Journal of Informatics and Web Engineering

Vol. 5 No. 2 (June 2026) eISSN: 2821-370X

Performance Analysis of Faster R-CNN and YOLOv8 Model for Mango Fruits Detection

Mohd Haris Lye[1*], Marawan Ashraf Eldeib[2]**

[1]Faculty of Artificial Intelligence and Engineering, Multimedia University, Jalan Multimedia, 63100 Cyberjaya, Malaysia

[2]Faculty 5: Computer Science, Electrical Engineering and Information Technology, University of Stuttgart, 70569 Stuttgart, Germany.

**corresponding author: (haris.lye@mmu.edu.my; ORCiD: 0000-0001-9248-6607)*

***corresponding author: (marawandeep13@gmail.com; ORCiD: 0009-0008-5285-424X)*

Abstract - Mango cultivation is a vital agricultural sector in many regions, providing significant economic benefits and contributing to environmental sustainability. Traditional mango detection and harvesting methods are prone to error and are labour-intensive. Recent advances in aerial imagery and Artificial Intelligence (AI) offer innovative solutions to these challenges. An automated yield estimation and fruit harvesting system will require an accurate fruit object detection model. This study explores the application of deep learning models for mango detection on two diverse datasets. This research focuses on evaluating two primary deep learning models: Faster Region-based Convolutional Neural Network (Faster R-CNN) and You Only Look Once (YOLO) variants. Experiments were conducted using datasets from the ACFR mango dataset and a locally collected dataset. The ACFR dataset is acquired from a moving ground vehicle while the local custom dataset is taken from a low flying drone. The Faster R-CNN model was tested with ResNet-50 backbones. YOLOv8 with simple training image augmentation demonstrated superior performance on both datasets, achieving a mean Average Precision with 0.5 Intersection over Union (mAP@0.5) of 0.959 on the ACFR dataset and 0.756 on the custom local mango image dataset. The YOLOv8 model outperforms Faster R-CNN by a large margin on both datasets. The advantage of simple image augmentation for improving mango detection has also been demonstrated. The YOLOv8 model is found to be able to detect mango fruits effectively in both dark and bright lighting conditions.

Keywords – Object Detection, Smart Farming, Deep Neural Network, Computer Vision, Mango.

Received: 26 August 2025; Accepted: 20 March 2026; Published: 16 June 2026

1. INTRODUCTION

Mango cultivation is a vital component of the agricultural sector in many tropical regions, including Malaysia. This crop is not only economically significant but also contributes as an important source of food. Despite its importance, traditional mango detection and harvesting methods remain prone to errors and labour-intensive. These inefficiencies can lead to suboptimal yield estimation and increased production costs, particularly in large-scale farming operations.

Journal of Informatics and Web Engineering
https://doi.org/10.33093/jiwe.2026.5.2.17

Published by MMU Press. URL: https://journals.mmupress.com/jiwe

Recent technological advancements, particularly in aerial imagery and Artificial Intelligence (AI), offer promising solutions to these challenges. When Unmanned Aerial Vehicles (UAV) are integrated with high-resolution cameras and sophisticated AI algorithms, it can open new avenues for precision agriculture. These technologies can potentially revolutionize the management and harvesting of mango orchards by providing accurate, efficient, and automated detection systems [1], [2] . In addition, many orchards have also started using ground moving robots to inspect fruits and predict yield.

This paper uses two different datasets for evaluating the popular object detection models based on deep learning methods. The first dataset contains mango fruits images acquired from a moving ground vehicle and the second dataset contains mango images from aerial imagery. Specifically, we investigate the performance of two advanced object detection models: the Faster Region-based Convolutional Neural Network (Faster R-CNN) [3] and You Only Look Once (YOLO) [4] variants. Faster R-CNN employs a two-stage detection process known for accuracy, while YOLO uses a single-stage approach optimized for real-time processing. The YOLO object detection model has gone through a number improvement over time aimed at improving detection accuracy and faster speed [5], [6], [7], [8]. This paper focuses on evaluating the YOLOv8 model with the Faster R-CNN model as the baseline for mango fruit detection task. A comparison of the mango fruit detection accuracy for the evaluated models will be presented. Two datasets are used, namely the ACFR mango dataset [9] and a locally created custom dataset using drone to capture aerial images. The model's performance is compared with the mean Average Precision measure.

2. LITERATURE REVIEW

Deep learning models are currently the standard approach for fruit detection in the orchard [1], [9]. Most of the models are assessed on the ACFR mango detection dataset provided by [9]. Models such Faster- RCNN and YOLO are commonly used in various applications. Some common application of fruit detection task is in yield estimation [10], [11] and automatic harvesting by robots [12], [13]. The image can be acquired by moving robot [9], [12] or UAV [1], [2]. The advances in automatic fruit detection have led to improvement in various aspects of precision agriculture and reduction of labour work required.

In the mango fruit detection problem, several object detection models have been proposed. The popular models are YOLO and Faster R-CNN. In the work [13], the Faster R-CNN model is used for mango detection with impressive F1 score of 0.881. The result is combined with image mask predicted by LIDAR estimation to predict the mango yield. The work by [11] applies the encoder decoder architecture for mango detection with weakly supervised learning. Training involves unsupervised learning via image reconstruction followed by image classification on the presence or absence of the fruit. Finally, the model is trained like other conventional object detection methods. The proposed method named MangoDetNet achieves F1 score of 0.861 when all the labelled training samples are used and 0.856 when labelled samples are halved. This led to significant reduction in annotation cost without affecting much the detection accuracy.

Various modifications of YOLO models have been proposed to improve its performance in mango detection. The work by [12] use multi-scale aggregation and image enhancement technique to attain F1 score of 0.977 and perform better than the standard YOLOv3 with 5 times faster speed. The work by [14] propose a modified YOLO architecture from YOLOv3 and YOLOv2 (tiny) named as MangoYOLO. Model architecture change is done by combining features from multiple intermediate layers. This customized model is compared to the standard YOLOv3, YOLOv2 and the Faster R-CNN models. MangoYOLO achieves the best performance in terms of F1 and average precision (AP). It also outperforms the competing models with lower processing speed. Recent work by [15] improvise YOLOv6 model with residual and attention structure. Evaluation on the ACFR Mango dataset yields the mAP0.5 of 0.961. The YOLO model has also been used on aerial images taken from UAV. In the work of [1], green mangoes are detected with YOLOv2 model with precision of 0.961. The detected mangoes are used in predicting the number of mango fruits in an orchard for yield estimation. In addition to detecting mango fruits, the YOLO model has also been used in a number of agriculture related applications such as detecting plant diseases with good performance [16], [17] .

The recent model YOLOv8 is proposed for mango fruit detection, considering its reported improved accuracy and speed [6], [8]. However, the model has not been tested on the mango dataset. The aerial image dataset has also not been tested on YOLOv8 model. This paper aims to study the capability of YOLOv8 in such aspects.

3. RESEARCH METHODOLOGY

3.1 Dataset Preparation

In this paper, two deep learning models for mango detection are selected. The models are the YOLOv8 and Faster R-CNN model. Both the models are evaluated on two different datasets. The first dataset known as ACFR mango dataset is publicly available and is prepared by [9]. The images are collected by a moving ground vehicle that move between the rows of trees at a large-scale mango orchard. A total of 1964 images are used. The second dataset is locally prepared with the use of a small drone DJI Mini 3 UAV. The UAV flows around a local neighbourhood and university area to collect images of mango trees where fruits are observed. The images are collected at UPM International Mango Grove located at Universiti Putra Malaysia Serdang Selangor and one residential area known as Taman Amanputra at Puchong Selangor, Malaysia. A total of 118 images is annotated with bounding box. Table 1 shows a comparison of the 2 datasets used for mango fruit detection with the selected models. Figure 1 shows the sample image from each dataset used.

Data augmentation for increasing training images and its diversity were used during training. The aim is to prevent model overfitting and improve generalisation. The augmentation techniques used included random horizontal and vertical flips with a probability of 0.5, random rotation of ±10 degrees, and random adjustments to hue (±0.015), saturation (±0.7), and brightness (±0.4). Additionally, random scaling was applied, ranging from 0.5 to 1.5 of the original size, and mosaic augmentation, which involves combining four images, was also utilised.

Table 1. The Split of the Datasets Used in the Project

Dataset	Total Images	Training Images	Validation Images	Test Images
Local	118	83	24	11
ACFR	1964	1464	250	250

Figure 1. Sample Image from the Local (Right) and ACFR Mango Dataset (Left)

3.2 Model Description

YOLOv8 is chosen for this study as they represent the recent advancements in the YOLO architecture, offering improved speed and accuracy over previous versions. The model is implemented using the Ultralytics framework [18]. Specifically, YOLOv8n or nano version is chosen where the smaller version helps balance performance and computational efficiency. The YOLOv8 architecture, as shown in Figure 2, consists of 3 major components, the backbone, neck and head. The backbone of the model is used for feature extraction from input image. It contains multiple convolutional blocks connected in series for obtaining features at different scales. The neck section contains layers used to combine the outputs of different layers from the backbone. The head section predicts the bounding of the detected object. Three different heads are used to predict bounding boxes of the object at 3 different scales. The YOLOv8 model utilizes several C2f (cross stage partial fusion) block for better feature extraction. This is achieved by splitting input feature maps and processing them separately with convolutional and bottleneck block. The results are then merged back to form the C2f block output [6]. The YOLOv8 model is available in 5 different sizes. This includes the Extra-large (x), Large (l), Medium (m), Small (s) and Nano (n). The sizes are determined by the depth and width of the layers in the block used [19]. Smaller models run faster but suffer from lower detection accuracy.

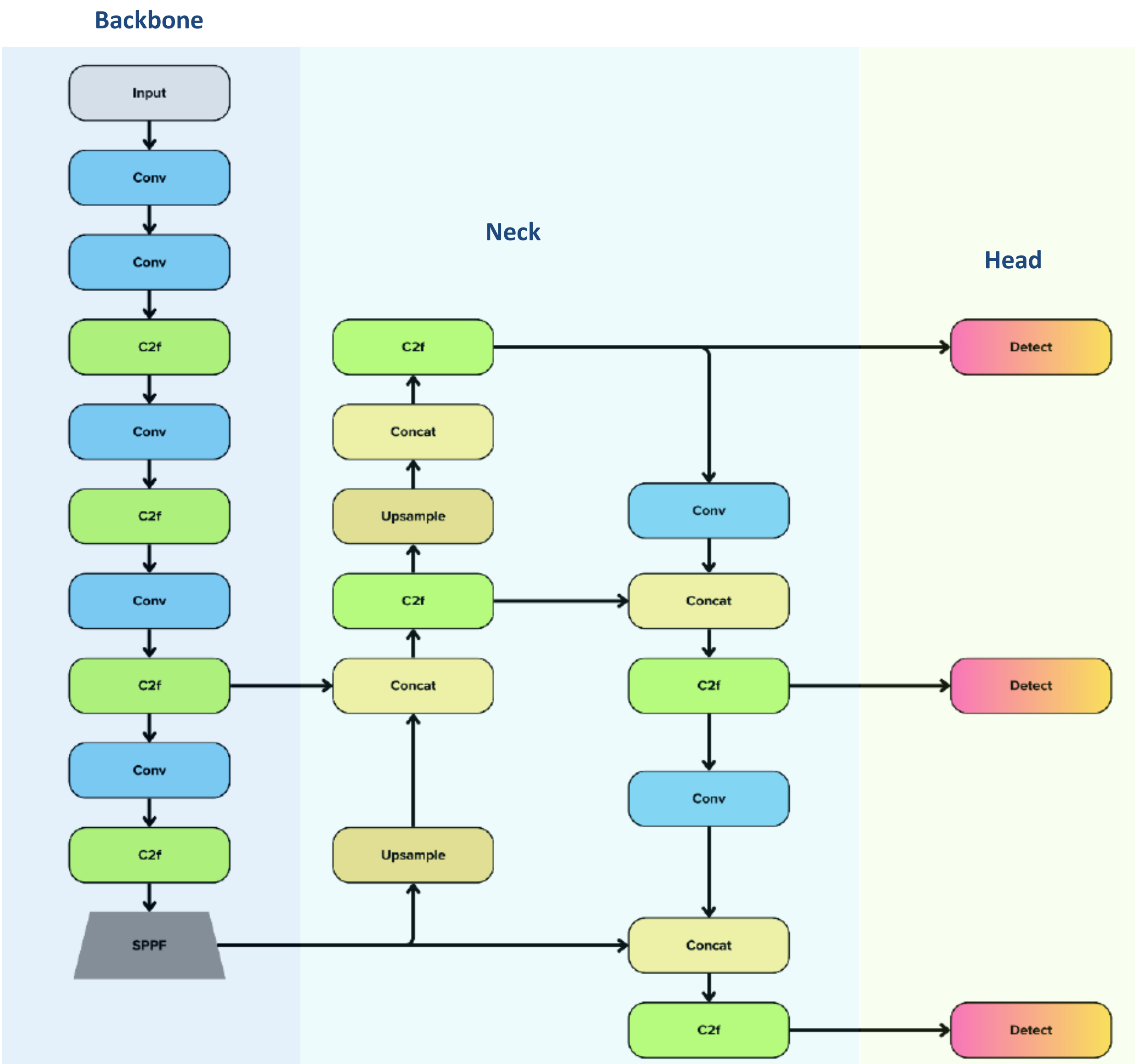

Figure 2. YOLOv8 Architecture [20]

The Faster R-CNN model [3] consist of two major parts, namely the Region Proposal Network (RPN) and Detection network. The RPN section identifies potential regions that may contain objects of interest, and the Detection block predict the type of object in the identified candidate regions. The model architecture is shown in Figure 3. The model was developed using PyTorch programming framework with Resnet-50 backbone architectures. ResNet-50, is a 50-layer deep residual network that utilizes skip connections to address the vanishing gradient problem in deep networks [21]. The result of Faster R-CNN that uses VGG-16 backbone is used as comparison [22]. The VGG-16 model strength lies in its simplicity and depth, which allows it to learn increasingly complex features in a series on convolutional layers with small kernel size.

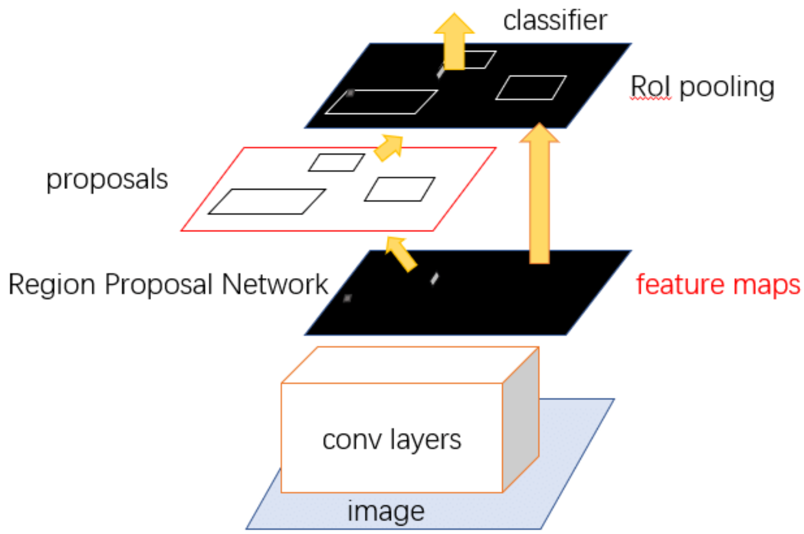

Figure 3. Faster R-CNN [3]

3.3 Model Training

The YOLOv8n model is used due to the fast inference speed and high detection accuracy as well. The model was trained for a maximum of 200 epochs with a batch size of 16 and an image size of 640x640 pixels. The Stochastic Gradient Descent (SGD) optimizer was used. Early stopping was implemented to prevent overfitting, with the criterion set to a patience of 10 epochs. The loss function for YOLO models combined Binary Cross-Entropy (BCE) and Complete Intersection Over Union (CIOU) to optimize both classification and localization accuracy. Faster R-CNN models were also trained for 200 epochs using an SGD optimizer, with learning rate starting at 0.005 and the momentum value of 0.9. The Detectron2 programming framework is used [23].

4. RESULTS AND DISCUSSIONS

4.1 Performance Metrics

The YOLOv8 and Faster R-CNN models are evaluated on the ACFR public mango dataset and a custom created local dataset taken from a low flying drone. To measure the alignment between the ground truth and predicted bounding boxes, the Intersection over Union (IoU) measure is used. The IoU is measured by taking the area of overlap between the ground truth box and the predicted bounding box divided by union of their area. Figure 4 shows how IoU is computed. The larger IoU value shows a more accurate prediction, as this represents a higher degree of overlap between the ground truth and the predicted bounding box. For our experiment the IoU ratio that exceeds 50% is used for assigning correct prediction to the predicted box.

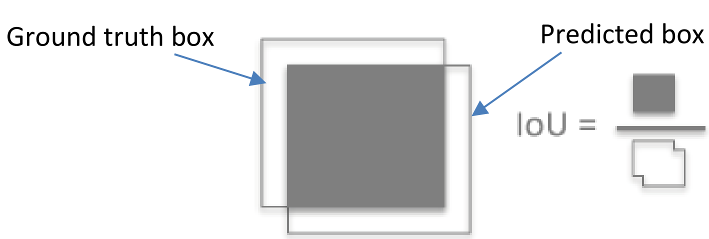

Figure 4. Computation of IoU

To measure the mango detection performance, the Recall, Precision and mean Average Precision (mAP@0.5) are used. Precision measures the accuracy of positive predictions and is calculated as given in Equation 1. It is computed by taking ratio of total correct positive prediction (TP) over total positive prediction (TP + FP). The total false positive is denoted FP. The Recall measures in Equation (2) gives the total ground truth boxes that are correctly predicted. For a more comprehensive measure, the mAP@0.5 is used where IoU is 0.5. Different precision and recall values are obtained by varying the model detection threshold. The average precision to obtain mAP is computed based on the area over the precision recall curve.

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

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

4.2 Results

The first result in Table 2 compares YOLOv8n with Faster R-CNN models on the ACFR mango dataset. The use of augmentation is evaluated as well. The YOLOv8n model with augmentation performs the best with mAP@0.5 value of 0.959. The result for Faster R-CNN (VGG-16) model is obtained from [9]. Table 3 shows the performance of YOLOv8n on our locally prepared dataset that contains aerial imagery images. Similar observation is obtained where model trained with image augmentation on the training set leads to improved performance in terms of mAP@0.5. Figure 5 shows sample YOLOv8n detection result on the test set for ACFR mango dataset while Figure 6 same the model detection results on selected images for our local custom dataset. Even in a dark environment used in ACFR dataset, most of the mangoes are correctly detected. The best mAP performance is bolded in both Table 2 and 3.

Table 2. Model Performance on ACFR Mango Dataset

Model	**Augmentation**	**mAP@0.5**
YOLOv8n	Yes	**0.959**
YOLOv8n	No	0.899
Faster R-CNN (ResNet-50)	Yes	0.931
Faster R-CNN (ResNet-50)	No	0.654
Faster R-CNN (VGG-16) [9]	Yes	0.890

Table 3. Models Performance on Local Mango Dataset.

Model	**Augmentation**	**mAP@0.5**
YOLOv8n	Yes	**0.756**
YOLOv8n	No	0.685
Faster R-CNN (ResNet-50)	Yes	0.5624

Based on the results tabulated in Table 2 and 3, the YOLOv8n model outperforms the Faster R-CNN model by a large margin. This is unsurprising, since YOLOv8 has been designed with many innovations in its architecture design, especially for detecting small objects. Unlike earlier YOLO versions, YOLOv8 effectively integrates feature maps from different scales. This allows fine-grained detail from high resolution maps and high-level context to be jointly used. The aggregated signal is then passed to the detection head to predict the accurate bounding box for the mango object.

Figure 5. Example of Mango Detections Using the YOLOv8 Model, with Correct Detections (Green) and False Positives (Red) on the ACFR Dataset

Figure 6. Example of Mango Detections Using the YOLOv8 Model on the Local Custom Dataset

5. CONCLUSION

This paper investigated the application of deep learning models, specifically Faster R-CNN and YOLO variants, for mango fruit detection in two diverse datasets. The first dataset is acquired by a ground moving vehicle and the second dataset from a low flying drone. On both datasets, the YOLOv8 model performs significantly better than the Faster R-CNN model that uses the Resnet-50 backbone. It is also shown that the use of simple image augmentation helps to improve detection accuracy for both the YOLO and Faster R-CNN model as well. Future research should focus on incorporating more advanced architecture, adding datasets to include more diverse conditions, and creating a dataset for mango ripeness detection. Deploying these models in real-world scenarios, such as integrating them with drones for automated mango detection and harvesting, would provide practical insights and applications. Speculatively, these advancements could lead to more comprehensive solutions for precision agriculture, improving crop monitoring and management.

ACKNOWLEDGEMENT

The author is grateful for the support from the Faculty of Artificial Intelligence and Engineering, Multimedia University in this paper.

FUNDING STATEMENT

The authors received no funding from any party for the research and publication of this article.

AUTHOR CONTRIBUTIONS

Mohd Haris Lye: Project Administration, Writing – Review & Editing;
Marawan Ashraf Eldeib: Conceptualization, Data Curation, Methodology, Validation, Writing – Original Draft Preparation.

CONFLICT OF INTERESTS

No conflict of interest was disclosed.

ETHICS STATEMENTS

Our publication ethics follow The Committee of Publication Ethics (COPE) guideline. https://publicationethics.org/

DATA AVAILABILITY

Derived data supporting the findings of this study are available from the corresponding author on request.

REFERENCES

[1] J. Xiong *et al.*, "Visual detection of green mangoes by an unmanned aerial vehicle in orchards based on a deep learning method," *Biosystems Engineering*, vol. 194, pp. 261–272, June 2020, doi: 10.1016/j.biosystemseng.2020.04.006.

[2] Y. Xiong *et al.*, "Detecting and Mapping Individual Fruit Trees in Complex Natural Environments via UAV Remote Sensing and Optimized YOLOv5," *IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing*, vol. 17, pp. 7554–7576, 2024, doi: 10.1109/jstars.2024.3379522.

[3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, June 2017, doi: 10.1109/tpami.2016.2577031.

[4] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, June 2016, pp. 779–788. doi: 10.1109/cvpr.2016.91.

[5] X. Wang, H. Li, X. Yue, and L. Meng, "A comprehensive survey on object detection YOLO", Proceedings http://ceur-ws. org ISSN 1613 (2023): 0073

[6] M. Hussain, "YOLOv1 to v8: Unveiling Each Variant–A Comprehensive Review of YOLO," *IEEE Access*, vol. 12, pp. 42816–42833, 2024, doi: 10.1109/access.2024.3378568.

[7] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed Tools Appl*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.

[8] J. Terven, D.-M. Cordova-Esparza, and J.-A. Romero-Gonzalez, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *MAKE*, vol. 5, no. 4, pp. 1680–1716, Nov. 2023, doi: 10.3390/make5040083.

[9] S. Bargoti and J. Underwood, "Deep fruit detection in orchards," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, Singapore: IEEE, May 2017, pp. 3626–3633. doi: 10.1109/icra.2017.7989417.

[10] P. Maheswari, P. Raja, and S. Natarajan, “MangoYieldNet: Fruit yield estimation for mango orchards using DeepLabv3 + with ResNet18 architecture,” *Multimed Tools Appl*, Apr. 2025, doi: 10.1007/s11042-025-20791-5.

[11] A. R. Denarda, F. Crocetti, G. Costante, P. Valigi, and M. L. Fravolini, “MangoDetNet: a novel label-efficient weakly supervised fruit detection framework,” *Precision Agric*, vol. 25, no. 6, pp. 3167–3188, Dec. 2024, doi: 10.1007/s11119-024-10187-0.

[12] Z.-F. Xu, R.-S. Jia, H.-M. Sun, Q.-M. Liu, and Z. Cui, “Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots,” *Appl Intell*, vol. 50, no. 12, pp. 4670–4687, Dec. 2020, doi: 10.1007/s10489-020-01818-w.

[13] M. Stein, S. Bargoti, and J. Underwood, “Image Based Mango Fruit Detection, Localisation and Yield Estimation Using Multiple View Geometry,” *Sensors*, vol. 16, no. 11, pp. 1915, Nov. 2016, doi: 10.3390/s16111915.

[14] A. Koirala, K. B. Walsh, Z. Wang, and C. McCarthy, “Deep learning for real-time fruit detection and orchard fruit load estimation: benchmarking of ‘MangoYOLO,’” *Precision Agric*, vol. 20, no. 6, pp. 1107–1135, Dec. 2019, doi: 10.1007/s11119-019-09642-0.

[15] Z. Zhong, L. Yun, F. Cheng, Z. Chen, and C. Zhang, “Light-YOLO: A Lightweight and Efficient YOLO-Based Deep Learning Model for Mango Detection,” *Agriculture*, vol. 14, no. 1, pp. 140, Jan. 2024, doi: 10.3390/agriculture14010140.

[16] U. Ali, M. A. Ismail, R. A. A. Habeeb, and S. R. A. Shah, “Performance Evaluation of YOLO Models in Plant Disease Detection,” *Journal of Informatics and Web Engineering*, vol. 3, no. 2, pp. 199–211, June 2024, doi: 10.33093/jiwe.2024.3.2.15.

[17] C. Pabitha, B. Vanathi, K. Revathi, and S. Benila, “Enhancing Citrus Plant Health through the Application of Image Processing Techniques for Disease Detection,” *Journal of Informatics and Web Engineering*, vol. 4, no. 2, pp. 53–63, June 2025, doi: 10.33093/jiwe.2025.4.2.4.

[18] Ultralytics, “Ultralytics YOLO: State-of-the-art AI for computer vision.” [Online]. Available: https://www.ultralytics.com/

[19] M. Yaseen, “What is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector,” 2024, *arXiv*: arXiv:2408.15857, doi: 10.48550/arXiv.2408.15857.

[20] R. E. Saragih, A. R. Purnajaya, I. Syafrinal, and Y. Pernando, “Mango and Banana Ripeness Detection based on Lightweight YOLOv8”, Jurnal Buana Informatika, 15(2), pp.79-88, 2024.

[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[22] M.D. Zeiler, and R. Fergus, Visualizing and Understanding Convolutional Networks,” in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2014, pp. 818–833. doi: 10.1007/978-3-319-10590-1_53.

[23] Meta AI, “Detectron2: Object detection and segmentation tool.” [Online]. Available: https://ai.meta.com/tools/detectron2/

BIOGRAPHIES OF AUTHORS

	**Mohd Haris Lye** received the B.Eng (Electrical and Electronics) degree from University of Science Malaysia in 1996, and the M.S. (Information Technology) degree from Multimedia University in 2004. His research interests include deep learning, computer vision, and artificial intelligence. Currently, he is a lecturer at the Faculty of Engineering, Multimedia University, Malaysia. He can be contacted at email: haris.lye@mmu.edu.my.
	**Marawan Ashraf Fawzy Eldeib** completed his Bachelor of Engineering (Hons.) in Electronics, with a major in Computer Science, from Multimedia University, Cyberjaya, Malaysia, in 2024. His final year project, titled *Mango Fruit Detection From Aerial Images*, involved developing a machine learning model that achieved high accuracy and was awarded Best Research Project. He is currently pursuing a Master of Science in Software Engineering at the University of Stuttgart, Germany. His research interests include Artificial Intelligence (AI) and Software Engineering. He can be contacted at email: marawandeep13@gmail.com.