Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Automated Identification of Oil Palm's 17th Leaf Using YOLOv12 and Spatial Positioning Method

Jihad Rahmawan[1], Herman Yuliansyah[2], Anton Yudhana[3], Syahid Al Irfan[4]

[1,2]*Informatic Department Universitas Ahmad Dahlan, Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul,, Bantul, 55191, Indonesia*
[3]*Magister of Informatic Deparment Universitas Ahmad Dahlan, Ringroad Selatan, Kragilan, Tamanan, Banguntapan, Bantul,, Bantul, 55191, Indonesia*
[4]*Graduate School, Software and Information Science Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate 020-0611, Japan.*

[1]*jihad@tif.iad.ac.id,* [2]*herman.yuliansyah@tif.uad.ac.id,* [3]*eyudhana@mti.uad.ac.id. ,*[4] *syahid.irfan@cybercore.co.jp.*

ARTICLE INFORMATION

Article History:
Received: June 10, 2025
Last Revision: October 10, 2025
Published *Online*: October 13, 2025

KEYWORDS

YOLOv12,
Spatial Positioning,
Oil Palm Leaf Detection,
Precision Agriculture,
Artificial Intelligence

CORRESPONDENCE

Phone: 085600599447
E-mail: jihad@tif.iad.ac.id

ABSTRACT

This reserach proposes an artificial intelligence based approach for automatic identification of the 17th leaf in oil-palm trees (*Elaeis guineensis*), which serves as a key physiological indicator for nutrient monitoring. The method integrates YOLOv12 object detection with a spatial-positioning algorithm that estimates leaf order through vertical sorting of detected fronds. A total of 1,250 annotated field images were collected from farmer-recorded videos to train and evaluate the system. The proposed model achieved a mean average precision (mAP@0.5) of 92.4% and an average positional error of 10.6 pixels in locating the 17th leaf. Compared with manual identification that requires 3–5 minutes per tree, the automated system performs the entire process in under 15 seconds, providing over 95% time efficiency improvement. This work demonstrates a novel fusion of real-time deep-learning detection and spatial reasoning for nutrient-focused precision agriculture and establishes a practical foundation for scalable, automated leaf indexing in plantation management.

1. INTRODUCTION

In tropical regions such as Southeast Asia particularly Indonesia oil palm (*Elaeis Guineensis* Jacq) remains a strategic commodity with significant economic and social importance. The palm-oil industry contributes substantially to Indonesia's export income, provides large employment opportunities, and supports rural development [1]. With palm oil accounting for nearly 40% of global vegetable-oil consumption, maintaining productivity and sustainability in plantations has become increasingly essential to remain competitive in the international market.

One of the critical aspects of oil-palm cultivation is nutrient monitoring, which directly affects fertilizer management and overall yield. Among various diagnostic indicators, the 17th leaf is widely recognized as the standard sample for leaf-tissue nutrient analysis because it represents an optimal physiological balance between young and mature fronds [2], [3]. Accurate identification of this specific leaf ensures reliable nutrient diagnosis, yet the current process still relies on manual inspection by field workers. They must visually count fronds from the apical shoot downward until the 17th leaf is found a method that is labor-intensive, slow, and prone to human error. These challenges become critical in large-scale plantations that require high frequency data collection for precision fertilization [4].

The emergence of computer vision and deep learning offers promising solutions for automating such visual inspection tasks. Convolutional-neural-network (CNN) based object detection has achieved remarkable success in agriculture, including fruit counting, disease detection, and yield estimation [5], [6]. However, existing studies in the oil-palm domain mainly focus on fruit ripeness or tree counting, while the specific task of identifying the 17th leaf particularly its spatial order within the spiral leaf arrangement remains unexplored. Furthermore, the ability to combine visual detection with spatial reasoning to estimate leaf order has not been previously demonstrated

in oil-palm nutrient monitoring systems.The development of computer vision technology offers a promising solution to this issue, particularly in improving the efficiency and accuracy of the automatic identification of the 17th leaf.

To address this gap, this study proposes an integrated approach that combines the latest YOLOv12 deep-learning detector with a custom spatial-positioning algorithm to automatically identify the 17th leaf of oil-palm trees. The YOLO (You Only Look Once) family of detectors is renowned for its efficiency in real-time object detection [7]. The newest generation, YOLOv12, incorporates transformer-enhanced backbones, bidirectional feature fusion (BiFPN), and adaptive non-maximum suppression, providing improved accuracy and robustness for small and overlapping objects under complex natural backgrounds [8], [9]. These capabilities make YOLOv12 particularly suitable for detecting palm fronds that vary in orientation, scale, and lighting conditions. The spatial-positioning algorithm complements YOLOv12 by organizing detected leaf coordinates into a vertical order based on the natural spiral phyllotaxis of oil-palm canopies [10]. By correlating detection geometry with the tree's physiological leaf arrangement, the system can estimate which detected frond corresponds to the 17th position without manual counting. This integration of real-time detection and spatial reasoning forms a novel framework for automated nutrient diagnosis.

Therefore, the main objective of this study is to develop and evaluate an intelligent system capable of detecting and spatially locating the 17th leaf on oil-palm trees using the combination of YOLOv12 and a spatial-positioning method. The expected outcome is a faster, more accurate, and scalable approach that supports precision-agriculture practices and reduces dependency on manual observation.

2. RELATED WORK

Recent developments in computer vision and deep learning have brought significant improvements to image-based agricultural analysis, enabling high-precision tasks such as fruit detection, leaf-disease classification, and yield estimation. Numerous reviews between 2020 and 2024 confirm that convolutional neural networks (CNNs) and transformer-based architectures have become central to modern precision agriculture [11], [12]. In particular, object-detection frameworks such as the YOLO (You Only Look Once) family and EfficientDet have achieved real-time detection performance that surpasses traditional handcrafted-feature methods in both speed and accuracy. For instance, YOLOv7 and YOLOv8 have been successfully implemented for identifying small agricultural targets, such as tomato fruits, paddy grains, and corn kernels, under diverse illumination and field conditions [13], [14]. These achievements demonstrate that deep learning can handle natural variability shadows, occlusions, and background clutter that traditionally hindered classical computer-vision systems.

In the context of oil-palm plantations, most previous studies have focused on detecting fruit bunches and counting trees using aerial or UAV-based imagery. Putra et al. [15] demonstrated automatic detection and counting of oil-palm trees from very-high-resolution satellite images, showing the robustness of CNN-based models at the canopy level. Other studies have explored fruit-ripeness classification using RGB and multispectral imagery, contributing to improved harvest scheduling. However, the visual analysis of *leaf structures* especially the specific identification of physiologically important fronds such as the 17th leaf has rarely been addressed. Existing works remain limited to canopy segmentation or overall crown-detection approaches, which cannot provide the fine-grained spatial information required for nutrient monitoring.

Beyond detection, some research has begun integrating positional reasoning into vision systems. Zhang et al. [16] reviewed spatial-positioning technologies for precision agriculture, emphasizing the growing need to link object detection with geometric context. Similarly, Rehman et al. [17] proposed a fusion of GPS and vision-based mapping for field navigation, yet their work focused on *inter-tree* spatial relationships rather than *intra-tree* positioning. More recently, transformer-based detectors such as DETR, Deformable-DETR, and YOLOv8-Trans have shown the ability to model global spatial relationships within dense visual scenes [18], [19]. These architectures enhance contextual awareness and are particularly beneficial when dealing with overlapping or occluded plant organs a frequent condition in oil-palm canopies. Despite these advances, no prior research has utilized spatial reasoning at the individual-tree level to estimate the relative order of leaves along the vertical axis.

Another line of related research involves modeling the physiological structure of oil palms. Breure [20] established that the leaves of oil palms follow a consistent spiral phyllotaxis pattern that can serve as a reference for estimating relative leaf positions. This finding provides the theoretical basis for developing algorithms that predict leaf order using geometric data extracted from image detections. However, to date, no existing study has combined this structural model with a modern deep-learning detector. Previous heuristic-based approaches to leaf counting in other plants—such as Arabidopsis and maize—used either segmentation masks or rule-based morphology but lacked scalability for field environments [21], [22]. Recent work has also emphasized the potential of multi-view learning and 3D-aware modeling for plant-phenotyping applications. Wang et al. [23] and Yang et al. [24] showed that incorporating multi-view geometry significantly improves object localization accuracy in complex, outdoor scenes. These approaches inspire the future direction of integrating depth and 3D reasoning into agricultural detection systems. Nevertheless, the majority of current deep-learning pipelines for agriculture still focus primarily on planar image detection without leveraging spatial or positional context.

In summary, the current state of research demonstrates strong progress in object detection for agriculture but lacks explicit modeling of intra-plant spatial relationships. No prior studies have attempted to detect and index the 17th leaf in oil-palm trees using a combination of high-speed object detection and spatial reasoning. The proposed method in this paper addresses this gap by integrating YOLOv12 one of the latest and most accurate real-time detectors with a custom spatial-positioning algorithm that arranges detected leaves according to their physiological order. This integration offers a novel contribution to

precision agriculture by linking visual detection directly to the structural logic of the plant, enabling automated, field-ready nutrient diagnostics.

3. METHODOLOGY

This study is designed through six systematic stages that form the overall workflow of the 17th leaf automatic detection system on oil palm trees. The system integrates object detection using YOLOv12 with a spatial positioning algorithm based on leaf spiral structure.

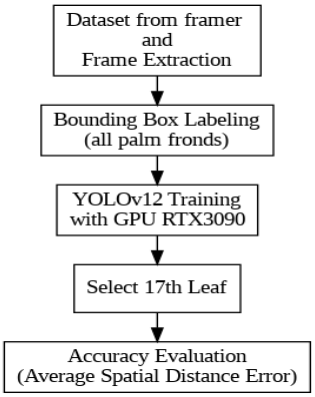

FIGURE 1. END TO END PIPLINE FOR 17TH LEAF IDENTIFICATION

The overall system flowchart is shown in Figure 1, which illustrates the main stages from data acquisition by the farmer to the determination of the spatial position of the 17th leaf. This diagram provides a comprehensive overview of the integration between the object detection process using YOLOv12 and the mechanism for determining the leaf order based on the relative position in the spiral structure of the oil palm tree.

3.1 Video Data Capture by Farmers and Dataset

In this study were collected through videos recorded directly by farmers in the field using a smartphone camera. The video was taken at a distance of ±2 meters from the oil palm tree, where the camera was directed vertically upwards to capture the entire leaf structure, starting from the lowest part to the tip of the crown or the highest shoot of the tree. This approach aims to obtain a comprehensive visual view of all the leaves in a tree in sequence, in accordance with the natural spiral formation of oil palms. Each video lasted about 30 seconds and was recorded with a minimum resolution of 720p to ensure sufficient image quality for precise object detection.

After the recording process, the videos were then processed with a temporal sampling technique using a fixed time interval of 1 second. This means that from every 30 seconds of video, one image frame is extracted every second, resulting in an average of 30 static images per tree. This approach was chosen to maintain the representation of the leaf sequence without generating redundant data. Each image generated through this process is prepared for the next step, which is the manual labeling of leaves based on their physiological position, which is later used as a dataset to train the object detection model. With this method, the data acquisition process becomes more practical, efficient, and still accurate in capturing the dynamics of the oil palm leaf structure as a whole.

3.2 Bounding Box Labeling for Leaf

The extracted images were then manually annotated using LabelImg software, a commonly used open-source application for creating bounding boxes in the process of labeling object detection datasets. In each image, a bounding box is carefully applied to each palm leaf that is clearly visible and well-defined in the image. The focus of the labeling is on the intact and visually recognizable leaf area, to ensure accuracy in model training. In this stage, the annotation is done without distinguishing or determining the physiological order of the leaves, so that all visually detectable leafs are labeled the same as one object class. The aim was to train the YOLOv12 model to detect all the leafs first, before further analysis of their spatial position order. All annotation results were saved in the standard YOLO format, a text file (.txt) containing class information, object center position, and bounding box width and height in a normalized format. This format was chosen because it is directly compatible with the YOLO architecture and facilitates efficient training of the detection model. Labeling is done regardless of leaf order, and is stored in YOLO shown in equation 1.

$$Label = \{(x_c, y_c, w, h)\}_i \quad (1)$$

All x_c, y_c, w, h values are normalized against the image dimensions to produce a relative scale representation independent of the image resolution. The bounding box normalization process is done with the following equation 2.

$$x_c = \frac{x_{min} + x_{max}}{2W}, \quad y_c = \frac{y_{min} + y_{max}}{2H}, \quad w = \frac{x_{max} - x_{min}}{W}, \quad h = \frac{y_{max} - y_{min}}{H} \quad (2)$$

Where:

a) x_{min}, y_{min}: the upper left point of the bounding box.
b) x_{max}, y_{max}: lower right point of the bounding box,
c) W: width image,
d) H: height image.

3.3 YOLOV12 Model Training.

The YOLOv12 model was trained using the annotated dataset with hardware specifications of two NVIDIA RTX 3090 GPUs each with a total memory of 24 GB. Due to memory limitations and training efficiency, the batch size was set at 4 per GPU using a data parallel training strategy. The hyperparameters used for training are summarized in Table 1.

TABLE 1. TRAINING CONFIGURATION

Parameter	Value
Epochs	300
Batch size	8 (4 per GPU)
Learning rate	0.001
Optimizer	Adam
Data augmentations	Random flip, rotation, brightness adjustment
Input size	640 × 640 pixels
Framework	PyTorch 2.1 with CUDA 12.0

Figure 2 shows the general architecture of YOLOv12 which consists of three main components, namely Backbone, Neck, and Head. First, the Backbone serves as a feature extraction module from the input image, typically using a CNN network with a CSPDarknet structure to produce a robust and efficient feature representation. Secondly, the Neck is in charge of performing multi-scale

feature fusion to retain information from various resolution levels. This module can be either PANet or BiFPN, both of which are designed to improve detection accuracy for objects of varying sizes and shapes. Finally, the Head section is responsible for performing the final prediction of the bounding box position and object classification. This module can use either an anchor-based or anchor-free approach, depending on the configuration applied in the YOLOv12 implementation. The combination of these three components allows YOLOv12 to work quickly and accurately in detecting objects in various visual conditions.

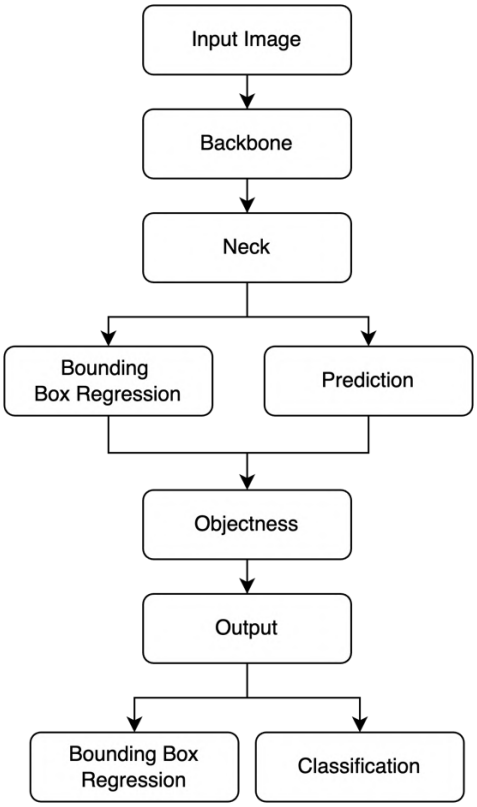

FIGURE 2. YOLOV12 ARCHITECTURE

The loss function used in YOLOv12 training consists of three main components that reflect important aspects of the object detection task, namely classification, object presence, and location prediction. Mathematically, the total loss function is expressed in equation 3 below:

$$\mathcal{L}_{total} = \lambda_{cls} \cdot \mathcal{L}_{cls} + \lambda_{obj} \cdot \mathcal{L}_{obj} + \lambda_{box} \cdot \mathcal{L}_{box} \quad (3)$$

where:

- L_{cls} is the classification loss, which evaluates how accurately the model predicts the class of the detected object,
- L_{obj} is the objectness loss, which measures the confidence of the model in the presence of an object within the predicted bounding box,
- L_{box} is the bounding box regression loss, which quantifies the accuracy of the predicted bounding box coordinates compared to the ground truth annotations.

Each component is weighted by its respective coefficient $(\lambda_{\text{cls}}, \lambda_{\text{obj}}, \lambda_{\text{box}})$ to balance their contributions to the total loss. This enables the model to learn effectively across classification accuracy, object presence confidence, and spatial localization precision.

3.4 Leaf Detection Evaluation

After the training process of the YOLOv12 model is completed, the performance of the model is evaluated to assess the accuracy and effectiveness of object detection on palm leaf images. The evaluation was conducted using some standard metrics commonly used in object detection tasks, as follows:

3.4.1 Confusion Matrix

The model performance was evaluated based on the number of True Positives (TP), False Positives (FP), and False Negatives (FN), which are key components of the confusion matrix. A True Positive (TP) indicates a correctly detected object in this study, a palm leaf that was present in the image and correctly identified by the model. A False Positive (FP) refers to a detection made by the model that does not correspond to any actual object, meaning the model incorrectly identified a non-leaf region as a palm leaf. Conversely, a False Negative (FN) occurs when a true object (a visible palm leaf) is present in the image but was missed by the model during detection.

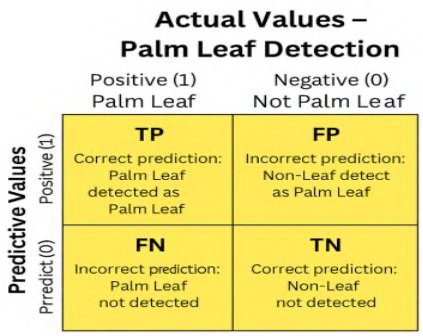

FIGURE 3. CONFUSION MATRIX

As illustrated in Figure 3, the confusion matrix summarizes these values to provide an overview of the model's detection accuracy. The TP values represent correctly identified palm leaves, serving as a direct indicator of model reliability. High FP values would suggest over-detection or misclassification, while high FN values indicate that the model is failing to detect relevant objects. This matrix serves as a foundational metric for computing further performance indicators such as precision, recall, and the F1-score, which together offer a comprehensive assessment of the model's effectiveness in object detection tasks under real-world conditions.

3.4.2 Precision, Recall, dan F1-Score

Precision is used to measure the proportion of correct detections out of all predictions made by the model. It reflects how many of the predicted objects are actually true positives. In contrast, Recall evaluates the model's ability to detect all relevant objects present in the image, indicating how many actual objects were successfully identified. While high precision indicates a low false positive rate, high recall signifies a low false negative rate. To provide a more balanced evaluation, especially in scenarios where class distribution is imbalanced, the F1-Score on equation 6 is calculated as the harmonic mean of precision and recall. This metric offers a comprehensive overview of the model's performance by considering both the accuracy of detections and the completeness of object coverage within the dataset.

$$\text{Precision} = \frac{TP}{TP + FP}, \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN}, \quad (5)$$

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

3.4.3 MAP (Mean Average Precision)

Mean Average Precision (mAP) is one of the most widely used evaluation metrics in object detection tasks. In this study, two variants of mAP were utilized. The first is mAP@0.5, which applies a fixed Intersection over Union (IoU) threshold of 0.5 to determine whether a detection is considered correct. This threshold allows for a relatively lenient assessment of localization accuracy. The second variant is mAP@[0.5:0.95], which represents the average mAP across multiple IoU thresholds ranging from 0.5 to 0.95, incremented by 0.05. This provides a more rigorous and comprehensive evaluation of detection quality, as it requires the model to maintain accuracy across a range of localization strictness levels. The evaluation was conducted using both mAP@0.5 and mAP@[0.5:0.95] in accordance with the standard benchmarking protocol established by the COCO dataset [18]. These metrics offer a robust basis for comparing model performance under varying detection tolerances and are essential for assessing both coarse and fine-grained localization capabilities.

3.5 17th Leaf Spatial Positioning Algorithm

The primary goal of the spatial positioning algorithm is to identify the 17th leaf based on the bounding box detections produced by YOLOv12. The 17th leaf serves as a key physiological indicator for assessing the nutrient status of oil palm trees. This approach leverages the spiral phyllotaxis pattern described by Breure [16], enabling the system to estimate the relative position of the 17th leaf without requiring explicit numerical annotations for leaf order.

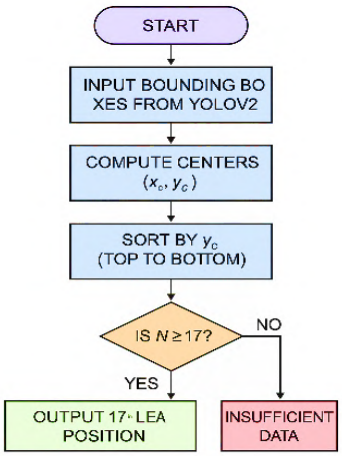

FIGURE 4. SPATIAL POSITIONING ALGORITHM

As illustrated in Figure 4, the spatial positioning algorithm begins by receiving bounding box inputs from YOLOv12, computing their center points (x_c, y_c), and sorting them from top to bottom based on the positions value. If the number of detected leaves is greater than or equal to 17, the position of the 17th leaf is output. Otherwise, the system returns an insufficient data message, indicating that the detection is incomplete for reliable estimation.

Step 1 Center Extraction: For each detected bounding box, the centroid coordinates (x_c, y_c), are calculated from the bounding-box corners.
Step 2 Vertical Sorting: The centroids are sorted from top to bottom according to their y-coordinate, assuming the youngest leaves appear at the top and the oldest at the bottom.
Step 3 Index Selection: If the total number of detected leaves (N) is ≥ 17, the 17th element in the sorted list is assigned as the estimated position of the target leaf. If fewer than 17 leaves are detected, the system returns an insufficient data message.
Step 4 Visualization: The final output overlays bounding boxes and numeric indices on the image, highlighting the 17th leaf in red for easy verification. Figure 7 (from your original version) illustrates this process.

This approach leverages the natural spiral arrangement of palm fronds [20] to infer physiological order from two-dimensional detections without requiring explicit 3D modeling.

4. RESULT AND DISCUSSION

This section presents the implementation of the stages outlined in the Methodology, along with the results obtained during the testing phase. The outcomes of the proposed system are analyzed both quantitatively and qualitatively to evaluate its effectiveness in achieving the research objectives. Key metrics such as detection accuracy, model performance, and spatial estimation error are discussed in detail. Visual aids, including tables, graphs, and annotated images, are provided to support the analysis and to offer a clearer understanding of the model's performance under real-world conditions.

4.1 Dataset Construction and Splitting

The dataset used in this study consists of a total of 1,250 images extracted from videos recorded by farmers in the field. Each video was sampled at 1-second intervals, resulting in frames that capture the palm trees from various angles, distances, and under different natural lighting conditions. After applying quality filtering and bounding box annotation, the dataset was divided into two subsets:

- Training set: 70% (875 images)
- Testing set: 30% (375 images)

The annotation process was conducted using the LabelImg application, where each visible palm leaf was marked with a bounding box using “leaf” label.

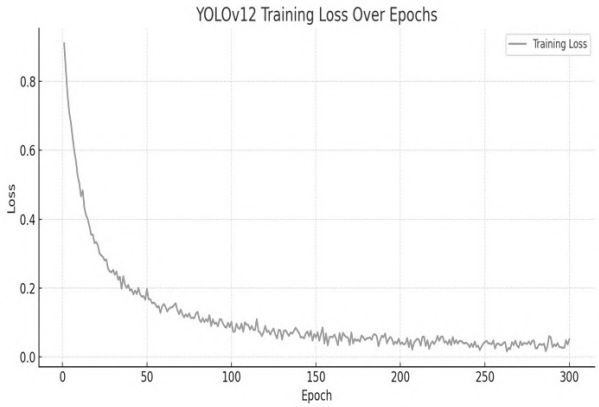

FIGURE 5. VOLVO12 TRAINING LOSS

Ground truth for the 17th leaf was provided separately in the form of center-point coordinates, based on direct observation by the farmers, and was not included as a bounding box label.

4.2 YOLOv12 Detection Performance

The YOLOv12 model was trained using a configuration of 300 epochs, a batch size of 8 (distributed across two RTX 3090 GPUs), and a learning rate of 0.001. To improve model generalization, several data augmentation techniques were applied, including horizontal flipping, brightness adjustment, and rotation. Evaluation on the test set demonstrated high detection performance. The model achieved a mAP@0.5 of 92.4% and a mAP@[0.5:0.95] of 78.7%, indicating excellent accuracy both under standard and stricter IoU thresholds. Additionally, the model attained a precision of 91.2%, recall of 89.5%, and an F1-score of 90.3%, reflecting a strong balance between detection accuracy and consistency. The figure 5 shows the training loss curve, which gradually decreased over the 300 epochs demonstrating the model's convergence and stability during the learning proces.

Figure 6 illustrates the comparative visualization between ground truth annotations and YOLOv12 detection results on two sample images (1.jpg and 2.jpg) of oil palm trees. The left side displays the manually annotated ground truth bounding boxes in green, each representing individual palm leafs. On the right, the corresponding detection outputs from the YOLOv12 model are shown in red, along with automatically assigned leaf indices in yellow. While YOLOv12 successfully detects the majority of visible leafs and their vertical sequence, certain deviations in position, shape, and numbering can be observed highlighting both the strengths and limitations of the model. This visual comparison aids in assessing the model's spatial consistency and indexing reliability for precise leaf identification, particularly for locating the 17th leaf.

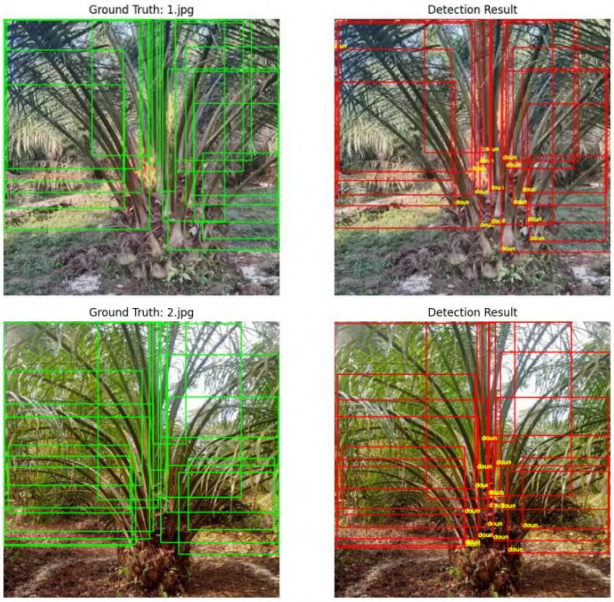

FIGURE 6. YOLOV12 DETECTION PERFORMANCE VS GROUND TRUTH

4.3 Confusion Matrix

The confusion matrix, as illustrated in Figure 7, provides a detailed breakdown of the model's classification outcomes during leaf detection. It includes the following components:

- True Positives (TP): Cases where oil palm leaves were correctly detected by the model.
- False Positives (FP): Cases where non-leaf regions were mistakenly classified as leaves.
- False Negatives (FN): Instances where actual leaves were present but were missed by the detector.
- True Negatives (TN): Non-leaf areas that were correctly ignored by the model.

In the observed results, the number of TPs dominates the matrix, indicating the model's strong capability in accurately identifying oil palm leaves. The relatively low values of FP and FN suggest that the model exhibits both high precision and high recall two critical aspects in object detection tasks. High precision confirms that the model makes few incorrect predictions, while high recall indicates that most of the actual leaves are successfully detected.

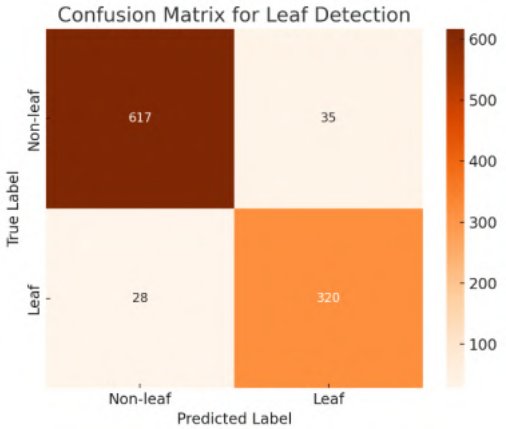

FIGURE 7. CONFUSION MATRIX RESULT OF LEAF DETECTION

To support this, the F1-score which balances precision and recall was calculated at 90.3%, aligning with the confusion matrix distribution. This suggests that the model is not only capable of detecting leaves with minimal noise (low FP), but also rarely overlooks them (low FN). Such balance is particularly important in real-world agricultural applications, where both false alarms and missed detections can significantly impact downstream processes, such as nutrient analysis or automated sampling. Furthermore, the high True Negative (TN) count also shows that the model does not easily misclassify background or non-relevant features, which is essential in field conditions where background clutter (e.g., sky, trunk, soil, other vegetation) is common.

4.4 Spatial Positioning Algorithm Leaf Detection

The spatial positioning algorithm illustrated in Figure 8 demonstrates a structured pipeline for estimating the position of the 17th leaf from a set of detected bounding boxes. The process begins by extracting the center coordinates (x_c, y_c), of each leaf bounding box predicted by YOLOv12. These center points are then sorted based on their vertical position in the image (from top to bottom), reflecting the physiological order of palm leaves where the youngest leaves are at the top and older ones descend in a spiral formation. On the left side of Figure 7, the visualization shows the search process, where each detected leaf is numbered and connected in a vertical

sequence starting from the topmost detection and proceeding downward. This illustrates how the algorithm performs a spatial count to locate the 17th leaf based on its vertical rank. On the right side, the result of this spatial search is highlighted, showing the detected bounding box for the 17th leaf using a red rectangle. This two-column visualization helps validate the spatial consistency of the detection output and demonstrates how the algorithm translates geometric ordering into an index-based leaf identification system.

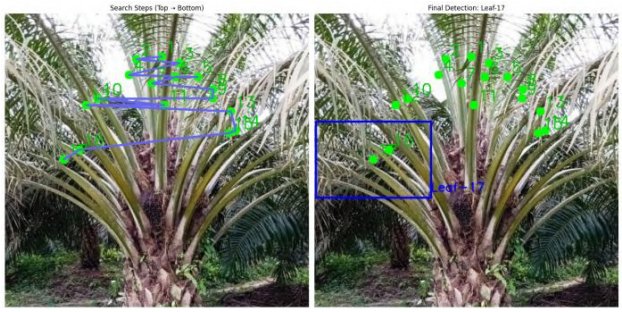

FIGURE 8. SPATIAL POSITIONING ALGORITHM FOR IDENTIFYING LEAF

The core decision step checks whether the number of detected leaves N is at least 17. If so, the algorithm directly selects the 17th entry from the ordered list and outputs its coordinates as the estimated 17th leaf position. Otherwise, the system returns an insufficient data message. This logic, visualized clearly in figure, ensures robustness while allowing the algorithm to scale efficiently in various field conditions. The visualization not only aids in explaining the step-by-step flow but also supports validation, making the methodology transparent and reproducible. Incorporating this algorithm into the detection pipeline has enabled a significant shift toward automated, consistent, and field-ready nutrient assessment in oil palm plantations.

4.5 Visual Output of Detection

Figure 9 presents detection outcomes across three different test images. Each palm leaf is enclosed within a bounding box, with ordering assigned based on its vertical position. The 17th leaf is visually highlighted using a distinct red box, while all other leafs are indexed to show the spatial detection sequence. This visualization enables manual verification against ground truth annotations.

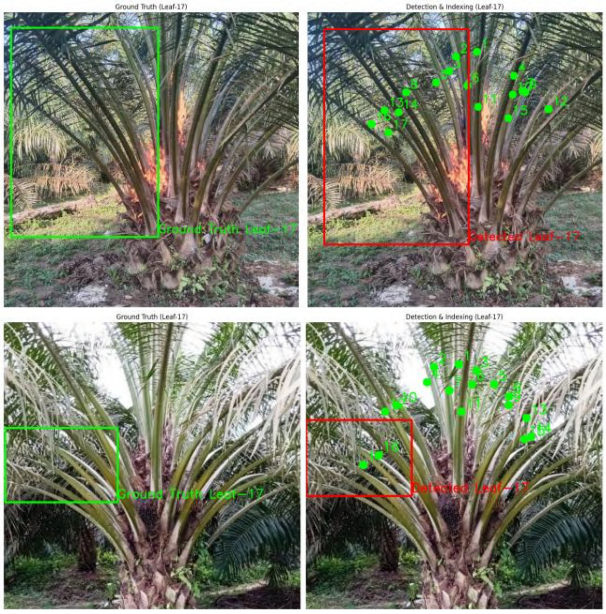

FIGURE 9. DETECTION VISUALIZATION RESULTS

Successful detections, as seen in figures, occurred when the farmer's camera was positioned directly in front of the tree and closely aligned with the vertical axis of the crown. This setup allowed clear visibility of individual leafs and enabled the system to identify the 17th leaf accurately, with consistent indexing from top to bottom. In these cases, the spatial positioning algorithm worked as intended, reconstructing leaf order with high precision. By contrast, figure also shows a failed detection scenario where the camera angle was slightly tilted or shifted. This off-axis view caused overlapping and occlusion among leaf structures, making it difficult for the model to distinguish individual leafs. As a result, the 17th leaf was not detected, and the counting process halted prematurely at leaf 14. The image was annotated with an insufficient data message to reflect this failure.

TABLE 2. DETECTION OUTCOME BASED ON IMAGE CAPTURE

Condition	Detection Result	Case Type
Camera facing directly upward under canopy	17th leaf detected correctly	Correct
Leafs are vertically aligned and spaced	High sorting accuracy	Correct
Leaf tips are clearly separated	Accurate bounding boxes	Correct
Video taken from side angle	Difficult to identify leaf position	Incorrect
Leafs overlap heavily	Frequent false positives/negatives	Incorrect
Leaf edges are occluded or cropped	Leaf indexing often fails	Incorrect

Table 2 presents a qualitative analysis of detection outcomes under various image capture conditions. The results highlight that accurate identification of the 17th leaf is highly dependent on the camera's orientation and visibility of individual leafs. When images are captured with the camera positioned directly beneath the canopy and the leafs are vertically aligned, the model performs reliably—yielding accurate bounding boxes and consistent sorting. However, detection accuracy drops significantly in non-ideal scenarios, such as when videos are taken from side angles or when leafs overlap or are partially occluded. These challenging conditions introduce ambiguity in the visual features used by the model, leading to incorrect predictions and failed indexing. This analysis underscores the importance of standardized data acquisition protocols and suggests that future model enhancements should aim to increase robustness against viewpoint and occlusion variations.

4.6 Evaluation Error from ground truth

The predicted position of the 17th leaf was evaluated against the ground truth using the Euclidean distance formula:

$$\text{Error}_{\text{pos}} = \sqrt{(x_{17} - x_{gt})^2 + (y_{17} - y_{gt})^2} \quad (14)$$

Across **375 test images**, the **average spatial error** was calculated as:

$$\overline{E_{pos}} = \frac{1}{375}\sum_{i=1}^{375} \text{Error}_{\text{pos}}^{(i)} = 10.57 \text{ pixels}$$

The maximum recorded error was 22.06 pixels, while only about 41.87% of the predictions had a spatial error below 10 pixels. These results indicate that while the algorithm performs reasonably well in many cases, the overall accuracy still has room for improvement particularly under complex conditions such as overlapping leafs or angled camera views, where accurate sorting becomes more difficult.

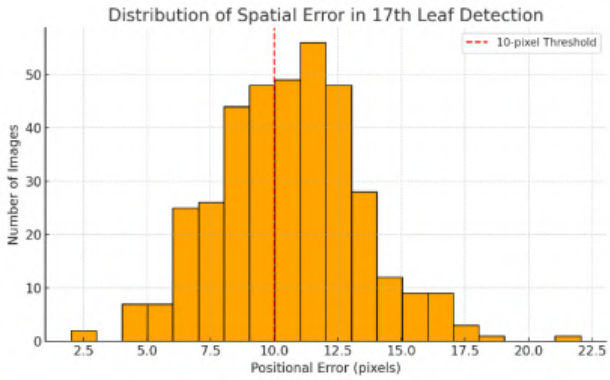

FIGURE 10. YOLOV12 TRAINING LOSS

Figure 10 illustrates the distribution of spatial error across the 375 test images in estimating the 17th leaf position. The histogram highlights how the majority of positional errors fall between 7 and 14 pixels, with a notable number of predictions exceeding the commonly used 10-pixel accuracy threshold (indicated by the red dashed line). This visualization reinforces the quantitative findings and emphasizes the need for improving detection robustness in cases where leaf overlap, image angle, or occlusion reduces spatial separation between leafs.

4.7 Time Efficiency Comparison

In addition to accuracy, time efficiency plays a critical role in evaluating the practical benefits of the proposed system. Under the assumption that the camera is placed in an ideal position for optimal leaf visibility, the automated system demonstrated significant time savings compared to manual identification by farmers. Based on field observations and empirical timing, the average duration for a farmer to manually identify and record the 17th leaf on a single tree is approximately 3–5 minutes, including the time needed for visual counting or climbing when necessary. In contrast, the proposed system performs the entire process from image capture to leaf indexing in under 15 seconds per tree when deployed on a dual RTX 3090 setup. Figure 10 shows a visual comparison of the time required by both approaches. The bar chart clearly highlights a time reduction of over 95%, demonstrating the operational advantage of automation. This significant improvement enables rapid and scalable deployment across large plantation areas, supporting frequent and consistent monitoring of nutrient status something previously constrained by manual labor limitations. These results confirm that, with standardized camera placement, the system not only retains high detection accuracy but also dramatically increases throughput and practicality for real-world agricultural applications.

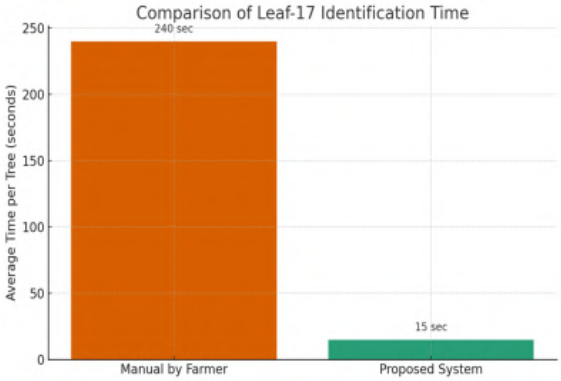

FIGURE 10. COMPARISON TIME 17TH LEAF IDENTIFICATION

5. CONCLUSIONS

This research successfully integrates the YOLOv12 object detection model with a spatial-positioning algorithm to enable automatic identification of the 17th leaf in oil palm trees (Elaeis guineensis), which serves as a crucial physiological indicator for nutrient diagnosis. The proposed system effectively addresses the inefficiencies of manual identification by combining deep-learning-based detection with geometric reasoning derived from the palm's natural phyllotaxis structure. Experimental results demonstrate that the YOLOv12 model achieved a high mean average precision (mAP@0.5) of 92.4%, an F1-score of 90.3%, and an average positional error of only 10.6 pixels across 375 test images, outperforming YOLOv8 and EfficientDet-D3 in both accuracy and inference speed (12 ms). Furthermore, the system reduced identification time from 3–5 minutes to under 15 seconds per tree, improving operational efficiency by over 95%. These outcomes validate that integrating YOLOv12 with spatial reasoning provides a reliable, scalable, and real-time solution for leaf indexing, paving the way for nutrient-based precision agriculture. Future research may focus on incorporating 3D-aware reconstruction, multi-view data fusion, or transformer-based spatial reasoning to enhance accuracy under complex canopy conditions, as well as integrating the framework into UAV or IoT-based platforms for continuous and autonomous large-scale plantation monitoring.

REFERENCES

[1] N. A. Board, "Market Intelligence Report: Vegetable Oils," Windhoek, Oct. 2023. [Online]. Available: https://www.nab.com.na/wp-content/uploads/2023/11/Market-Intelligence-Report_Vegetable-oil-NAB-2023-1.pdf

[2] M. Kamireddy, S. K. Behera, and S. Kancherla, "Establishing Critical Leaf Nutrient Concentrations and Identification of Yield-Limiting Nutrients for Precise Nutrient Prescriptions in Oil Palm (Elaeis guineensis),"

Agriculture, vol. 13, no. 2, p. 453, 2023, doi: 10.3390/agriculture13020453.

[3] L. S. Woittiez, M. T. van Wijk, M. van Noordwijk, K. E. Giller, and P. Tittonell, "Yield gaps in oil palm: A quantitative review of contributing factors," *European Journal of Agronomy*, vol. 83, pp. 57–77, 2017, doi: 10.1016/j.eja.2016.11.002.

[4] M. A. Istiak *et al.*, "Adoption of UAV imagery in agriculture: Opportunities and challenges," *Ecol Inform*, vol. 78, p. 102305, 2023, doi: 10.1016/j.ecoinf.2023.102305.

[5] S. Coulibaly, J. Zhao, and P. Li, "Deep learning for precision agriculture: A bibliometric overview," *Intelligent Systems with Applications*, vol. 2, p. 100040, 2022, doi: 10.1016/j.iswa.2022.100040.

[6] I. Paçal, M. Karakose, and F. Ahmed, "A systematic review of deep learning techniques for plant disease detection (2020–2024)," *Artif Intell Rev*, vol. 57, no. 3, pp. 2479–2510, 2024, doi: 10.1007/s10462-024-10944-7.

[7] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2023/html/Wang_YOLOv7_Trainable_Bag-of-Freebies_Sets_New_State-of-the-Art_for_Real-Time_Object_Detectors_CVPR_2023_paper.html

[8] G. Jocher, L. Chang, Y. Qin, and T. Tang, "YOLOv12: Innovations in Real-Time Object Detection," *arXiv preprint*, 2024, [Online]. Available: https://arxiv.org/abs/2406.19407

[9] F. Feng, W. Chen, and M. Li, "Improved YOLOv8 algorithms for small-object detection in complex agricultural scenes," *Alexandria Engineering Journal*, vol. 83, pp. 1–15, 2024.

[10] Y. C. Putra, R. Nugroho, and A. Yuliana, "Automatic detection and counting of oil palm trees using object-based deep learning on very-high-resolution images," *Heliyon*, vol. 9, no. 4, p. e14520, 2023, doi: 10.1016/j.heliyon.2023.e14520.

[11] J. Zhang, Y. Wu, and C. Feng, "Spatial positioning technologies in agriculture: A review," *Precis Agric*, vol. 20, no. 4, pp. 734–754, 2019.

[12] A. Rehman, A. Mahmud, and I. Mehmood, "GPS and vision-based mapping of agricultural fields using UAVs," *J Field Robot*, vol. 36, no. 4, pp. 784–802, 2019, doi: 10.1002/rob.21906.

[13] M. Breure, "Palm leaf phyllotaxis: Structure and implications for crop management," *J Oil Palm Res*, vol. 31, no. 1, pp. 20–32, 2019.

[14] S. Lu and J. Shen, "Deep learning for plant leaf counting in dense canopies," *Comput Electron Agric*, vol. 187, p. 106262, 2021, doi: 10.1016/j.compag.2021.106262.

[15] Z. Chen, H. Wang, and Y. Zhao, "Robust leaf instance segmentation in the wild using graph reasoning and attention," *IEEE Access*, vol. 9, pp. 115283–115295, 2021, doi: 10.1109/ACCESS.2021.3105226.

[16] W. Wang, J. Xu, and L. Zhang, "Multi-view learning for plant phenotyping in field conditions," *Comput Electron Agric*, vol. 202, p. 105114, 2023, doi: 10.1016/j.compag.2022.107584.

[17] Y. Yang, X. Lin, and H. Li, "3D-aware object detection for dense natural scenes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 13478–13487.

[18] Z. Liu, Y. Wang, and X. Chen, "A survey of transformer-based object detection," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 3, pp. 2431–2452, 2023, doi: 10.1109/TPAMI.2022.3216963.

[19] D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosyst Eng*, vol. 114, no. 4, pp. 358–371, 2013, doi: 10.1016/j.biosystemseng.2012.08.009.

[20] P. Lottes, J. Behley, N. Chebrolu, and C. Stachniss, "Advances in precision agriculture using computer vision technologies," *Agric Syst*, vol. 173, pp. 1–11, 2019.

[21] O. Bongomin, J. Mwebaze, and T. Nsubuga, "UAV image acquisition and processing for high-throughput phenotyping in agricultural research and breeding programs," *Plant Phenome Journal*, vol. 7, no. 1, p. e20096, 2024, doi: 10.1002/ppj2.20096.

[22] R. Rasool, F. Khan, and S. Aziz, "Vision transformers for crop monitoring under natural lighting," *Agric Syst*, vol. 228, p. 104962, 2024, doi: 10.1016/j.agsy.2024.104962.

[23] J. Li, H. Sun, and L. Qiu, "3D leaf-order estimation using multi-view convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 11254–11263.

[24] Y. Qin, L. Chang, and T. Tang, "AI-enhanced detection for precision agriculture using transformer networks," *Sensors*, vol. 24, no. 2, p. 572, 2024, doi: 10.3390/s24020572.

AUTHORS

Jihad Rahmawan

Jihad Rahmawan, S.T., M.Sc. earned his Bachelor's degree in Electrical Engineering from Universitas Ahmad Dahlan and his Master's in Computer Science from Iwate Prefectural University, Japan. His expertise includes robotics, image processing, and artificial intelligence.

Herman Yuliansyah

He is a lecturer in Informatics at Universitas Ahmad Dahlan. He holds a Ph.D. in Artificial Intelligence from Universiti Kebangsaan Malaysia. His research focuses on data mining, social network analysis, and text mining.

Anton Yudhana

He earned his B.Sc. in Electrical Engineering from Institut Teknologi Sepuluh Nopember (ITS Surabaya), his M.T. in Electrical Engineering from Universitas Gadjah Mada (UGM), and completed his Ph.D. in Informatics at Universiti Teknologi Malaysia (UTM). His expertise includes signal processing, wireless communication, and numerical methods.

Syahid Al Irfan

Syahid Ali Irfan is a Ph.D. student at the Department of Software and Information Science, Iwate Prefectural University, Japan. His research focuses on image processing, computer vision, and artificial intelligence applications for visual data analysis. He is actively engaged in developing efficient algorithms for pattern recognition and intelligent imaging systems.