5807	98447	2023	0	5391	93.00	0.964

Table 5. Evaluation results of YOLOv8 + Hungarian matching

YOLOv8+Hungarian	Number of Frames	Ture Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)	Accuracy (%)	F1 Score
nano	5807	105818	2574	0	4706	93.56	0.966
small	5807	105060	1982	0	3787	94.79	0.972
medium	5807	98447	2023	0	5391	93.00	0.964

In summary, Table 5 consolidates the performance evaluation across all YOLOv8 variants. The YOLOv8s model consistently outperformed both YOLOv8n and YOLOv8m, achieving the highest overall accuracy (94.79%) and F1-score

(0.972). This superior capability in detecting and matching palm oil fruits, coupled with lower false positive and false negative rates, indicates better generalization and fewer erroneous associations. While YOLOv8n showed the lowest overall performance and YOLOv8m had variable results, YOLOv8s consistently offered the most optimal balance for this particular task.

3.2 Distance estimation

Actual distances between the camera and the loose fruits were measured using a calibrated measuring tape with 1 mm resolution. Loose fruits were placed at predefined distances (20–120 cm in 10 cm increments) along a straight line from the camera position. Each measurement was repeated three times and averaged to minimize measurement error, and the camera's pushing device was aligned along a marked straight path to ensure consistent positioning. Analysis of the distance estimation results at a camera height of 20 cm, as detailed in Table 6, revealed a system performance characterized by a MAPE of 2.84% and an MAE of 1.45 cm. The system demonstrated a good capacity for estimating distance, with deviations generally within 2 cm and below 3% relative error. Notably, at a distance of 110 cm, the error was minimal, with an absolute error of 0.05 cm and a relative error of 0.05%. Conversely, the highest error occurred at a distance of 20 cm, showing an absolute error of 2.74 cm and a relative error of 13.70%. This increased error at closer ranges is attributed to the larger disparity values inherent in stereo vision at short distances, making the system more sensitive to minor errors. As the distance increased, the error generally decreased, suggesting a less stable performance at closer proximity.

Table 6. Calculation of error at a height of 20 cm

Data	Actual Distance (cm)	Estimated Distance (cm)	Mean Absolute Percentage Error (MAPE) (%)	Mean Absolute Error (MAE) (cm)
t20j20	20	22.74	13.70	2.74
t20j30	30	29.46	1.80	0.54
t20j40	40	40.6	1.50	0.6
t20j50	50	49.39	1.22	0.61
t20j60	60	61.3	2.17	1.3
t20j70	70	70.79	1.13	0.79
t20j80	80	80.92	1.15	0.92
t20j90	90	91.66	1.84	1.66
t20j100	100	106.26	6.26	6.26
t20j110	110	109.95	0.05	0.05
t20j120	120	120.53	0.44	0.53
	Average		2.84	1.45

The analysis of distance estimation at a camera height of 30 cm, presented in Table 7, indicated a significant improvement in performance. The system achieved a MAPE of 1.89% and an MAE of 1.18 cm. These results signify a very good estimation capability, with errors generally less than 2 cm and below 2% relative error. However, some fluctuations were observed; for instance, at 100 cm distance, the relative error reached 4.63% with an absolute error of 4.63 cm. Conversely, at 120 cm, the relative error was very low at 0.06% with an absolute error of 0.07 cm. The non-linear and fluctuating error rates observed at this height suggest potential inconsistencies attributed to non-uniform lighting conditions and minor instability in the camera movement apparatus, which may have introduced excessive vibrations during data collection.

At a camera height of 40 cm, the distance estimation system demonstrated improved stability and accuracy, as detailed in Table 8. The analysis showed a MAPE of 1.44% and an MAE of 0.97 cm. The system exhibited excellent performance, with errors typically below 1 cm in absolute terms and under 2% relatively. While the closest tested distance (40 cm) showed the highest relative error (4.08%) and absolute error (1.63 cm), performance significantly improved at 60 cm, with an error of only 0.02% relative and 0.01 cm absolute. The average error figures below 1 cm absolute and below 1.5% relative validate the system's good performance across the measured range of 30–120 cm. Notably, the errors at 40 cm height remained relatively stable across all tested distances, with minimal deviations from the actual values.

Table 7. Calculation of error at a height of 30 cm

Data	Actual Distance (cm)	Estimated Distance (cm)	Mean Absolute Percentage Error (MAPE) (%)	Mean Absolute Error (MAE) (cm)
t30j20	20	19.3	3.50	0.7
t30j30	30	29.34	2.20	0.66
t30j40	40	40.24	0.60	0.24
t30j50	50	51.72	3.44	1.72
t30j60	60	61.78	2.97	1.78
t30j70	70	70.08	0.11	0.08
t30j80	80	81.03	1.29	1.03
t30j90	90	90.58	0.64	0.58
t30j100	100	104.63	4.63	4.63
t30j110	110	111.49	1.35	1.49
t30j120	120	120.07	0.06	0.07
	Average		1.89	1.18

Table 8. Calculation of error at a height of 40 cm

Data	Actual Distance (cm)	Estimated Distance (cm)	Mean Absolute Percentage Error (MAPE) (%)	Mean Absolute Error (MAE) (cm)
t40j30	30	29.23	2.57	0.77
t40j40	40	41.63	4.08	1.63
t40j50	50	50.14	0.28	0.14
t40j60	60	60.01	0.02	0.01
t40j70	70	71.56	2.23	1.56
t40j80	80	80.72	0.90	0.72
t40j90	90	90.14	0.16	0.14
t40j100	100	99.15	0.85	0.85
t40j110	110	112.05	1.86	2.05
t40j120	120	121.79	1.49	1.79
	Average		1.44	0.97

The analysis of distance estimation at a camera height of 50 cm, presented in Table 9, revealed the system best performance. The results showed a MAPE of 0.86% and an MAE of 0.62 cm. Even at the distance with the highest relative error (70 cm, 1.76% relative error and 1.41 cm absolute error), the performance was still strong. The lowest errors were observed at 110 cm, with a relative error of 0.15% and an absolute error of 0.17 cm. Overall, the system at this height demonstrated the most robust performance, with errors

consistently close to zero and not significantly deviating across the tested distances. The very small average error values highlight 50 cm as the most ideal height for accurate distance measurements within the 40 cm to 120 cm range.

Table 9. Calculation of error at a height of 50 cm

Data	Actual Distance (cm)	Estimated Distance (cm)	Mean Absolute Percentage Error (MAPE) (%)	Mean Absolute Error (MAE) (cm)
t50j40	40	39.17	2.08	0.83
t50j50	50	49.9	0.20	0.1
t50j60	60	60.21	0.35	0.21
t50j70	70	69.25	1.07	0.75
t50j80	80	81.41	1.76	1.41
t50j90	90	89.12	0.98	0.88
t50j100	100	100.69	0.69	0.69
t50j110	110	109.83	0.15	0.17
t50j120	120	120.54	0.45	0.54
	Average		0.86	0.62

Table 10. Overall distance estimation results of the scenario

Camera Height (cm)	Camera Movement	Mean Absolute Percentage Error (MAPE) (%)	Mean Absolute Error (MAE) (cm)
20	progress	2.84%	1.45
30	progress	1.89%	1.18
40	progress	1.44%	0.97
50	progress	0.86%	0.62

An overall summary of the distance estimation performance at different camera heights, as presented in Table 10, indicates a clear trend. The system consistently improved in accuracy with increasing camera height. Specifically, the 50 cm camera height yielded the best performance, with the lowest MAPE of 0.86% and MAE of 0.62 cm. Conversely, the 20 cm height showed the highest MAPE (2.84%) and MAE (1.45 cm), largely due to the increased disparity and sensitivity to minor errors at closer ranges. This consistent reduction in error with greater camera height suggests that a higher perspective provides a more stable and accurate stereo view for distance measurement. While errors across all heights were generally small, the 50 cm height is identified as the most optimal for estimating distances from the base of the camera to the loose fruits palm oil.

An example of the detection and distance estimation results of loose fruits palm oil using the ZED 2 stereo camera is presented in Figure 9.

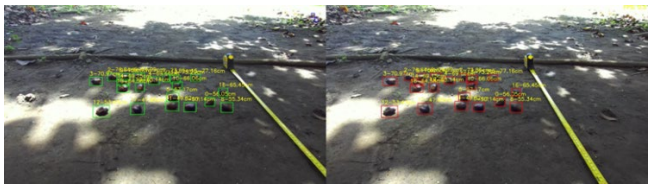

Figure 9. Detection and distance estimation results

The empirical results obtained from this study provide significant insights into the performance of a stereo vision pipeline for detecting and estimating the distance of loose palm oil fruits, with a focus on its potential application in automated harvesting systems. The investigation into different YOLOv8 variants, coupled with Hungarian Matching, revealed distinct performance characteristics influenced by model complexity and camera height.

The comparative analysis of detection and matching performance, as detailed in Section 3.1 and summarized in Tables 2-5, clearly indicates that YOLOv8s offers the most optimal balance of accuracy, F1-score, and robustness for this task. The superiority of YOLOv8s can be attributed to its architectural balance (11.2M parameters, 28.6 GFLOPs), providing sufficient representational capacity without overfitting risks. YOLOv8n (3.2M parameters) exhibited limited feature extraction capability, particularly at closer ranges where fruits exhibit greater intra-class variability. YOLOv8m's unstable performance, despite having 25.9M parameters and the highest general-purpose mAP (50.2), can be explained by the mismatch between model complexity and dataset scale (3,998 images). Furthermore, YOLOv8m's larger receptive field may aggregate excessive contextual background information (soil, debris, shadows), leading to confusion with small, irregularly shaped loose fruits. YOLOv8s consistently delivered high performance across various conditions. The peak accuracy and F1-score for all variants were observed at a camera height of 30 cm, suggesting that this elevation provides an advantageous perspective for capturing sufficient detail for accurate detection and matching of palm kernels. The lower incidence of false positives and false negatives with YOLOv8s further supports its selection as the preferred model.

Regarding distance estimation, the application of the Pythagorean Theorem, utilizing depth data from the ZED stereo camera, demonstrated the system's capability to estimate distances with reasonable accuracy. As discussed in Section 3.2 and supported by Tables 6-9, camera height has a direct and significant impact on estimation accuracy. The performance at a 20 cm camera height was the least stable, with higher errors at closer ranges attributed to increased disparity noise. Improvements were observed at 30 cm and 40 cm, with the latter showing notably more consistent results. The most compelling outcomes were achieved at a camera height of 50 cm, which yielded the lowest MAPE (0.86%) and MAE (0.62 cm). The improvement from 20 cm (MAPE 2.84%) to 50 cm (MAPE 0.86%) is not attributable to changes in the physical stereo baseline, which remains fixed at 120 mm. Rather, the improvement stems from more favorable stereo geometry at greater heights: the increased camera-to-ground angle reduces the sensitivity of triangulated depth to pixel-level disparity errors. At 30 cm, the camera captures loose fruits at closer range with higher spatial resolution per fruit, improving detection reliability. For practical deployment, a dual-height or adjustable-height camera mount could be employed, or alternatively, the system could operate at an intermediate height (40 cm, MAPE 1.44%, accuracy 94.42%) that balances both objectives. The observed errors, although generally low, were influenced by external factors such as lighting variations and the mechanical stability of the camera platform, highlighting areas for potential refinement in system implementation.

The findings of this study contribute to the development of intelligent agricultural robotics by providing empirical evidence for an effective stereo vision approach to palm kernel detection and distance estimation. The demonstrated superiority of YOLOv8s for detection and the identified

optimal camera height of 50 cm for distance estimation offer critical design guidelines for automated harvesting robots in the palm oil industry. By accurately identifying and locating loose palm fruits, such a system has the potential to significantly enhance harvesting efficiency, reduce manual labor dependency, minimize fruit loss, and ultimately boost productivity in palm oil plantations. Future work could focus on integrating this pipeline into a robotic system and further optimizing performance by addressing the identified environmental error factors.

4. CONCLUSIONS

This study successfully developed and validated an integrated stereovision pipeline for the autonomous detection and distance estimation of loose fruits palm oil, addressing critical inefficiencies in manual harvesting. Our findings conclusively demonstrate that the YOLOv8s model, when combined with Hungarian Matching for stereo correspondence, offers the most robust performance for detecting loose fruits palm oil, achieving a high overall accuracy of 94.79% and an F1-score of 0.972 across various conditions. This configuration significantly outperformed other YOLOv8 variants, particularly in terms of minimizing false positives and negatives, crucial for reliable robotic operation. Furthermore, the research identified distinct optimal camera heights for different tasks: approximately 30 cm proved most favorable for stable loose fruits palm oil detection and matching, achieving up to 97.63% accuracy. Conversely, a height of 50 cm yielded the highest accuracy in distance estimation, with the lowest average MAPE of 0.86% and MAE of 0.62 cm, highlighting the necessity of balancing these operational parameters. The system's ability to provide centimeter-level distance accuracy and robust detection, even amidst environmental variations and close-range challenges, establishes a strong foundation for its integration into autonomous harvesting robots. This work significantly contributes to the body of knowledge in precision agriculture by demonstrating a practical and effective solution that surpasses the limitations of previous monocular approaches. Future research could focus on further enhancing robustness to extreme lighting variations, investigating adaptive algorithms for dynamic environmental conditions, and exploring advanced object tracking capabilities for improved loose fruits palm oil collection efficiency.

ACKNOWLEDGMENT

The researchers would like to thank the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for funding this research under the Fundamental Research Scheme in 2025 with Grand Number 02209/UN4.22.2/PT.01.03/2025.

REFERENCES

[1] Austin, K.G., Kasibhatla, P.S., Urban, D.L., Stolle, F., Vincent, J. (2015). Reconciling oil palm expansion and climate change mitigation in Kalimantan, Indonesia. PLOS One, 10(5): e0127963. https://doi.org/10.1371/journal.pone.0127963

[2] Rachmadona, N., Harada, Y., Amoah, J., Quayson, E., Aznury, M., Hama, S., Kondo, A., Ogino, C. (2022). Integrated bioconversion process for biodiesel production utilizing waste from the palm oil industry. Journal of Environmental Chemical Engineering, 10(3): 107550. https://doi.org/10.1016/j.jece.2022.107550

[3] Gheewala, S.H., Jaroenkietkajorn, U., Nilsalab, P., Silalertruksa, T., Somkerd, T., Laosiripojana, N. (2022). Sustainability assessment of palm oil-based refinery systems for food, fuel, and chemicals. Biofuel Research Journal, 9(4): 1750-1763. https://doi.org/10.18331/BRJ2022.9.4.5

[4] OECD and Food and Agriculture Organization of the United Nations, OECD-FAO Agricultural Outlook 2025-2034.

[5] de Vos, R.E., Nurfalah, L., Tenorio, F.A., et al. (2023). Shortening harvest interval, reaping benefits? A study on harvest practices in oil palm smallholder farming systems in Indonesia. Agricultural Systems, 211: 103753. https://doi.org/10.1016/j.agsy.2023.103753

[6] Sugianto, H., Donough, C.R., Monzon, J.P., Pradiko, I., et al. (2025). Improving yield and profit in smallholder oil palm fields through better agronomy. Agricultural Systems, 224: 104269. https://doi.org/10.1016/j.agsy.2025.104269

[7] Yusoff, M.Z.M. (2019). Loose fruit collector machine in Malaysia: A review. International Journal of Engineering Technology and Sciences, 6(2): 65-75. https://doi.org/10.15282/ijets.v6i2.2909

[8] Hassan, M.S., Hasan, W.Z.W., Mustafa, M.F., Kadir, M.A., Azis, N., Yusoff, M.Z.M. (2019). Loose fruit recognition system with implementation of SURF feature extraction method. In 2019 IEEE International Circuits and Systems Symposium (ICSyS), Kuantan, Malaysia, pp. 1-4. https://doi.org/10.1109/ICSyS47076.2019.8982441

[9] Irfan, M.I., Hasan, W.Z.W., Harun, H.R., Mustafa, M.F., Japar, A.F., Kadir, K. (2022). Development of loose fruit collector system for palm oil. In 2022 IEEE 8th International Conference on Smart Instrumentation, Measurement and Applications (ICSIMA), Melaka, Malaysia, pp. 196-199. https://doi.org/10.1109/ICSIMA55652.2022.9928934

[10] Yusoff, M.Z.M., Zamri, A., Abd Kadir, M.Z.A., Hassan, W.W., Azis, N. (2020). Development of integrated loose fruit collector machine for oil palm plantations. Bulletin of Electrical Engineering and Informatics, 9(2): 500-506. https://doi.org/10.11591/eei.v9i2.2087

[11] Nubun, C.A., Lihan, S. (2022). Potential pathogenic bacteria in loose oil palm fruit (LOPF) from smallholdings in Serian, Sarawak. Asian Journal of Medicine and Biomedicine, 6(S1): 143-146. https://doi.org/10.37231/ajmb.2022.6.S1.566

[12] Forest, I.L., Wan Yusuf, S.M., Saili, A.R., Mahdian, S., Rizieq, R. (2024). Discover the challenges faced by the smallholder in the loose fruit collection process. In IOP Conference Series: Earth and Environmental Science, 1397(1): 012034. https://doi.org/10.1088/1755-1315/1397/1/012034

[13] Warni, E., Achmad, A., Syahsir, A.R.R. (2025). Harnessing YOLO for loose fruits detection: Boosting productivity in palm oil plantations. In 2025 International Conference on Advancement in Data Science, E-learning and Information System (ICADEIS), Bandung,

Indonesia, pp. 1-6. https://doi.org/10.1109/ICADEIS65852.2025.10933352
[14] Opara, I.K., Opara, U.L., Okolie, J.A., Fawole, O.A. (2024). Machine learning application in horticulture and prospects for predicting fresh produce losses and waste: A review. Plants, 13(9): 1200. https://doi.org/10.3390/plants13091200
[15] Khalid, M.R., Shuib, A.R., Kamarudin, N. (2021). Mechanising oil palm loose fruits collection-A review. Journal of Oil Palm Research, 33(1): 1-11. https://doi.org/10.21894/jopr.2020.0069
[16] Teo, Y.X., Chan, Y.S., Gouwanda, D., Gopalai, A.A., Nurzaman, S.G., Thannirmalai, S. (2021). Quantification of muscles activations and joints range of motions during oil palm fresh fruit bunch harvesting and loose fruit collection. Scientific Reports, 11(1): 15020. https://doi.org/10.1038/s41598-021-94268-4
[17] Radzi, M.K.F.M., Khalid, M.R.M., Azaman, M.I.H., Mohamed, A., Thadeus, D.J., Bakri, M.A.M. (2023). The initiative to further enhance technology adoption in the Malaysian oil palm industry. Advances in Agricultural and Food Research Journal, 4(2). https://doi.org/10.36877/aafrj.a0000413
[18] Nor, A.M., Saad, M.S., Baharudin, M.E., Zakaria, M.Z. (2024). Addressing labour ergonomics through automation in oil palm plantation activities: A necessity for sustainable agriculture. Malaysian Journal of Ergonomics (MJEr), 6: 1-10. https://doi.org/10.58915/mjer.v6.2024.1288
[19] Chaithra, N., Jha, J., Sayal, A., Gupta, V., Gupta, A. (2023). A paradigm shift towards computer vision. In 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT), Dehradun, India, pp. 54-58. https://doi.org/10.1109/DICCT56244.2023.10110300
[20] Xiang, A.J., Huddin, A.B., Ibrahim, M.F., Hashim, F.H. (2021). An oil palm loose fruits image detection system using Faster R-CNN and Jetson TX2. In 2021 International Conference on Electrical Engineering and Informatics (ICEEI), Kuala Terengganu, Malaysia, pp. 1-6. https://doi.org/10.1109/ICEEI52609.2021.9611111
[21] Ghazal, S., Munir, A., Qureshi, W.S. (2024). Computer vision in smart agriculture and precision farming: Techniques and applications. Artificial Intelligence in Agriculture, 13: 64-83. https://doi.org/10.1016/j.aiia.2024.06.004
[22] Tian, H., Wang, T., Liu, Y., Qiao, X., Li, Y. (2020). Computer vision technology in agricultural automation—A review. Information Processing in Agriculture, 7(1): 1-19. https://doi.org/10.1016/j.inpa.2019.09.006
[23] Laad, M., Maurya, R., Saiyed, N. (2024). Unveiling the vision: A comprehensive review of computer vision in AI and ML. In 2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS), Chennai, India, pp. 1-6. https://doi.org/10.1109/ADICS58448.2024.10533631
[24] Shi, X., Wang, S., Zhang, B., Ding, X., Qi, P., Qu, H., Li, N., Wu, J., Yang, H. (2025). Advances in object detection and localization techniques for fruit harvesting robots. Agronomy, 15(1): 145. https://doi.org/10.3390/agronomy15010145
[25] Septiarini, A., Hamdani, H., Hatta, H.R., Anwar, K. (2020). Automatic image segmentation of oil palm fruits by applying the contour-based approach. Scientia Horticulturae, 261: 108939. https://doi.org/10.1016/j.scienta.2019.108939
[26] Bellou, E., Pisica, I., Banitsas, K. (2024). Aerial inspection of high-voltage power lines using YOLOv8 real-time object detector. Energies, 17(11): 2535. https://doi.org/10.3390/en17112535
[27] Han, T., Dong, Q., Wang, X., Sun, L. (2024). BED-YOLO: An enhanced YOLOv8 for high-precision real-time bearing defect detection. IEEE Transactions on Instrumentation and Measurement, 73: 1-13. https://doi.org/10.1109/TIM.2024.3472791
[28] Bilous, N., Malko, V., Frohme, M., Nechyporenko, A. (2024). Comparison of CNN-based architectures for detection of different object classes. AI, 5(4): 2300-2320. https://doi.org/10.3390/ai5040113
[29] Li, M., Zhang, Z., Lei, L., Wang, X., Guo, X. (2020). Agricultural greenhouses detection in high-resolution satellite images based on convolutional neural networks: Comparison of faster R-CNN, YOLO v3 and SSD. Sensors, 20(17): 4938. https://doi.org/10.3390/s20174938
[30] Vajgl, M., Hurtik, P., Nejezchleba, T. (2022). Dist-yolo: Fast object detection with distance estimation. Applied Sciences, 12(3): 1354. https://doi.org/10.3390/app12031354
[31] Shu, F., Lesur, P., Xie, Y., Pagani, A., Stricker, D. (2021). SLAM in the field: An evaluation of monocular mapping and localization on challenging dynamic agricultural environment. In 2021 IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, pp. 1760-1770. https://doi.org/10.1109/WACV48630.2021.00180
[32] Lin, C., Sun, G., Wu, D., Xie, C. (2023). Vehicle detection and tracking with roadside LiDAR using improved ResNet18 and the Hungarian algorithm. Sensors, 23(19): 8143. https://doi.org/10.3390/s23198143
[33] Li, H., Li, L., Lv, X., Zhao, R. (2025). Intelligent monocular visual dynamic detection method for safe distance of hot work operation. In 2025 IEEE 5th International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, pp. 374-380. https://doi.org/10.1109/ICPECA63937.2025.10928789