System Detection Ripeness of Fresh Fruits Bunch Palm Oil with YOLOv4

Ahmad Sopian
Student in Computer Science
Faculty of Math and Sciences
IPB University
Bogor, Indonesia
amsopian22ahmad@apps.ipb.ac.id

Kudang Boro Seminar
Professor in Mechanical & Biosystem Engineer
Faculty of Agriculture and Technology
IPB University
Bogor, Indonesia
kseminar@apps.ipb.ac.id

Sudradjat
Profesor in Agronomy and Holticulture
Faculty of Agriculture
IPB University
Bogor , Indonesia
sudradjat@apps.ipb.ac.id

Abstract

The harvest of oil palm fresh fruit bunches (FFB), especially in Indonesia, is determined manually, mainly based on the color and number of fallen fruit. Information technology innovation can help determine harvest time quite accurately and consistently. Computer visuals are one method that predicts can detect oil palm fruit maturity. The importance of the ability to detect objects directly, especially in oil palm plantations, is certainly a challenge. YOLOv4 can observe any changes indirect objects properly and quickly and even has a very deep model structure making it very suitable for detecting bunch maturity. FFB has a maturity level which is referred to as a fraction. There are 3 fractions analyzed in YOLOv4 to fraction 1, fraction 2, and fraction 3. Based on the study results, the YOLOv4 method was able to detect the maturity level of bunches of fraction 1, fraction 2, and fraction 3 with an accuracy of mAP@0.50 of 99.17% mAP@0.75 of 97.08% at the checkpoint weight of 6000. In study has disadvantaged unbalanced data set due to the difficult terrain in the field and the lack of resources due to the pandemic. The result model learning can develop to predict ripeness fruits bunch real-time with smartphone or raspberry.

Keywords
Deep Learning, Object Detection, Palm Oil, Ripeness Bunches, YOLOv4

1. Introduction (12 font)

Palm oil is one of the promising crops to boost the Indonesian economy. The area of oil palm plantations in Indonesia in 2021 is 12 million hectares, and the production is 44 million tons (USDA 2021). Indonesia will become the first country in 2021 to export 29 million tons of palm oil to various countries (USDA 2021). Oil palm produces fresh fruit bunches (FFB), then produced into crude oil (CPO). These crude oil products can be found in everyday life, such as butter, cooking oil, soap, etc. (USDA 2021). The harvesting process is one of the main factors for the success of oil palm production. Because oil palm contains Free Fatty Acids (FFA), which are dangerous if consumed, therefore, oil palm must be harvested according to predetermined criteria (USDA 2021). Oil palm has seven stages of maturity which are referred to as fractions, where the fractions is level ripeness from fractions 00, fractions 0, fractions 1 - 5 (Darmosarkoro et al., 2008) and highly recommended harvested in fractions 1-3. The worker harvesters usually look at the colour of the fruit and the number of fruits that fall to the ground while ignoring the ripeness criteria. Although, the harvester has different knowledge, making the harvesting process inconsistent, inaccurate, and unmeasurable. The fault in harvesting bunches can lose the company's operations. Therefore, to increase oil palm productivity, especially in the bunch harvesting process, innovative technology is needed to support the criteria for bunch maturity so that the harvesting process is consistent and accurate.

Computer visual (CV) is a technological approach that can improve harvest quality to be more consistent and accurate because this technique can allow computers to see like humans. This computer vision technology has been widely applied in agriculture to increase crop productivity. Many researchers have developed computer vision systems to help improve the quality of crop yields, such as determining the ripeness of apples (Tian et al., 2019) selecting the type of citrus fruit (Chen et al., 2020) and providing information on the kind of flower (Cheng and Zhang, 2020). Advances in supercomputers and the development of image processing make CV easier to classify and detect objects directly on images or videos (Tripathi and Maktedar, 2020). The critical thing that computer visuals can do is detect and classify objects in photos or videos.

Object detection techniques have been widely used to help facilitate fieldwork, particularly in oil palm plantations. For example, Prasetyo and Santoso (2020) detected and counted a collection of objects harvested using the Faster R-CNN method, with data collected via the internet. Then, Junos and Khairuddin (2021) conducted a study to detect the number of loose fruits that fell to the ground using the YOLOv3 method, and the data was collected using a drone. Both authors have successfully applied object detection methods in their research. YOLO has the latest YOLOv4, which can detect objects faster and more accurately than the previous version (YOLOv3) (Bochkovskiy et al., 2020). The training process is faster even though it only uses one GPU, and the application for detecting fresh fruit bunches has not been found. Then, the YOLOv4 method used is Pascal VOC (Everinghma et al., 2010). This method in Redmon and Farhadi (2018) explained that YOLO could provide the best accuracy evaluation results using Pascal VOC, while the COCO metric has poor accuracy. So, the main objective of this research is to use the YOLOv4 network to detect the ripeness of bunches based on the level of fruit maturity, namely fractions. The quality of the data is also improved, with data collection carried out directly in the field and accompanied by a field foreman to help determine the maturity level of the bunches.

Based on the description above, this research focuses on detecting the maturity of oil palm fresh fruit bunches using the YOLOv4 structure. We used ripe sets are determined by three classes, that is fraction 1, fraction 2, and fraction 3; this is because the ripening level has the best quality of CPO for harvest time, according to Darmosarkoro et al (2008). Oil palm has several varieties; in this study, we used PPKS SP 540. This developed system can assist workers field in determining the level of fruit maturity based on the class of fractions that are ready to harvest so that the decision to quote fruit is consistent and accurate. Furthermore, with the application of the YOLOv4 method, the results of this study are very appropriate if directly used for robotic visual abilities because of the superior ability of YOLOv4 in detecting objects directly and quickly on moving objects.

2. Data Pre-Processing

2.1 Image Acquisition

The collection of fresh fruit bunches image data was carried out directly at the location of the oil palm plantations of PT. Kalimantan Agro Nusantara (Kalianusa), Rantau Pulung, East Kutai, East Kalimantan, Afdeling 8 Block 1 with coordinates (0° 44,520'N, 117° 13.338'E). We choose Afdeling 8 block 1 because this plantation has the main criteria (trees) of oil palm at the height of 5 m. Tree age reaches > 8 years. These criteria make it easier to shoot fruit bunches and decide the maturity of the bunches to be better because observations can be made without tools. The planting age of more than 8 years aims to find out that oil palm trees have produced fresh fruit bunches well. The oil palm variety is PPKS SP 540, in which the maturity levels meet the requirement of this study (Darmosarkoro et al., 2008). The coordinates of Afdeling 8 Block 1 can be seen in Figure 1 and the research stages in Figure 2. Figure 1 the area in red is a location for the acquisition process. In collecting the images, we were accompanied by a field foreman, who helped determine the ripeness fruits level to be photographed. The data acquisition process is carried out on foot by entering the area of oil palm plantations. Image data acquisition was carried out for two days from 20th to 21st of January 2021, at 09.00-12.00, with sunny weather. For mobility reasons, while in the field, we used the smartphone camera of Samsung A7 with a resolution of 4240x4240 (18 MP). The number of images collected was 600 pieces. The sample bunches used were 200 photographed directly on the tree, then we shot each bunch three times from different angles (left, right, and front). Distance from the camera to the fruit bunches is approximately 30cm – 60cm by adjusting the position of the fruit on the tree.

For choosing the three criteria for this fraction because we were guided by the PPKS 2008 book (Darmosarkoro et al., 2008), which explained that the bunches were ready to be harvested. The collection of images is carried out with a foreman with 10 years of experience in palm oil. Before taking pictures of the bunches, we first discussed with the foreman to determine the maturity of the bunches in the field. The results fractions 1, 2, and 3 are presented in Table 1 based on the knowledge of the foreman. From Table 1, we can see that each bunch fraction has specific criteria and a different colour. During collecting pictures, the foreman helped provide directions and explanations regarding the fruit fractions that we would photograph. Suppose the foreman says the bunch falls within the criteria for fraction 1, fraction 2, or fraction 3. In that case, we will immediately take pictures of the bunch. An example of the image acquisition results can be seen in Figure 3.

Figure 1. Location of Image Acquisition Ripeness Bunch Palm Oil

The acquisition process went smoothly but still had some obstacles in the field. The first obstacle and the main problem is direct tracing in oil palm plantations, which is tiring. Hence, the process of taking pictures becomes less than optimal and hard to get dataset balance class. Second, access and terrain in oil palm plantations are very extreme. The road access to location Afdeling 8 Block 1 is difficult with motorbike cause water and muddy. In addition, the terrain in oil palm plantations is quite challenging to pass on foot because the weeds are high. The dry midrib covering the hole is quite deep, so getting a sample of bunches with a proportional category is very difficult. We have limited resources, so we maximize the available data for this research.

Table 1. The Criteria Ripeness of FFB Based on Fractions

Class	Criteria
Fraction 1	The color is reddish, and there is one fruit bunch fall.
Fraction 2	The color is shiny red, and there are 2 – 5 fruits a bunch fall.
Fraction 3	The color is orange, and there are 6 – 10 fruits a bunch fall.

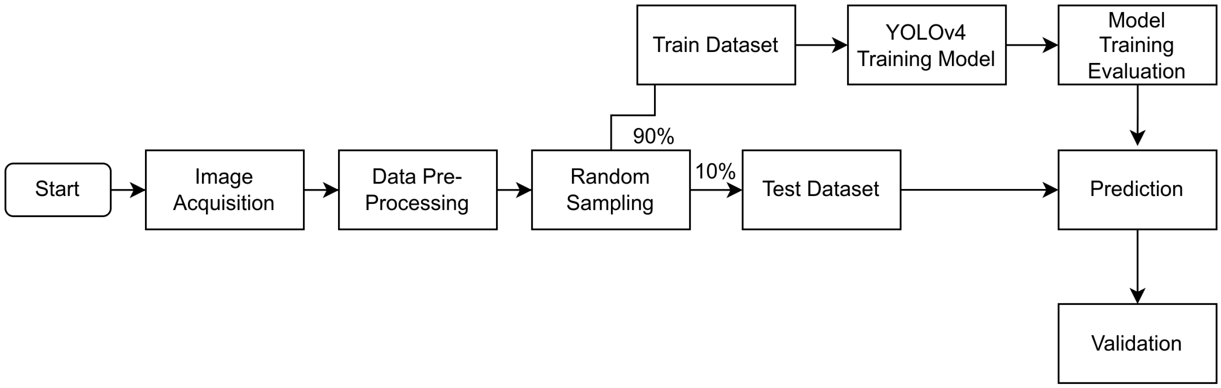


Figure 2. Research Stage

2.1 Image Annotations and Data Production

This study uses image data of fresh fruit bunches with ripe criteria in fractions 1, fraction 2, and fraction 3. A total of 600 images were collected. The photos from the camera are then transferred to a computer. We have explained previously that the acquisition process in this study has not been maximized due to several conditions and obstacles in the field. The quality of the images obtained is not all in good condition. Here, we found that some of the bunch objects were distorted around the bunch objects, such as the midrib and fibre covering almost the entire bunch, and the photos were blurry. These disturbances affect the detection model training process. Therefore, we removed the image that has annoyance poor quality. An example of an image that causes interference is presented in Figure 4. After the selection process, we finally obtained 309 cluster images with good quality and without interference. Table 2 show information on the amount before and after manual image quality checks.

From the 309 images we used for this research, we lowered the resolution quality from 4240 x 4240 to 600 x 600. The purpose of reducing the image resolution was to lighten the load on the computer during the training process. The images were divided into two separated samples with a ratio of 90:10, where 90% of images were used as training data and 10% as test data, as shown in Table 3.

(a) (b) (c)

Figure 3. The example Image Capture Bunch Plam Oil. (a) fraction 1, (b) fraction 2, (c) fraction 3

The last step, we annotated the images. Image annotation aims to create a class of bunch maturity based on the notes that have been made during the acquisition process. At this stage, we used tool annotation for make a class labelling that is LabelImg application (Tzutalin, 2015)). The Labelling application itself supports class formats for YOLO methods. Image annotation is performed on test data and training data, respectively. Figure 5 is an example using the software LabelImg. it is essential to determine the location of the target object and class. Fraction class is a maturity fruits level consisting of fraction_1, fraction_2, and fraction_3.

Table 2. The total acquisition images & screening images

Class	Acquisition Result	Manual Screening Result
Fraction 1	297	155
Fraction 2	165	82
Fraction 3	138	72
Total	600	309

Table 3. The result random sampling images

Class	Training Images	Testing Images
Fraction 1	139	16
Fraction 2	74	8
Fraction 3	65	7
Total	278	31

Figure 4. The bad quality shoot bunch in object class target

Figure 5. The appearance of LabelImg application

3. You Only Look Once version 4 (YOLOv4)

YOLOv4 [13] is the latest version of the YOLO family (Redmon et al., 2016) (Redmon and Farhadi, 2017) (Redmon and Farhadi, 2018). YOLO recognizes objects using a regression approach to get each class's bounding box coordinates and probability directly. YOLOv4 has been improved so that the algorithm can work with high precision and precision in real-time and with minimal computational effort. The YOLOv4 model has a complex structure, as shown in Figure 6.

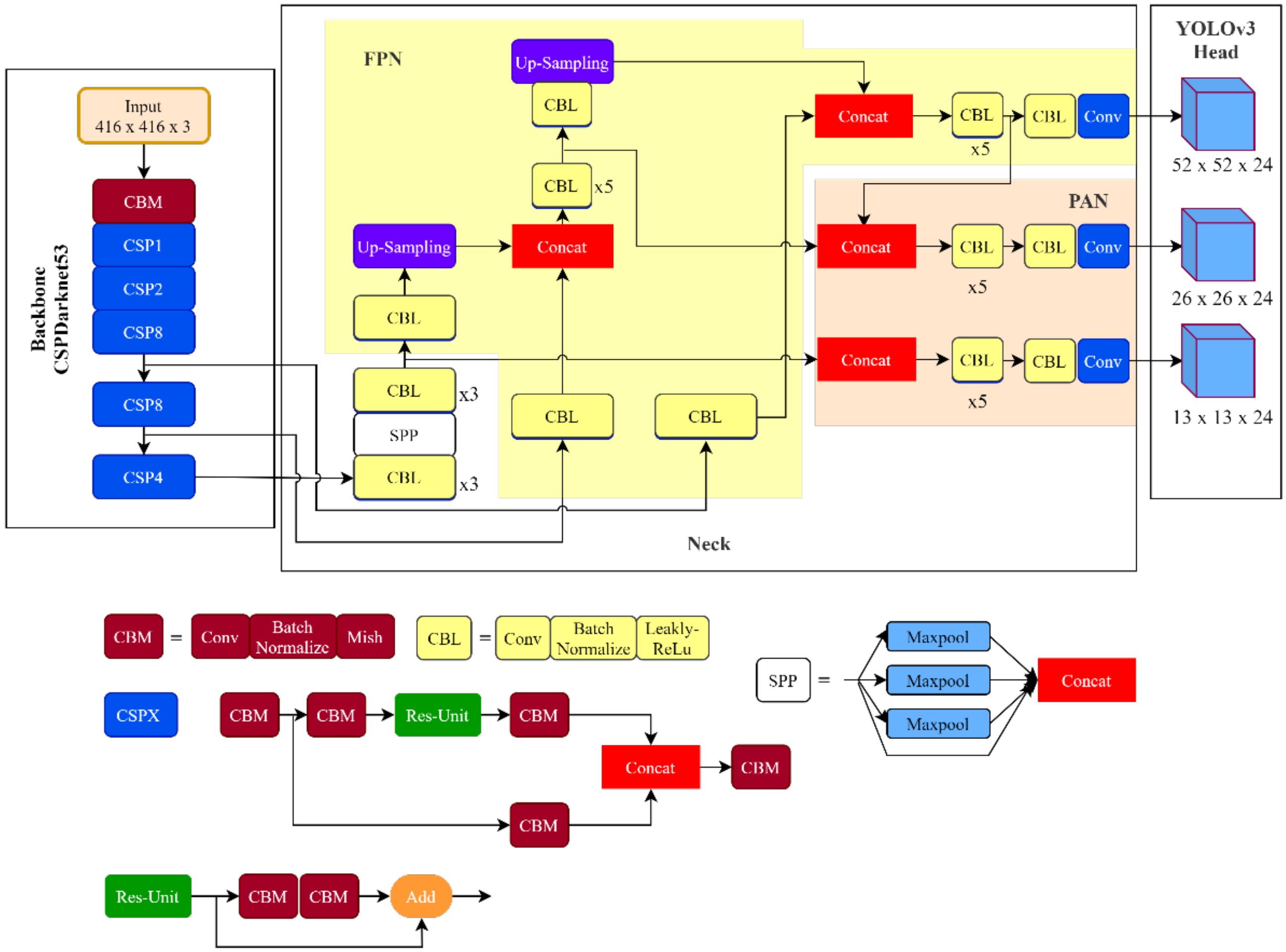


Figure 6. YOLOv4 Structure (Bockovskiy et al., 2020)

YOLOv4 uses a brand-new backbone referred to as CSPDarknet53 (Redmon et al., 2016); this is often because of the mix of the Darknet53 formula (Redmon and Farhadi, 2018) and Partial Cross-Stage (CSP) (Wang et al., 2020) . Darknet53 is inspired by ResNet, where the remaining modules are saved within themselves. Then, CSP can think about the model's learning ability. The feature layer is the input within the remaining modules, and the higher-level feature info is the output. The feature suggests that training objectives from the model in the ResNet module are the distinction between output and input. Therefore, realizing residual learning while reducing model parameters and strengthening learning features. The neck consists of SPPNet and PANet. In SPPNet, first, the feature layer is convolved 3 times. The input layer is maximally collected using the maximum collected cores of various sizes. The results are collected first and then combined three times, increasing network reception. PANet combines the layers when the operation of Backbone and SPPNet create a sample, making the initial feature double its height and width and then combining the convolution and sampling feature. The layers with the feature layers obtained by CSPDarknet53 into a feature fusion form then perform down-sampling, press the peak and width, and at last stacking with the previous layer to appreciate a lot of feature fusion (five times). The prediction module will build predictions by features extracted from the network. The network prediction results can alter the position of the three previous frames. Therefore, the last one will be filtered with a non-maximum-suppression (NMS) formula to induce the ultimate frame prediction.

YOLOv4 proposes a new augmentation with mosaic data augmentation technique to expand the information set and introduce CIOU as a loss function, making the network optimize within the direction of accelerating the overlapping area, effectively increasing accuracy.

3.1 Object Localization and Prediction Process

YOLOv4 gets the bounding box's coordinates using the YOLOv3 algorithm and is often stated as a prediction module. The prediction module can be used when extracted into 3 feature layers. The valuable feature is 13 x 13 x 24, which is equivalent to dividing the input image into 13 x 13 grids. Every grid is accountable for detecting objects within the space corresponding to this grid. Once the item centre has lied within this area, it is necessary to use this grid to detect the object. Images can plan every grid of the previous three boxes (anchor). The predicted network can also modify the position of model parameters of the last 3 squares to urge the ultimate prediction result. Similarly, the 26 x 26 x 24 and 52 x 52 x 24 feature layer prediction method is the same as the 13 x 13 x 24 feature layer process.

In Figure 6, the feature layer is split into a 13 x 13 grid to explain the item location process and the class prediction. The input images with 3-channel colours have a size of 416 x 416 x 3 that has passed the backbone and neck process. Feature extraction gets the input image through the network. It represents the effective feature layer with 13 x 13 x 24 within the prediction module (head detector). The feature layer is divided into 13 x 13 grids, and every grid has three anchor boxes. The network predicts 4 coordinates for every bounding box t_x, t_y, t_h, t_w. If the cell is offset from the highest left corner of the image by (c_x, c_y) and the previous bounding box containing a dimension width and height of p_w, p_h, then the prediction corresponds to:

$$b_x = \sigma(t_x) + c_x \tag{1}$$
$$b_y = \sigma(t_y) + c_y \tag{2}$$
$$b_w = p_w e^{(t_w)} \tag{3}$$
$$b_h = p_h e^{(t_h)} \tag{4}$$

In the training process, the network continuously learns for the four bounding box parameters, i.e. t_x, t_y, t_w, t_h such that systematically regulating the previous box's position to approach the expected work and eventually obtaining the ultimate prediction result t_x, t_y, respectively, denote that a sigmoid operation delimitst_x, t_y to make sure that the prediction center is within the grid. The network prediction results can adjust the position of the 3 previous boxes. Then the final box predictions will be filtered by ranking the arrogance level and NMS to induce detection results.
Confidence score describes the accuracy of the model predicting that an object is of a specific class. The following equation is the confidence score equation:

$$confidence = p_r(Class_i \vee Object) \times p_r(Object) \times IoU_{pred}^{truth} \tag{5}$$

In equation (5), $p_r(Class_i \vee Object)$describes the probability of type object when it is known as an object. $p_r(Object)$represents the probability of whether the prediction box contains an object. If there is an object then $p_r(Object) = 1$, otherwise is 0. IoU_{pred}^{truth}tells the value of the overlap ratio between predicted square and the actual square.

4. Experiment Result and Analysis

The YOLOv4 detection method used in this study is the Darknet framework (Bochkovskiy et al., 2020) . Experiments to train and test the model were carried out on a google collab virtual machine with a Tesla K80 GPU. The application of parameters of YOLOv4 is presented in Table 4.

To improve the accuracy of model detection and to adapt the input required for the Darknet framework, the input image changes into 416 x 416 pixels. Network for training model set the batch size to 64, and a total of 9000 iterations was used to analyze the training process. Parameters such as momentum, initial learning speed, weight decay regulation, and other parameters refer to the original parameters in the YOLOv4 model. This study applies transfer learning techniques to improve model training performance. The pre-training model used for transfer learning is yolov4.conv.137.

Table 4. The parameters used for model training process

Hyperparameter	**Initialization**
Input images	416 x 416 x 3
Batch	64
Momentum	0.9490
Learning Rate	0.0010
Decay	0.0005
Training Steps	9000

4.1 Model Testing

After the model training is completed, the checkpoint weights obtained are used to examine the model. The model is evaluated from several aspects. To assess the categorical data from fresh fruit bunches, we can classify the measurement into 3 categories. TP (True Positive) category describes the class results obtained according to the target class. FP (False Positive) implies that the number of samples in the detected object category does not match the target object category. FN (False Positive) indicates that the accurate piece is the opposite or in the undiscovered category. The quantity is (TP + FP) for all positive cases, i.e., the proportion of true cases (TP) is called the degree of precision, which represents the proportion of samples of confirmed cases in positive cases among the samples detected by the model, as shown in Equation (6).

$$Precision = \frac{TP}{(TP+FP)} \tag{6}$$

The total positive samples within one set are equal to (TP + FN). Therefore, the recall rate is employed to measure the model's ability to notice real cases in the test set, as shown in Equation (7).

$$Recall = \frac{TP}{(TP+FN)} \tag{7}$$

To characterize the precision of the model from YOLOv4, we use the indicators AP (Average precision) and mAP (mean average precision) to evaluate the accuracy of the model, as shown in Equations (8) and (9).

$$P = \int_0^1 P\,(R)dR \tag{8}$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{N} \tag{9}$$

Where P, R, N each represent the level of precision, recall, and the number of objects in all categories respectively. F1 represents the harmonic average of the precision level and the recall rate, as shown in Equation (10):

$$F_1 = \frac{2 \times P \times R}{(P+R)} \tag{10}$$

If F1 is higher, the test model will be more effective. We used 31 images with a total of 31 objects as the test set. In this research, this evaluation model is carried out for all weighted checkpoints. Where there are nine checkpoint weights generated during the model training process, checkpoint weights are automatically stored when steps reach multiples of 1000. Each weight can highlight TP, FP, FN, precision, recall, F1, AP, and mAP values. Table 5 shows the evaluation results for the importance of TP, FP, FN, precision, recall, and F1 with IoU = 0.5. Table 5 that the maximum TP and FP or minimum FN values are at checkpoint weights of 4000. So, checkpoint weights of 4000 explain that the detection model in step 4000 has more effective object detection capabilities than other weights. In the same case, precision, recall, and F1 get the most optimal values. However, the evaluation results have not provided a reasonable conclusion for the level of model accuracy in detecting objects. Therefore, we use the AP and mAP indicators to see the status of model accuracy in detecting the maturity task of oil palm bunches.

Table 5. The comparison result

Weight Checkpoint	Size	Object	TP	FP	FN	*Precision*	*Recall*	F_1
1000	416	31	27	17	4	0.613	0.871	0.720
2000	416	31	28	4	3	0.880	0.900	0.890
3000	416	31	26	5	5	0.840	0.840	0.840
4000	**416**	**31**	**28**	**3**	**3**	**0.900**	**0.900**	**0.900**
5000	416	31	27	4	4	0.870	0.870	0.870
6000	416	31	26	5	5	0.840	0.840	0.840
7000	416	31	28	4	3	0.880	0.900	0.890
8000	416	31	28	4	3	0.880	0.900	0.890
9000	416	31	26	6	5	0.810	0.840	0.830

Next, we evaluate the accuracy predict detection by looking at average precision value for each class of bunch. Result of this evaluation with threshold IoU = 0.50 and 0.75. Table 6 shown the results of the assessment of the detection model against the test data set. AP@0,50 with checkpoint weights of 4000 and 6000 have the same and better AP value performance than other weighted values. Then, the comparison of accuracy results at AP@0.75 at checkpoint weight of 6000, faction_1 has a value of 2%, faction_2 has 19%, and faction_3 has 29% better than the accuracy at checkpoint weight of 4000. Checkpoint weights from 5000 to 9000 have 100% accuracy at AP@0.50 and AP@0.75 for the faction_3 category. Faction_3 has the most dominant color characteristics compared to other fractions.

Table 6. The comparison result average precision for each class

Weight Checkpoint	Size	IoU	fraction_1	fraction_2	fraction_3
1000	416	AP@.50 AP@0.75	0.97 0.51	0.79 0.42	0.81 0.14
2000	416	AP@.50 AP@0.75	0.99 0.93	0.81 0.69	0.98 0.75
3000	416	AP@.50 AP@0.75	0.98 0.88	0.78 0.78	0.98 0.98
4000	416	AP@.50 AP@0.75	**1.00** 0.91	**0.97** 0.78	**1.00** 0.71
5000	416	AP@.50 AP@0.75	0.99 0.89	0.82 0.82	**1.00** **1.00**
6000	416	AP@.50 AP@0.75	**1.00** **0.93**	**0.97** **0.97**	**1.00** **1.00**
7000	416	AP@.50 AP@0.75	0.98 0.92	0.88 0.88	**1.00** **1.00**
8000	416	AP@.50 AP@0.75	0.97 0.90	0.78 0.78	**1.00** **1.00**
9000	416	AP@.50 AP@0.75	0.96 0.88	0.78 0.78	**1.00** **1.00**

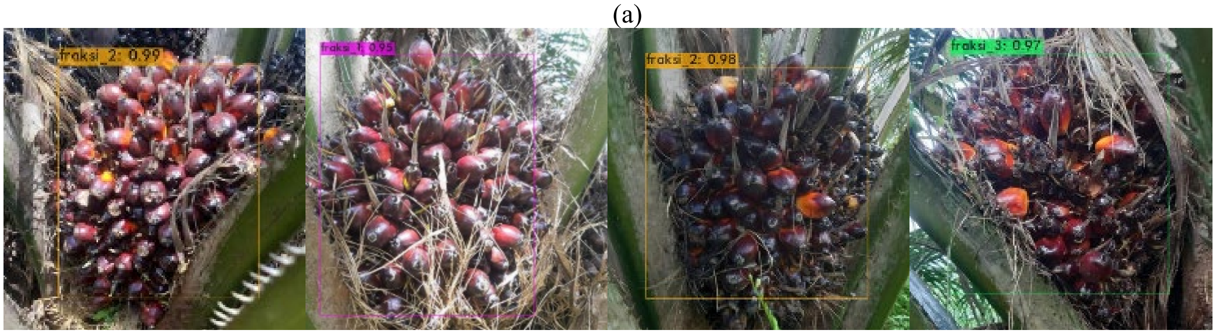
Measure object detection capability for all categories of mature bunches its important to. We tested them at IOU = 0.5 and IOU = 0.75 thresholds further to evaluate the detection capability of the comprehensive model, as shown in Table 7. From the comparison results in Table 7, we can see that the accuracy value of mAP@0.50 on checkpoint weights 4000 and 6000 has the same accuracy value. In contrast, the accuracy of mAP@0.75 is 17% higher than checkpoint weight 4000. The difference in accuracy in the IOU threshold aims to determine which model can detect cluster objects more accurately (confidence score). Based on Table 6 and Table 7, the model's most significant confidence value is the 6000 weights. Because this model can detect objects with an accuracy rate above 75% (IOU = 0.75), we use a checkpoint weight of 6000 as the model used in the test data.

Table 7. The comparison result mean average precision

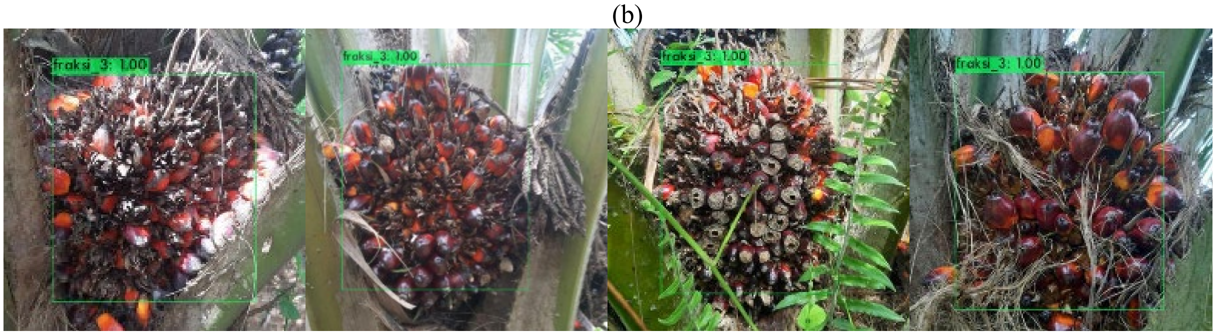
Weight checkpoint	**mAP@0.50**	**mAP@0.75**
1000	0.85	0.36
2000	0.92	0.79
3000	0.91	0.88
4000	**0.99**	0.80
5000	0.94	0.90
6000	**0.99**	**0.97**
7000	0.95	0.93
8000	0.91	0.89
9000	0.91	0.88

After knowing which weight can detect bunches of maturity objects based on the AP and mAP values, we conducted an experiment to display the object detection results on the test data. Figure 7 shows the results of object detection. We use training data and see the detection results. The detection model we use is a checkpoint weight of 6000 because it has superior and effective performance. The image that we show is an example of the detection results obtained from the test data. Each class of mature fractions is taken four images as an example of the detection results.

(a)



(b)



(c)

Figure 7. The result of detection. (a) target class fraction 1, (b) target class fraction 2, (c) target class fraction 3

5. Conclusion and Future Direction

In this paper, the YOLOv4 is used to assist in detecting the maturity of bunches based on levels in fraction 1, fraction 2, and fraction 3 because it has the most optimal level of fruit maturity for harvesting. The YOLOv4 method produces checkpoint weights as a model that can be used for cluster object detection. Judging from the AP and mAP accuracy results, the checkpoint weights at 6000 get better accuracy than other checkpoints. For AP@0.50 mature category, fraction_1 and fraction_3 have 100% accuracy, and fraction_2 has 97% accuracy with a mAP@0.50 value of 99%. Meanwhile, at AP@0.75 mature category, fraction_1 has 93% accuracy, fraction_2 has 97% accuracy, and fraction_3 has 100% accuracy with a mAP@0.75 value of 97%. These results explain that YOLOv4 can detect class objects' ripeness level of bunches accurately and adequately.

However, in this study, many shortcomings still need to be corrected:

1. Get proportional image data by proper digital cameras (DSLR) to improve the quality of images. Therefore, enriching features in the model training process because the results of DSLR cameras are sharper than smartphone cameras.
2. The need for consideration in data acquisition for bunches that have disturbances, such as midribs and fibres that cover bunches. It is essential because if this model is developed in a visual robotic fruit picker, real cases such as this are designed in a visual, mechanical fruit picker. In the future, it must satisfactorily resolve this disorder.
3. Improvements need to be made by trying several other detection methods such as Faster R-CNN, YOLOv1-v3, SSD, RetinaNet, and so on to compare the performance of the detection model.

The issue of the imbalanced dataset is not easy to avoid. Like a case in this research, where collected image very difficult of the terrain and limited resources make the data unbalanced. Therefore, it is necessary to make improvements in the image acquisition process. Alternatively, can use image augmentation techniques to make the image balanced.

References (12 font)

[USDA]United State Departemen of Agriculture, “Palm Oil Area Harvested by Country in 1000 HA,” 2021. https://www.indexmundi.com/agriculture/?commodity=palm-oil&graph=production (accessed Jun. 10, 2021).

[USDA]United State Departemen of Agriculture, “Palm Oil Exports by Country in 1000 MT,” 2021. https://www.indexmundi.com/agriculture/?commodity=palm-oil&graph=exports (accessed Jun. 10, 2021).

[USDA]United State Departemen of Agriculture, “Palm Oil Production by Country in 1000 MT,” 2021. https://www.indexmundi.com/agriculture/?commodity=palm-oil&graph=production (accessed Jun. 10, 2021).

Bochkovskiy, A., Wang, C.,Y., and Liao, H., Y., M., “Yolov4: Optimal speed and accuracy of object detection,” arXiv Prepr. arXiv2004.10934, 2020.

Chen, W., Lu, S., Liu, B., Li, G., and Qian, T., “Detecting Citrus in Orchard Environment by Using Improved YOLOv4,” Sci. Program., vol. 2020, p. 8859237, 2020, doi: 10.1155/2020/8859237.

Cheng, Z., and Zhang, F., “Flower End-to-End Detection Based on YOLOv4 Using a Mobile Device,” Wirel. Commun. Mob. Comput., vol. 2020, p. 8870649, 2020, doi: 10.1155/2020/8870649.

Darmosarkoro, W., Akiyat, and Edy, S., Pembibitan Kelapa Sawit, Medan: PPKS, 2008

Everingham, M., Van Gool, L., Williams, C., K., I., Winn, J., and Zisserman, A., “The pascal visual object classes (voc) challenge,” Int. J. Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010.

Junos, M., H., Mohd Khairuddin, A., S., Thannirmalai, S., and Dahari, M., “Automatic detection of oil palm fruits from UAV images using an improved YOLO model,” Vis. Comput., 2021, doi: 10.1007/s00371-021-02116-3.

Prasetyo, N., A., and Santoso, A., J., “Automatic Detection and Calculation of Palm Oil Fresh Fruit Bunches using Faster R-CNN,” Int. J. Appl. Sci. Eng., vol. 17, no. 2, pp. 121–134, 2020.

Redmon, J., and Farhadi, A., “Yolov3: An incremental improvement,” arXiv Prepr. arXiv1804.02767, 2018.

Redmon, J., and Farhadi, A., “YOLO9000: better, faster, stronger,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263–7271, 2017

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016

Richey, B., and Shirvaikar, M., V., "Deep learning based real-time detection of Northern Corn Leaf Blight crop disease using YoloV4," in Real-Time Image Processing and Deep Learning 2021, vol. 11736, p. 1173606. 2021

Setyamidjaja, D., Budidaya kelapa sawit, Kanisius, 1991.

Suo, R. et al., "Improved multi-classes kiwifruit detection in orchard to avoid collisions during robotic picking," Comput. Electron. Agric., vol. 182, p. 106052, 2021, doi: https://doi.org/10.1016/j.compag.2021.106052.

Tian, G., Yang, Z., Wang, H., Wang, E., Li, and Liang Z., "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," Comput. Electron. Agric., vol. 157, pp. 417–426, 2019.

Tripathi, M., K., and Maktedar, D., D., "A role of computer vision in fruits and vegetables among various horticulture products of agriculture fields: A survey," Inf. Process. Agric., vol. 7, no. 2, pp. 183–203, 2020.

Tzutalin, "LabelImg," 2015. https://github.com/tzutalin/labelImg. (accessed Mar. 21, 2021)

Wang, D., Wang, Y., Li, M.,. Yang, X., Wu, J., and Li, W., "Using an Improved YOLOv4 Deep Learning Network for Accurate Detection of Whitefly and Thrips on Sticky Trap Images," Trans. ASABE, vol. 64, no. 3, pp. 919–927, 2021.

Wang,C.,-Y., Liao, H.,-Y., M., WuY.,-H., Chen, P.-Y., Hsieh, J.,-W., and Yeh, I.-H., "CSPNet: A new backbone that can enhance learning capability of CNN," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 390–391. 2020,

Biographies

Ahmad Sopian. is a master's student in Computer Science at IPB University, Bogor. Previously, he was an undergraduate student at Mulawarman University, Samarinda. It is based on this educational background that he has an interest in Data Science and Computer Visuals.

Kudang B. Seminar is a professor in the Department of Mechanical & Biosystem Engineering & the Computer Science Department, Bogor Agricultural University (IPB). He received his master and doctoral degrees in the field Computer Science at the University of Brunswick, Canada. He was the president of and is currently is an honorary board member of Asia-Pacific Federation for Information Technology in Agriculture (APFITA) http://www.apfita.org/ and that of Indonesian Society of Agriculture Informatics (HIPI). He is also a member of International Society of Precision Agriculture (ISPA) https://www.ispag.org/. His research interests include Digital & Smart Agriculture, Precision Agriculture, Traceability Systems, Knowledge Management System for Agriculture, Distributed Artificial Intelligent System (Multi-agent Systems). He is an author of national and international articles, a reviewer of International Agricultural Engineering Journal (IAEJ) and Journal of Biosystems Engineering. Currently, he is the Dean of Agricultural Engineering and Technology Faculty at IPB University. He is also the coordinator research team of Precipalm (Precision Agriculture Platform for Oil Palm), the satellite-based system for oil palm nutrition estimator and recommender.

Sudradjat is a professor in the Department of Agronomy and Horticulture, Bogor Agricultural University (IPB). He received his master in the field Agronomy at IPB and doctoral degrees in the field Agronomy (ENSA) at the University of Rennes, France. He research interest at Satellite / drone based precision agriculture for plantation (palm oil) and food. He has a invention fisrt Co-invenor PreciPalm: Program for determining satellite fertilizer-based fertilization recommendations (main inventor of Prof. Dr. Kudang B. Seminar), second determine the recommendations for location-specific Palm of fertilization, and third build jonggol palm oil research and education plantation. The appreciation satyalancana 20 and 30 years from the president of the republic of indonesia and award from the 2019 legal entity academic senate council state council. The history in organization like deputy chairperson of the indonesian civil society (maksi) in 2015-2020, secretary of the academic senate (sa) of ipb 2014-2019, deputy secretary of ptnbh academic senate assembly 2016-2017, and secretary of ptnbh academic senate assembly 2017-2018.