International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV

Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

INTERNATIONAL JOURNAL OF INNOVATION AND INDUSTRIAL REVOLUTION (IJIREV)
www.ijirev.com

GLOBAL ACADEMIC EXCELLENCE

ENHANCING OIL PALM FRUIT DETECTION: A COMPARATIVE ANALYSIS OF YOLO ALGORITHMS

Ainul Hakim Fizal Sabillah[1], Hafizal Mohamad[2*], Sagaya Sabestinal Amalathas[3], Seung Hwan Won[4], Muhammad Hakim Ahmad Sobri[5]

[1] Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, Nilai, Malaysia
Email: ainulhakimfizam@raudah.usim.edu.my
[2] Faculty of Engineering and Built Environment, Universiti Sains Islam Malaysia, Nilai, Malaysia
Email: hafizal@usim.edu.my
[3] Connected Intelligence Research Group, University of Southampton Malaysia, Iskandar Puteri, Malaysia
Email: s.s.amalathas@soton.ac.uk
[4] Connected Intelligence Research Group, University of Southampton Malaysia, Iskandar Puteri, Malaysia
Email: sw6f12@ecs.soton.ac.uk
[5] Aerodyne Technology Sdn. Bhd., Cyberjaya, Malaysia
Email: hakim.sobri@aerodyne.group
[*] Corresponding Author

Article Info:

Article history:
Received date: 05.01.2025
Revised date: 19.01.2025
Accepted date: 10.02.2025
Published date: 03.03.2025

To cite this document:

Sabillah, A. H. F., Mohamad, H., Amalathas, S. S., Seung, H. W., & Sobri, M. H. A. (2025). Enhancing Oil Palm Fruit Detection: A Comparative Analysis Of Yolo Algorithms. *International Journal of Innovation and Industrial Revolution, 7* (20), 01-18.

DOI: 10.35631/IJIREV.720001

Abstract:

The oil palm industry is pivotal in agricultural research because of its importance. The central focus of this study, however, revolves around elevating cutting-edge intelligent techniques in agriculture, specifically for the improved detection of Fresh Fruit Bunches within oil palm plantations. Moreover, Malaysia's economic impact on oil palm production is explored, emphasizing its position as a leading producer and exporter of palm oil. The study compares and corroborates the performances among a series of YOLO algorithm models, namely YOLOv3, YOLOv4, YOLOv7, and YOLOv8, by exploiting diverse essential metrics that embrace mean Average Precision, precision, recall, and F1-score. Through the rigorous evaluations of these models, the research contributes to the precision agricultural field, underscoring the superior performance of YOLOv8 in accurately detecting FFBs and facilitating its realization of advanced computer vision techniques for optimal oil palm plantation management and enhanced productivity.

Keywords:

Object Detection, Computer Vision, Yolo, Ffb, Palm Oil

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

Introduction

Scientifically known as Elaeis Guineensis Jacq., the oil palm tree is a well-known tree with a rich history that dates back thousands of years. It is native to West Africa (Hai, 2002). In the early 1870s, the British introduced oil palm to Peninsular Malaysia as an ornamental plant (Malaysia Palm Oil Council, n.d.). In 1917, the first commercial planting of oil palm occurred at Tennamaran Estate in Batang Berjuntai, Selangor (Wahid, Weng, & Masri, 2009). Inspired by his friend's success in Sumatra, Henri Fauconnier replaced his coffee plantation with oil palm at Tennamaram Estate, laying the foundation for Malaysia's palm oil industry (Otieno et al., 2016). After gaining independence in 1957, Malaysia implemented an agricultural diversification program to reduce dependence on rubber and tin, and oil palm emerged as an ideal crop due to its high economic returns, faster maturity, and labor-intensive nature. Three hundred seventy-five hectares of oil palm was allocated by the Federal Land Development Authority (FELDA) for cultivation in the 1960s, surpassing rubber as Malaysia's main commodity crop in 1989(Otieno et al., 2016).

The oil palm is a versatile crop that produces two types of vegetable oil: palm oil from the fruit's mesocarp and palm kernel oil from the seed (Malaysia Palm Oil Council, n.d.). Malaysia's oil palm industry has grown significantly, covering 5.9 million hectares of land and producing 19.14 million tons of palm oil and 2.20 million tons of palm kernel oil in 2020 (Parveez et al, 2021). Global demand for palm oil has increased its value, with worldwide sales reaching US$48.7 billion in 2021. Indonesia and Malaysia are the largest exporters, accounting for 83% of the total value of palm oil exports (Alfatah, Mistar, Syabriyana, & Supardan, 2022). Table 1 manifests the top 5 countries in 2021 that exported the highest dollar value worth of palm oil (Workman, 2023). The industry has played a crucial role in countries' socioeconomic development, particularly Malaysia, recognized as one of the largest palm oil producers and exporters (Feintrenie, Schwarze, & Levang, 2010), (Mahat, 2012). However, ensuring the quality of palm oil requires effective monitoring and control of the production process, including targeting the ripeness of the oil palm for optimal oil extraction as a high Oil Extraction Rate (OER), low levels of free fatty acids, and ripeness of the oil palm fruit, greatly influence the overall quality of palm oil (Khamis, Selamat, Ghazalli, Md Saleh, & Yusoff, 2022). Specific standards and guidelines are established to classify and grade FFBs based on their maturity levels in oil palm ripeness identification. These criteria primarily focus on the color of the bunch and the presence of loose fruits. According to Malaysia Palm Oil Board (MPOB) standards as exhibited in Table 2 (Malaysian Palm Oil Board, 2003).

Table 1. Top Exporters of Palm Oil Globally in 2021

Country	Value (US$ billion)	Percentage (%)
Indonesia	26.7	54.7
Malaysia	14.2	29.2
Netherlands	1.2	2.5
Papua New Guinea	0.783	1.6
Thailand	0.713	1.5

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

Table 2. Ripeness Standard Guidelines Set by MPOB

Ripeness	Color	Loose Fruits
Unripe	Purplish Black	0
Underripe	Redish Purple	<10
Ripe	Redish Orange	>10
Overripe	Dark Redish	~50%

Oil palm harvesting is a pivotal process within plantations, involving the crucial task of identifying and counting Fresh Fruit Bunches (FFBs) through a manual approach, where harvesters identify and count FFBs from row to row, observing only 10% of the palm population and counting all palms every 10th palm row (Black bunch census (BBC) at 4-month intervals, n.d.). However, this conventional method is plagued by various challenges, including time consumption, high energy expenditure, cost-ineffectiveness, and inaccuracies. To decide on the correct period for the harvest, the harvesters look at the black bunches of each tree, which is called the black bunch census (Black bunch census (BBC) at 4-month intervals, n.d.). The information gathered from this process is then taken to the plantation owners, who make use of such important information to formulate the best strategy for harvesting and organizing their resources.

Despite the significance of oil palm harvesting, the industry has relied heavily on a large workforce of relatively unskilled labor operating in the demanding conditions of a hot and humid tropical climate year-round (Crowley, 2020). Based on the findings, low wages and competition from other employment opportunities have contributed to the use of temporary migrant labor and the shortage of staff over the years, worsened by the COVID-19 pandemic disruptions (Crowley, 2020), (Raghu, 2021). Moreover, the industry has also suffered from issues regarding poor labor standards investigation, and many important commercial companies began to be blacklisted, and many imports were banned by the US Customs and Border Protection in 2020-21 (Jamal, 2021). These problems point to the fact that there is a need for the improvement of labor practices in the oil palm industry in the efforts aimed at making positive changes in the workers' health and increasing the sustainability of the industry.

The oil palm industry has a historical background and has expanded its position to the world market, with Malaysia holding a reputation as one of the largest producers and exporters of palm oil and its products. The main purpose of this study is to advance the existing efforts towards improving the quality of palm oil production. There is an attempt to automate the process of FFB ripeness assessment with the help of Deep Learning (DL) and Computer Vision (CV) (Stieg, n.d.), (Simplilearn, n.d.). In precision agriculture, automated fruit identification or recognition through image processing is vital, especially in object detection and in locating more pertinent targets, especially on large tracts of naturally productive land and image or video data (Syazwani, Asraf, Amin, & Dalila, 2022), (Karthikeyan, Subashini, Srinivasan, Santhanakrishnan, & Ahilan, 2023). It will be crucial to develop this technology to improve farming practices and increase productivity. The proposed model, through the integration of these advanced technologies, is bound to transform the current techniques and practices in the industry while improving both the efficiency and accuracy of the production system. Moreover, this automation can greatly decrease the amount of effort required by people whilst enhancing the quality of the final product (Marinoudi, Sørensen, Pearson, & Bochtis, 2019), (Srivastava, n.d.).

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

This research also presents an approach for identifying the Oil Palm FFBs with the aid of a more advanced YOLOv8 model that is more accurate and efficient compared to its earlier versions. Thus, our work through the utilization of CV enhanced by DL offers innovation for precision agriculture in oil palm plantations, leading to an enhancement in productivity and labor force efficiency. Thus, the proposed system has high performance in terms of mean Average Precision (mAP), precision, recall, the F1-score, and established data enhancement policies show the viability of the proposed system for practical application.

Methodology

Site Visit

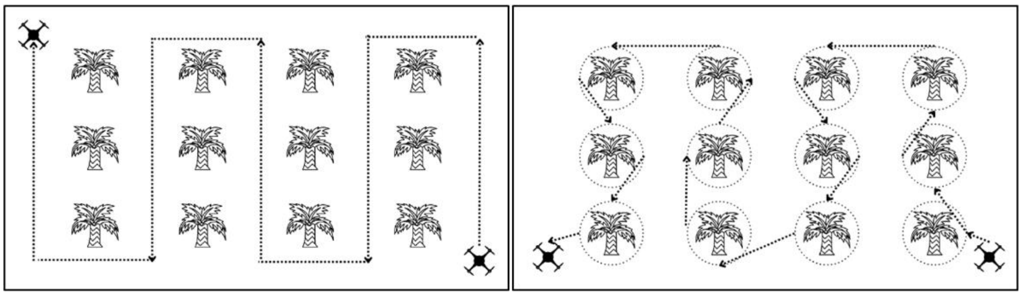
With the purpose of identifying and gathering FFBs as data samples, a site visit to Zenxin's Oil Palm Estate in Johor, Malaysia, took place on 5th November 2022. Preparations and planning regarding the visit were made and agreed upon a month before the visit, to ensure that all processes ran smoothly. On-site, the team was divided into two groups. The first group included the drone team from Aerodyne and researchers from the University of Southampton Malaysia. They employed a UAV for aerial mapping and shooting videos. The second group used smartphones to capture images and videos of Universiti Sains Islam Malaysia teams. We were able to capture 119 pictures and video approximately 8 minutes of video within two and a half hours. The videotaping was done in a circular manner to capture all round angles of every tree. Based on the physical assessment of the estate, it had an average of approximately 22 trees that can be assessed for this site visit purpose. Figure1 shows one of the collection datasets for the site visit.

Figure 1: Sample Picture of FFB On the Oil Palm Tree During the Site Visit At Zenxin's Plantation

Data Collection Method

Through our discussion with the Aerodyne team, we identified two approaches to collect FFB datasets from oil palm trees. These methods are the straight-line and circular paths, as shown below in Figure 2. The straight-line path mentioned has its own advantages. Firstly, it can take a longer distance and take less time as compared to circular motion because the drone can just fly directly in a linear manner. The present method is also safe and effective because it requires few corners to turn. On the other hand, the circular path the drone takes to capture more detail on the datasets makes it easy for the analysis to be exhaustive. This method might have to be used where more information pertaining to the FFB datasets is required.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001



(a) (b)

Figure 2: The Drone Maneuver Configuration of Straight Line (a) and Circular Path (b)

Data Preparation

Image Annotation

For image data to be efficiently used in the algorithms of CV and DL, then the images have to be annotated. Annotation is a process of tagging the target data and further dividing them into categories, which gives the algorithm an opportunity to perform a correct analysis of the visual information. This process is crucial for functional datasets because it informs the training model about the essential parts of the image, which can then identify those classes in new and never-before-seen images. LabelImg software was adopted to perform image annotation. This lightweight and user-friendly tool is designed to label object-bounding boxes in images. As stated in the objective, the decision to utilize the YOLO series, comprising YOLOv3, YOLOv4, YOLOv7, and YOLOv8, as our initial algorithms. The choice of previous YOLOs, which are YOLOv3 and YOLOv4, is due to the most used version in agriculture and image detection, as the model is mature and stable. With their well-documented performance and suitability for our project's requirements, it is essential to save the annotation in a *.txt file format as the models is trained to process data in that specific format, ensuring seamless integration and efficient data handling. By employing LabelImg, we can verify that our image datasets are accurately annotated and ready to be applied in DL and CV algorithms. The annotated data is crucial for the algorithm's performance, and we need to ensure that the annotation is done with precision and accuracy.

Data Augmentation

Data augmentation, a widely used regularization strategy that involves duplicating and transforming the original data in various ways to create more data for the DL model to train on, can effectively reduce overfitting. A state in which the model performs well on given datasets but poorly on new and unseen data. This technique will improve the robustness of object detection models, especially in occlusion situations (Deng, Zhao, Zhang, Zhang, & Mei, 2023). It encompasses simple operations such as flipping, scaling, random cropping, and rotating images, generating new similar but not identical examples. By exposing the model to a broader range of variation in the input data, data augmentation helps make the model more robust to generalize to new data, as demonstrated in Figure 3, which compares original and augmented images.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

Figure 3: An Example Of Data Augmentations, Which Create Variations In The Dataset, Enabling The Model To Better Handle Different Environmental Conditions And Perspectives. From Top Left To Top Right: Flip Left-Right, Flip Upside-Down, Rotate -45o, Rotate 45o. From Bottom Right To Bottom Left: Scale Down (0.8), Scale Up (1.2), Gaussian Blur (10), Motion Blur (50)

Result and Discussion

Data augmentation is a critical component of training deep learning models, and its benefits have been proved in the enhancement in performance of image classification problems. However, its full potential in object detection has not been thoroughly explored (Ghazali, Samad, Arshad, & Karim, 2009). Considering the additional resources required for annotating images specifically for object detection, data augmentation becomes even more crucial for this computer vision task. By generating derivative images from the base training dataset, data augmentation reduces the time spent on manual labeling and allows more efficient object detection model utilization.

For data augmentation evaluation, we are focusing on applying data augmentation using YOLOv3, a popular object detection algorithm. The goal is to compare the performance of YOLOv3 when trained with and without data augmentation. This experiment aims to assess the impact of data augmentation on the model’s accuracy, robustness, and overall detection capabilities. Figure 4 shows the effects of data augmentation in the object detection model. The graph illustrates the mAP at 0.5 confidence with each iteration and concludes that the model with data augmentation is suppressing the model without data augmentation at best iteration 71.16% and 32.54%, respectively.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

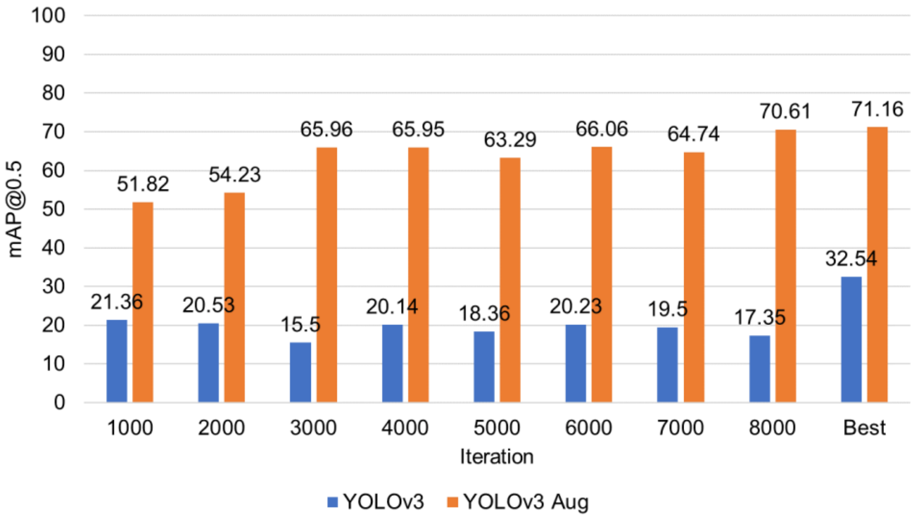


Figure 4: Comparison of YOLOv3 Without Data Augmentation (Blue) and With Data Augmentation (Orange)

Input Image Comparison

The size of the input image can also affect the performance of object detection models. Larger input images can provide more information and improve object detection accuracy. However, increasing the input image size can also increase the computational cost and reduce the model's processing speed. We will analyze the impact of different input image sizes among YOLOv3, YOLOv4, YOLOv7, and YOLOv8 and determine each model's optimal input image size in precision agriculture.

YOLOv3 Parameters

YOLOv3 (Redmon, & Farhadi, 2018), published in ArXiv in 2018, marked a significant advancement in the YOLO series, boasting a more extensive architecture while maintaining real-time performance, solidifying its status as the most popular and stable version. It has been a benchmark and favored choice for researchers seeking to integrate deep learning into their systems. YOLOv3, like its predecessors, excels in object detection, employing a single Convolutional Neural Network (CNN) to simultaneously predict bounding boxes and object classes, with Darknet53 as its notable feature extraction network – a robust 53-layer CNN backbone with three detection heads. This architecture differs from YOLOv2 by utilizing convoluted layers for feature map down-sampling, introducing a residual structure to mitigate gradient-related issues, and accommodating multi-scale training. While finer grids enhance detection accuracy, it's important to note that the network's speed and accuracy trade-offs must be carefully considered (Song, Gao, & Chen, 2021).

Adopting the Darknet framework, the YOLOv3 configuration file consists of various parameters crucial in determining the model's architecture and training settings. Table 3 presents the hyperparameter list and its value. These parameters define the neural network's layers, filters, batch size, learning rate, and other hyperparameters. One of the essential parameters in the configuration file is the learning rate. The learning rate determines the step size at which the model updates its parameters during training. It has been seen that the higher learning rate facilitates faster learning for the model, but it may cause instability in the training process, whereas a lower learning rate causes slow convergence, or the model gets stuck in local minima. The learning rate in this configuration file is fixed at 0.001, which is a common

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

norm in most DL applications. Another important factor is the batch size, which means the number of input images seen by the model in one pass through the network during learning. In this study, the batch size is taken to be 64, and the number of splits is also taken to be 16 for each batch so that this model can take less GPU memory.

Table 3: YOLOv3 Hyperparameter Configuration and Values

Hyperparameters	Value
Batch	64
Subdivision	16
Learning rate	0.001
Max. batches	8000
Steps (80%, 90%)	6400, 7200

The performance of the YOLOv3 model can be assessed by utilizing metrics such as mAP at each iteration. The mAPs are a widely adopted metric in object detection that quantifies the accuracy of a model in identifying objects within images. The first part of this project involves one class detection, which is detecting FFB only. A resolution of 250 x 250 and 416 x 416 pixels is employed as the image size in this network. Utilizing smaller image sizes can considerably expedite and stabilize the training process but may also reduce the model's accuracy. Figure 5 illustrates the output graph obtained during the training process for 1728 images. The maximum batch size utilized in this training is set to 8000. Additionally, the graph displays the mAP values achieved at the best iteration in each resolution. Both illustrate conclude that applying the higher resolution of the input image for the YOLOv3 model improves significantly at the best of 71.16% at 416 resolution and 48.76% at 250 resolutions, improving 22.4%.

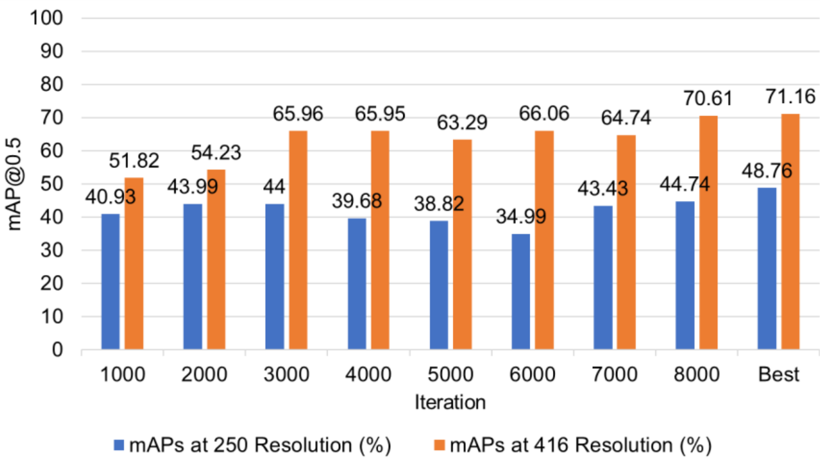


Figure 5: Comparison of Yolov3 with Different Resolution Input Images

YOLOv4 Parameters

YOLOv4 (Bochkovskiy, Wang, & Liao, 2020) represents an enhanced iteration of the YOLOv3 object detection algorithm that employs deep learning techniques to identify objects within images or videos. Similar to its predecessors, YOLOv4 utilizes a solitary CNN to make predictions regarding bounding boxes and class labels for objects detected within an image. However, YOLOv4 introduces several improvements and innovations to the original YOLO

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

algorithm. For example, it utilizes a more extensive network architecture with a deeper backbone network called CSPDarknet53, designed to improve feature extraction efficiency and accuracy.

The YOLOv4 object detection algorithm also introduces several improvements against the YOLOv3 configuration file. In Table 4, the YOLOv4 model implements the same darknet framework as YOLOv3, but the configuration file includes new and updated parameters that determine the model's architecture and training settings as it introduces the CSPDarknet53. One critical parameter in the YOLOv4 configuration file is the learning rate, which determines the step size at which the model updates its parameters during training. In YOLOv4, the learning rate is set to 0.001, which has been shown to improve training stability and convergence. Another essential parameter is the batch size, which is set to 64, but the number of subdivisions is set to 16, allowing for more efficient memory usage. The max batches parameter is 8000. The steps parameter is also changed in YOLOv4 to reduce the learning rate after 6400 and 7200 iterations during the training phase. Additionally, the configuration file for YOLOv4 also includes the changes in the filters, sizes, and strides of the convolutional layers, the numbers of anchor boxes, and classes for more accurate object detection. The following changes to the YOLOv4 configuration file can enhance object detection accuracy, reduce the time required to train the model, and improve the stability of the model.

Table 4: YOLOv4 Hyperparameter Configuration and Values

Hyperparameters	Value
Batch	64
Subdivision	16
Learning rate	0.001
Max. batches	8000
Steps (80%, 90%)	6400, 7200

While evaluating the performance of the applied YOLOv4 object detection model in each iteration, it is possible to use indicators such as mAP. In this project, only one class is considered, and that is FFB. The YOLOv4 model is trained in two different sizes, with image resolutions of 250 x 250 and 416 x 416 pixels, to enhance the disparity in object detection precision as much as possible while taking a reasonable amount of time to train in. The training process improves by setting the max batch to 8000 while each step to 6400 and 7200, which reduces the time taken to train and increases accuracy. It is also clear from Figure 6 that the training has been done for 1728 images. The mAP values are recorded and shown in the graph, indicating the accuracy of the YOLOv4 model in detecting FFB objects in the best iteration in each resolution. Overall, the YOLOv4 model demonstrates improved object detection accuracy compared to YOLOv3, making it a powerful tool for object detection tasks. It also shows that the increase in input image resolution does not improve significantly; iteration for 416 resolutions, representing 76.92%, and for 250 resolutions at 73.6% mAP, which is only a 3.32% improvement.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

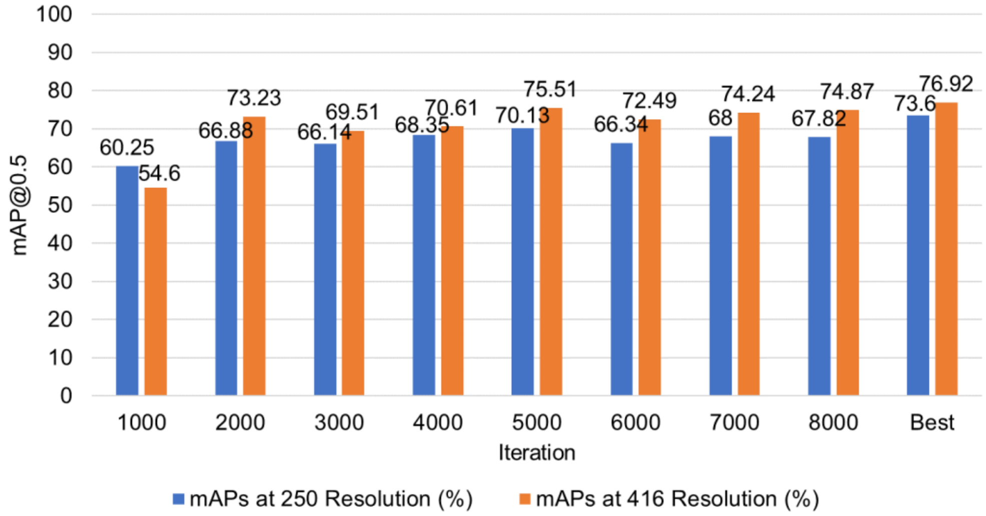


Figure 6: Comparison of YOLOv4 with Different Resolution Input Images

YOLOv7 Parameters

The authors of YOLOv4 also published YOLOv7 (Wang, C.-Y., Bochkovskiy, & Liao, 2023) on ArXiv in July 2022. YOLOv7 demonstrated superior performance to existing object detectors, achieving remarkable speed and accuracy ranging from 5 frames per Second (FPS) to 160 FPS. Like YOLOv4, it was trained exclusively on the Microsoft - Common Object in Context (MS-COCO) dataset without utilizing pre-trained backbones. YOLOv7 introduced several architectural changes and incorporated a series of bag-of-freebies techniques, improving accuracy while maintaining the same inference speed. However, it is essential to note that these modifications primarily impacted the training time rather than the speed at which the model made predictions during inference.

Table 5 displays the hyperparameter configuration and their corresponding values for the YOLOv7 model. The batch size is set to 4, meaning the model processes four images in each iteration during training. The momentum value is 0.937, which influences the speed and direction of weight updates during optimization. The learning rate is 0.01, determining how much the model adjusts its parameters based on the computed gradients. The maximum number of batches is 8000, indicating the total number of iterations the model will undergo during training. Lastly, the model is trained for 2000 epochs, representing the number of times the entire dataset has passed through the network.

Table 5: YOLOv7 Hyperparameter Configuration and Values

Hyperparameters	Value
Batch	4
Momentum	0.937
Learning rate	0.01
Max. batches	8000
Epochs	2000

To assess the performance of the YOLOv7 object detection model, evaluation metrics such as mAP can be utilized in each iteration. This project focuses explicitly on detecting a single class, the FFB. The YOLOv7 model is trained using images of two sizes, 250 x 250 pixels and 416

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

x 416 pixels. This enables a more precise comparison of object detection while still maintaining reasonable training time. The model is set to 4 batch sizes and 2000 epochs to create a consistent experiment environment. To make the maximum number of batches 8000 initially set for YOLOv3 and YOLOv4. Fig 7 manifests the comparison performance of YOLOv7 leveraging 250 and 416 input image resolution. As recorded on the graph, the improvement of YOLOv7 compared to previous YOLO series, such as YOLOv3 and YOLOv4, is significant. The best iteration and mAP of 250 and 416 resolution are 92.31% and 98.87%, respectively

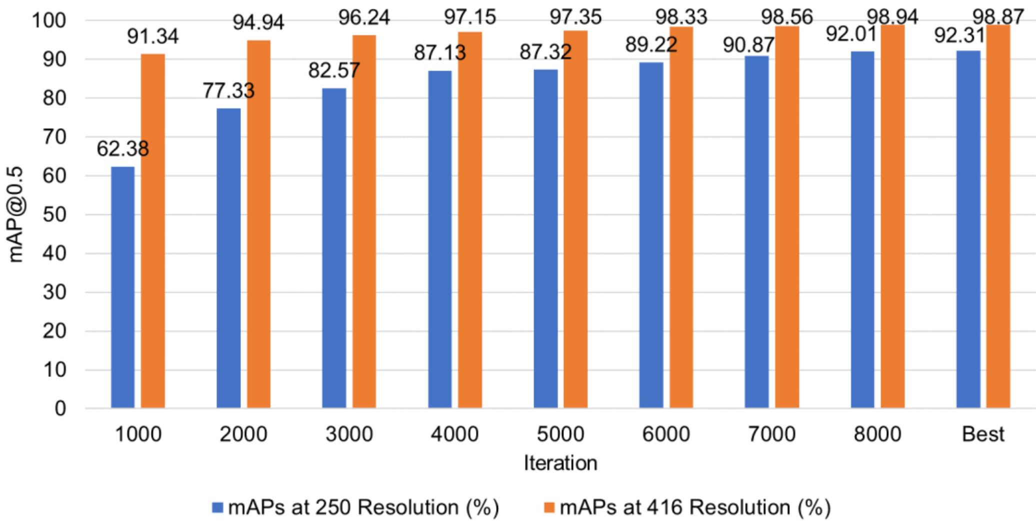


Figure 7: Comparison of YOLOv7 With Different Resolution Input Images

YOLOv8 Parameters

YOLOv8, introduced in January 2023 by Ultralytics (Jocher, Chaurasia, Waxmann, & Laughing, n.d.), the same company behind YOLOv5, offers a range of five scaled versions: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large). This latest iteration of the YOLO series is designed to support various vision tasks, including object detection, segmentation, pose estimation, tracking, and classification. With its versatile capabilities and different scaled versions, YOLOv8 aims to provide flexibility and improved performance across various computer vision applications.

In YOLOv8, the backbone architecture is similar to YOLOv5, but there are notable changes in the CSPLayer, now called the C2f module. The C2f module, which stands for the cross-stage partial bottleneck with two convolutions, is crucial in enhancing detection accuracy. It achieves this by integrating high-level features with contextual information. Combining these elements, the C2f module effectively improves the model's ability to detect and classify objects accurately.

Table 6 provides the hyperparameter configuration and their respective values for the given model. The batch size is set to 4, meaning the model processes four images in each iteration during training. The maximum number of batches is 8000, indicating the total number of iterations the model will undergo during training. Additionally, the model is trained for 2000 epochs, representing the number of times the entire dataset has passed through the network. Due to the model being released early this year, there is less research and papers about the YOLOv8 model. There is limited information regarding the configuration file to tune the hyperparameter, so there is not much to modify.

International Journal of Innovation and Industrial Revolution
EISSN: 2637-0972
IJIREV
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

Table 6: YOLOv8 Hyperparameter Configuration and Values

Hyperparameters	Value
Batch	4
Max. batches	8000
Epochs	2000

Decisions like mAP can be made in every iteration to assess the effectiveness of the YOLOv8 object detection neural network model. This project focuses on detecting one class, which is FFB. The YOLOv8 model is trained on images of 250 x 250 and 416 x 416 pixels, allowing for more precise object detection comparison while maintaining a reasonable training time. The model is set to 2 batch sizes and 4000 epochs to create a consistent experiment environment. To make the maximum number of batches 8000 initially set for YOLOv3 and YOLOv4. Figure 8 illustrates the comparison performance of YOLOv8 leveraging 250 and 416 input image resolution. As recorded, the improvement of YOLOv8 compared to previous YOLO series, such as YOLOv3 and YOLOv4, is significant, but YOLOv7 is almost on par. The best iteration and mAP of 250 and 416 resolution are 97.43% and 98.98%, respectively.

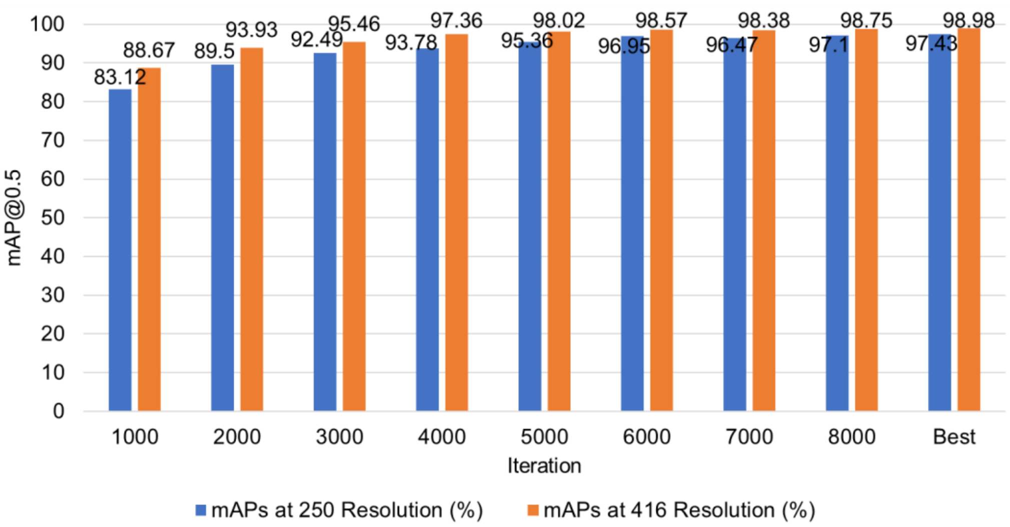


Figure 8: Comparison of YOLOv8 With Different Resolution Input Images

Performance Comparison for YOLOv3, YOLOv4, YOLOv7 and YOLOv8

The YOLO series, encompassing YOLOv3 and YOLOv4, has gained significant popularity in precision agriculture for its exceptional object detection precision and accuracy. These models are highly effective, easy to implement, and exhibit robustness, making them ideal for agricultural applications. However, to explore the latest advancements in the field, we also compare the selected YOLO series, which comprises the cutting-edge YOLOv7 and YOLOv8 models. The addition of YOLOv7 and YOLOv8 allows for a comprehensive evaluation of the advancements within the YOLO series. These newer models have surpassed the previous versions regarding speed and accuracy, reducing real-time object detection parameters by 40% and computation power by 50% (Boesch, 2023).

Consequently, they achieve faster inference speeds and higher detection accuracy. Furthermore, several sub-files require modifications in the associated file to accommodate the selected datasets. These modifications include adjusting parameters such as the number of classes (NC), class names, and directories of images in the train, validation, and test folders.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

By incorporating these updates, we ensure that the comparison encompasses the most up-to-date and state-of-the-art advancements in the YOLO series, providing valuable insights into the latest improvements, features, and performance enhancements introduced.

The experiment compared different versions of YOLO with varying image resolutions, focusing specifically on a 250 x 250 input image resolution, as present in Figure 9. The graph demonstrates that YOLOv8 consistently outperforms the other versions regarding mAP scores across all iterations. YOLOv7 also exhibits competitive performance, particularly in the later iterations. On the other hand, YOLOv4 and YOLOv3 show lower mAP scores compared to the newer versions. From the “Best” column, which refers to the best mAP score of all the versions of each YOLO, YOLOv8 yields the highest mAP of 97.43%, YOLOv7 has 92.31%, YOLOv4 has 73.6%, and YOLOv3 has 71.16%. These results suggest that YOLOv8 is superior in this respect, which confirms the fact that YOLOv8 is more accurate in object detection.

Figure 10 shows the evaluation of the various YOLO versions with the input image size of 416 x 416. The graph shows that YOLOv8 surpasses all the other versions in the number of mAP scores during every iteration. In general, YOLOv7 also performs rather well, especially in its later iterations. YOLOv4 also gives reasonably good results, although it has slightly less performance than other versions of YOLO, while YOLOv3 gives comparatively fewer mAP scores than all other versions of YOLOs. From these outcomes, it can be concluded that when operating in the 416 x 416 image resolution, the YOLOv8 version has the highest results in object detection among all the YOLO versions. YOLOv7 also has a good performance, and YOLOv4 has moderately good results. However, YOLOv3 is slower in mAP scores than the other versions introduced.

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

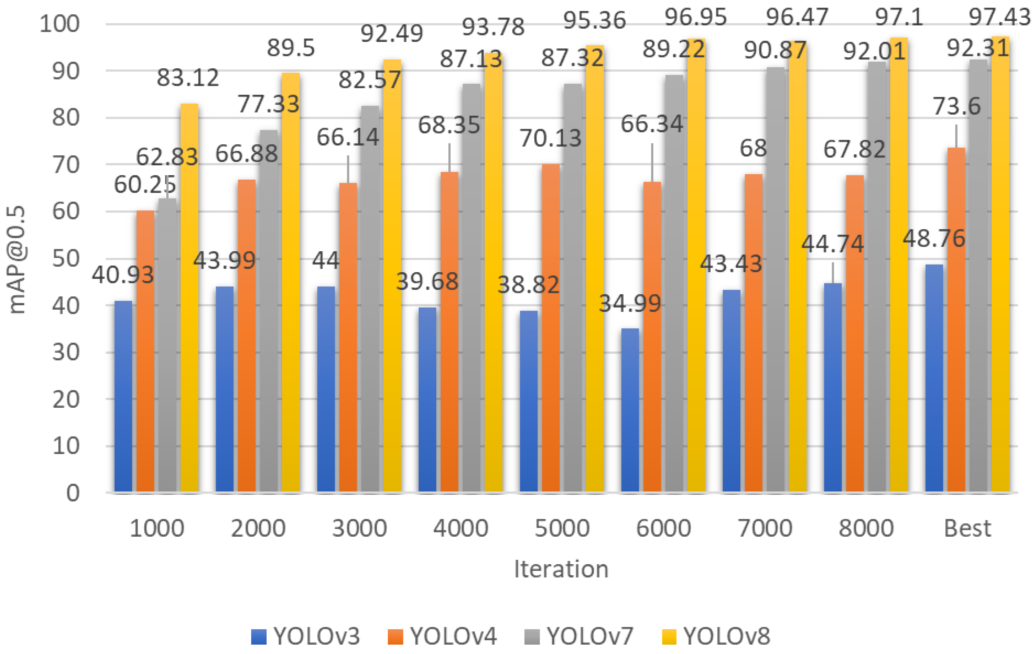


Figure 10: Comparison of Selected YOLO Series with Input Images At 250 Resolution

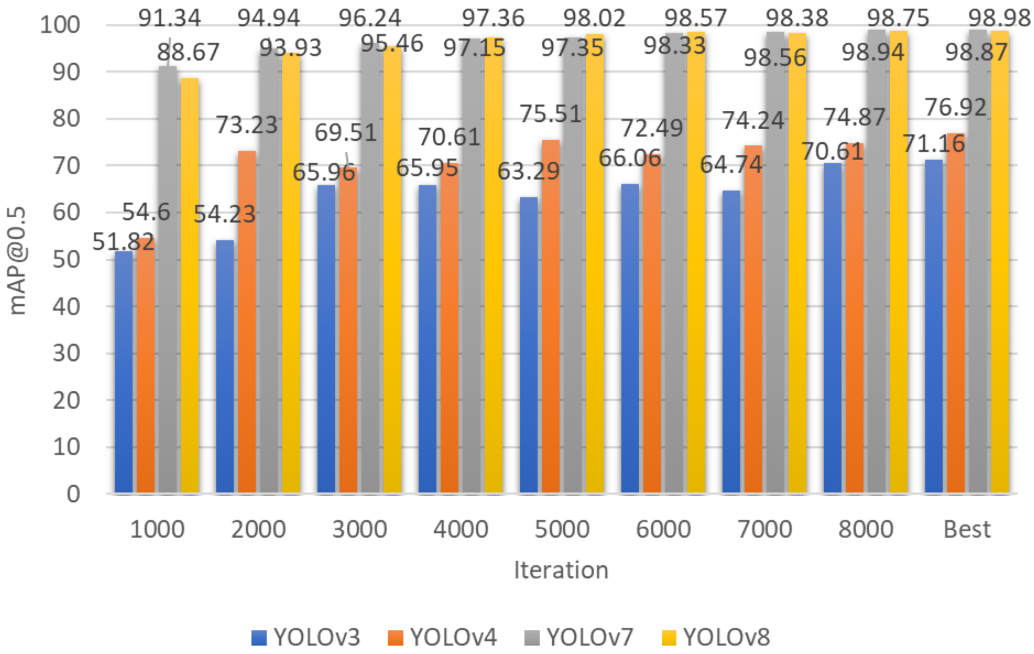


Figure 9: Comparison of Selected YOLO Series with Input Images At 416 Resolution

Precision, Recall and F1-Score Performance

The commonly used evaluation metrics that are typically implemented in ML, including DL models for binary classification tasks, are precision, recall, and F1-Score. These metrics give a detailed analysis of the performance of the given model, which is the ability to accurately

International Journal of Innovation and Industrial Revolution IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001
diagnose the positive cases while avoiding false positives and negatives. In Table 7 are the performance matrices of each YOLO model with respect to the input image resolutions.

Table 7: The Performance Matrix of YOLOv3, YOLOv4, YOLOv7, And YOLOv8 With Different Resolutions: 250 And 416

Resolution	**250**			**416**		
Model	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
YOLOv3	0.87	0.37	0.52	0.84	0.68	0.75
YOLOv4	0.73	0.74	0.71	0.82	0.74	0.78
YOLOv7	0.91	0.84	0.87	0.96	0.96	0.96
YOLOv8	0.94	0.94	0.94	0.96	0.97	0.97

For the 250 resolutions, YOLOv3 has obtained a precision of 0.87, a recall of 0.37, and an F1-score of 0.52. For the YOLOv4, the accuracy was 0.73, the corresponding recall was 0.74, and an F1 score of 0.71. With regard to the chosen metrics, the precision of YOLOv7 turned to 0.91 while recall was 0.84, and the F1-score was 0.87. At last, with all designs, YOLOv8 got a precision of 0.94, then a recall of 0.94, and an F1-score of 0.94.

This became better when the resolution was increased to 416; YOLOv3 scored a precision of 0.84, a recall of 0.68, and an F1-score of 0.75. The YOLOv4 also increased with a precision of 0.82, a recall of 0.74, and an F1-score of 0.78. As for the evaluation of the model, YOLOv7 maintained accuracy at 0.96, recall attained 0.96, and F1-score also remained constant at 0.96. Like the previous YOLO series, YOLOv8 also had a performance of a precision of 0.96, a recall of 0.97, and an F1-score of 0.97.

From the metrics, we able obtain a relevant assessment of the models' performance. A higher precision means that the class is less likely to be false positive, while a higher recall means that the class is less likely to be false negative. Thus, there is a single percentage number – termed F1-score – that gives a balanced analysis of the precision and recall. By comparing the two models, it becomes quite clear that the performance changes with the density of the grid resolution. The four architectures indicate that YOLOv3 and YOLOv4 perform worse than YOLOv7, and YOLOv7 performs worse than YOLOv8 in terms of precision, recall, and F1 scores.

Conclusion

Conclusively, the study has effectively acquired qualified datasets of oil palm FFBs and has also benchmarked and verified the efficacies of the selected YOLO series models, including YOLOv3, YOLOv4, YOLOv7, and YOLOv8. The comparison also included the consideration of what the change of the iteration from 1000 to 8000 mAP scores means to the data augmentation techniques, as well as the examination of the effect of the input image resolution. Our study also highlights the pivotal role of data augmentation in enhancing object detection, particularly evident in the substantial performance gap between YOLOv3 models with and without data augmentation. Furthermore, we demonstrate that the higher input image resolutions significantly boost the detection accuracy across YOLOv3, YOLOv4, YOLOv7, and YOLOv8 models. Especially by utilizing both YOLOv7 and YOLOv8, consistently outperforming their predecessors, compared with the legacy models evaluated, we have achieved impressive mAP values of 92.31% and 98.87% at 250 and 416 resolutions, respectively. It is worth noting that precision, recall, and F1-score evaluations further

International Journal of Innovation and Industrial Revolution
IJIREV
EISSN: 2637-0972
Volume 7 Issue 20 (March 2025) PP. 01-18
DOI 10.35631/IJIREV.720001

underscore the robustness of YOLOv7 and YOLOv8 across resolutions. By particularly emphasizing the significance of data augmentation and careful consideration of input image resolutions in achieving superior detection performance, these findings manifest invaluable insights into optimizing object detection models for diverse applications, which could be adopted from simply computer vision to more complicated real-world scenarios such as autonomous vehicles and surveillance systems.

Acknowledgements
This research was financially supported by MOSTI-TeD1. The grant number and details of the funding are as follows: TDF08211458.

References

Alfatah, T., Mistar, E. M., Syabriyana, M., & Supardan, M. D. (2022). Advances in oil palm shell fibre reinforced thermoplastic and thermoset polymer composites. *Alexandria Engineering Journal, 61*(6), 4945–4962. https://doi.org/10.1016/J.AEJ.2021.09.061

Black bunch census (BBC) at 4-month intervals. (n.d.). *One Oil Palm*. Retrieved September 14, 2023, from http://www.oneoilpalm.com/reviews/black-bunch-census-at-4-month-intervals/

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. *Computer Vision and Pattern Recognition, 10934*. Retrieved June 27, 2023, from https://github.com/AlexeyAB/darknet

Boesch, G. (2023). YOLOv7: The fastest object detection algorithm. *Viso.ai*. Retrieved June 27, 2023, from https://viso.ai/deep-learning/yolov7-guide/

Crowley, M. Z. (2020). Foreign labor shortages in the Malaysian palm oil industry: Impacts and recommendations. *Asian Journal of Agriculture and Development, 17*(2), 1–18. https://doi.org/10.22004/AG.ECON.307975

Deng, X., Zhao, H., Zhang, H., Zhang, J., & Mei, B. (2023). Grid self-occlusion: A grid self occlusion data augmentation for better classification. *Signal, Image and Video Processing, 17*(3), 705–713. https://doi.org/10.1007/s11760-022-02278-0

Feintrenie, L., Schwarze, S., & Levang, P. (2010). Are local people conservationists? Analysis of transition dynamics from agroforests to monoculture plantations in Indonesia. *Ecology and Society, 15*(4). Retrieved September 13, 2023, from https://www.jstor.org/stable/26268223

Ghazali, K. H., Samad, R., Arshad, N. W., & Karim, R. A. (2009). Image processing analysis of oil palm fruits for automatic grading. In *Proceedings of the International Conference on Instrumentation, Control & Automation*. Retrieved September 13, 2023, from https://www.academia.edu/download/4808957/ica2009-a22.pdf

Hai, T. C. (2002). The palm oil industry in Malaysia: From seed to frying pan. WWF Switzerland.

Jamal, U. (2021). US ban on palm oil from Malaysia's Sime Darby shows abuses are now common knowledge. *ASEAN Today*. Retrieved July 25, 2023, from https://www.aseantoday.com/2021/01/us-ban-on-palm-oil-from-malaysias-sime-darby-shows-abuses-are-now-common-knowledge/

Jocher, G., Chaurasia, A., Waxmann, S., & Laughing. (n.d.). Ultralytics YOLOv8 docs. *Ultralytics*. Retrieved July 4, 2023, from https://docs.ultralytics.com/

Karthikeyan, M., Subashini, T. S., Srinivasan, R., Santhanakrishnan, C., & Ahilan, A. (2023). YOLOAPPLE: Augment Yolov3 deep learning algorithm for apple fruit quality detection. *Signal, Image and Video Processing.* https://doi.org/10.1007/s11760-023-02710-z

Khamis, N., Selamat, H., Ghazalli, S., Md Saleh, N. I., & Yusoff, N. (2022). Comparison of palm oil fresh fruit bunches (FFB) ripeness classification technique using deep learning method. *ASCC 2022 - 13th Asian Control Conference Proceedings*, 64–68. https://doi.org/10.23919/ASCC56756.2022.9828345

Mahat, S. B. A. (2012). The palm oil industry from the perspective of sustainable development: A case study of Malaysian palm oil industry. Graduate School of Asia Pacific Studies, Ritsumeikan Asia Pacific University. Retrieved September 12, 2023, from https://core.ac.uk/download/pdf/60541187.pdf

Malaysia Palm Oil Council. (n.d.). About palm oil. Malaysia Palm Oil Council. Retrieved June 21, 2023, from https://mpoc.org.my/about-palm-oil/

Malaysian Palm Oil Board (MPOB). (2003). *Oil palm fruit grading manual* (2nd ed.). Malaysian Palm Oil Board.

Malaysia Palm Oil Council. (n.d.). *The oil palm tree*. Malaysia Palm Oil Council. Retrieved June 21, 2023, from https://mpoc.org.my/the-oil-palm-tree/

Marinoudi, V., Sørensen, C. G., Pearson, S., & Bochtis, D. (2019). Robotics and labour in agriculture: A context consideration. *Biosystems Engineering, 184*, 111–121. https://doi.org/10.1016/J.BIOSYSTEMSENG.2019.06.013

Otieno, N. E., et al. (2016). Palm oil production in Malaysia: An analytical systems model for balancing economic prosperity, forest conservation, and social welfare. *Agricultural Sciences, 7*(2), 55–69. https://doi.org/10.4236/AS.2016.72006

Parveez, G. K. A., et al. (2021). Oil palm economic performance in Malaysia and R&D progress in 2020. *Journal of Oil Palm Research, 33*(2), 181–214. https://doi.org/10.21894/jopr.2021.0026

Raghu, A. (2021). Palm oil output at risk as virus lockdown worsens labor shortage. *Bloomberg.* Retrieved July 25, 2023, from https://www.bloomberg.com/news/articles/2021-06-14/palm-oil-output-at-risk-as-virus-lockdown-worsens-labor-shortage#xj4y7vzkg

Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. *Computer Vision and Pattern Recognition, 1804*, 1–6. Retrieved June 27, 2023, from https://pjreddie.com/yolo/

Simplilearn. (n.d.). What is computer vision: Applications, benefits, and how to learn it. *Simplilearn.* Retrieved July 25, 2023, from https://www.simplilearn.com/computer-vision-article

Song, X., Gao, S., & Chen, C. (2021). A multispectral feature fusion network for robust pedestrian detection. *Alexandria Engineering Journal, 60*(1), 73–85. https://doi.org/10.1016/J.AEJ.2020.05.035

Srivastava, N. (n.d.). Computer vision and deep learning for AI-driven smart agriculture. *E2E Cloud.* Retrieved July 26, 2023, from https://www.e2enetworks.com/blog/computer-vision-and-deep-learning-for-agriculture

Stieg, C. (n.d.). Difference between AI, machine learning, deep learning. *Codecademy.* Retrieved July 25, 2023, from https://www.codecademy.com/resources/blog/difference-between-ai-machine-deep-learning-examples/

Syazwani, R. W. N., Asraf, H. M., Amin, M. M. S., & Dalila, K. N. (2022). Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning. *Alexandria Engineering Journal*, *61*(2), 1265–1276. https://doi.org/10.1016/J.AEJ.2021.06.053

Wahid, M. B., Weng, C. K., & Masri, R. (2009). Palm oil: Nature's gift to Malaysia and Malaysia's gift to the world. *Oil Palm Industry Economic Journal*, *9*(1).

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7464–7475). Retrieved July 4, 2023, from https://github.com/WongKinYiu/yolov7

Workman, D. (2023, June 22). *Palm oil exports by country 2021*. World's Top Export. Retrieved June 22, 2023, from https://www.worldstopexports.com/palm-oil-exports-by-country/