IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 27 July 2022, accepted 29 August 2022, date of publication 6 September 2022, date of current version 16 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3204762

APPLIED RESEARCH

Real-Time Detection of Ripe Oil Palm Fresh Fruit Bunch Based on YOLOv4

JIN WERN LAI, HAFIZ RASHIDI RAMLI, (Member, IEEE), LUTHFFI IDZHAR ISMAIL, AND WAN ZUHA WAN HASAN, (Senior Member, IEEE)

Department of Electrical and Electronic Engineering, Faculty of Engineering, Universiti Putra Malaysia (UPM), Serdang, Selangor 43400, Malaysia

Corresponding author: Hafiz Rashidi Ramli (hrhr@upm.edu.my)

This work was supported by the Universiti Putra Malaysia Research Grant (Putra Grant, GP-GPB/2021/9699100).

ABSTRACT Fresh Fruit Bunch (FFB) is the main ingredient in palm oil production. Harvesting FFB from oil palm trees at its peak ripeness stage is crucial to maximise the oil extraction rate (OER) and quality. In current harvesting practices, misclassification of FFB ripeness can occur due to human error, resulting in OER loss. Therefore, a vision-based ripe FFB detection system is proposed as the first step in a robotic FFB harvesting system. In this work, live camera input is fed into a Convolutional Neural Network (CNN) model known as YOLOv4 to detect the presence of ripe FFBs on the oil palm trees in real-time. Once a ripe FFB is detected on the tree, a signal is transmitted via ROS to the robotic harvesting mechanism. To train the YOLOv4 model, a large number of ripe FFB images were collected using an Intel Realsense Camera D435 with a resolution of 1920 × 1080. During data acquisition, a subject matter expert assisted in classifying the FFBs in terms of ripe or unripe. During the testing phase, the result of the mean Average Precision (mAP) and recall are 87.9 % and 82 % as the detection fulfilled the Intersect over Union (IoU) with more than 0.5 after 2000 iterations and the system operated at the real-time speed of roughly 21 Frame Per Second (FPS).

INDEX TERMS Object detection, oil palm, fresh fruit bunch, fruit maturity, YOLO.

I. INTRODUCTION

Malaysia is one of the biggest palm oil-producing countries in the world. The palm oil industry is a significant contributor to the country's Gross Domestic Product (GDP). Palm oil companies have more than a million hectares of plantation land to produce Fresh Fruit Bunches (FFBs) which will be harvested when it is ripe to extract their valuable oil. Therefore, several rules and guidelines were developed to achieve the maximum oil extraction rate (OER) according to the guideline of the Malaysian Palm Oil Board (MPOB). In the oil palm estate, the FFBs can only be harvested once the trees reach maturity at three years old. The field workers will harvest the FFBs on the 10th-14th days of the harvesting interval. The harvesters will search for oil palm trees with a certain number of detached fruitlets that have dropped to the ground. According to the current guideline, this indicates that there are FFBs on the trees that have ripened and should be harvested. The ripe FFBs are usually identified by their colour which is a bright red and yellow, in contrast to the brown and black of unripe FFBs. The harvested FFBs will then be collected and transported to the palm oil mill for oil extraction [1]. The general rule is that the FFBs are to be delivered to the mill within 24 hours after harvesting to ensure the quality of the fruit is at the highest level. However, this implementation is not guaranteed due to factors such as rain during the harvesting process and other unforeseen logistical issues.

The problem of labour shortage has had a tremendous impact on the economic growth of the oil palm industry which is traditionally very labour intensive [2]. Oil palm estates have reported labour shortages of approximately 20-30 %, which affected the potential yield to decrease by around 15 % due to post-harvest losses [3]. Therefore, the plantation industry should resolve this problem by implementing the latest technology in the FFB harvesting process.

Since the last decade, numerous methods of identifying the maturity of FFBs have been proposed by researchers to

The associate editor coordinating the review of this manuscript and approving it for publication was Utku Kose.

help ensure that the FFBs harvested have reached the proper maturity stage. Generally, there are two types of techniques to analyse the ripeness of FFBs which are destructive methods and non-destructive methods. Destructive testing methods require physical contact in such a way that affects the integrity of the FFB and severely reduces the amount of oil that can be extracted from the tested FFB such as fruit battery [4]. This technique uses a copper and zinc electrode to prick the oil palm fruitlet for measurement. These metal electrodes prick through the surface of the oil palm fruitlet and an ionization occurred. The electrolyte reaction indicates the element contents in the fruitlet and that varies according to the different maturity stages. The characteristic of moisture and lipid content become an indicator to justify the maturity stages.

Non-destructive techniques can involve non-contact features that are either visual or non-visual. Visually, the most popular technique was to use colour-based image analysis [5], [7], [8]. Others have also explored using near-infrared (NIR) spectroscopy [9], [10], multispectral analysis [11], [12], LIDAR [13], [14], fluorescence sensing[15], and laser light backscattering [16]. In terms of non-visual features, previous work has been done on using microwave sensors [17], inductive sensors [18], and the fruit battery method [19].

Although several methods are proposed to identify the maturity of FFB, most of these are still in the experimental stage and more studies are needed to determine their effectiveness for FFB ripeness detection. Most methods have only been tested in a laboratory environment. Among these are the fluorescence technique[15], near-infrared spectroscopy [10], laser backscattering imaging [16], and fruit battery method [20]. Those methods could potentially be suited for detection applications in the mill after harvesting as a mechanism for quality control. From the literature, it is found that the methods featuring colour images [21], LiDAR [22], and inductive sensors [23] are being developed for the application phase currently.

In the past, conventional methods were applied to classify the ripeness of FFBs, but the computational complexity of these methods did not allow for real-time operation. This has changed in the last decade due to three factors. Firstly, advancements in deep learning algorithms have given rise to a large variety of Convolutional Neural Networks (CNN) models that are more powerful than ever. Secondly, the existence of an established artificial intelligence (AI) development environment that provides software tools and guidance to use deep learning techniques to solve a multitude of problems ranging from facial recognition to cancer cell detection. Thirdly, the production of high-powered Graphical Processing Units (GPUs) and GPU-enabled Single Board Computers (SBC) that have become widely available make deep learning accessible to more researchers and encourage its development and implementation into real-world systems.

In this paper, we proposed using a deep learning model to called YOLO (an acronym for 'You Only Look Once') to detect the ripe FFBs [24]. The advantages of the YOLO model are that it is able to perform detection with high accuracy and in real-time due to its speedy processing technique. In the past, the YOLO series model had been implemented to detect agricultural fruits such as apples [25], tomatoes [26], and pears [27]. In the oil palm sector, Junos *et al.* [28] developed an automatic detection system that included the YOLO model to detect FFBs. The authors compared the performance between YOLOv3 series models and the result show the model is feasible for object detection in the oil palm sector.

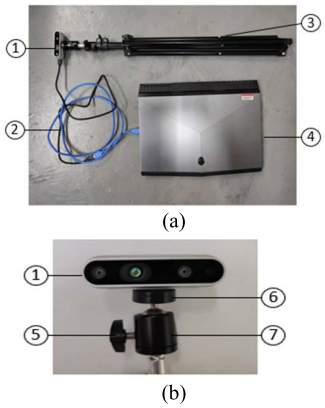
In this project, the objective is to develop a system to automatically detect ripe, unharvested FFBs in real-time using a combination of computer vision and artificial intelligence (AI). Real-time operation is considered an important feature of this system because it is meant for on-field application as part of a robotic harvesting mechanism. Firstly, an RGB camera (Intel Realsense D435) captures the view of the oil palm tree and transmits the data to a Single Board Computer (Nvidia Jetson NX) that is loaded with an inference model based on YOLOv4. The trained algorithm would identify the target object, record its coordinates, and send its positional information to the robotic harvesting mechanism within the Robot Operating System (ROS). Based on the coordinate received from the detection module, the positional information of the FFB was obtained by using a kinematic equation. Since it has been released for some time, YOLOv4 has become highly compatible with ROS which is a core component of the robotic harvesting system that is being developed separately. This was the main factor in the selection of YOLOv4 for this work. Also, although it was not the latest iteration of the YOLO model, it is still able to perform object detection with high accuracy and speed.

II. MATERIALS AND METHOD

In order to develop an AI-based vision system to detect ripe FFBs on the oil palm trees, the algorithm must be trained to do so using visual data or samples of ripe FFBs on the trees. In this section, the work done for data acquisition, preparation and training will be explained in detail.

A. DATA ACQUISITION

The oil palm trees selected for data capture range from 8 to 13 years old because this is when the trees produce the most FFBs. The sample data was recorded from November to December 2021 at an oil palm plantation in the state of Selangor, Malaysia. During data acquisition, a subject matter expert assisted in identifying ripe and unripe FFBs on the trees. Fig. 1 shows the hardware setup used for this work. An Intel Realsense D435 Camera was used to capture the FFB visual data. The camera was mounted onto an adjustable platform to enable it to capture images at the same height or level as the FFBs on the trees. The captured images have a resolution of 1920×1080 pixels which are stored in a laptop computer connected via a USB-C cable. The laptop computer used for capturing and storing data is mounted with an Intel



1. Intel Realsense Camera D435
2. USB cable
3. Adjustable platform
4. Laptop
5. Fixing screw
6. Screw mount
7. Ball and cage

FIGURE 1. (a) Hardware setup and (b) mounted camera used for training data collection and testing detection performance.

Core i7-8750H processor and GeForce DTX 1070 graphic card.

In the beginning, the camera is lifted at the same elevation and the distance is 3 meters away from the targeted FFB. Next, the camera is moved toward the FFB and positioned at different angles to ensure the video captures many varied views of the FFB on oil palm trees. Then, image data was extracted from a total of 49 oil palm trees at different time intervals over the course of 3 days to allow for variation in the environment and lighting in the training data. From the 49 oil palm trees sampled, 24 oil palm trees had at least one ripe FFB on them whereas 25 oil palm trees did not have any ripe FFBs.

B. DATA PREPARATION AND TRAINING

In the workflow for preparing the training data, the videos captured of the oil palm trees are first extracted into images. These images can be categorized into positive images and negative images. Positive images are images of trees that contain the desired object for detection, which is the ripe FFB. In contrast, negative images are images of trees without any ripe FFBs. There are a total of 240 positive images and 250 negative images. After that, 10 images were systematically chosen for each tree among a large number of images after extraction. The reason only 10 images were taken is that we wanted to avoid redundancy in the training images. These 10 images chosen show the subject in a variety of angles, backgrounds, positions, and lighting. This variation is very important in the training process of the deep learning algorithm as it requires this information for it to be able to learn and detect patterns in the images.

Next, the positive images undergo a manual labelling process using the labelImg software. The positive and negative images are further separated into two datasets which are the training set and validation set. The positive images are divided into 210 and 30 images for the training set and validation set. Meanwhile, the negative images are divided into 220 and 30 images for the training set and validation set. The images extracted from each tree are unique, which means that the images from the same tree would not be included in both the training set and validation set. Table 1 shows the separation of the images for YOLOv4 training.

TABLE 1. Separation of images for training the YOLOv4 model.

Dataset	Training set	Validation set	Number of images
Positive	210	30	240
Negative	220	30	250
Total	430	60	490

TABLE 2. The values of the parameter in the YOLOv4 model.

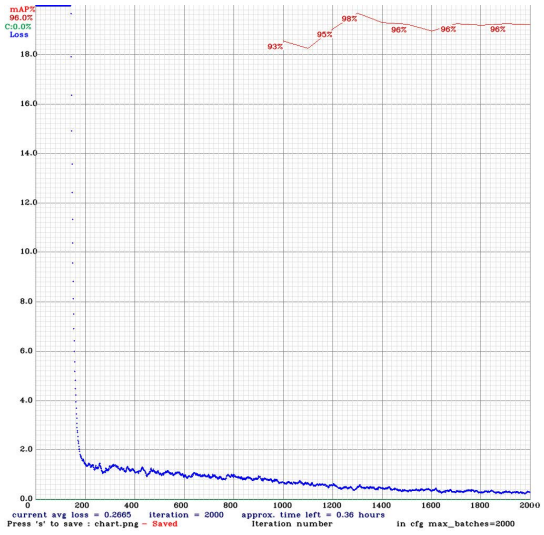
Parameter	Value applied
Width	608
Height	608
Momentum	0.949
Decay	0.0005
Angle	0
Saturation	1.5
Exposure	1.5
Hue	0.1
Learning rate	0.001
Maximum batch	3000

In this study, the parameters applied in the YOLOv4 model are the default values provided by Bochkovskiy *et al.*[29]. Table 2 shows the values of the parameters used in the training. The data augmentation includes the parameters exposure and saturation with a factor of 1.5 to randomly change of intensity and brightness of colour present in the image while during training and the parameter hue is set as 0.1. In addition, another method of data augmentation, Mosaic, was included during training, which contributes to the overall improvement in the YOLOv4 model. The Mosaic method mixes 4 training images to generate a new image that is contextually different in context from the original. However, the size of the anchor box, S is required to calculate according to the number of object detection. The formula to calculate the size of the anchor box is shown as:

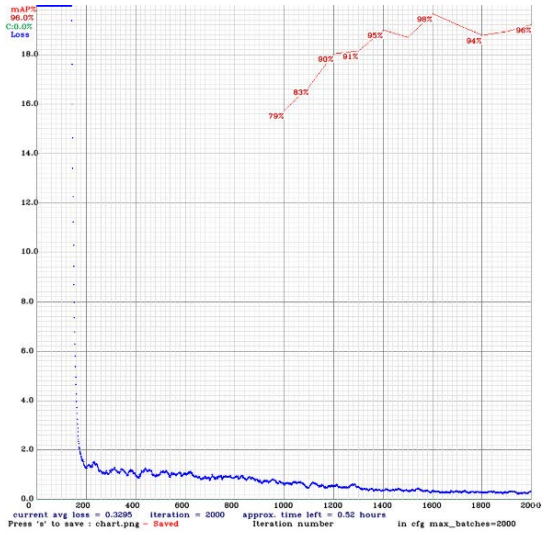
$$S = (N + C) \times A \tag{1}$$

where N represents the number of object detection, C represents the prediction of coordinates, and A represents the number of anchor boxes per grid.

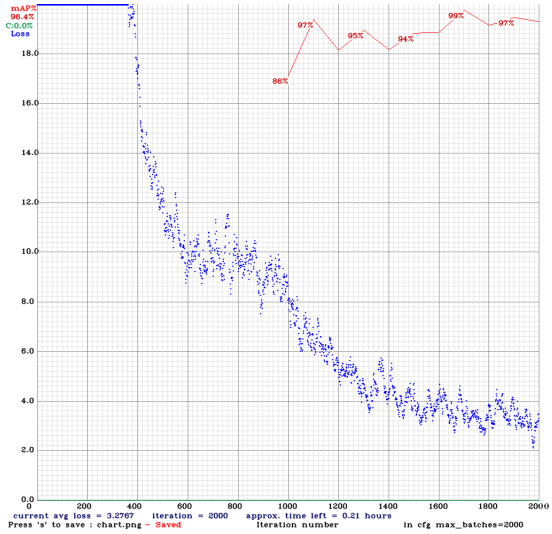
In this study, a few variations of the YOLOv4 model are evaluated to compare their performance in the task of FFB detection. The additional model applied in this study is Scaled-YOLOv4 which was specially designed to suit various GPUs during operation. For example, the YOLOv4-CSP and YOLOv4-tiny architectures are both categorised under Scaled-YOLOv4. The YOLOv4-CSP was designed with an emphasis on balancing between the execution speed and accuracy rather than the general YOLOv4 which is more focused on fast operating speed and optimization for parallel computation. On the other hand, YOLOv4-tiny was designed for implementation in low-spec devices as the amount of computational complexity and model size have reduced [30]. Thus, the performance of FFB detection using these three types of models which are YOLOv4, YOLOv4-CSP, and YOLOv4-tiny will be examined. The result is analysed after



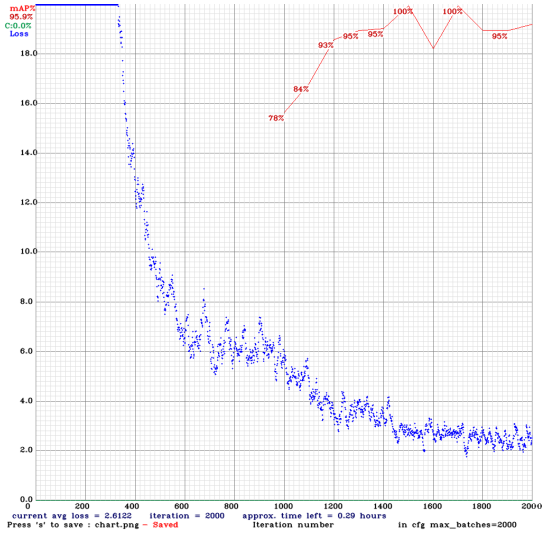
(a)



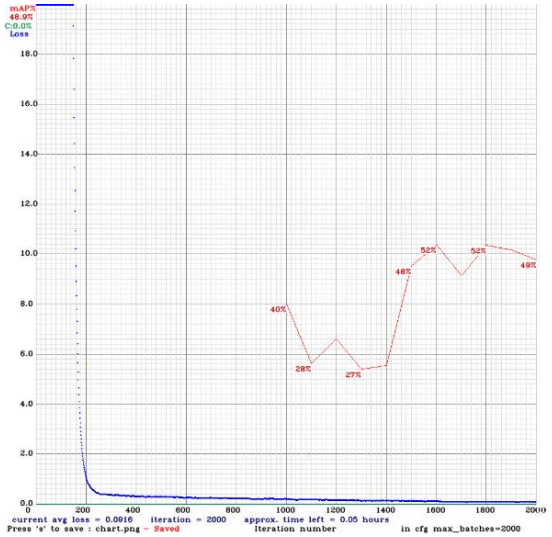
(b)



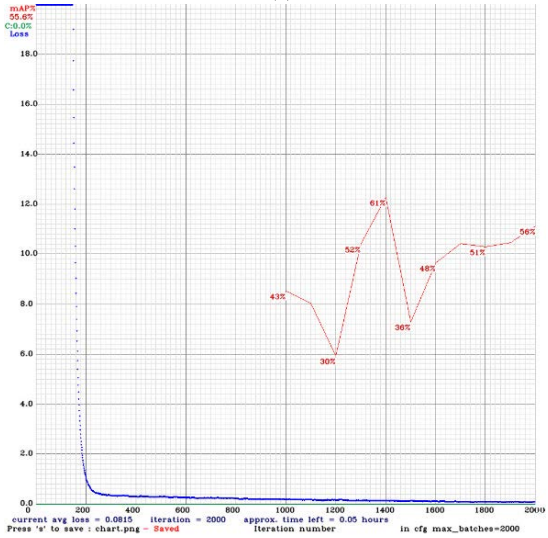
(c)



(d)



(e)



(f)

FIGURE 2. (a) YOLOv4-512, (b)YOLOv4-608, (c) YOLOv4-CSP-512, (d)YOLOv4-CSP-608, (e) YOLOv4-tiny-512, (f) YOLOv4-tiny-608.

TABLE 3. Analysis of the YOLOv4 model by every 1000 iteration.

Iteration	1000	2000	3000
Precision	86 %	100 %	100 %
Recall	80 %	97 %	100 %
F1-score	83 %	98 %	100 %
Average IoU	64.24 %	77.93 %	79.75 %
mAP	87.88 %	99.89 %	100 %

2000 iterations of training by using 608 × 608 and 512 × 512 resizing input network images in every model.

III. RESULT AND DISCUSSION

A. TRAINING PERFORMANCE OF YOLOV4 MODEL

After the YOLOv4 model is trained, the performance of the model is analysed every 1000 iterations. Table 3 shows the performance analysis of the YOLOv4 model, where it is observed that the detection performance is directly proportional to the number of training iterations. The precision reaches 86 % when 1000 iterations are performed and this increases to 100 % when the number of iterations is 2000 and above. The result of the average IoU on detecting the ripe FFBs, which represents the coverage of the bounding box on the target object, also increases in percentage every 1000 iterations. However, the excellent performance of using the model with 3000 iterations is actually due to overfitting. This model detects exactly the particular data it was trained with and it may fail to predict future observations reliably. Therefore, the trained algorithm with 2000 iterations for the YOLOv4 model is suitable for this proposed detection system under 490 images of the dataset.

B. COMPARISON WITH DIFFERENT YOLOV4 MODELS

Fig. 2 shows the learning curve for the YOLOv4 models. The blue and red lines represent the average loss and mean average precision to conduct the improvement trend when training iteration increases. The comparison of the average loss is similar between the same models. Also, 512 × 512 input network sizes in the YOLOv4 and YOLOv4-CSP have maintained consistent high mAPs compare with 608 × 608 input network sizes. It shows that 512 × 512 input network size produces sufficient accuracy at the early stages of training in YOLOv4 and YOLOv4-CSP. Besides, YOLOv4-CSP experienced slow declines in average loss compared to other YOLOv4 models. It may be due to the model scaling method corresponding changes in the scaling size, depth, and width during training which increases the computation complexity. On the other hand, the average loss of YOLOv4-tiny decreases massively at the beginning of the iteration. This is because a higher learning rate was applied compared to other YOLOv4 models.

During the training process, the YOLO-tiny models spend the least time completing the training among other models. YOLOv4 and YOLOv4-CSP need approximately 6 to 7 hours to train the dataset for 2000 iterations. The YOLOv4-tiny models take less than 2 hours to end the entire process

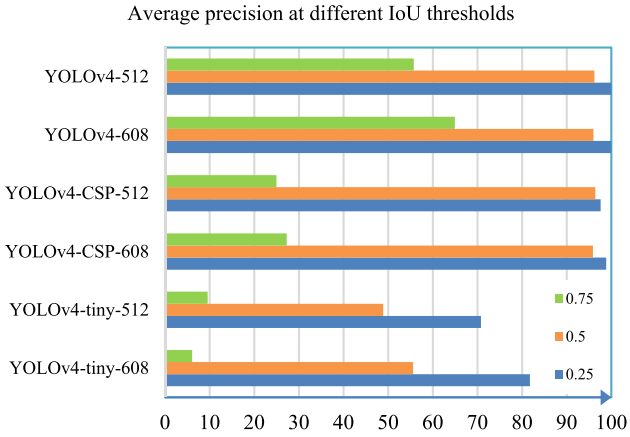


FIGURE 3. The average precision of YOLOv4 models at 25 %, 50 %, and 75 % IoU thresholds on validation set.

which is three times faster than YOLOv4 and YOLOv4-CSP. This is due to the low number of convolutional layers in the model architecture of YOLOv4-tiny, allowing it to be trained with low computational cost. Table 4 shows the size of the weight files in different models. The YOLO-tiny weight file is remarkably small at 23.5 MB. It verifies that weight size is correlated with computational speed. As the original YOLOv4 models have the largest weight files and also the longest training time.

From the data, it can be classified into four types: true positive (TP), false positive (FP), true negative (TN), and false negative (FN), based on the combinations of the true class and predicted class of the learner. The precision, recall, and F1-score are used to examine the reliability of the models. Precision is the number of positive class predictions that are genuinely positive class predictions. Recall is the number of correct positive class predictions made out of all correct positive cases in the dataset. The F1-score provides a single score that addresses both precision and recall concerns in a single number. The equation of precision (P), recall (R), and the F1-score are stated below:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{4}$$

The confusion matrix of the detection performance for several YOLOv4 models is shown in Table 5. The original YOLOv4 models have the greatest performance regarding different perspectives. The precision, recall, and F1-score have a high percentage of 97 %. The average IoU and mAP achieved above 75 % and 96 % in both 512 × 512 and 608 × 608 input network sizes. When analysing the YOLOv4-CSP model, the mAP has similar accuracy to the original YOLOv4 models. However, even though YOLOv4-CSP models have higher mAP than YOLOv4, it is less predictive and sensitive as the result shows the precision, recall, F1-score, and average IoU are slightly lower than YOLOv4. Next, the precision,

FIGURE 4. Evaluation of detection system on FFB indifferent environments: (a) Bright sunlight, (b) Shady lighting, (c) Close up view, (d) Far distance view, (e) Obstructed by frond, (f) Motion blur and (g) Unripe FFB.

recall, and F1-score are much lower in the YOLOv4-tiny models which were expected since these models were optimized for speed and not accuracy. It affects the mAP of the YOLOv4-tiny-512 and YOLOv4-tiny-608 models and they scored 48.89 % and 55.60 % respectively, which are lower than other models.

TABLE 4. Size of weight files in different models.

Model name	Weight size (MB)
YOLOv4-512	256.0
YOLOv4-608	256.0
YOLOv4-CSP-512	210.2
YOLOv4-CSP-608	210.2
YOLOv4-tiny-512	23.5
YOLOv4-tiny-608	23.5

TABLE 5. Analysis of different YOLOv4 models.

Model name	Precision	Recall	F1-score	Average IoU	mAP
YOLOv4 -512	97 %	97 %	97 %	75.85 %	96.00 %
YOLOv4 -608	97 %	97 %	97 %	77.13 %	96.22 %
YOLOv4 -CSP-512	90 %	87 %	88 %	67.48 %	95.89 %
YOLOv4 -CSP-608	84 %	90 %	87 %	63.24 %	96.43 %
YOLOv4 -tiny-512	57 %	57 %	57 %	38.74 %	55.60 %
YOLOv4 -tiny-608	48 %	77 %	59 %	33.76 %	48.89 %

TABLE 6. Evaluation of the YOLOv4 model in on-site testing dataset.

Evaluation	Percentage
Precision	95 %
Recall	82 %
F1-score	88 %
Average IoU	70.19 %
mAP	87.9 %

The effects of different IoU thresholds on the performance of average precision in YOLOv4 models are analysed and shown in FIGURE 3. Through the observation, YOLOv4 and YOLOv4-CSP have outstanding results on 0.25 and 0.5 IoU thresholds. However, YOLOv4-CSP models do not achieve a better result than YOLOv4 when the IoU threshold is increased to 0.75. This may be because the YOLOv4 models include the Panet path-aggregation neck and SPP block over the CSPDarknet53 which increases the receptive field to separate the most symbolic context features. For the YOLOv4-tiny, it is unable to compete with the other models in terms of average precision. Generally, the 608 × 608 input network size has better overall performance compared to the 512 × 512 input network size. This shows that the resizing of the input network images has diminished important features that were present at the beginning of the training.

C. REAL-TIME ON-SITE TESTING OF YOLOV4 MODEL

The detection system was tested in the oil palm estate at Carey Island but in a different area from where training data was collected. The equipment used in the data acquisition was also used in this testing stage. The goal of the testing was to evaluate the performance of the ripe FFB detection system. The model selected for this purpose was the YOLOv4 due to our findings during training that it has the highest average precision compared to YOLOv4-tiny and YOLOv4-CSP. The weight file applied in the model is the 2000 iteration version. Testing was conducted at 9.00 am when the weather was sunny, and the temperature is around 32 °C. In the field, a total of 20 trees were selected for the testing where 10 trees have unripe FFBs only and 10 trees have at least 1 ripe FFB. The targeted trees are similar in age and height to the trees that were used to train the model. Fig. 4 shows the output of the detection system on ripe FFB during on-site testing. In addition, the model operated at the real-time speed of an average of 21 frames per second. Table 6 shows the evaluation of the YOLOv4 model from the on-site testing. During the testing phase, the YOLOv4 model recognises the ripe FFB most of the time, with the mAP reaching 87.9 %. Moreover, the average IoU achieved was 70.19 %. This study confirms that this autonomous detection system can detect the ripe FFB in real-time.

IV. CONCLUSION

The work presented in this paper is the first example of a real-time ripe oil palm FFB detection system that is based on the YOLOv4 model. Although no changes were made to the architecture of the model, it is shown that with the selected methodology and hyperparameters for training, the detection output is very encouraging. Based on the analysis, the trained YOLOv4 model obtained a mAP of 87.9 % in detecting the ripe FFB. The results of the recall and F1-score were 82 % and 88 % respectively as the detection fulfilled the IoU with more than 0.5 after 2000 iterations. During testing in the oil palm estate, the system operated at the real-time speed of roughly 21 FPS and achieved a mAP of 87.9 %. The performance can potentially be improved further with improvements in the training data, model architecture, and hyperparameter optimization. For future work, the system will be expanded to assist in the harvesting process by identifying the palm fronds surrounding the FFB and the FFB stalk that needs to be cut. This information is then relayed to a robotic arm through ROS that will proceed to harvest the FFB autonomously using a fitted cutting mechanism.

ACKNOWLEDGMENT

This research was fully funded by the Universiti Putra Malaysia (UPM) Research Grant (Putra Grant UPM, GP-GPB/2021/9699100).

REFERENCES

[1] C. L. Chew, C. Y. Ng, W. O. Hong, T. Y. Wu, Y.-Y. Lee, L. E. Low, P. S. Kong, and E. S. Chan, "Improving sustainability of palm oil production by increasing oil extraction rate: A review," *Food Bioprocess Technol.*, vol. 14, no. 4, pp. 573–586, Apr. 2021.
[2] R. Abdullah, A. Ismail, and A. K. A. Rahman, "Labour requirements in the Malaysian oil palm industry in 2010," *Oil Palm Ind. Econ. J.*, vol. 11, no. 2, pp. 1–12, 2011.
[3] A. Ibragimov, S. F. Sidique, and Y. S. Tey, "Productivity for sustainable growth in Malaysian oil palm production: A system dynamics modeling approach," *J. Cleaner Prod.*, vol. 213, pp. 1051–1062, Mar. 2019.
[4] K. Minakata, K. Tashiro, H. Wakiwaka, K. Kobayashi, N. Misrom, N. A. Aliteh, and H. Nagata, "Proposal of fruit battery method for estimating oil palm ripeness," in *Proc. 12th Int. Conf. Sens. Technol. (ICST)*, Dec. 2018, pp. 399–402.
[5] M. S. M. Alfatni, A. R. Mohamed Shariff, O. M. Ben Saaed, A. M. Albhbah, and A. Mustapha, "Colour feature extraction techniques for real time system of oil palm fresh fruit bunch maturity grading," in *Proc. IOP Conf. Ser. Earth Environ. Sci.*, 2020, vol. 540, no. 1, Art. no. 012092.

[6] A. Septiarini, A. Sunyoto, H. Hamdani, A. A. Kasim, F. Utaminingrum, and H. R. Hatta, "Machine vision for the maturity classification of oil palm fresh fruit bunches based on color and texture features," *Scientia Horticulturae*, vol. 286, Aug. 2021, Art. no. 110245.
[7] N. Fadilah and J. Mohamad-Saleh, "Color feature extraction of oil palm fresh fruit bunch image for ripeness classification," in *Proc. 13th Int. Conf. Appl. Comput. Appl. Comput. Sci.*, 2014, pp. 51–55.
[8] A. Septiarini, H. Hamdani, H. R. Hatta, and A. A. Kasim, "Image-based processing for ripeness classification of oil palm fruit," in *Proc. 5th Int. Conf. Sci. Inf. Technol. (ICSITech)*, Oct. 2019, pp. 23–26.
[9] D. D. Silalahi, C. E. Reaño, F. P. Lansigan, R. G. Panopio, and N. C. Bantayan, "Using genetic algorithm neural network on near infrared spectral data for ripeness grading of oil palm (Elaeis guineensis Jacq.) fresh fruit," *Inf. Process. Agricult.*, vol. 3, no. 4, pp. 252–261, Dec. 2016.
[10] D. D. Silalahi, C. E. Reaño, F. P. Lansigan, R. G. Panopio, N. C. Bantayan, F. Davrieux, J. P. Caliman, Y. Y. Yuan, and Sudarno, "Near infrared spectroscopy: A rapid and non-destructive technique to assess the ripeness of oil palm (Elaeis guineensis Jacq.) fresh fruit," *J. Near Infr. Spectrosc.*, vol. 24, no. 2, pp. 179–190, Apr. 2016.
[11] A. W. Setiawan and O. E. Prasetya, "Palm oil fresh fruit bunch grading system using multispectral image analysis in HSV," in *Proc. IEEE Int. Conf. Informat., IoT, Enabling Technol. (ICIoT)*, Feb. 2020, pp. 85–88.
[12] A. W. Setiawan, R. Mengko, A. P. H. Putri, D. Danudirdjo, and A. R. Ananda, "Classification of palm oil fresh fruit bunch using multi-band optical sensors," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 4, pp. 2386–2393, 2019.
[13] F. H. Hashim, Z. M. Zulkifli, T. Raj, and A. B. Huddin, "A rapid and non-destructive technique in determining the ripeness of oil palm fresh fruit bunch (FFB)," *Jurnal Kejuruteraan*, vol. 30, no. 1, pp. 93–101, Jun. 2018.
[14] A. D. I. M. Seri, M. S. M. Kassim, S. R. A. Rahman, and A. A. B. Sajak, "Development of virescens fresh fruit bunch ripeness prediction using LiDAR for smart agriculture," in *Proc. IEEE Region Symp. (TENSYMP)*, Aug. 2021, pp. 1–8.
[15] M. H. M. Hazir, A. R. M. Shariff, M. D. Amiruddin, A. R. Ramli, and M. I. Saripan, "Oil palm bunch ripeness classification using fluorescence technique," *J. Food Eng.*, vol. 113, no. 4, pp. 534–540, Dec. 2012.
[16] M. M. Ali, N. Hashim, and A. S. A. Hamid, "Combination of laser-light backscattering imaging and computer vision for rapid determination of oil palm fresh fruit bunches maturity," *Comput. Electron. Agricult.*, vol. 169, Feb. 2020, Art. no. 105235.
[17] K. Y. You, F. H. Wee, Y. S. Lee, Z. Abbas, K. Y. Lee, E. M. Cheng, C. S. Khe, and M. F. Jamlos, "A review of oil palm fruit ripeness monitoring using microwave techniques in Malaysia," in *Proc. IOP Conf. Ser. Mater. Sci. Eng.*, 2020, vol. 767, no. 1, Art. no. 012007.
[18] N. A. Aliteh, N. Misron, I. Aris, R. M. Sidek, K. Tashiro, and H. Wakiwaka, "Triple flat-type inductive-based oil palm fruit maturity sensor," *Sensors*, vol. 18, no. 8, p. 2496, 2018.
[19] N. Misron, N. S. K. Azhar, M. N. Hamidon, I. Aris, K. Tashiro, and H. Nagata, "Effect of charging parameter on fruit battery-based oil palm maturity sensor," *Micromachines*, vol. 11, no. 9, p. 806, Aug. 2020.
[20] N. A. Aliteh, K. Minakata, K. Tashiro, H. Wakiwaka, K. Kobayashi, H. Nagata, and N. Misron, "Fruit battery method for oil palm fruit ripeness sensor and comparison with computer vision method," *Sensors*, vol. 20, no. 3, pp. 1–14, 2020.
[21] N. Sabri, Z. Ibrahim, S. Syahlan, N. Jamil, and N. N. A. Mangshor, "Palm oil fresh fruit bunch ripeness grading identification using color features," *J. Fundam. Appl. Sci.*, vol. 9, no. 4S, p. 563, Jan. 2018.
[22] H. S. Husin, N. Amar, A. A. B. Sajak, and M. S. M. Kassim, "Distribution map of oil palm fresh fruit bunch using LiDAR," in *Proc. 12th Int. Conf. Inf. Commun. Syst. (ICICS)*, 2021, pp. 4–9.
[23] R. Sinambela, T. Mandang, I. D. M. Subrata, and W. Hermawan, "Application of an inductive sensor system for identifying ripeness and forecasting harvest time of oil palm," *Scientia Horticulturae*, vol. 265, Apr. 2020, Art. no. 109231.
[24] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
[25] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," *Comput. Electron. Agricult.*, vol. 157, pp. 417–426, Feb. 2019.
[26] M. O. Lawal, "Tomato detection based on modified YOLOv3 framework," *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, Dec. 2021.
[27] A. I. B. Parico and T. Ahamed, "Real time pear fruit detection and counting using YOLOv4 models and deep SORT," *Sensors*, vol. 21, no. 14, p. 4803, Jul. 2021.
[28] M. H. Junos, A. S. M. Khairuddin, S. Thannirmalai, and M. Dahari, "An optimized YOLO-based object detection model for crop harvesting system," *IET Image Process.*, vol. 15, no. 9, pp. 2112–2125, Jul. 2021.
[29] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[30] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13024–13033.

JIN WERN LAI received the B.Sc. degree (Hons.) in electrical and electronic engineering from the Univerisiti Malaysia Sabah, Sabah, Malaysia, in 2020. He is currently pursuing the master's degree in robotics and automation with the Universiti Putra Malaysa, Serdang, Malaysia. His research interests include robotics, image processing, machine learning, and deep learning.

HAFIZ RASHIDI RAMLI (Member, IEEE) received the B.Eng. degree in electrical and electronic engineering and the M.Sc. degree in control and automation engineering from the Universiti Putra Malaysia (UPM), in 2007 and 2010, respectively, and the Ph.D. degree in biomedical engineering from Imperial College London, in 2017. His current research interests include artificial intelligence, image processing, haptics, and surgical simulation.

LUTHFFI IDZHAR ISMAIL received the Ph.D. (Doctor of Computer Science Engineering) degree from Ghent University, Belgium, in 2020. His Ph.D. research focused on helping children with cognitive impairment in improving their quality of life through human–robot interaction. He is currently a Senior Lecturer with the Department of Electrical and Electronics, Faculty of Engineering, Universiti Putra Malaysia. His current research interests include robotics for medical application, education, and smart agricultural.

WAN ZUHA WAN HASAN (Senior Member, IEEE) received the degree in electrical and electronic engineering from the Universiti Putra Malaysia, in 1997, and the Ph.D. degree in microelectronic engineering from the Universiti Kebangsaan Malaysia, in 2010. He is currently a Senior Lecturer at the Department of Electrical and Electronic Engineering, Universiti Putra Malaysia. His research interests include memory testing, MEMS sensor, and robotic and automation.