MEKATRONIKA
e-ISSN: 2637-0883
VOL. 5, ISSUE 1, 79 – 87
DOI: https://doi.org/10.15282/mekatronika.v5i1.9415

ORIGINAL ARTICLE

Deep Learning for Medium-Scale Agricultural Crop Detection through Aerial View Images

Lau Wei Hong[1], Mohd Azraai Mohd Razman[1,*], Mohd Izzat Mohd Rahman[1] and Muhammad Nur Aiman Shapiee[1]

[1]Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

ABSTRACT – This research project focuses on utilizing two state-of-the-art YOLOv4-based deep learning models, for large-scale agricultural crop detection using Unmanned Aerial Vehicles (UAVs). The objective is to develop an accurate and efficient crop detection system capable of identifying chili crops, eggplant crops, and empty polybags in agricultural fields. Crops detection is important for the development of a robotic vision in maximize the productivity and efficiency in agriculture associate with the development of concept Industry 4.0. This study seek to explore the comparison between YOLOv4 and YOLOv4 tiny model in term of mean average precision (mAP), precision, recall, F1-score, detection time and memory consumption. A custom dataset with 300 images was collected and annotated into total bounding boxes of 23335 with 6969 chili tree, 15402 eggplant tree and 964 empty polybag. The dataset was separated into train, validation and test set with the ratio of 70:20:10. The dataset was trained into YOLOv4 and YOLOv4 tiny with 2000 iterations. The result has shown that the YOLOv4 has the higher mean AP of 91.49% with 244.2mb memory storage consumption while YOLOv4 tiny achieve lower mean AP of 71.83% with 22.4mb. In summary, this research has significated the implementation of deep learning models to perform large-scale agricultural crop detection and can be further develop into automation industrial 4.0 of local agricultural sector.

ARTICLE HISTORY
Received: 24th March 2023
Revised: 26th April 2023
Accepted: 10th May 2023
Published: 19th May 2023

KEYWORDS
Deep Learning
Agriculture
Crops
Object Detection
Classification

INTRODUCTION

In Malaysia, agriculture is one of the commodities that have high economic value and have very high potential to development. Malaysia's gross agricultural output in 2015 was RM73,853.6 million, a rise of 6.7% annually from RM53,452.1 million in 2010 [1]. The significance of the agriculture sector in Malaysia's economic development was highlighted by this statistic. Hence, it is important to strictly control the harvest quality and quantity of crops.

In order to adapt to the development of technologies in computer vision, an accurate and timely identification of crop types and their distribution is essential for effective crop management, yield prediction, and resource allocation. Traditional methods of crop monitoring, such as ground surveys and satellite imagery, are limited in their ability to provide high-resolution and real-time data. Unmanned Aerial Vehicles (UAVs) equipped with advanced sensing technologies and deep learning models have emerged as a promising solution for large-scale agricultural crop detection. Drones are able to deliver helpful information on soil conditions, disease, plant maturity, and others effectively in real-time with the adaption of Industry 4.0 applications in agriculture [2].

There are various approaches in the development of agriculture crop detection, methods as in feature extraction using Normalized Difference Vegetation Index (NDVI) also one of the method to be apply in crop detection and health classification [3]. Other than that, neural network backbone of ResNet-18 and MobileNetV1 also implemented in chili, eggplant, and potato crop detection for the agriculture industry, from application in the farm, orchard, or greenhouse to the consumer, through the use of drones, automatic machinery, or autonomous systems [4].

Among several deep learning models, the object detection framework You Only Look Once (YOLO) stands out excellently. By accurately predicting bounding boxes and class probabilities in a single forward pass, YOLO models provide real-time processing capabilities [5]. YOLO has been created in a number of different iterations, including YOLOv4, YOLOv4 tiny, YOLOv3, and YOLOv2. These models are excellent candidates for crop detection in agricultural landscapes because of their great performance in a variety of item detection tasks. For this research, a detail comparison research done between YOLOv4 and YOLOv4 tiny model in their performance of large-scale agricultural crops detection and classification.

The remainder of this paper is organized as follows: Section 2 briefly provides a review on the recent works related to the topics of application YOLO-based deep learning object detection model in agricultural sector. Section 3 explains the methodology in custom dataset preparation. Section 4 provides the experimental settings and discusses the experimental results. Section 5 concludes the finding of the research.

*CORRESPONDING AUTHOR | Mohd Azraai Mohd Razman | ✉ mohdazraai@ump.edu.my

RELATED WORK

In the research proposed by Bochkovskiy et al. [6], The best object detection accuracy and speed are examined by the authors. The study uses the MS COCO dataset to compare various models. Comparison of performances in term of average precision between different model such as YOLOv4, EfficientDet, ASFF, HarDNet, RetinaNet, SM-NAS, NAS-FPN, ATSS, RDSNet, and CenterMask has been done. YOLOv4 are found to be on the Pareto optimality curve and to be faster and more accurate than the fastest and most accurate detectors as a result of the comparison.

Referring to a research paper presented by Chen et al. [7], to identify weeds and crops related to sesame, the authors created a new YOLO model called YOLO-sesame. The YOLOv4 model is the framework of the YOLO-sesame model. Authors compared the precision, recall, AP, F1-score, mAP, and FPS of YOLO-sesame to various architecture models for crop and weed recognition, such as Fast R-CNN, SSD, EfficientDet-d0, YOLOv3, YOLOv4, and YOLOv4-tiny. The experiments show that the YOLO-sesame model significantly enhances detection performance while preserving detection speed. The mAP value for sesame and weed detection is 96.16 percent and the FPS is 36.8. It performs better than popular models like YOLOv4, Fast R-CNN, SSD, and others.

According to a study done by Bao et al. [8], they suggested a training model based on DDMA-YOLO architecture for efficiently detecting and monitoring the tea leaf blight disease by unmanned aerial vehicle (UAV) remote sensing. For their study, 132 UAV tea photos were gathered in four plots. As a result, when compared to Faster R-CNN, SSD, Retinanet, YOLOv3, YOLOv4, and YOLOv5, the AP of the DDMA-YOLO attained the maximum value of 76.8 percent.

Based to a research done by MacEachern et al. [9], Six YOLO-based deep learning artificial neural network models were developed in order to estimate yield and detect the ripeness stage in wild blueberries. YOLOv3, YOLOv3-SPP, YOLOv3-Tiny, YOLOv4, and YOLOv4-Tiny were the models. The models were created, with YOLOv4 performing best with mean average precisions of 79.79 percent and 88.12 percent in categorising berry colours and ripeness in both cases. YOLOv4 also achieved the highest F1-score of 0.82 compared to others.

In compliance with the research paper from Fan et al. [10], the researcher proposed a real-time apple defects inspection using YOLOv4-based deep learning model. The outcome shown that pruning-based YOLO V4 networks outperformed state-of-the-art YOLOv4-based models like YOLO V4-tiny and YOLO V4-P. With a 93.74 percent mAP, 0.91 F1-score, 8.82 MB model size, and 8.36 ms of inference time, YOLO V4 performed exceptionally well.

Custom Image Dataset For Agricultural Crops Detection

Image acquisition

The data used in this study was collected from two farms: a chili farm and an eggplant farm. The chili farm was located at latitude 3° 34' 20.0" N, longitude 103° 23' 52.8" E, while the eggplant farm was located at latitude 3° 34' 58.1" N, longitude 103° 24' 36.4" E. The farms were situated in the area Kampung Bentan, Pekan, Pahang and were both owned and operated by local farmers.

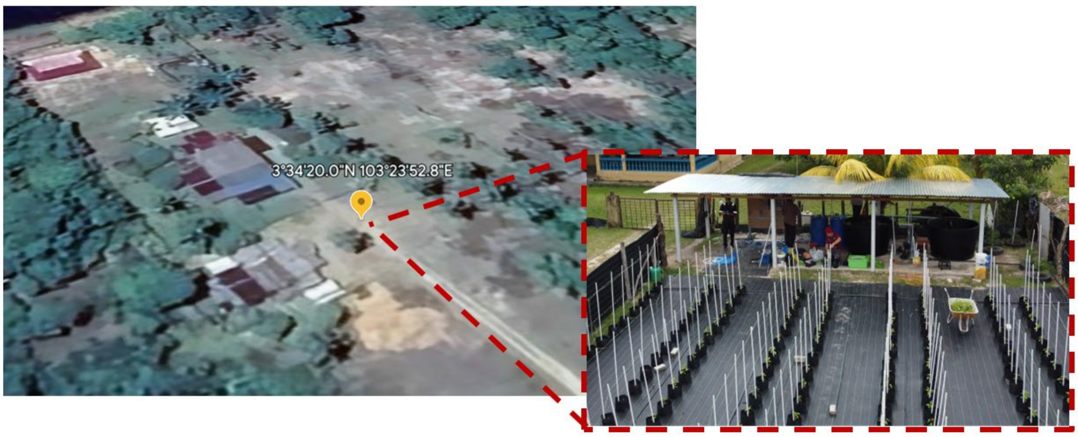

Figure 1. Geographical location of the chili farm

Figure 2. Geographical location of the eggplant farm

Data was collected over a three-month period, between February 2023 and April 2023. During this time, video footage was collected from both farms using the UAV device as mentioned in the previous section. A total of 12 min 47 sec of video clips were collected from the eggplant farm, while 19 min 33 sec of video clips were collected from the chili farm.

To ensure a comprehensive dataset, video footage was collected at different stages of the crop's growth, including the seedling, growth, and fruiting stages. By collecting footage at different stages, the dataset can provide a comprehensive view of the crop's development and operations at the chili and eggplant farms. From the obtained video clip, 300 of static frames were exported as png file. The exported image files were then filtered and then proceeded to the image annotation.

Figure 3. Physical appearance of crops in different stage

Image annotation

The images are labelled using an image labelling or annotation tool before being utilised for bounding box object recognition and segmentation. The image labelling tool is used, as its name implies, to identify the items in a picture. The main objective of the programme is to give users the ability to draw attention to or capture a particular object in a photo. There is various type of image annotation such as bounding box, polygonal segmentation, semantic segmentation, 3D cuboid segmentation, keypoint-landmarks, and line-splines [11].

In this study, “LabelImg” [12] was applied as the image annotation tool. For the type of image annotation, we used the most versatile method, bounding boxes as proposed by other researchers. We labelled the crops using the bounding boxes and named the classes as Chili, Empty, and Eggplant correspondingly. The annotation files outputted from the annotation tool was saved in the format of YOLO or txt file.

Performance valuation

Average Precision (AP) is a common metric for evaluating the precision of object detectors such as YOLO, Faster R-CNN, SSD, and others. When the recall value is more than or equal to 0, the average precision value is calculated. [13] “Precision” evaluates the precision of your predictions. in other words, what percentage of the prediction model were accurate.“Recall” evaluates how well the model find all the true positive detection. For instance, our object detection model was able to find how much percentage of the potential positive cases among all. While F1 score is the combination of metric precision and recall scores to evaluate the overall performance. Intersection Over Union (IoU) is a number that indicates how much the two boxes overlap. The overlap of the actual bounding box and prediction bounding box is calculated by IoU in the context of object identification and segmentation [14].

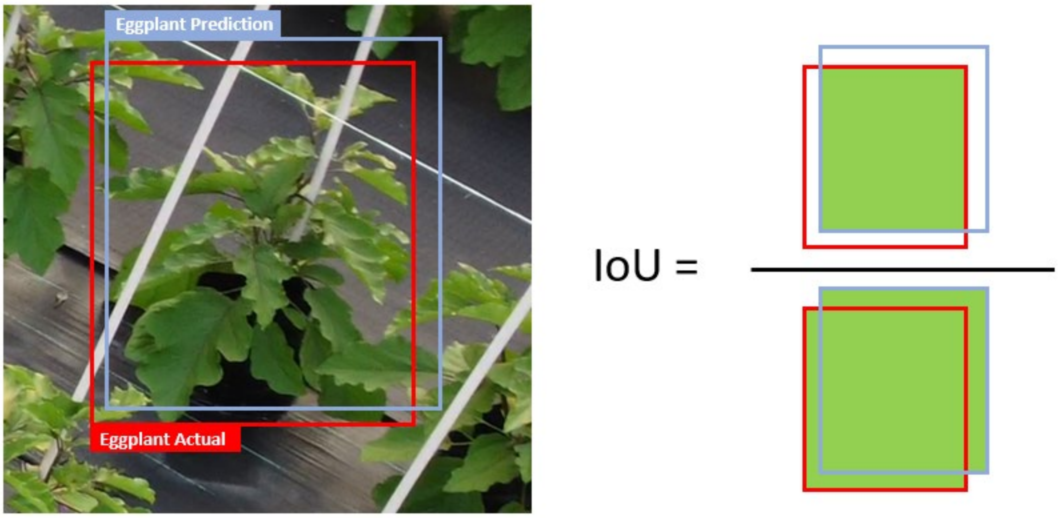

Figure 4. Visualization of IoU

EXPERIMENTAL RESULTS

This experiment investigate the performance of YOLOv4 in comparison with another state-of-the-art YOLOv4-based deep learning model as YOLOv4 tiny. The experiment was done on Google Colab [15], online GPU service with 12GB Nvidia K80, Ubuntu 20.04 LTS, Python 3.9, Cuda 12, and Cudnn 11.8.

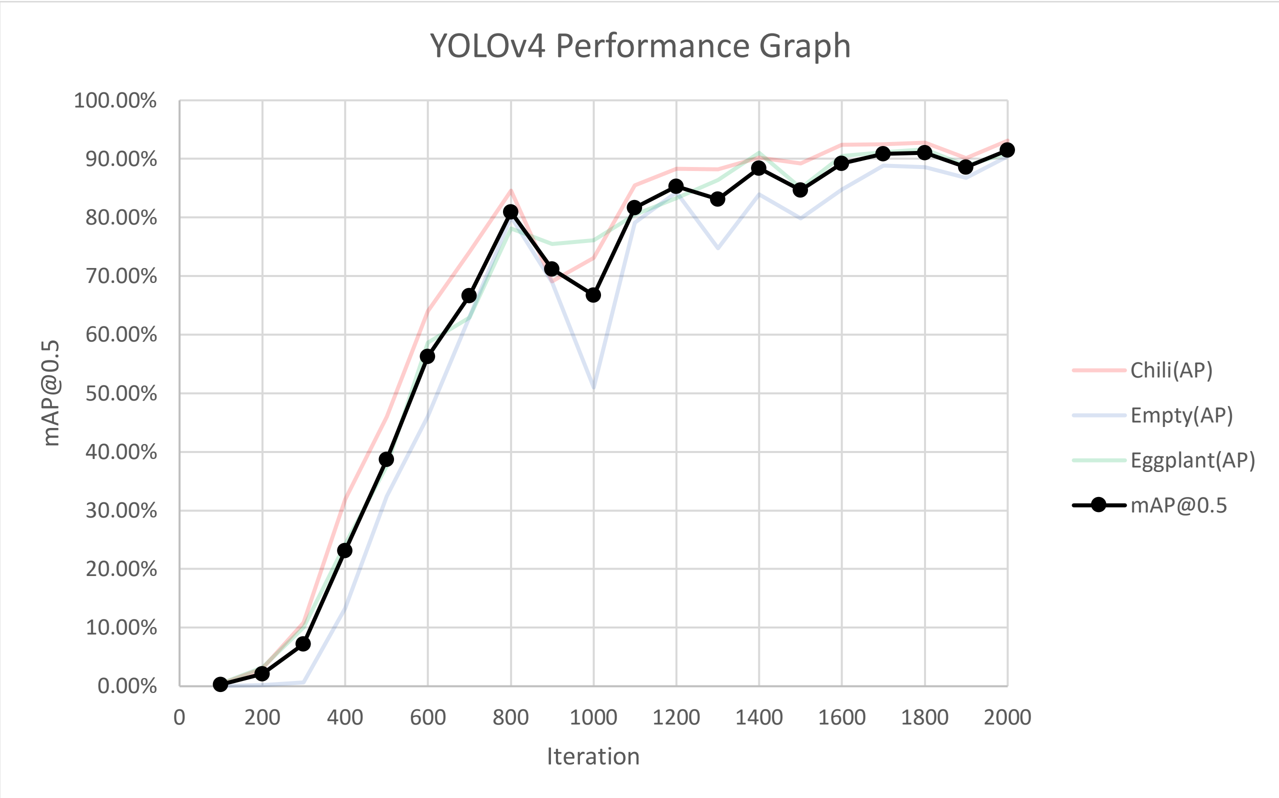

Figure 5. YOLOv4 Validation Performance (mAP@0.5)

Figure 6. YOLOv4 Tiny Validation Performance (mAP@0.5)

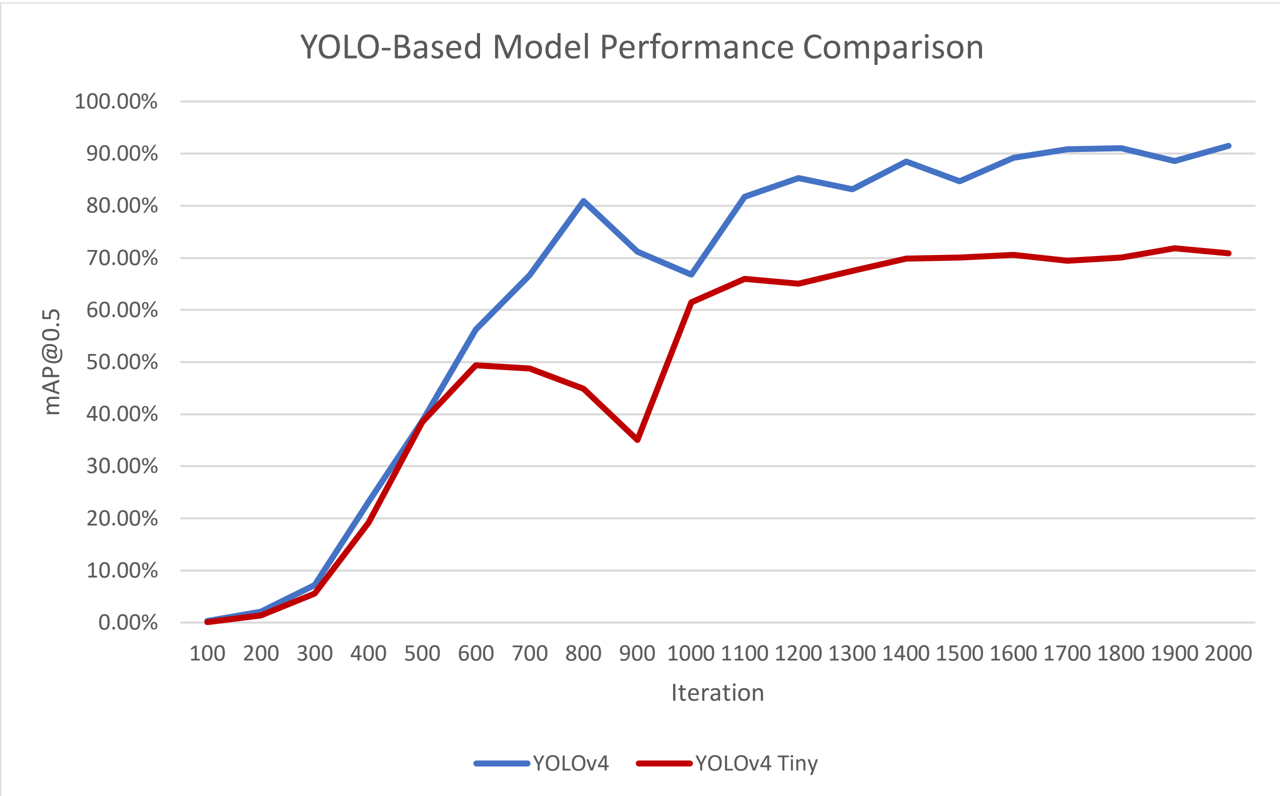

Figure 7. Mean AP comparison of YOLOv4-based models with 0.5 IOU

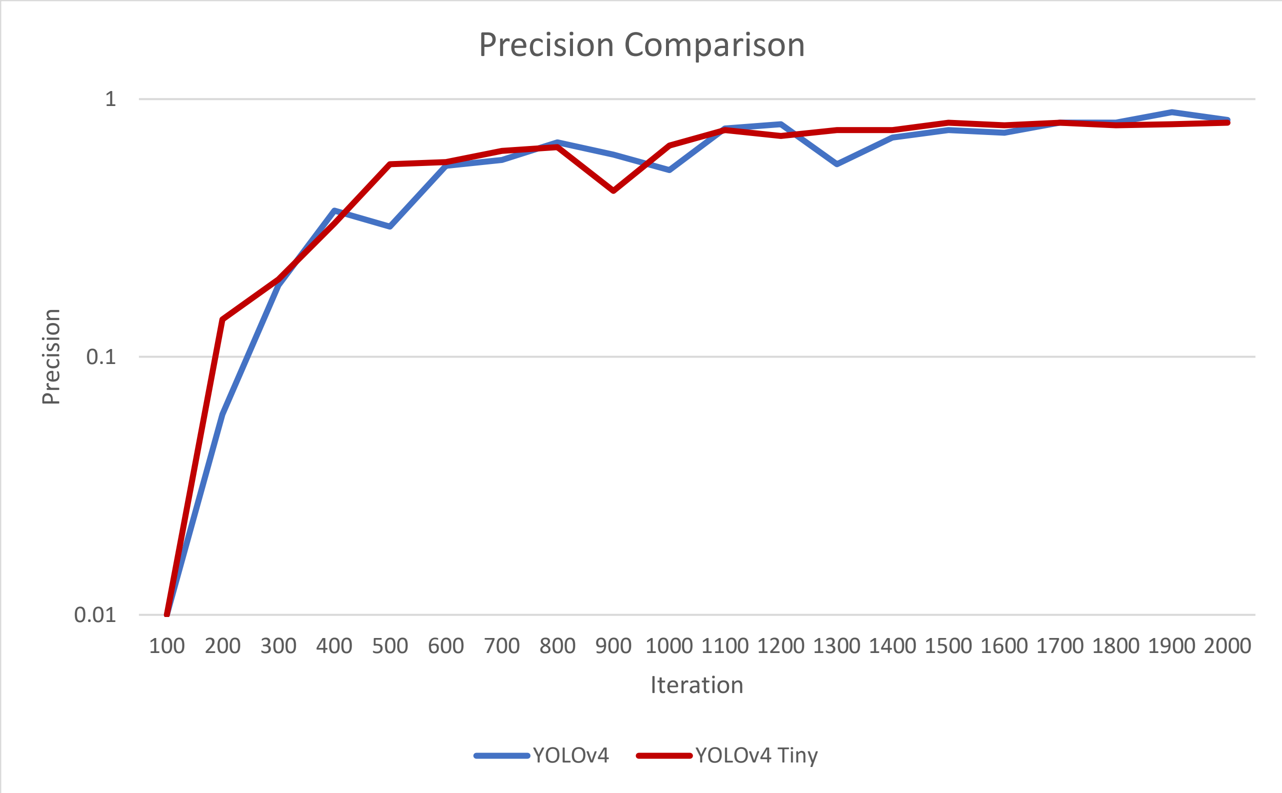

Figure 8. Precision Graph

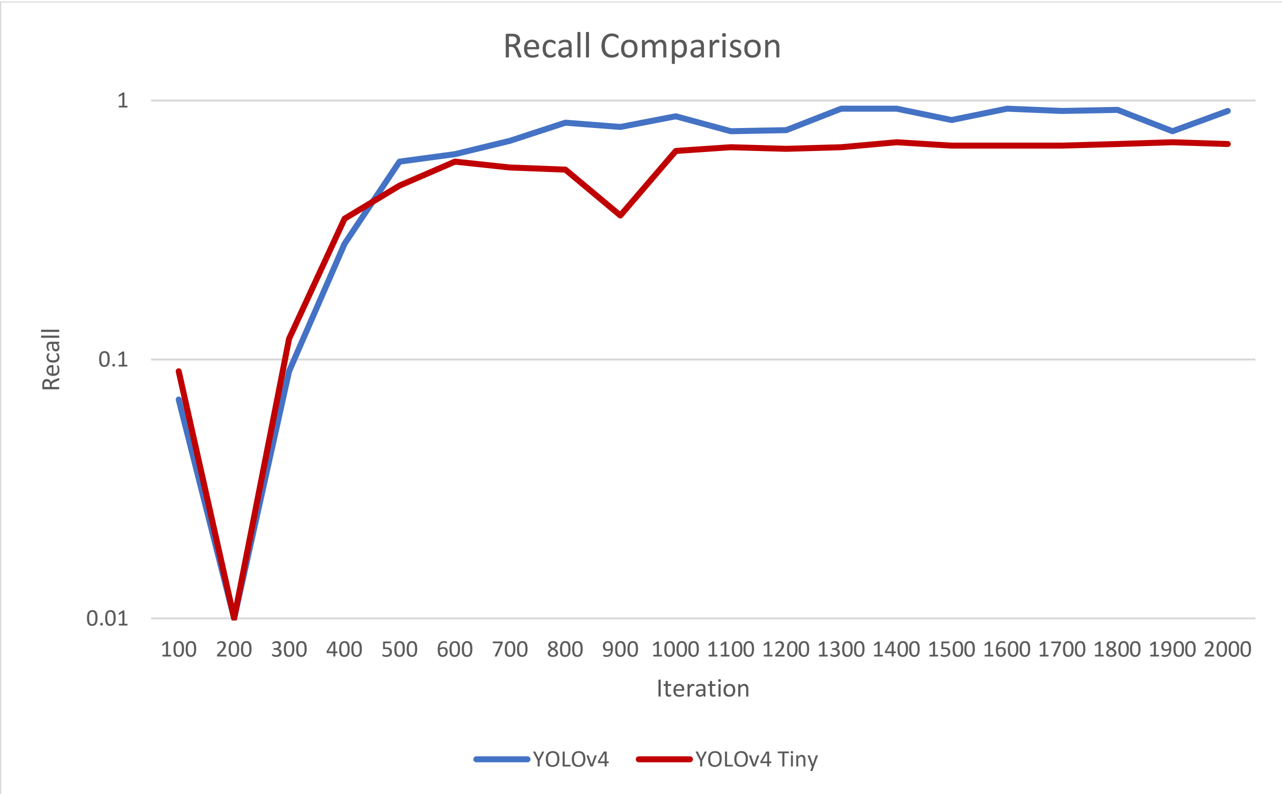

Figure 9. Recall Graph

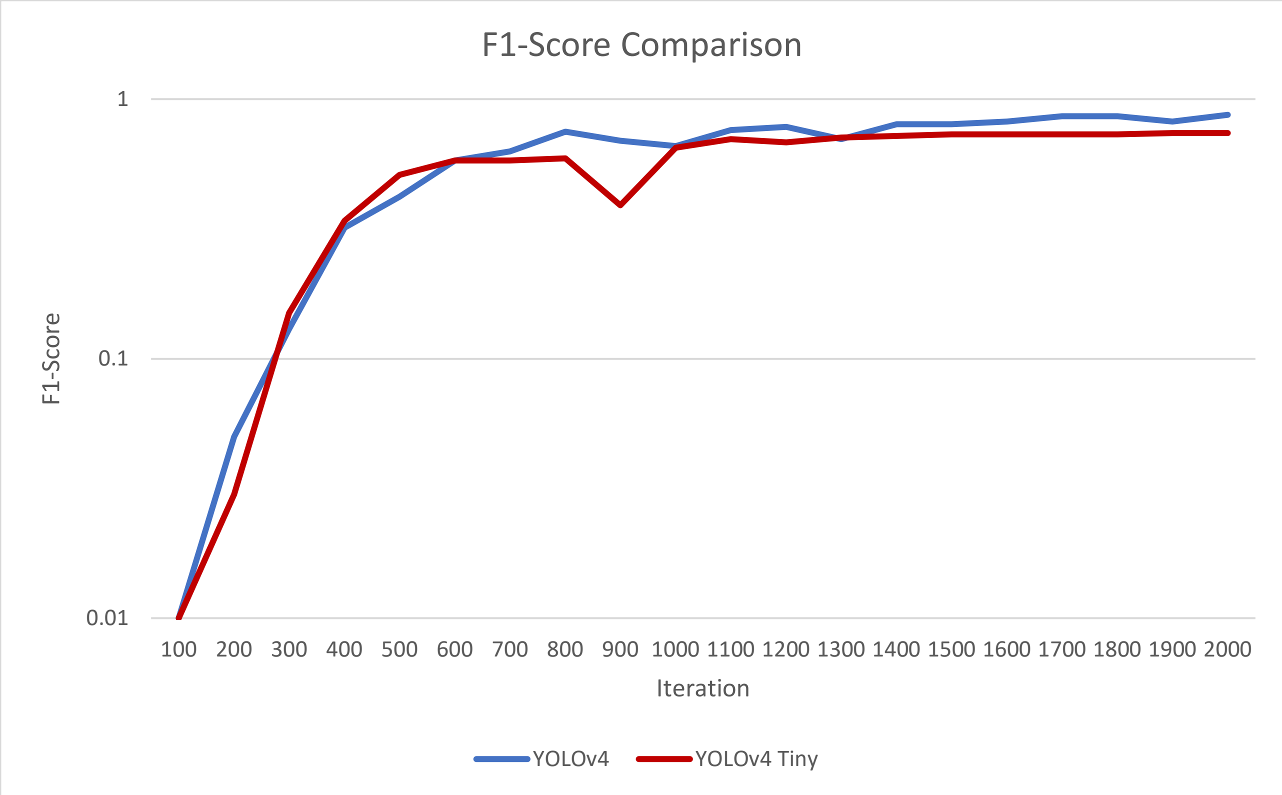

Figure 10. F1-Score Graph

Based on the observation from all graphs above for model YOLOv4 and YOLOv4 tiny, a graph trend of having relatively high increasing rate can be observed in the early training iteration. After certain numbers of iterations, the average precision tend to slow down the improvement rate. There were several sudden downward spikes observed within the increment over iterations. The decrement of the performance highly related to the backpropagration process during neural network training process. The size or frequent of the sharp spikes occurances highly related to the learning rate of the training configuration. After reaching the second half of the graph, the performance trend tend to increase gradually closer to 100% of average precision. Similar trend of increment also portraited on the parameters of precision, recall and F1-score. At the end of 2000 iterations, YOLOv4 reached the highest mAP of 91.49%, which is 27.36% higher than YOLOv4 tiny with the highest mAP of 71.83%.

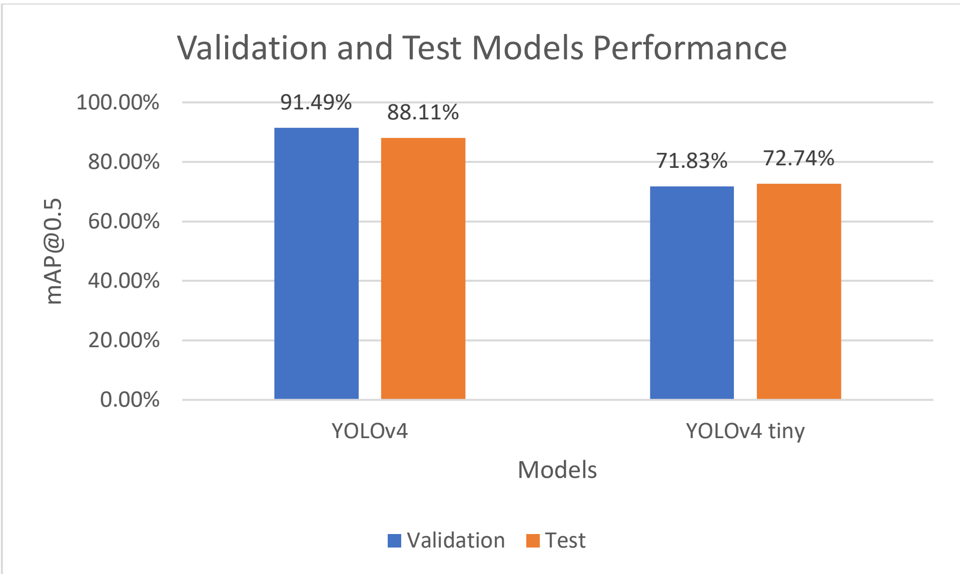

Figure 11. Validation and Test Models Performance Chart

Table 1. Performance details of YOLOv4-based model

Models	Model size	Average detection speed
YOLOv4	244.2 MB	794.5 ms
YOLOv4 tiny	22.4 MB	689.9 ms

The bar graph compares the performance of the YOLOv4 and YOLOv4-tiny models according to their test and validation results using the evaluation metric of Mean Average Precision (mAP) at a 0.5 level of confidence. The validation mAP@0.5 for YOLOv4 is stated to be 91.49 %, which indicates a high degree of accuracy in identifying and localizing objects in the validation dataset. YOLOv4-tiny, on the other hand, obtained a validation mAP@0.5 of 71.83 %, showing a relatively lower but still remarkable performance.

When based on environmental test results, YOLOv4 attained a mAP@0.5 of 88.11 %. This indicates that the model maintained a level of accuracy when used with unseen test data. YOLOv4-tiny demonstrated a slightly better test performance with a mAP@0.5 of 72.74 % despite achieving a lower mAP in validation.

The table compares YOLOv4 with YOLOv4 tiny in terms of model size and average detection speed. The model size of YOLOv4 is reported in the table as 244.2 MB, indicating a comparatively greater model size. In contrast, YOLOv4 tiny has a model size that is 22.4 MB, which makes it lighter and easier to deploy on platforms or devices with limited resources. According to the table's average detection speed, YOLOv4 has an average detection speed of 794.5 ms. On the other hand, displays a 15.2% faster average detection speed of 689.9 ms. These results were based on the performance using GPU 12GB Nvidia K80 from online virtual machine, Google Colab.

The result highlights the importance of particular parameters based on the requirements and limitations when deciding between YOLOv4 and YOLOv4 tiny. Due to its smaller model size and faster detection performance, YOLOv4 tiny may be preferred if computational resources are constrained or real-time processing is essential. On the other hand, YOLOv4 can give superior precision in object recognition tasks even with its larger model size and slightly longer processing time if accuracy is of the utmost significance and computational resources are not a limiting constraint.

CONCLUSION

In summary, this study investigated the use of YOLOv4 and YOLOv4 tiny deep learning models for large-scale agricultural crop detection using UAV imagery. The outcomes showed that both models were capable of effectively detecting and categorising eggplant, chilli, and empty polybags in agricultural fields. While YOLOv4 tiny offered a compromise between accuracy and computational efficiency and was suited for real-time crop detection on resource-constrained UAV platforms, YOLOv4 shown higher accuracy but required more processing resources. The results show the feasibility of YOLOv4-based models for precision agriculture, allowing local farmers to allocate resources and monitor crops effectively. The study's findings help crop detection systems get better, and they give farmers and agronomists useful information for improving crop management techniques. The models may need to be improved more in the future, and their applicability to various crop varieties and environmental factors may also be investigated.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Malaysia Pahang for funding this work under an internal grant RDU200332.

REFERENCES

[1] Department of Statistics Malaysia Official Portal, "Economic Census 2016 - Crops," *Source Malaysia's Off. Stat.*, no. July, 2017, [Online]. Available: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=404&bul_id=d01JVFg2UzYwaW5iZnpjK1ZBbEVxUT09&menu_id=Z0VTZGU1UHBUT1VJMFlpaXRRR0xpdz09

[2] Agmatix, "The role of industry 4.0 in agriculture," 2023. https://www.agmatix.com/blog/the-role-of-industry-4-0-in-agriculture/ (accessed Jun. 01, 2023).

[3] S. Puteh *et al.*, "Features Extraction of Capsicum Frutescens (C . F) NDVI Values using Image Processing," vol. 2, no. 1, pp. 38–46, 2020.

[4] M. I. M. Rahman, M. A. M. Razman, I. M. Khairuddin, A. P. P. A. Majeed, M. A. Abdullah, and W. H. M. Isa, "Various Type of Crops and Trees Detection Using Clustering Technique Through Image Processing," *Lect. Notes Electr. Eng.*, vol. 988, pp. 325–332, 2023, doi: 10.1007/978-981-19-8703-8_28/COVER.

[5] A. Sharma, "Achieving Optimal Speed and Accuracy in Object Detection (YOLOv4) - PyImageSearch," 2022. https://pyimagesearch.com/2022/05/16/achieving-optimal-speed-and-accuracy-in-object-detection-yolov4/ (accessed Jun. 01, 2023).

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," Apr. 2020, Accessed: Jan. 16, 2023. [Online]. Available: http://arxiv.org/abs/2004.10934

[7] J. Chen *et al.*, "Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion," *Comput. Electron. Agric.*, vol. 202, p. 107412, Nov. 2022, doi: 10.1016/J.COMPAG.2022.107412.

[8] W. Bao, Z. Zhu, G. Hu, X. Zhou, D. Zhang, and X. Yang, "UAV remote sensing detection of tea leaf blight based on DDMA-YOLO," *Comput. Electron. Agric.*, vol. 205, p. 107637, Feb. 2023, doi: 10.1016/J.COMPAG.2023.107637.

[9] C. B. MacEachern, T. J. Esau, A. W. Schumann, P. J. Hennessy, and Q. U. Zaman, "Detection of fruit maturity stage and yield estimation in wild blueberry using deep learning convolutional neural networks," *Smart Agric. Technol.*, vol. 3, p. 100099, Feb. 2023, doi: 10.1016/J.ATECH.2022.100099.

[10] S. Fan *et al.*, "Real-time defects detection for apple sorting using NIR cameras with pruning-based YOLOV4 network," *Comput. Electron. Agric.*, vol. 193, p. 106715, Feb. 2022, doi: 10.1016/J.COMPAG.2022.106715.

[11] S. Pokrhel, "Image Data Labelling and Annotation — Everything you need to know | by Sabina Pokhrel | Towards Data Science," *towards data science*, 2020. https://towardsdatascience.com/image-data-labelling-and-annotation-everything-you-need-to-know-86ede6c684b1 (accessed Jun. 02, 2023).

[12] T. Lin, "labelImg · PyPI," 2021. https://pypi.org/project/labelImg/1.4.0/ (accessed Jun. 02, 2023).

[13] J. Hui, "mAP (mean Average Precision) for Object Detection | by Jonathan Hui | Medium," 2018. https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173 (accessed Jun. 01, 2023).

[14] Kukil, "Intersection over Union (IoU) in Object Detection and Segmentation," *learnopencv.com*, 2022. https://learnopencv.com/intersection-over-union-iou-in-object-detection-and-segmentation/ (accessed Jun. 01, 2023).

[15] Prabanjan Raja, "What is Google Colab?," *Scaler Topics*, 2022. https://www.androidpolice.com/google-colab-explainer/ (accessed Jun. 02, 2023).