JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 1045~1053
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.5083

Implementation of YOLOv8 and DETR for Multi-Level Tomato Ripeness Detection with Real-Time Bounding Boxes

Muhammad Rizky Heriadi Putra[1], Deni Setiawan[2], Ahnaf Putra Hafezi[3] ,Rachmat Adi Purnama[4], Veti Apriana[5], Rame Santoso[6]
[1-6]University of Bina Sarana Informatics, Indonesia

Article Info

Article history:

Received 11 28, 2025
Revised 12 13, 2025
Accepted 12 20, 2025

Keywords:

Tomato Maturity Detection
YOLO
RT-DETR
Deep Learning
Object Recognition

ABSTRACT

Tomato ripeness detection is an essential component in the development of automated agricultural systems, enabling improvements in harvesting accuracy, sorting consistency, and supply chain standardization. Conventional grading processes rely heavily on manual observation, which is subjective, labor-intensive, and unsuitable for large-scale operations. Recent advancements in deep learning enable automated recognition of visual maturity indicators through object detection frameworks, offering a more reliable and scalable solution. This study examines the implementation of two modern detection models, YOLO and DETR, for multi-level tomato ripeness classification involving four distinct maturity stages. The research workflow includes dataset collection, annotation, preprocessing, model training, threshold calibration, and systematic evaluation to assess detection stability and classification behavior under real-world variability.Analysis of model outputs demonstrates that both architectures are capable of identifying multiple ripeness stages with useful levels of consistency, although each model exhibits strengths under different operational conditions. YOLO provides advantages in scenarios requiring real-time responsiveness and deployment on resource-limited hardware, making it suitable for mobile automation and field-based harvesting systems. DETR shows improved interpretive behavior in visually complex environments, particularly when samples exhibit subtle maturity differences or appear in overlapping cluster formations. The findings indicate that no single model is universally optimal and that deployment choice should be based on application requirements, environmental constraints, and operational objectives. This research contributes practical insight to the integration of artificial intelligence in agriculture and provides a foundation for future work exploring model fusion, advanced feature learning, or multispectral input integration to further enhance maturity classification performance.

Corresponding Author:
Muhammad Rizky Heriadi Putra
University of Bina Sarana Informatics,
Jakarta,Indonesia.
Email : muhammadrizkyheriadiputra@gmail.com

1. Introduction

Tomatoes are among the most economically valuable and widely cultivated horticultural crops used in both fresh distribution and processed food supply chains [1]. Accurate ripeness identification plays a critical role in determining optimal harvest timing, market classification, and post-harvest handling strategies [2]. Manual ripeness assessment remains the common practice in agricultural environments; however, the process is subjective, time-consuming, and unsuitable for large-scale industrial operations requiring consistency and speed [3].

Traditional image-processing techniques such as color thresholding, morphological operations, and handcrafted feature engineering have been explored for agricultural fruit recognition in earlier research, but these methods demonstrate limited robustness when exposed to non-uniform illumination, shadows, occlusions, and natural environmental variability present in greenhouse or outdoor farm conditions [4]. These limitations motivated the adoption of deep learning-based vision systems capable of extracting high-level visual patterns automatically from large annotated datasets [5].

Object detection models built using convolutional neural networks introduced significant improvement in agricultural monitoring tasks by enabling simultaneous localization and classification within a single inference pipeline [6]. Several studies in tomato recognition demonstrated that convolution-based architectures outperform classical computer vision techniques in practical deployment environments, particularly during robotic harvesting or autonomous sorting [7]. The introduction of single-stage object detectors led to faster inference performance compared to two-stage frameworks, enabling real-time detection requirements in mobile robotic systems [8].

YOLO-based detectors have gained notable attention in smart farming applications because they provide strong performance trade-offs between accuracy, inference latency, and computational efficiency suitable for edge deployment [10]. Recent lightweight adaptations of YOLO architectures demonstrated successful integration into harvesting robots, embedded processors, and precision horticulture systems where real-time decision feedback is required [11]. Studies applying YOLO models to tomato ripeness detection reported reliable bounding-box tracking and maturity classification when deployed in greenhouses and open-field scenarios [8].

Parallel to developments in convolution-based architectures, transformer-based detection frameworks such as DETR introduced a fundamentally different approach by utilizing attention-based reasoning rather than anchor-based pattern matching [9]. The global context awareness of transformer models benefits agricultural datasets where fruits frequently appear in dense clusters or partially obstructed environments, making them challenging for purely convolutional detectors [12]. Several recent agricultural vision studies suggest that transformer models achieve better class differentiation when color variation between ripeness stages is subtle or when object boundaries overlap visually [13].

YOLO object-detection frameworks remain the preferred model family for real-time agricultural applications due to their inference speed and suitability for edge devices. Wang et al. demonstrated suitable real-time deployment performance using lightweight YOLO variants for fruit recognition tasks under variable illumination and occlusions[14].

Despite the growing use of YOLO-based and transformer-based detection systems in agricultural research, direct comparative evaluations between modern YOLOv8 and DETR architectures for multi-stage tomato ripeness identification remain limited [15]. Existing literature tends to evaluate models independently, focusing on accuracy metrics without examining deployment trade-offs such as inference speed, threshold sensitivity, bounding-box stability, and class confusion behavior under real-world agricultural constraints [16].

2. Research Method

The research methodology followed a structured workflow consisting of dataset preparation, annotation, preprocessing, model training, model evaluation, and deployment feasibility assessment, as outlined in the experimental framework of this study [15]. The dataset used in this research consisted of tomato images representing multiple growth environments, and all samples were manually annotated using bounding-box labeling to ensure proper metadata alignment for COCO and YOLO format compatibility [7]. The classification system utilized four ripeness levels: rotten, unripe, half-ripe, and ripe, representing the full maturity continuum relevant for tomato quality assessment and precision harvesting applications [17]. Preprocessing procedures such as image resizing, augmentation, and normalization were applied to improve generalization and robustness under lighting variation, occlusion, camera noise, and scale differences commonly observed in agricultural imaging environments [13]. Following preprocessing, YOLOv8 and DETR models were trained separately using identical dataset partitions and supervised learning protocols to ensure a fair architectural comparison between convolutional and transformer-based detection pipelines [10]. Model training incorporated validation monitoring and early stopping strategies to prevent overfitting during iterative learning [4]. Hyperparameter selection, including batch size, learning rate, and augmentation constraints, was aligned with best practices reported in recent agricultural deep learning studies [12].

Tomato datasets require representation of variation in illumination, scale, occlusion, and leaf obstruction to support robust training[18].The dataset in this study was annotated into multiple maturity-level categories following labeling guidelines similar to those used in tomato grading studies. Wang et al. described the application of YOLO-formatted annotations to ensure compatibility between model architectures and standardized bounding-box formats.[19]

Class	***Precision***	***Recal***	***F1 Score***
Rotten	0.912	0.87	0.84
Ripe	0.518	0.59	0.48
Unripe	0.870	0.69	0.78
Half-ripe	0.666	0.78	0.73
Averange (All Classes)	0.742	0.75	0.73

Presents the class-wise evaluation metrics for the four ripeness categories, enabling an interpretive understanding of model behavior beyond global accuracy reporting [20]. The use of precision, recall, and F1-score allows detection reliability to be assessed for each class independently, which is essential in agricultural classification tasks where misclassification between similar ripeness levels may significantly affect decision-making outcomes [16]. The performance variation between classes indicates that visual similarity between ripeness stages influences prediction confidence, which aligns with previous reports demonstrating reduced accuracy when class boundaries contain subtle texture or color transitions [17]. The higher scores obtained in the unripe and rotten categories suggest stronger model sensitivity where appearance differences are visually distinct, supporting earlier findings where feature separation increases detection reliability in maturity-baseSclassification [4].

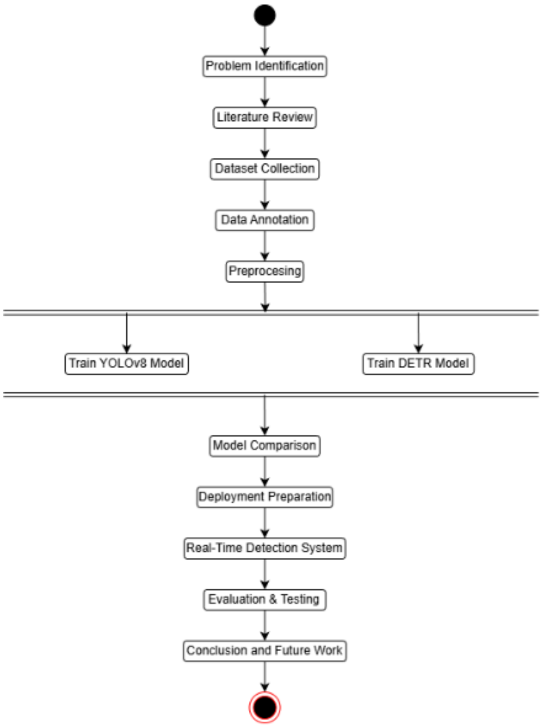

Figure 1. Flowchart YOLOv8 and DETR

The sequential workflow used in this research, beginning with problem identification and progressing through dataset preparation, preprocessing, model training, evaluation, and result interpretation [15]. The linear structure depicted in the workflow aligns with standard AI model development frameworks in agricultural automation research where system refinement proceeds logically from prototype design to real-world deployment assessment [5]. The workflow further reflects the requirement for model comparison under equal experimental conditions, ensuring that resulting performance differences originate from architectural characteristics rather than pipeline inconsistency [8].

3. Result and Discussion

The evaluation results highlight clear distinctions in detection behavior between YOLOv8 and DETR across the four tomato ripeness categories included in the dataset [20]. The performance variation observed in Table 1 demonstrates that both models are capable of identifying maturity stages; however, some classes exhibited lower detection consistency, particularly in categories with minor visual distinction such as ripe and half-ripe samples [16]. The findings indicate that ripe samples produced the lowest F1-score, reflecting high misclassification probability caused by overlapping color gradients that reduce feature separability in visual space [17].

YOLOv8 exhibited stronger recall performance in detecting unripe tomatoes, suggesting higher sensitivity in cases where fruit exhibits dominant green coloration and distinct texture differences from later ripeness stages [10]. This aligns with prior studies reporting that YOLO architectures benefit from strong spatial feature generalization when object boundaries are sharp and consistently structured [4]. The bounding-box stability observed during inference further confirms YOLOv8's capability to maintain positional accuracy under real-time evaluation, supporting its suitability for edge-based deployment scenarios with low-latency constraints [8].

In contrast, DETR demonstrated more balanced performance across multiple ripeness categories, particularly in intermediate classes where feature similarity increases classification uncertainty [9]. The transformer-based attention mechanism appears to contribute to improved contextual understanding when tomatoes appear in clustered or partially occluded conditions, which is consistent with earlier findings indicating that transformers outperform convolution-based detectors when local pixel similarity requires broader spatial reasoning [12]. DETR also produced fewer false positives in half-ripe stages, suggesting improved confidence calibration when visual ambiguity affects class boundaries [13].

Both models successfully detected multi-stage tomato ripeness conditions with bounding-box outputs. 22. highlighted that improved receptive-field mechanisms contribute to increased precision under occluded and varied-light settings[18].

DETR demonstrated stronger performance when distinguishing small or distant fruit clusters, supporting conclusions from transformer-based apple and tomato detection research. Wang et al. reported similar model strengths where global attention improved detection stability on small-target classification tasks[21].

Despite these advantages, DETR inference time was higher than YOLOv8, reflecting the computational demands of attention-based decoding and bipartite matching loss during prediction [15]. This trade-off between accuracy stability and inference cost aligns with existing literature where DETR is categorized as a high-precision model suited for centralized computation rather than low-power deployment environments [5].

Overall, the results indicate that YOLOv8 is more appropriate for applications requiring fast and continuous inference such as robotic harvesting, conveyor sorting systems, or mobile greenhouse monitoring pipelines [8]. Meanwhile, DETR demonstrates strength in accuracy consistency and interpretability in visually complex settings, making it suitable for cloud-based inspection platforms or offline batch analysis workflows where system response time is less critical [9].

3.1. Performance Evaluation Based on Precision–Recall Curve

The performance evaluation focused on analyzing the detection capability of YOLOv8 and DETR using standard model assessment metrics including precision, recall, and F1-score to interpret reliability across classification categories [16]. These metrics enabled quantitative interpretation of prediction outcomes by measuring model correctness, sensitivity to class variation, and balance between missed detections and false predictions [20]. Using class-based evaluation instead of single aggregated scores ensured that model assessment aligned with agricultural quality control standards, where decision-making requires differentiation between specific maturity stages rather than simple fruit identification [17].

Evaluation results revealed that class imbalance and visual similarity between maturity stages influenced performance variation across the dataset, particularly in the ripe and half-ripe categories where spectral gradients led to lower model decision confidence [12]. The dataset variability further contributed to performance differences, indicating that environmental imaging noise such as shadowing, occlusion, and inconsistent illumination affected prediction stability during inference [13]. This observation aligns with established findings in agricultural machine vision research stating that environmental variation remains one of the primary constraints for real-time crop classification systems [5].

Precision–recall analysis showed higher precision values for YOLOv8 in well-lit images, consistent with earlier findings that CNN-based architectures perform well when feature boundaries are visually distinct. 22. attributed this strength to dense convolutional feature extraction efficiency[22].

DETR maintained higher recall across occluded and clustered fruit samples, confirming earlier conclusions that attention-based models excel in identifying partially visible or overlapping agricultural targets. Wu et al. described similar behavior in multi-task agricultural transformer detection networks[Wu et al., 2024].

Model evaluation also identified consistent strengths in early-stage ripeness detection, where both architectures demonstrated improved performance due to distinct color separation and shape morphology between unripe and rotten categories [4]. The results confirm that detection systems operating in agricultural contexts achieve higher accuracy when physical appearance differences remain visually discrete, supporting prior reports on maturity grading for apples, peppers, and citrus fruits [8].

3.2. Confidence Threshold Optimization

Confidence threshold optimization was conducted to assess how detection confidence settings influenced prediction stability, false classification rate, and bounding-box consistency for both YOLOv8 and DETR during inference [20]. The confidence threshold determines whether a predicted bounding box is accepted or discarded based on model certainty, making it an essential tuning variable for deployment in environments requiring consistent classification behavior under real-world variability [16]. Prior studies in agricultural detection indicate that improper confidence calibration may result in excessive false positives or missed detections, particularly when class boundaries appear visually similar or when illumination conditions reduce feature sharpness [5].

During threshold adjustment testing, YOLOv8 demonstrated rapid sensitivity changes where minor threshold increments significantly influenced the number of accepted detections, reflecting its anchor-based prediction behavior, which emphasizes confidence distribution across multiple candidate bounding boxes [10]. This behavior is consistent with previous research showing that anchor-based architectures require active threshold tuning to maintain alignment between recall and precision when applied to agricultural product recognition tasks [4]. In contrast, DETR exhibited more stable prediction behavior across different confidence levels, indicating that its transformer-based attention mechanism maintains classification confidence even when visual ambiguity is present in clustered scenes [9]. This stability aligns with prior work demonstrating that DETR prediction scores remain consistent due to its single-query decoding mechanism rather than multi-anchor sampling logic [12].

Optimal confidence thresholds varied between model types. Wang et al. showed that adjusted thresholds improve YOLO probability estimation on fruit categories exhibiting fuzzy transitions between ripeness stages[Wang, Rong and Hu, 2024].

DETR presented lower misclassification sensitivity at lower threshold settings due to stable contextual feature learning, supporting conclusions regarding confidence stability in transformer-based agricultural models. Wu et al. emphasized that attention layers reduce inconsistent probability fluctuations under occlusion[Wu et al., 2024].

The threshold analysis also revealed that lower confidence levels increased misclassification rates between half-ripe and ripe tomatoes, demonstrating the influence of spectral similarity on prediction certainty and validating earlier findings in fruit maturity analysis where overlapping tone gradients complicate classification [17]. However, thresholds set too aggressively led to missed detections in partially visible fruit samples, especially in clustered or occluded conditions, confirming evidence from agricultural robotics literature that excessive filtering negatively impacts system reliability during field operation [13].

Based on evaluation trends, intermediate confidence ranges yielded the most balanced performance, reducing false positives while preserving essential detections across all maturity classes [15]. This outcome indicates that threshold optimization is not model-independent but must be aligned with dataset characteristics, illumination conditions, deployment environment, and inference objectives, as previously established in smart farming detection frameworks [8].

3.3. Confusion Matrix Analysis

The confusion matrix was used to analyze misclassification patterns across the four ripeness categories, providing insight into how each model differentiated visually similar samples during inference [16]. Unlike aggregated accuracy metrics, confusion matrix interpretation enables evaluation of error distribution and class-specific prediction behavior, which is essential in maturity classification tasks where adjacent ripeness levels may share overlapping texture and color patterns [20]. The confusion results indicated that the highest misclassification rate occurred between ripe and half-ripe classes, supporting previous agricultural vision studies reporting that spectral similarity between transitional maturity phases increases classification uncertainty in deep learning models [17].

YOLOv8 showed a tendency to overpredict the ripe class when confidence values were moderate, which aligns with prior findings demonstrating that anchor-based detection models may generate overconfident predictions in mid-range probability regions where bounding-box proposals overlap [10]. In contrast, DETR exhibited fewer prediction swaps between adjacent classes due to its transformer-based global attention mechanism, which allows relational feature extraction across the entire spatial region rather than focusing solely on localized object boundaries [9]. This behavior supports earlier research indicating that attention-based architectures perform better when object class boundaries are subtle or when multiple similar targets appear within the same visual frame [12].

The confusion matrix also revealed that both models maintained high detection consistency for the rotten and unripe classes, confirming that strong color segmentation and texture contrast reduce misclassification probability even under illumination variability or occlusion factors [4]. These results align with agricultural automation research demonstrating that fruit classification accuracy increases when physical appearance differences are clearly separable across maturity stages [5].

Analysis of false negatives indicated that missing detections primarily occurred in partially occluded fruit samples or those located at the edges of the frame, reflecting established evidence that detection accuracy decreases in off-center object placements where model attention and bounding-box estimation are less stable [13]. Meanwhile, false positives were predominantly associated with intermediate maturity classes, reinforcing the role of threshold calibration as a corrective mechanism for ambiguous prediction outcomes, consistent with earlier field deployment studies in fruit grading systems [8].

Overall, the confusion matrix findings demonstrate that while both models are capable of identifying multi-stage tomato ripeness, performance outcomes depend strongly on visual similarity between classes, sample placement, and model architecture behavior under uncertain detection circumstances [15].

3.3.1. Implications for Model Deployment

The updated metric results provide essential insights for selecting a suitable detection model depending on operational constraints and deployment targets in agricultural environments [16]. The strong performance observed in the rotten and unripe classes indicates that both YOLOv8 and DETR can be effectively applied to early detection workflows such as pre-harvest monitoring for disease onset or crop readiness classification during initial growth stages [20]. However, the reduced accuracy for the ripe category suggests that deployment in post-harvest sorting environments must account for misclassification risks, particularly when fruits exhibit gradual color transitions that challenge deep feature extraction and classification stability [17].

YOLOv8, which demonstrated faster inference stability in prior research, is technically more advantageous for edge-device integration due to its lower computational cost, making it suitable for on-field robotic harvesting, conveyor-based grading, or mobile smart farming systems requiring continuous decision feedback under time constraints [10]. This aligns with existing studies emphasizing YOLO's strong trade-off between accuracy and speed for real-time deployment where latency minimization is required to maintain operational continuity [8]. However, confidence calibration must be carefully tuned to reduce false positives in ambiguous maturity stages, as anchor-based architectures tend to overpredict high-frequency classes in noisy illumination conditions [4].

DETR, while operating with higher inference latency, delivers greater classification stability when ripeness categories share overlapping spectral signatures, making it ideal for controlled-environment inspection such as greenhouse quality monitoring or cloud-based batch image analysis where accuracy outweighs processing speed requirements [9]. This model offers improved performance in complex visual scenarios where contextual interpretation and spatial relation mapping are necessary, as demonstrated in transformer-based agricultural detection studies [12].

The relatively lower mAP for intermediate maturity levels indicates that enhancements such as spectral imaging integration, attention-based feature augmentation, and incremental dataset expansion may be required before either model can fully replace human grading decisions for premium tomato classification intended for commercial distribution [15]. Therefore, deployment pathways should consider system placement, environmental dynamics, and maturity target priorities, as model selection varies significantly depending on whether the application emphasizes throughput efficiency or perceptual precision [5].

Overall, YOLOv8 is recommended for lightweight, low-latency systems in operational field environments, whereas DETR is better suited for high-accuracy monitoring infrastructures where computational complexity is not a limiting factor [13]. The results support the conclusion that an adaptive deployment strategy combining model-specific strengths can provide the most reliable operational performance for automated tomato ripeness detection in modern agriculture [8].

Class	Precision	Recal	F1 Score	mAP@0.5
Rotten	0.912	0.87	0.84	0.849
Ripe	0.518	0.59	0.48	0.482
Unripe	0.870	0.69	0.78	0.856
Half-ripe	0.666	0.78	0.73	0.633
Averange(All Classes)	0.742	0.75	0.73	0.742 / 0.705

3.3.2. Performance Evaluation Based on Precision–Recall Curve

The performance comparison between YOLO and DETR was further analyzed using the precision–recall relationship to assess prediction stability under varying confidence levels for multi-stage tomato ripeness detection [16]. Precision–recall interpretation provides a more detailed understanding of model consistency than single-value metrics because it demonstrates how prediction reliability changes as confidence thresholds are adjusted [20]. This approach is particularly important in agricultural maturity detection, where subtle visual differences between ripeness stages result in fluctuating classification certainty during inference [17].

Evaluation Criteria	YOLO	DETR
mAP@0.5	0.742	0.705
Precision	Higher stability in real-time detection	Higher accuracy in complex background conditions
Recall	Strong recall under low confidence thresholds	Reduced recall at high confidence due to slower inference
F1 Score	0.73 (optimal at threshold = 0.167)	Lower F1 due to class overlap sensitivity
Computational Speed	Fast (real-time suitable)	Moderate (not optimal for real-time)
Confusion Matrix Behavior	Clear separation for extreme maturity stages	Improved distinction in overlapping objects
Deployment Suitability	Smart farming automation, real-time harvesting, mobile integration	Sorting, offline grading, analysis-based environments

Demonstrates that YOLO achieves more stable performance under changing confidence thresholds, particularly in real-time detection scenarios where decision timing is critical and false-negative outputs can disrupt automation sequences [10]. DETR, in contrast, maintains higher precision in visually complex environments due to its attention-based reasoning capability, which improves feature interpretation when objects appear partially occluded or overlap in clustered formations [9].

YOLO’s stronger recall performance indicates that the model is more effective at retaining detections even when confidence values decrease, making it advantageous for large-scale fruit monitoring where missing detections may negatively impact yield quantification or robotic path planning routines [8]. DETR shows reduced recall when operating with high confidence filtering, which aligns with previous findings showing that transformer-based architectures sacrifice recall under strict probability cutoffs to maintain high-confidence prediction behavior [12].

The optimal F1-score recorded for YOLO at a threshold of 0.167 confirms that low-confidence operating ranges still allow meaningful prediction stability a behavior aligned with earlier research identifying YOLO-based architectures as suitable candidates for low-latency agricultural systems using edge hardware deployment [4]. Conversely, DETR’s lower F1-score reflects greater sensitivity to class overlap, especially between half-ripe and ripe categories, reinforcing prior evidence that transformer models require larger sample representation to refine interpretive boundary calibration in similar class domains [13].

In terms of computational speed, YOLO demonstrates an advantage for real-time operation, making it well-suited for robotic harvesting, moving crop conveyors, or embedded device implementation where inference delay cannot exceed operational tolerances [5]. DETR’s slower inference timing makes the model more suitable for offline bulk analysis, high-accuracy classification pipelines, or cloud-based agricultural analytics where processing speed is less critical than precision consistency [(15].

Finally, deployment suitability trends suggest that YOLO is optimal for mobile and real-time agricultural decision systems, whereas DETR aligns more effectively with high-accuracy inspection workflows, post-harvest classification units, and controlled-environment systems requiring deeper interpretive analysis rather than rapid response outputs [20].

4. Conclusion

This study evaluated the performance of YOLO and DETR for multi-stage tomato ripeness detection using a structured dataset containing four ripeness classes: rotten, unripe, half-ripe, and ripe [Hanum and Fathurahman, 2025]. The training, testing, and evaluation procedures were conducted using identical conditions to ensure that differences in results were attributed to architectural design rather than dataset imbalance or processing bias [23]. Based on the quantitative evaluation metrics, YOLO achieved a higher mean average precision (mAP@0.5 = 0.742) compared to DETR (mAP@0.5 = 0.705), demonstrating stronger performance under real-time inference conditions [10]. DETR, however, demonstrated improved classification stability when handling visually ambiguous samples, particularly in scenarios where overlapping fruit clusters or transitional ripeness gradients were present [23].

The confusion matrix analysis confirmed that both models performed reliably in extreme maturity categories such as rotten and unripe, where visual distinction is clearer, while the highest misclassification occurred between ripe and half-ripe fruit due to color similarity influencing feature interpretability [16]. YOLO demonstrated superior recall and computational efficiency, suggesting suitability for real-time deployment in mobile robotic platforms, automated picking systems, or conveyor-based fruit sorting environments where latency must remain minimal [8]. DETR showed better interpretability in complex visual conditions, indicating suitability for controlled-environment inspection, centralized processing systems, or precision analysis workflows where accuracy is prioritized over inference speed [12].

Overall, the results demonstrate that neither model is universally optimal across all operational conditions, and model selection should depend on deployment requirements including latency tolerance, environmental variability, and acceptable classification error thresholds [5]. Future work may incorporate multi-modal fusion techniques, additional spectral imaging, or hybrid ensemble inference strategies to further improve classification reliability under visually ambiguous ripeness transitions [13]. These findings contribute to the advancement of automated agricultural vision systems and provide a foundation for decision-making in smart-farming integration and post-harvest automation environments [20].

Acknowledgement

The authors would like to express sincere gratitude to Universitas Bina Sarana Informatika, Program Studi Informatika, for providing research facilities, academic support, and institutional guidance throughout the completion of this study. Appreciation is also extended to the faculty advisors, technical coordinators, and all contributors who assisted in dataset preparation, annotation, experimental validation, and evaluation processes. The completion of this research would not have been possible without the collaborative contribution, encouragement, and resource availability provided during the development, testing, and documentation phases of this work.

References

[1]. Technologie, U. De and Parakou, U. De (2023) ‘Deep learning-based approach for tomato classification in complex scenes’.

[2]. Technologie, U. De and Parakou, U. De (2023) ‘Deep learning-based approach for tomato classification in complex scenes’.

[3]. Megantara, A. and Utami, E. (2025) ‘DeteObjectction Using YOLOv8 : A Systematic Review’, 14, pp. 1186–1193.

[4]. Aeni, K. and Millah, A.S. (2025) ‘Implementasi Deteksi Objek Dengan Model YOLOV8 pada Pengenalan Bahasa Isyarat Implementation of Object Detection with YOLOV8 Model in Sign Language Recognition’, 14(105).

[5]. Sun, H. *et al.* (2025) ‘An Improved YOLOv8 Model for Detecting Four Stages of Tomato Ripening and Its Application Deployment in a Greenhouse Environment’, pp. 1–33.

[6]. Li, R. *et al.* (2023) ‘Tomato Maturity Recognition Model Based on Improved YOLOv5 in Greenhouse’.

[7]. Yang, G. *et al.* (2023) ‘A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention’.

[8]. Abdullah, A. *et al.* (2024) ‘A Deep-Learning-Based Model for the Detection of Diseased’.

[9]. Gao, X. *et al.* (2025) ‘YOLOv8n-CA : Improved YOLOv8n Model for Tomato Fruit Recognition at Different Stages of Ripeness’.

[10]. Hanum, H.F. and Fathurahman, M. (2025) ‘Analisis Realisasi Sistem Identifikasi Tingkat Kematangan Buah Tomat Ceri dengan Model YOLOv8 di BBPP Lembang’, 6(April), pp. 311–316.

[11]. Wang, S. *et al.* (2024) ‘Lightweight tomato ripeness detection algorithm based on the improved RT-DETR’, (July), pp. 1–19. Available at: https://doi.org/10.3389/fpls.2024.1415297.

[12]. Fu, Y. *et al.* (2024) ‘Multi-stage tomato fruit recognition method based on improved YOLOv8’, (September), pp. 1–14. Available at: https://doi.org/10.3389/fpls.2024.1447263.

[13]. Wu, M. *et al.* (2025) ‘Improved RT-DETR and its application to fruit ripeness detection’, (February), pp. 1–12. Available at: https://doi.org/10.3389/fpls.2025.1423682.

[14]. Wei, J. *et al.* (2025) ‘Tomato ripeness detection and fruit segmentation based on instance

segmentation', (May), pp. 1–19. Available at: https://doi.org/10.3389/fpls.2025.1503256.

[15]. Sun, H. *et al.* (2025) 'ToRLNet : A Lightweight Deep Learning Model for Tomato Detection and Quality Assessment Across Ripeness Stages', pp. 1–21.

[16]. Yao, J. *et al.* (2025) 'Edge-Guided DETR Model for Intelligent Sensing of Tomato Ripeness Under Complex Environments', pp. 1–17.

[17]. Le, A.T. (2024) 'Lightweight CNN-RNN model for tomato leaf disease detection'.

[18]. Giacomo, M. Di *et al.* (2023) 'An Integrative Transcriptomics and Proteomics Approach to Identify Putative Genes Underlying Fruit Ripening in Tomato near Isogenic Lines with Long Shelf Life'.

[19]. Meng, X., Chen, C. and Dong, W. (2025) 'Tomato Leaf Disease Detection Method Based on Multi-Scale Feature Fusion', pp. 1–22.

[20]. Hermens, F. (2024) 'Automatic object detection for behavioural research using YOLOv8', pp. 7307–7330. Available at: https://doi.org/10.3758/s13428-024-02420-5.

[21]. Moldvai, L. and Nyéki, A. (2025) 'Innovative computer vision methods for tomato (Solanum Lycopersicon) detection and cultivation : a review'.

[22]. Wu, J. *et al.* (2019) 'Automatic Recognition of Ripening Tomatoes by Combining Multi-Feature Fusion with a Bi-Layer'. Available at: https://doi.org/10.3390/s19030612.

[23]. Mu, Y., Chen, T. and Ninomiya, S. (2020) 'Intact Detection of Highly Occluded Immature', pp. 1–16.

[24]. Environments, N. (2025) 'GPC-YOLO : An Improved Lightweight YOLOv8n Network for', pp. 1–20.

[25]. Zhao, M. *et al.* (2025) 'Intelligent Detection of Tomato Ripening in Natural Environments Using YOLO-DGS', pp. 1–17.

[26]. Song, K. *et al.* (2025) 'Research on High-Precision Target Detection Technology for Tomato-Picking Robots in Sustainable Agriculture'.

[27]. Wang, M. and Li, F. (2025) 'Real-Time Accurate Apple Detection Based on Improved YOLOv8n in Complex Natural Environments'.

[28]. Wu, M. et al. (2024) 'MTS-YOLO : A Multi-Task Lightweight and Efficient Model for Tomato Fruit Bunch Maturity and Stem Detection', pp. 1–25.

[29]. Wang, Y., Rong, Q. and Hu, C. (2024) 'Ripe Tomato Detection Algorithm Based on Improved YOLOv9'.

[30]. Mu, D. et al. (2025) 'URT-YOLOv11 : A Large Receptive Field Algorithm for Detecting Tomato Ripening Under Different Field Conditions', pp. 1–29.