Journal of Deep Learning, Computer Vision and Digital Image Processing
E-ISSN: 2986-8939 P-ISSN: 2986-8920
Vol. 4, No. 2, June 2026
DOI. https://doi.org/10.61255/decoding.v4i2.1281

RT-DETR-Based Computer Vision System for Real-Time Detection and Classification of Oil Palm Fruit Maturity Levels in Plantations

Nur Hafiqah Rambe
(Institut Teknologi sawit Indonesia, Medan, Indonesia)

Ratu Mutiara Siregar✉
(Institut Teknologi sawit Indonesia, Medan, Indonesia)

Raden Aris Sugianto
(Institut Teknologi sawit Indonesia, Medan, Indonesia)

OPEN ACCESS

ARTICLE HISTORY
Received: April, 30 2026
Revised: May, 24 2026
Accepted: June, 18 2026

KEYWORDS
Computer vision;
Deep learning;
Oil palm fruit maturity;
RT-DETR;
Smart agriculture

ABSTRACT

Purpose – This study aimed to develop an automated oil palm fruit maturity level detection system using the real-time detection transformer (RT-DETR) algorithm to overcome the limitations of conventional visual inspection methods, which are often subjective and inconsistent. This study evaluated the effectiveness of the RT-DETR in detecting and classifying oil palm fruit maturity levels to support quality control processes in plantation operations.

Method – A computer vision-based approach was implemented using the RT-DETR-L object detection model. The dataset consisted of 14,620 annotated oil palm fruit images categorized into four maturity levels: unripe, underripe, ripe, and overripe. The research process included data collection, image annotation, preprocessing, model training, and evaluation of the model. The model performance was assessed using precision, recall, mean Average Precision (mAP@50), and inference speed metrics.

Findings – The experimental results show that the RT-DETR-L model achieved a precision of 93.2%, 95.6%, and mAP@50 of 96.9%, respectively. The model successfully detected and classified oil palm fruit maturity levels across all categories with high accuracy. Furthermore, the model achieved an inference time of 25–28 ms per image and a processing speed of 10–14 FPS on an NVIDIA RTX 3050 4GB GPU, demonstrating its capability for real-time applications.

Research Implications – The findings indicate that RT-DETR-L can improve the efficiency, consistency, and accuracy of oil palm fruit sorting and quality control processes. However, this study was limited to the available datasets and testing scenarios used. Future research should evaluate the model under diverse environmental conditions, lighting variations, and field deployment settings to improve its generalizability and robustness.

Originality – Unlike previous studies that primarily employed CNN-based detectors or focused on binary maturity classification, this study investigated the application of a transformer-based RT-DETR-L architecture for detecting four oil palm fruit maturity categories. The results demonstrate that RT-DETR-L can provide high detection accuracy while maintaining real-time performance in smart agriculture applications.

Correspondence Author: ✉ratu_ms@itsi.ac.id
To cite this article : Rambe, N. H. , Siregar, R. M., Sugianto, R. A. (2026). RT-DETR-Based Computer Vision System for Real-Time Detection and Classification of Oil Palm Fruit Maturity Levels in Plantations. *Journal of Deep Learning, Computer Vision and Digital Image Processing,* 4(2), 43-57. https://doi.org/10.61255/decoding.v4i2.1281

INTRODUCTION

Oil palm (*Elaeis guineensis*) is one of the most important plantation commodities and plays a strategic role in Indonesia's economy. According to [1], the total area of oil palm plantations in Indonesia has exceeded 15 million hectares, positioning Indonesia as one of the world's largest producers of crude palm oil (CPO). The continuous expansion of plantation areas has increased the demand for more efficient harvesting, sorting, and processing systems that can maintain product quality while supporting large-scale production [2]. Among the factors affecting palm oil quality, the maturity level of harvested fruit is one of the most critical determinants of oil yield and overall production efficiency of palm oil.

The maturity level of oil palm fruit directly influences both the oil extraction rate and the free fatty acid (FFA) content. Harvesting fruit before or after its optimal maturity stage may reduce oil yield and negatively affect the quality of the resulting CPO [3], [4]. Therefore, an accurate maturity assessment is essential for ensuring product quality and maximizing economic returns. However, maturity level determination in many plantations and palm oil mills is still performed manually through visual inspections. This conventional approach relies heavily on workers' experience and subjective judgment, often resulting in inconsistent classification [5]. Variations in lighting conditions, fruit color appearance, fruit bunch arrangements, and complex backgrounds further complicate the assessment process and increase the likelihood of classification errors [6]. In large-scale plantation operations, inaccurate maturity level identification may adversely affect sorting efficiency, processing performance, and the quality of the final product[7].

Recent advances in artificial intelligence and computer vision have created new opportunities for automating agricultural inspections. Computer vision is a branch of artificial intelligence that enables computers to automatically acquire, process, and interpret visual information from digital images and video. In agricultural applications, computer vision technologies have been increasingly utilized for crop monitoring, disease detection, yield estimation, and fruit quality assessment. In the oil palm sector, these technologies provide a promising solution for automatic maturity level detection based on visual characteristics such as color, texture, and fruit morphology [8].

Object detection is one of the most important areas of computer vision [9]. Unlike image classification, which only predicts the category of an image, object detection simultaneously identifies object classes and determines their locations within an image using bounding boxes [10]. This capability is particularly valuable in agricultural environments, where multiple objects may appear in a single image under varying environmental conditions. Real-time object detection systems can support automated sorting and monitoring processes by providing rapid and accurate information on the detected objects [11].

Several studies have explored the application of machine learning and deep learning techniques for oil palm fruit maturity classification. A previous study [12] employed a convolutional neural network (CNN) model to classify oil palm fruit maturity levels and reported an average accuracy of 76.52%, a best accuracy of 82.61% and an F1-score of 0.76. These findings demonstrate the potential of deep learning approaches for automating maturity identification. However, the proposed method focuses primarily on image-level classification and does not provide object localization capabilities. Consequently, the model may face limitations when applied to complex field environments containing multiple fruits or under varying visual conditions.

Another study [13] utilized the K-nearest neighbor (K-NN) algorithm based on manually extracted color, shape, and texture features of fruits. The highest classification accuracy of 96.6% was achieved using color features, whereas the shape and texture features achieved accuracies of 73.3% and 66%, respectively. Although the results demonstrated promising performance, this approach remained highly dependent on handcrafted feature extraction. These methods are generally sensitive to changes in illumination, camera angles, and background variations, limiting their robustness in real plantation environments. Furthermore, K-NN-based approaches do not support real-time object detection and may be less suitable for large-scale operational deployment [14].

The limitations identified in previous studies reveal several research gaps. First, many existing approaches focus solely on image classification without incorporating object localization capabilities [15], [16] [17]. For example, Shiddiq et al. [16] developed a multispectral imaging–based oil palm fruit bunch maturity detection system using the YOLOv4 algorithm and reported an mAP@0.50 of 99.47%, with precision values of 0.92 and 0.88 and recall values of 0.99 for ripe and unripe classes, respectively. The system was also evaluated on a moving conveyor and achieved processing speeds of 2.99–3.88 FPS. Although this study demonstrated excellent detection performance under a controlled multispectral imaging environment, the proposed system was limited to binary maturity classification (ripe and unripe) and relied on specialized imaging hardware. Such requirements may restrict the scalability and deployment in more diverse plantation environments, where multiple maturity levels must be identified using conventional RGB imagery. Second, traditional machine learning methods require manual feature engineering, which reduces their adaptability to diverse environmental conditions. Third, previous studies have reported challenges in achieving a balance between detection accuracy and computational efficiency, highlighting the need for more advanced object detection architectures that are suitable for operational environments. These gaps indicate the need for object detection architectures capable of providing accurate localization and multi-class maturity classification while maintaining computational efficiency under complex operating conditions.

To address these challenges, this study proposes the implementation of a real-time detection transformer large (RT-DETR-L) model for the automated detection of oil palm fruit maturity levels. The RT-DETR-L is a transformer-based object detection architecture designed to achieve high detection accuracy while maintaining computational efficiency, making it suitable for real-time object detection scenarios [18]. The model combines the feature extraction capabilities of convolutional neural networks (CNNs) with the advantages of global context modeling of transformer architectures, thereby enabling a more effective representation of the spatial relationships among the detected objects. In addition, RT-DETR offers improved computational efficiency and inference speed, making it suitable for automated object detection applications in real-world environments [19] [20].

The significance of this study lies in its potential contribution to the development of intelligent agricultural technologies in the palm oil industry. By enabling automated and accurate maturity level detection, the proposed system can support more efficient sorting operations, reduce human error, and improve quality control processes throughout the production chain. Furthermore, the implementation of real-time object detection technologies aligns with ongoing efforts to modernize agricultural production systems using artificial intelligence and digital transformation.

Therefore, this study aimed to evaluate the effectiveness of the RT-DETR-L model in detecting oil palm fruit maturity levels using computer vision techniques. Specifically, this study investigated the model's ability to classify oil palm fruits into four maturity categories: unripe, underripe, ripe, and overripe, and assessed its performance using precision, recall, and mean Average Precision (mAP) metrics. In addition, this study examined the suitability of RT-DETR-L for automated agricultural inspection tasks under varying visual conditions. We hypothesized that RT-DETR-L could achieve high detection accuracy and robust object localization performance, making it a suitable approach for automated oil palm fruit maturity assessment and smart agriculture applications.

METHOD

Research Design

This study employed a quantitative experimental research design to develop and evaluate an automated oil palm fruit maturity level detection system using a Real-Time Detection Transformer Large (RT-DETR-L) model. This study focused on assessing the capability of a transformer-based object detection architecture to identify, localize, and classify oil palm fruits according to their maturity levels using digital-image analysis. RT-DETR-L was selected because it combines

convolutional neural network (CNN)-based feature extraction with transformer-based global context modeling, enabling accurate object localization and classification while maintaining computational efficiency suitable for real-time object detection applications [21] [22].

Population and Sampling Method

The population in this study consisted of digital images of oil palm fruits representing various maturity stages. A purposive sampling method was applied to ensure that the dataset adequately represented the visual characteristics of each maturity category. The final dataset comprised 14,620 images divided into four classes: unripe, underripe, ripe, and overripe.

The dataset was obtained from two sources: primary data consisting of 1,000 images collected through direct documentation at the practice plantation of the Institut Teknologi Sawit Indonesia (ITSI) using a smartphone. The secondary data consisted of 13,620 images obtained from the publicly available Palm Oil 2 dataset hosted on the Roboflow Universe platform (Dataset ID: palm-oil-2-1gztp, Version 1). In addition, most of the images used in this study were obtained from the publicly available Roboflow dataset. Although the inclusion of primary images increased data diversity, the dataset may not fully represent the variability in plantation and palm oil mill environments. Consequently, domain shifts related to lighting conditions, camera specifications, fruit presentation, and operational settings may affect the performance of models when deployed in real-world applications. Data collection will be conducted between January and February 2026.

Prior to model training, all images were manually reviewed and annotated using bounding box labels corresponding to the four maturity categories. Annotation was performed by two researchers with domain knowledge of oil palm agronomy, using standardized annotation guidelines applied uniformly across all classes. Each bounding box was drawn to encompass the full fruit bunch and assigned one of four maturity labels: unripe, underripe, ripe, or overripe.

To ensure annotation consistency, a sample audit of approximately 10% of the annotated images was conducted by a third reviewer, and discrepancies were resolved through discussion before the images were included in the final dataset. Agreement among annotators was evaluated during the auditing process, and labeling inconsistencies were corrected prior to dataset finalization to ensure annotation reliability. This procedure was conducted to minimize labeling inconsistencies and improve the reliability of the dataset used for model development [23].

To support the model development and evaluation, the dataset was divided into training, validation, and testing subsets using a 70:20:10 ratio. Prior to partitioning, preprocessing and export were conducted on the Roboflow platform, during which images that did not meet quality standards, including those with missing annotations, corrupted files, or duplicate entries, were excluded from the final split. Following this filtering process, 12,690 images were retained for the model development. The final distribution was as follows: the training subset consisted of 8,883 images containing approximately 17,693 annotated instances; the validation subset consisted of 2,538 images containing approximately 5,054 annotated instances; and the testing subset consisted of 1,269 images containing 2,528 annotated instances. Each subset maintained a balanced class distribution across the four maturity categories. This partitioning strategy enabled the model performance to be assessed on previously unseen data while reducing the risk of overfitting and improving the reliability of the performance evaluation.

Table 1. Dataset Split Distribution

Dataset Type	Percentage	Number of Images
Training Data	70%	8,883
Validation Data	20%	2,538
Testing Data	10%	1,269

Dataset Type	Percentage	Number of Images
Total	100%	12,690

Table 1 presents the distribution of the dataset after preprocessing and quality filtering. A total of 12,690 images were retained and divided into training, validation, and testing subsets using a 70:20:10 ratio, respectively. This partitioning strategy was applied to ensure sufficient data for model learning while maintaining reliable validation and testing sets for performance evaluation.

Instrumentation and Research Instruments

The primary research instrument was a labeled image dataset consisting of oil palm fruit images with corresponding bounding-box annotations. Image annotation was performed using the Roboflow platform, where each fruit object was assigned a bounding box and classified according to its maturity. The annotation process was used as a reference standard for the model training and evaluation. Labels were assigned based on the visual maturity characteristics of each fruit bunch, including surface color, texture, and the presence of loose fruits, following the established agronomic criteria for oil palm maturity classification.

All images were stored in JPG format and maintained at a minimum resolution of 640 × 640 pixels. The RT-DETR-L model was implemented using the Ultralytics deep learning framework and trained on Google Colab A100 High-RAM, equipped with an NVIDIA A100 GPU. The experimental environment also included an Intel Core i5 processor (or equivalent), 8 GB RAM, and 512 GB of SSD storage. To evaluate the deployment performance under practical conditions, inference testing was conducted on a local computer equipped with an NVIDIA GeForce RTX 3050 GPU with 4 GB VRAM.

Research Procedure

This study utilized the RT-DETR-L architecture provided by the Ultralytics framework with pre-trained weights (rtdetr-l/.pt) and the default backbone configuration defined in the official implementation. The research procedure consisted of four main stages: dataset preparation, data pre-processing, model training, and evaluation. During the dataset preparation stage, oil palm fruit images were collected, categorized, and annotated based on their maturity levels. Each image was manually labeled to identify the location and class of the target objects [17].

The preprocessing stage was conducted to ensure compatibility with the RT-DETR architecture. First, all images were resized to 640 × 640 pixels to satisfy the model input requirements [24]. Pixel normalization was applied to improve the training stability and convergence. To increase data diversity and improve generalization performance, data augmentation techniques, including horizontal flipping, translation, scaling, mosaic augmentation, and RandAugment, were applied during training [25].

Following preprocessing, the RT-DETR-L model was initialized using pre-trained weights (rtdetr-l. pt) and trained using the prepared dataset. The model was trained for 150 epochs with a batch size of 32 and an input image resolution of 640 × 640 pixels. Training was performed using the AdamW optimizer with an initial learning rate of 0.0001 and a weight-decay coefficient of 0.0005. The data augmentation techniques included horizontal flipping, translation, scaling, mosaic augmentation, and RandAugment. During training, the model performed forward propagation to generate object predictions and backpropagation to update the network parameters based on the calculated loss function. Validation was conducted after each training epoch to monitor the performance and prevent overfitting. The model with the best validation performance was saved and used for the final testing.

Data Analysis Plan

The model performance was evaluated using standard object detection metrics, including precision, recall, F1-score, and mean Average Precision at an Intersection over Union threshold of 0.5 (mAP@50). These metrics are widely used to assess object detection performance and provide comprehensive information regarding the classification accuracy and localization quality [26].

Precision measures the proportion of correctly predicted positive detections among all positive predictions.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Recall measures the proportion of correctly detected objects among all actual positive objects, as follows:

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

The F1-score represents the harmonic mean of precision and recall as follows:

$$F1 - score = 2x \frac{Precision \text{ x } Recall}{Precision + Recall} \quad (3)$$

In addition to the detection accuracy metrics, the inference performance was evaluated using the inference time and processing speed. During validation on an NVIDIA A100 GPU, the RT-DETR-L model achieved an inference time of 2.4 ms per image, with a total processing time of approximately 2.8 ms per image, including preprocessing and postprocessing. Deployment testing on an NVIDIA GeForce RTX 3050 4 GB GPU achieved an inference time of 25–28 ms per image and a processing speed of 10–14 FPS.

Methodological Limitations

This study focused exclusively on the application of RT-DETR for oil palm fruit maturity level detection using image data. The model was trained and evaluated using four maturity categories under specific environmental conditions represented in the dataset. Consequently, the model performance may be influenced by factors such as image quality, illumination variation, camera specifications, annotation consistency, and dataset diversity. Although data augmentation was employed to improve robustness, further validation using larger datasets and more diverse plantation environments is required to strengthen the generalizability of these findings.

RESULTS AND DISCUSSION

Results

This section presents the experimental results of the Real-Time Detection Transformer Large (RT-DETR-L) model for detecting oil palm fruit maturity levels based on digital images. The evaluation results included a confusion matrix, precision–recall analysis, confidence-based performance curves, quantitative detection metrics, and visualization of the object detection outputs. The results are

presented through evaluation metrics, graphs, tables, and detection visualizations to demonstrate the model's ability to accurately classify and localize oil palm fruits across different maturity categories while maintaining real-time detection performance.

Confusion Matrix

A confusion matrix was employed to analyze the classification performance of the RT-DETR-L model by comparing the actual maturity labels with the results predicted by the model. This matrix presents the number of correct predictions and misclassifications for each category, enabling a more comprehensive evaluation of the per-class detection performance in terms of the precision and recall.

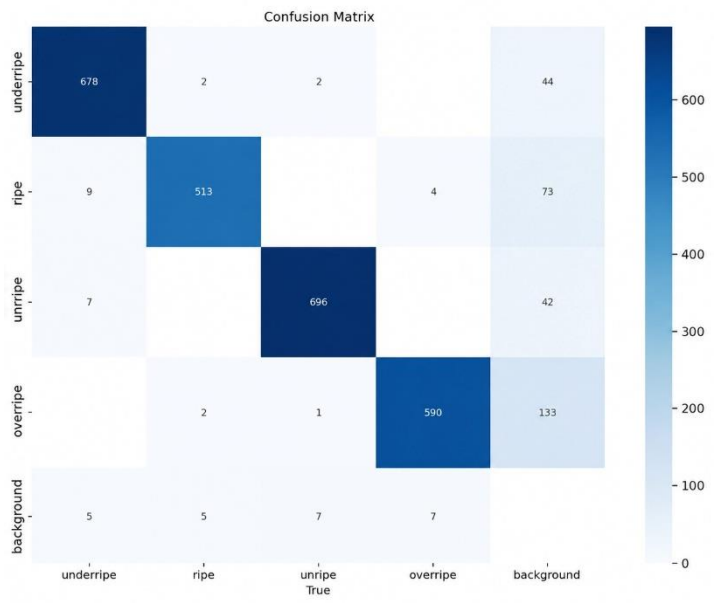

Figure 1. Confusion Matrix

As shown in Figure 1, most instances were correctly predicted, as indicated by the dominance of high values on the matrix diagonal. The unripe, underripe, ripe, and overripe classes exhibited strong true-positive counts, demonstrating the model's high mAP and reliable classification capability across all maturity categories. Some misclassifications were still observed, particularly between classes with similar visual characteristics, such as fruit surface color and texture during the maturity transition phases. In addition, a small number of background regions were incorrectly detected as oil palm-fruit objects, contributing to false-positive predictions. Overall, the RT-DETR-L model achieved a strong detection performance across all four maturity classes, with a relatively low misclassification rate, as reflected in the precision, recall, and mAP@50 scores reported in Table 2.

Model Performance Evaluation

The performance of the RT-DETR model was evaluated using precision, recall, F1-score, and mean Average Precision (mAP) metrics to assess its effectiveness in detecting and classifying the maturity levels of oil palm fruits.

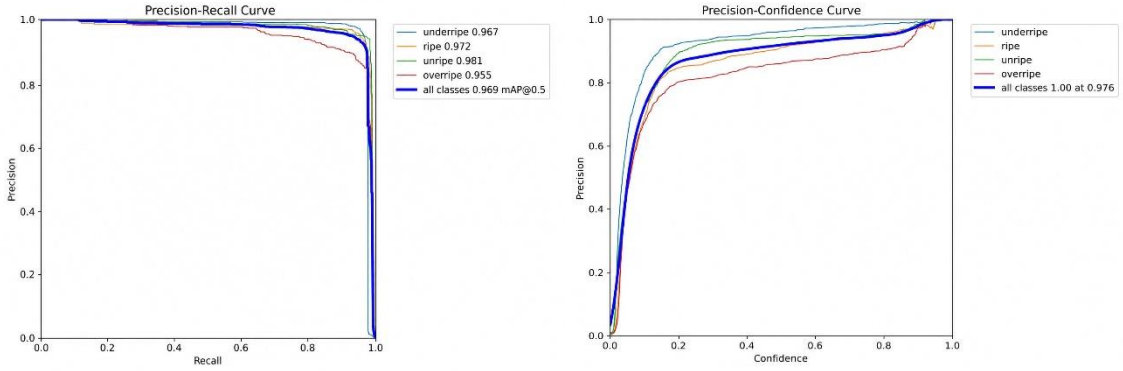

Figure 2. Precision-Recall Curve and Precision-Confidence Curve

The precision-recall curve was used to evaluate the quality of the model detection based on the relationship between precision and recall in each class of objects. Based on Figure 2, the Average Precision (AP) values for the underripe, ripe class was 0.972, unripe class was 0.981, and overripe classes were 0.967, 0.972, 0.981, and 0.955, respectively. Overall, the model obtained an mAP@0.5 value of 0.969 or 96.9%, indicating that the RT-DETR model has excellent detection and classification performance for recognizing the maturity level of oil palm fruit.

The precision–confidence curve shows the relationship between the confidence threshold and the precision model value for each palm fruit maturity class. As shown in Figure 2, the precision increases as the confidence threshold increases. At a high confidence level, all classes had a precision close to 1.00. The underripe and unripe classes showed the most stable precision performance compared to the other classes, whereas the overripe classes had slightly lower precision. Overall, the model achieved a maximum precision of 1.00 at a confidence threshold of approximately 0.976.

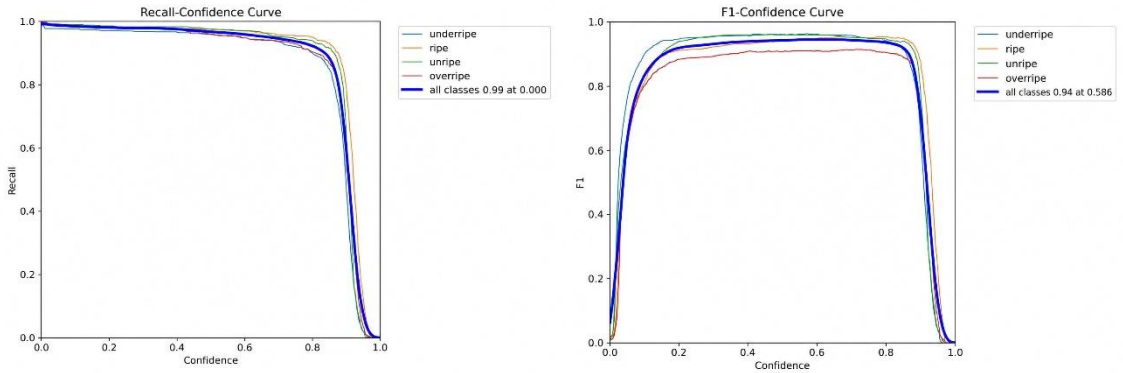

Figure 3. Recall-Confidence Curve and F-1 Confidence Curve

The recall-confidence curve was used to describe the ability of the model to detect all objects at various confidence threshold values. As shown in Figure 3, the model obtained the highest F1-score of 0.94 at a confidence threshold of 0.586. The recall value tends to be stable in the medium confidence range and then decreases when the confidence threshold is close to 0.9 because the model becomes more selective in detecting objects, and the number of objects that are successfully recognized decreases. These results demonstrate that the RT-DETR model has excellent object detection capabilities for recognizing different levels of oil palm maturity.

Meanwhile, the F1-Confidence Curve is used to show the balance between the precision and recall values at various confidence threshold levels. Based on Figure 3, the model obtained the highest F1-score of 0.94 at a confidence threshold of 0.586. The F1-score appears to be stable in the mid-

confidence range, indicating that the model can maintain a balance between detecting objects correctly and minimizing prediction errors. However, when the confidence threshold approached 1.0, the F1-score decreased significantly as the model became more rigorous, resulting in fewer identified objects. Overall, the results of this curve show that the RT-DETR model has a good and stable detection performance in classifying the maturity level of oil palm fruit.

Results of Final Training Model Evaluation

Table 2. Final Evaluation Results of RT-DETR Model Training

Class	Images	Instances	Precision (P)	Recall (R)	mAP50	mAP50-95
All	1269	2528	0,932	0,956	0,969	0,762
Unripe	298	706	0,949	0,97	0,981	0,758
Underripe	304	699	0,974	0,944	0,967	0,746
Ripe	383	522	0,932	0,967	0,972	0,812
Overripe	349	601	0,875	0,943	0,955	0,733

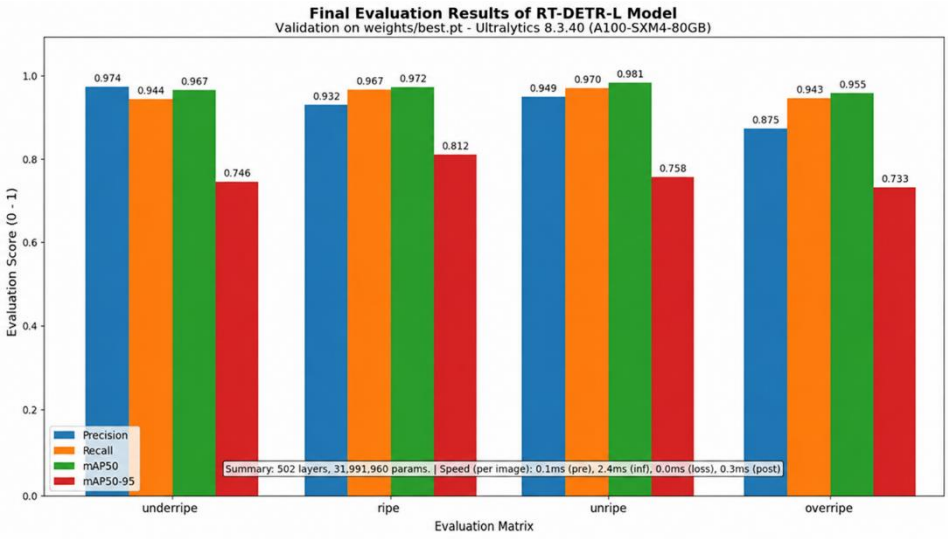

Figure 3. Diagram of the Final Evaluation Results of the RT-DETR Model

The evaluation results of the Real-Time Detection Transformer (RT-DETR) model demonstrated outstanding performance across all oil palm fruit maturity categories. Based on the testing outcomes, the model achieved an Average Precision of 0.932, recall of 0.956, and mean average precision (mAP@0.5) of 0.969. These results indicate that the model possesses a high level of prediction accuracy and strong object detection capability for the testing dataset.

From the class-wise evaluation, the unripe category achieved the highest performance, with an mAP value of 0.981, followed by the ripe and underripe classes, with values of 0.972 and 0.967, respectively. The overripe category obtained an mAP value of 0.955, which reflects excellent detection capability. The variation in performance across classes suggests that the model can effectively recognize the characteristics of objects, although slight performance differences remain in certain categories.

Among all the categories, the overripe class achieved the lowest precision, recall, and mAP values. This performance difference may be attributed to the visual similarities between overripe fruits and neighboring maturity stages, particularly under varying illumination conditions. Variations in fruit surface texture, color degradation patterns, and background complexity may also contribute to the increased classification difficulty for this category.

Furthermore, the model achieved an mAP@0.5:0.95 score of 0.762, indicating that its performance remained reliable, even under stricter evaluation criteria. The evaluation curves also showed that the precision, recall, and mAP values for each class were consistently high and stable. Overall, these findings confirm that the RT-DETR-L model demonstrated a strong detection performance across all four maturity categories. However, it should be noted that the performance was not uniform across classes. The overripe class consistently achieved the lowest precision (0.875), recall (0.943), and mAP@50 (0.955), suggesting that this category remains the most challenging to detect, likely because of its visual overlap with neighboring maturity stages. These results indicate that although the model performs reliably for most categories, certain classification boundaries remain difficult to resolve under real-world imaging conditions.

Although the model achieved a strong mAP@50 of 0.969, the mAP@50–95 score decreased to 0.762 when stricter localization criteria were applied. Notably, a high mAP@50 does not necessarily guarantee precise bounding box localization at stricter IoU thresholds. The gap between mAP@50 and mAP@50–95 indicates that although the model reliably detects and classifies oil palm fruit maturity levels, localization precision becomes more challenging as the IoU threshold increases. This is a common characteristic of object detection models and does not diminish the practical value of the results, particularly for agricultural inspection applications, where approximate localization is often sufficient. Future research may address this through larger annotated datasets, improved bounding box annotation quality, and architectural optimizations targeted at localization refinement.

DISCUSSIONS

Qualitative evaluation of the RT-DETR-L model demonstrated its ability to consistently detect and classify oil palm fruits across all maturity categories, namely unripe, underripe, ripe, and overripe. The model successfully recognized the visual characteristics of each category based on the differences in fruit color, texture, and surface appearance, while maintaining relatively stable confidence scores across most test images. In addition, the generated bounding boxes accurately localized fruit objects under various image conditions.

Nonetheless, some detection boundary deviations were still observed in images with uneven lighting conditions or complex backgrounds. This is especially true for categories that have color similarities, such as underripe and ripe fruits; therefore, the model requires a more detailed identification process to distinguish the visual characteristics between classes. Overall, the qualitative detection results show that the RT-DETR model has good object classification and localization capabilities under field-test conditions. The results of the visualization of the RT-DETR model detection in each maturity level category are shown in Figure 5 and Figure 6.

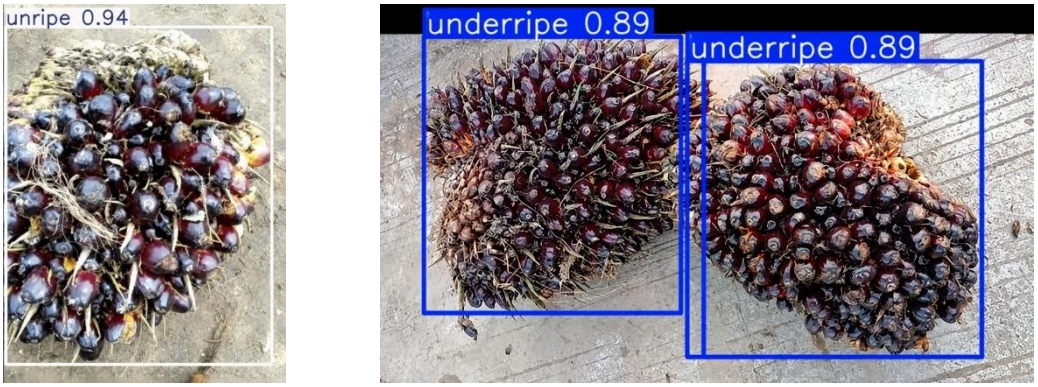

Figure 5. Detection Results of Oil Palm Fruits in the unripe and Underripe Categories

Figure 5 shows the results of detecting oil palm fruits in the unripe and underripe categories using the RT-DETR model. The model can identify both categories well based on differences in visual characteristics, particularly in terms of the color and texture of the fruit surface. In the unripe category, the objects were successfully detected with accurately localized bounding boxes and a *confidence* value of 0.94. The high confidence value is influenced by the characteristics of the dark purple color, which is more contrasting, making it easier for the model to extract the features of the object. Meanwhile, in the underripe category, the model was still able to detect stably with a *confidence* value of 0.89. However, the similarity of the visual characteristics between the underripe and ripe categories caused the model's confidence level to be slightly lower than that of the unripe category. Overall, the detection results show that the RT-DETR model has a good ability to classify and localize objects in both categories.

Figure 6. Detection Results of Oil Palm Fruits in the Ripe and Overripe Categories

Figure 6 shows the results of detecting ripe and overripe oil palm fruits using the RT-DETR model. The model can recognize both categories based on the color change and visual characteristics of the fruit surface. In the ripe category, the object was detected very well through the dominance of brownish-red color, which is the main characteristic of the maturity level of palm fruit, with a confidence value of 0.95. The resulting bounding box also appears stable and precise under various test image conditions.

Meanwhile, in the overripe category, the model was still able to detect quite well, with a confidence value of 0.83. The model recognizes objects based on the darker color of the fruit and changes in surface texture compared to other categories. Despite some suboptimal lighting conditions and visual similarities to the ripe category, the RT-DETR can effectively differentiate between the two

categories. This demonstrates that the model effectively learned discriminative visual features to differentiate between the oil palm fruit maturity levels.

The quantitative results obtained in this study are comparable to those reported in previous studies. However, these comparisons should be interpreted cautiously because the evaluated models were tested on different datasets and assessed using different performance metrics. A CNN-based oil palm fruit classification model reported by [12] achieved an average accuracy of 76.52%, a maximum accuracy of 82.61%, and an F1-score of 0.76. In comparison, the RT-DETR-L model achieved a precision of 93.2%, recall of 95.6%, and mAP@50 of 96.9%. These improvements can be attributed to the ability of transformer-based architectures to capture long-range contextual information while simultaneously performing object localization and classification tasks. Unlike image classification models, RT-DETR-L can identify object locations using bounding boxes, making it more suitable for practical deployment in automated sorting and monitoring systems. Therefore, the comparison presented in this study should be regarded as contextual rather than a controlled one. Direct conclusions regarding the superiority of RT-DETR-L over alternative approaches cannot be drawn without evaluation using identical datasets, experimental settings, and performance metrics.

Among all maturity categories, the overripe class consistently produced the lowest precision, recall, and mean average precision (mAP) values. This performance difference is likely caused by the visual similarities between overripe fruits and neighboring maturity stages, particularly ripe fruits. Overripe fruits often exhibit gradual color degradation and texture variations that overlap with the characteristics of the ripe category, increasing the difficulty of classification. From a practical perspective, this finding is important because the inaccurate identification of overripe fruits may affect harvesting decisions and quality control processes in palm oil production.

The model achieved an mAP@50 of 0.969 and an mAP@50–95 of 0.762. The decrease in performance under stricter Intersection over Union (IoU) thresholds indicates that object classification remained highly accurate, whereas precise localization became more challenging as the overlap requirements increased. Nevertheless, the obtained mAP@50–95 value demonstrates that the model maintained a reliable localization performance across a range of IoU thresholds. These findings suggest that the RT-DETR-L model is highly effective for oil palm fruit maturity classification, whereas further improvements in bounding box precision could enhance localization performance under more stringent evaluation criteria. Future studies may improve localization accuracy using larger datasets, enhanced annotation quality, and additional optimization of training strategies.

In addition to detection accuracy, the RT-DETR-L model demonstrated practical and real-time performance. Deployment testing on an NVIDIA GeForce RTX 3050 4 GB GPU achieved an inference time of approximately 25–28 ms per image, corresponding to a processing speed of 10–14 FPS. These results indicate that the proposed model can perform object detection within a time frame suitable for near-real-time agricultural inspection and sorting applications, supporting the rationale for selecting RT-DETR-L as the detection architecture.

Compared with the multispectral imaging–based approach proposed by Shiddiq et al. [16], which achieved a mAP@50 of 99.47%, the RT-DETR-L model produced slightly lower detection accuracy but offered practical advantages because it relied only on conventional RGB imagery. This characteristic reduces

hardware requirements and may facilitate broader deployment in plantation environments where multispectral imaging systems are not readily available.

CONCLUSION

Based on the results of this study, the Real-Time Detection Transformer Large (RT-DETR-L) model was successfully implemented for automated oil palm fruit maturity level detection. The model demonstrated a strong detection performance, achieving a precision of 93.2%, recall of 95.6%, mAP@50 of 96.9%, and mAP@50–95 of 76.2%. The model accurately classified oil palm fruits into

four maturity categories: unripe, underripe, ripe, and overripe, while providing reliable object localization through bounding box detection. In addition, deployment testing on an NVIDIA GeForce RTX 3050 4 GB GPU achieved an inference time of approximately 25–28 ms per image and a processing speed of 10–14 FPS, indicating its suitability for near-real-time agricultural inspection applications.

Despite these promising results, several limitations of this study should be acknowledged. The model was trained and evaluated using a dataset collected under limited environmental conditions and may not fully represent the variability encountered in large-scale plantation operations. In addition, this study focused solely on the RT-DETR-L architecture and did not include a direct comparison with other state-of-the-art object detection models under identical experimental settings. Therefore, further validation is required to assess the generalizability of the proposed approach across various environments and datasets.

Future research should focus on expanding the dataset with more diverse environmental conditions, lighting variations, and acquisition devices. Comparative evaluations with alternative object detection architectures, such as YOLO-based and other transformer-based models, are also recommended. Furthermore, integration with embedded hardware platforms and automated conveyor-based sorting systems should be investigated to support practical deployment in industrial oil palm processing environments while maintaining the real-time performance requirements.

ACKNOWLEDGMENT

The author would like to express sincere gratitude to the Institut Teknologi Sawit Indonesia for providing institutional support, research facilities, and assistance throughout the data collection and system testing processes, which contributed significantly to the successful completion of this study. We also extend our appreciation to all individuals and parties who assisted in dataset preparation, data processing, and model evaluation during the research.

This research did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors. The author declares that there are no conflicts of interest regarding the publication of this study.

AUTHOR CONTRIBUTION STATEMENT

NH conceived and designed the study, collected and prepared the dataset, performed data annotation and preprocessing, developed and trained the RT-DETR model, conducted the experiments, analyzed the results, and wrote the manuscript. RMS contributed to the research design, supervised the methodology, validated the experimental results, assisted with data interpretation, and critically reviewed and revised the manuscript. RAS contributed to the research supervision, provided technical guidance on the implementation of the RT-DETR model, evaluated the research findings, and reviewed and approved the final manuscript. All authors have read and approved the final manuscript.

AI DISCLOSURE STATEMENT

The authors used ChatGPT, developed by OpenAI, during the preparation of this manuscript for language editing, grammatical improvement, and manuscript refinement. After using this tool, the authors carefully reviewed, revised, and validated all the content to ensure its accuracy, originality, and scientific integrity. The authors take full responsibility for the content of this manuscript.

REFERENCES

[1] Direktorat Jenderal Perkebunan Kementerian Pertanian Republik Indonesia, "Direktorat Jenderal Perkebunan Kementerian Pertanian Republik Indonesia," in *Statistik Perkebunan Jilid 1 2023-2025*, 2024, p. 90.

[2] Kementerian Pertanian Republik Indonesia, "Outlook Komoditas Perkebunan Kelapa Sawit," *Pusat Data dan Sistem Informasi Pertanian Sekretariat Jenderal Kementerian Pertanian*, pp. 1–78, 2024.

[3] R. A. Sirait, G. Supriyanto, and Priyambada, "Pengaruh Kematangan Buah terhadap FFA dan Besarnya Kandungan Minyak di Dalamnya Di Pabrik Kelapa Sawit," *Journal Agroforetech*, vol. 1, no. 1, 2023, [Online]. Available: https://jurnal.instiperjogja.ac.id/index.php/JOM/article/view/394

[4] R. Triyogi, R. Magdalena, and B. Hidayat, "Mendeteksi Kematangan Buah Kelapa Sawit Menggunakan Convolution Neural Network Deep Learning," vol. 1, no. 1, pp. 22–27, 2023, doi: doi:http://doi.org/10.25124/logic.v1i1.6732.

[5] R. Kurniawan and Nurahman, "Klasifikasi Tingkat Kematangan Buah Sawit Berdasarkan Ekstraksi Fitur RGB dan GLCM Menggunakan Algoritma K-NN," vol. 22, pp. 457–466, 2023, doi: https://doi.org/10.32409/jikstik.22.4.3402.

[6] Z. Genoveva and R. D. Syah, "MODEL MACHINE LEARNING UNTUK DETEKSI TINGKAT KEMATANGAN TANDAN BUAH SEGAR KELAPA SAWIT MENGGUNAKAN METODE YOLOV8 Machine," vol. 8, pp. 121–136, 2024.

[7] M. Y. M. A. Mansour, K. D. Dambul, and K. Y. Choo, "Object Detection Algorithms for Ripeness Classification of Oil Palm Fresh Fruit Bunch," *International Journal of Technology*, vol. 13, no. 6, pp. 1326–1335, 2022, doi: 10.14716/ijtech.v13i6.5932.

[8] R. R. Dewi, I. Hermawan, and D. Abera, "Evaluation of YOLOv11 and RF-DETR Architectures for UAV-Based Detection of Healthy Oil Palm Trees," in *2025 8th International Conference of Computer and Informatics Engineering (IC2IE)*, IEEE, Sep. 2025, pp. 1–8. doi: 10.1109/IC2IE67206.2025.11283417.

[9] H. M. Hutajulu, A. Yanie, L. Adriana, and D. Safitri, "Rancang Bangun Deteksi Kematangan Buah Kelapa Sawit Dan Peringatan Berbasis Telegram," *Prosiding Seminar Nasional Teknik (Semnastek) UISU 2023: Peran Teknologi Berkelanjutan dalam Era Disrupsi*, pp. 207–213, 2023.

[10] D. S. Prasvita, A. M. Arymurthy, and D. Chahyati, "Deep Learning Model for Automatic Detection of Oil Palm Trees in Indonesia with YOLO-V5," in *Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology*, New York, NY, USA: ACM, Oct. 2023, pp. 39–44. doi: 10.1145/3626641.3626924.

[11] D. N. Huda, M. R. Romdoni, L. Safitri, A. Winarni, and A. Rahman, "Real-time Detection Transformer (RT-DETR) of Ornamental Fish Diseases with YOLOv9 using CNN (Convolutional Neural Network) Algorithm," vol. 8, no. 2, pp. 463–471, 2024, doi: https://doi.org/10.30871/jaic.v8i2.8561.

[12] N. R. R. Wati, A. Abdullah, and Sucipto, "Klasifikasi Tingkat Kematangan Buah Kelapa Sawit Menggunakan Metode Convolutional Neural Network," vol. 4, no. September, 2025, doi: https://doi.org/10.55606/jupti.v4i3.5841.

[13] S. I. Guslianto and S. 'Uyun, "Klasifikasi Kematangan Buah Sawit Berdasarkan Fitur Warna, Bentuk dan Tekstur Menggunakan Algoritma K-NN," vol. 9, no. 3, pp. 407–414, 2023, doi: https://doi.org/10.26418/jp.v9i3.64877.

[14] A. Elumalai and V. Jeganathan, "Real-Time Classifications of Pests from Agriculture Industry Based on Their Color and Texture Features by Using Machine Learning Models," vol. 5, no. 2, pp. 9–16, 2024, Accessed: Jun. 10, 2026. [Online]. Available: https://journal.ijprse.com/index.php/ijprse/article/view/1013

[15] R. P. Putra, J. Jumadi, and D. Lianda, "Pengolahan Citra Digital Untuk Mengidentifikasi Tingkat Kematangan Buah Kelapa Sawit Berdasarkan Warna Rgb Dan Hsv Dengan Menggunakan Metode Self Organizing Map (SOM)," vol. 20, no. 1, pp. 98–105, 2024.

[16] M. Shiddiq, R. Salambue, F. Wardana, V. V. Dasta, and I. Okta, "Multispectral imaging and deep learning for oil palm fruit bunch ripeness detection," vol. 13, no. 6, pp. 4168–4181, 2024, doi: 10.11591/eei.v13i6.8120.

[17] Q. P. Ramadhani and F. Dewanta, "Palm Fruit Ripeness and Quality Detection System Using YOLOv11," *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media*, vol. 2025, no. 1, pp. 163–168, Jan. 2026, doi: 10.20895/centive.v2025i1.537.

[18] Y. Zhao *et al.*, "DETRs Beat YOLOs on Real-time Object Detection," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 16965–16974, 2024, doi: 10.1109/CVPR52733.2024.01605.

[19] S. Wang, C. Xia, F. Lv, and Y. Shi, "RT-DETRv3: Real-time End-to-End Object Detection with Hierarchical Dense Positive Supervision", doi: https://doi.org/10.48550/arXiv.2409.08475.

[20] L. A. Bustamante and J. C. Gutierrez, "Enhancing Real-Time Detection Transformer (RT-DETR) for Handgun Detection on Nvidia Jetson," vol. 28, no. 3, pp. 1–19, 2025, doi: 10.1109/CLEI64178.2024.10700426.

[21] B. Ath Thariq Syams, R. Mutiara Siregar, and M. Akbar Syahbana Pane, "Classification of Plant Pests Using the Real-Time Detection Transformer (RT-DETR) Algorithm in Oil Palm Plants," *Sisfo: Jurnal Ilmiah Sistem Informasi*, vol. 10, no. 1, pp. 198–204, May 2026, doi: 10.29103/sisfo.v10i1.27029.

[22] T. Selvakumar, M. N. A. H. Sha'abani, M. S. M. Aras, and M. B. Bahar, "Development of a YOLOv11-Based Deep Learning System for Insect Pest Detection and Classification in Oil Palm Plantation," *Mod. Appl. Sci.*, vol. 19, no. 2, p. 32, Oct. 2025, doi: 10.5539/mas.v19n2p32.

[23] K. H. Ghazali, S. Riyadi, A. D. Andriyani, R. Muda, A. Lubis, and S. Yan, "RT-DETR-Palm A Transformer Based Approach for Oil Palm Tree Detection Using UAV Imagery," in *2025 International Conference on Advanced Technologies in Energy and Informatic (ICATEI)*, IEEE, Oct. 2025, pp. 447–452. doi: 10.1109/ICATEI67676.2025.11404965.

[24] E. Bagus Nugroho, F. Akhyar, and L. Novamizanti, "Metode Deteksi Obyek Berbasis Computer Vision dan Deep Learning untuk Sistem Inspeksi Cacat pada Permukaan Printed Circuit Board (PCB) ," *e-Proceeding of Engineering*, vol. 11, no. 2, p. 801, 2024.

[25] Rizky Delianngi, Ratu Mutiara Siregar, Nurliana, Muhammad Akbar Syahbana Pane, Phaklen Ehkan, and Andi Prayogi, "Performance Evaluation of YOLOv9, YOLOv10, and YOLOv11 for Real-Time Early Detection of Ganoderma Boninense in Oil Palm," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 10, no. 2, pp. 429–440, Apr. 2026, doi: 10.29207/resti.v10i2.7479.

[26] Y. Chen *et al.*, "YOLO-MS: Rethinking Multi-Scale Representation Learning for Real-Time Object Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 47, no. 6, pp. 4240–4252, Jun. 2025, doi: 10.1109/TPAMI.2025.3538473.