Recognition and counting of oil palm tree with deep learning using satellite image

I Nurhabib¹, K B Seminar¹ and Sudradjat²

¹Mechanical and Biosystems Engineering Department, IPB University, Bogor, Indonesia
²Agronomy and Horticulture Department, IPB University, Bogor, Indonesia

Corresponding author email: iqbalnurhabib@gmail.com

Abstract One of the challenges in oil palm plantation is to accurately count the number of palm trees in a unit of area. Knowing the number of trees will contribute to better managerial and operational tasks for fertirization. This requires a method that can be used to calculate oil palm accurately in a quick fashion. This study aims to discuss the method of identifying and counting oil palm trees using a deep learning system based on the YOLO algorithm to process the images captured by a satellite from Google Earth. The system prototype has been successfully built and tested to recognize oil palm trees and to count them in a case study area IPB-Cargill Oil Palm Education Garden, commonly called IPB-Cargill Oil Palm Teaching Farm, is in Singasari, Jonggol, Bogor West Java. Based on the experimental results, the used training dataset using YOLO with 2500 step iterations with a loss value of 0.6 obtained an accuracy values of 85.6%, 98.9% for precision, and 86.6% for recall. This shows that the model is adequate to correctly recognize and to count palm tree objects.

Keywords: deep learning, google earth, oil palm, population counting, precision agriculture, yolo algorithm.

1. Introduction
Agriculture is a sector that has a crucial role in economic activity in Indonesia [1]. This can be seen from the Gross Domestic Product (GDP) contribution, which is quite large, around 12.72 percent in 2019 [2]. The plantation is one part of the agricultural sector that has great potential. The contribution of plantations in 2019 was 3.27 percent of total GDP and 25.71 percent of the agricultural industry [2].

Oil palm has great potential, so technology is needed to analyze productivity. Oil palm productivity analysis looks at the factors that influence growth to increase productivity [3]. Analyzing the factors that affect the productivity of oil palm cannot be done easily considering the many factors that influence it. Technological innovation and the application of technological innovations such as precision agriculture is the solution. Precision agriculture reduces unwanted impacts on the environment [5]. This makes it a tremendous opportunity if precision agriculture can be applied in Indonesia, a developing country that controls the business chain, to improve the community's economy [5]. Remote sensing is a part of precision agriculture. Remote sensing can be used to see the status of oil palm development. The advantages of remote sensing include providing fast and real-time analysis results. The use of remote sensing includes land cover classification, change detection, age estimation, above-ground biomass estimation, carbon estimation, and pest and disease detection, and production yield estimation [6]. Estimated production yields can be used in calculating predictions on oil palm production, which are carried out manually by calculating directly on oil palm plantations. Measurement of plantation area can
also be done as analysis to calculate the amount and production estimates. This makes it quite difficult for farmers to identify trees quickly with a vast number of trees.

Entering the industrial revolution 4.0 is possible to develop technology to identify in automatic tree predictions. A method that can be developed using artificial intelligence, in which a computer can automatically recognize visible objects. It makes it easier for farmers to predict the number of trees automatically. Making this model is done so that the computer can learn and recognize the object to be identified. Learning on the computer can use a deep learning model. This research uses image processing with a deep learning method which can analyze the number of trees better. The research conducted this time uses the YOLO (You Only Look Once) deep learning algorithm. This algorithm has the advantage of faster object identification compared to other algorithms. So in this study, the author will test whether this algorithm can also be used to identify trees. The results of the deep learning model obtained can then be applied to calculate oil palm trees.

2. Methodologies
The material used in this study was the material used in this research is oil palm image data satellite the location. The software used is Anaconda Navigator, Spyder, Google Earth, Label.img, Microsoft Excel, Microsoft Word, and QGIS. For the algorithm using the Python programming language version 3.7.8. This research uses TensorFlow, NumPy, and OpenCV modules.

2.1. Research procedure
The research was carried out in several stages, including data collection, data preprocessing, CNN development, modelling results, and evaluation seen in figure 1.

![Research procedure](image)

**Figure 1.** Research procedure.

2.2. Images dataset collection
Dataset collection is carried out for the data training process. The dataset collected is from images from satellites taken on Google Earth on 02/08/2020 as data processing. The more datasets collected, the better the object recognition process to be detected will be. Because the model made will recognize more images to be identified [7].
Table 1. Satellite image information.

| Indeks | Citra satellite 1 | Citra satellite 2 | Citra satellite 3 |
|--------|-------------------|-------------------|-------------------|
| Source | Google Earth      | Google Earth      | Google Earth      |
| Latitude | -6°47'2717"       | -6°28'18.85"      | -6°472717"        |
| Longitude | 107°01'81.72"     | 107°01'31.30"     | 107.018172"       |
| Image capture area | 39,880 m²         | 29,047 m²         | 21,079 m²         |
| Oil palm area | 20,413 m²         | 17,084 m²         | 15,85 m²          |
| Spectral | RGB              | RGB              | RGB              |
| Pixel size | 4800 x 2916 pixel | 4800 x 2916 pixel | 4800 x 2916 pixel |
| Number of theoretical trees | 227               | 232               | 205               |
| The actual number of trees | 257               | 221               | 188               |
| Planting distance | 9.2 m            | 9.2 m            | 9.2 m            |
| Dead tree | 12               | 2                | 3                |
| Cut due to access to electricity | 1                 | 1                | 0                |
| Image resolution | 0.6 m x 0.6 m    | 0.6 m x 0.6 m    | 0.6 m x 0.6 m    |

The table shows that the area of images 1, 2, and 3 have vast differences. These differences result in the number of trees in the population unit is different and can be searched with the population number search formula.

\[
Total\ Population = \frac{ha}{a \times b} \tag{1}
\]

ha is the area captured from satellite imagery and is the actual area, a is the spacing between oil palm trees, and b is the distance between rows [4]. The following is an example of satellite image capture obtained according to table 1.

2.3. Preprocess data
Image data obtained from IPB-Cargill Oil Palm Education Garden, commonly called IPB-Cargill Oil Palm Teaching Farm, is in Singasari, Jonggol, Bogor, West Java with coordinates 06° 28.319’ South, 107° 01.103’ east and is at 116 m above sea level. Image data that has been obtained cut and resized the pixels to be smaller than the original. CNN, which is applied to classify images, receives input of various
sizes so that the characteristics of the original image can still be seen and maintain the spatial relationship between pixels. After collecting the dataset is to do the labelling. Labelling is done to determine the object to be recognized in an image. The image obtained has a variety of objects so that the oil palm object can be recognized in an image by CNN and correctly identified as oil palm using the label.img application. The object labelling is done by marking the oil palm tree and naming the resulting image object. The output of this object labelling process is a *.xml file. This *.xml file is then used in the training data process. Labelling can be seen in Figure 3.

![Image](image_url)

**Figure 3.** Data label data set.

3. Result and discussions

3.1. Model development

The model development is carried out to produce outputs following the research objectives, namely recognition and counting oil palm trees with the desired level of accuracy using classification. At this stage, the classification is done using YOLO. Classification aims to classify objects of oil palm. In making modelling, each modelling is done using random images with the amount of data divided into two parts, namely training data and datasets with a ratio of 80% for test data and 20% for training data. This study uses the YOLO (You Only Look Once) algorithm in the modelling process. The YOLO algorithm has the advantage of identifying multiple objects compared to other algorithms [7]. YOLO can read objects up to 45 FPS if used in real-time [8]. If the speed of the model used to identify is faster and more accurate, the model will be better [9] so that the YOLO algorithm can be used. The YOLO algorithm can identify many objects directly and simultaneously without identifying objects one by one [10] in identifying oil palm trees with high density. In this study, many labels are needed in one image so that the use of the YOLO algorithm is perfect for predicting many objects because the predictions do not use SoftMax. Data training is the process of training the algorithm model created to recognize the dataset images that have been collected. This data training process uses a *.xml file obtained from the labelling results on the dataset. The greater the number of datasets used in the training process, the better the training results obtained with a smaller loss value. The method of training data consists of two stages, namely forward propagation and backward propagation.

3.2. Data preprocessing

The YOLO model architecture has 106 convolutional layers, which can be seen in Fig. The architecture shows a pooling layer and a connection layer. The convolutional layer functions to find the position and value of the previous layer by convolution in performing dot product operations on the data entered. The pooling layer concludes data by reducing data and then taking the most significant value in some data. A connected layer is a layer contained in the last layer. The corresponding layer changes the matrix obtained from the previous layer into a one-dimensional matrix [11].

The output generated when data enters the convolutional layer is an activation map. At this stage, the incoming data undergoes a convolution process to convert the filter to all data. The activation map is the output of the filter process performed on the previous layer.

In the pooling layer, a filter will shift the activation map area [12]. The connected layer has neurons that are connected to all neurons in the next layer. Before connecting to the connected layer, the activation map generated in the previous stage needs to be converted into one-dimensional data. Then the data is used as input at the connected layer stage. This training process is carried out until a small training loss value is obtained close to zero. This indicates that the model has recognized the object to
be identified. The last layer in the convolutional layer is used to predict the probability and bounding box. The expected Bounding Box corresponds to the size of the detected object, so the Bounding Box prediction value ranges from zero to one. The activation function used in the last layer is linear, while it uses the ReLU activation function [13].

3.3. Training results test

After the training is complete, the model can be run to identify oil palms. Detection of objects with the YOLO algorithm applies an image dividing network into an \( S \times S \) grid. The grid has the task of detecting objects. In each grid, a Bounding box will be predicted along with the confidence value. This confidence value shows how accurately the Bounding box can predict the object.

The Bounding box value can be determined by getting five anchor boxes, namely \( x, y, w, h \), and \( c \). The \( x \) and \( y \) values are the coordinates of the midpoint of the Bounding box. The value of \( w \) and \( h \) is the ratio of the size to the relative width and height of the grid, and the value of \( c \) is the confidence value in the Bounding box. The grid can predict probability class values if there are expected objects in it. When using YOLO, the grid will be multiplied by the class probability value with the confidence value from the Bounding box to produce class confidence values specifically for each Bounding box. This value shows the probability class generated in the Bounding box and how accurately the confidence value predicts the object. To predict the Bounding box, YOLO will predict the coordinates of the midpoint of the Bounding box relative to the grid location.

![Figure 4. Bounding box prediction.](image)

The bounding box is obtained from the values of the anchor box that is formed. The value of \( bx \) is obtained from the value of sigma \( tx \), where \( t \) is the ordinate and \( x \) is a prediction about the x-axis and is added to the length value of \( cx \) (cell \( x \)). The by value is obtained from the sigma \( ty \) value where \( t \) is the ordinate and \( y \) is a prediction about the y-axis and is added to the length value of \( cy \) (cell \( y \)). The \( bw \) and \( bh \) values are obtained from the width and height values of the Bounding box. For object prediction, it is obtained from the sigma value \( t \) like figure 4.

Oil palm measurements are needed to determine the catch area per bounding box. Oil palm has a midrib length of 7.5 to 9 m [4]. The image has a resolution of 0.6 m x 0.6 m in the test image, and it is found that each detected tree image has an area of 225 m2 to 324 m2 so that there are 625 to 900 pixels in one bounding box, as shown in Figure 5.

![Figure 5. Reading objects to form a Bounding box.](image)

The number of datasets used for training data is 3100 oil palm objects (Table 2). In one picture, there can be one or more oil palm trees. The better the image used for training data, the better the results obtained [14].
Table 2. Test data and the number of labels in the study.

| Object       | Amount | Label |
|--------------|--------|-------|
| Palm oil     | 1200   | 3100  |
| Nonpalm oil  | 721    | 3100  |

In the dataset, there are images and non-oil palms with their function as supporting data. The output of the data output after running with the model is shown below.

Figure 6. The output of the calculation of oil palm image 1.

Figure 7. Output results from the calculation of oil palm image 2.
3.4. Value of training loss

The training process carried out at this research stage was carried out to obtain the weight or weight value used for identifying oil palm trees. A training loss graph is also generated during the training process, which shows the process of improving the error values obtained during the training dataset process. The training loss graph is obtained at each step of the training dataset. The value of training loss is obtained from the following equation [13].

\[
\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^b \left[ (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] \\
+ \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^b \left( \sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left( \sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \\
+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^b \left( C_i - \hat{C}_i \right)^2 + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^b \left( C_i - \hat{C}_i \right)^2 \\
+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \sum_{c \in \text{classes}} \left( p_i(c) - \hat{p}_i(c) \right)^2
\] (2)

The training process is carried out using 1200 images with labels of 3100 and the number of steps being 2500 steps. The number of steps is the number of iterations of the two processes in the training data, namely the forward propagation and backpropagation processes, to evaluate the model to recognize the object to be identified. The training process involves 2500 steps to obtain a weight value that can automatically identify oil palm trees. The loss value obtained in the step is 0.6. A loss value close to zero indicates that the model made has recognized the object to be identified.

Figure 8. Output results from the calculation of oil palm image 3.
The weight or weight values obtained are then used to identify oil palm trees in the image. The weight test on the picture is carried out by testing the model to determine the tree in the embodiment and knowing the accuracy value of the model made.

3.5. Validation results and model accuracy
Based on the results of the model test, the average value of the Confusion matrix is as follows

\[
\begin{array}{ccccccc}
\text{Image 1} & \text{Image 2} & \text{Image 3} \\
\hline
\text{Actual} & \text{Palm oil} & \text{Non-palm oil} & \text{Palm oil} & \text{Non-palm oil} & \text{Palm oil} & \text{Non-palm oil} \\
\text{Palm oil} & 224 & 33 & 187 & 34 & 162 & 26 \\
\text{Non-palm oil} & 5 & 0 & 1 & 0 & 1 & 0 \\
\text{Total} & 229 & 33 & 188 & 34 & 163 & 26 \\
\end{array}
\]

In Table 3, there is a confusion matrix generated from the test data. The table predicts the number of oil palm trees detected is 224, 187, and 162 objects of oil palm trees.

\[\text{Figure 9. Output tensorboard training data.}\]

\[\text{Figure 10. Results of image output analysis 1.}\]
Figure 1

Figure 2

Table 4. Accuracy, precision, and recall values.

| Image   | Accuracy | Precision | Recall |
|---------|----------|-----------|--------|
| Image 1 | 85.4 %   | 98 %      | 87.1 % |
| Image 2 | 84.2 %   | 99.4 %    | 84.6 % |
| Image 3 | 87.7 %   | 99.4 %    | 88.1 % |
| Average | 85.6 %   | 98.9 %    | 86.6 % |

Based on Table 4, in image 1, as many as 229 objects were detected, and as many as 33 objects were not detected. In image 2, there were 188 objects and 34 objects of oil palm that were not detected. In image 3 there were detected, and 162 objects and 26 oil palms were not detected.

Evaluation of the confusion matrix calculated the average value of accuracy, precision, and recall using the equation according to the method section. The calculation results obtained an average accuracy value of 85.6%, precision of 98.9%, and recall of 86.6%.

Based on these results, the accuracy, precision, and recall values of the model are obtained. The average accuracy value of the model got value of 85.6%. Accuracy describes how accurately the model can classify objects correctly. In machine learning, especially in classification, the model has high accuracy if it can precisely predict the output of several inputs [15]. The average value of precision obtained is 98.9%. A high precision value indicates that the machine learning model is good [15]. Precision shows the level of consistency of a model in classifying. The average recall value obtained was 86.6%. Like precision, the recall value indicates the consistency of the model in classifying. The difference is in the way it is calculated. Precision compares the total number of correct predictions with the total number of predictions aimed at. However, recall compares the number of correct predictions with the actual number of targeted classes [15]. The recall value is smaller due to many undetected objects. This is because the images are difficult to distinguish.
Based on these results, it can be said that the model is relatively stable in classifying objects correctly. This is indicated by the number of objects detected correctly than detected incorrectly. The model error in classifying objects is thought to be caused by an unclear image, so the model cannot distinguish.

4. Conclusion
Recognition and counting of oil palm tree with deep learning using satellite Image have been successfully built and tested. The value of the training loss graph is 0.6, which is close to zero. Indicates that the model adequately able to recognize oil palm tree objects. The confusion Matrix test results show that the average accuracy is 85.6%, the average precision is 98.9%, and the recall is 86.6%. Based on these results, it can be concluded that the model is also able to detect non oil palm tree objects. It is also suggested for future improvement is to improve the quality of satellite images using image correction methods and the choice of alternative satellites to obtain better image.

References
[1] Rusastra I W, Suryadi M 2004 Ekonomi tenaga kerja pertanian dan implikasinya dalam peningkatan produksi dan kesejahteraan buruh tani J. Litbang Pertanian 3 23
[2] BPS 2020 Indonesian Oil Palm Statistics 2019 (Jakarta:Badan Pusat Statistik) p 13
[3] Yohansyah W M, Lubis I 2014 Analisis Produktivitas Kelapa Sawit (Elaeis guineensis Jacq.) di PT. Perdana Inti Sawit Perkasa I, Riau. Bul Agrohorti 1 125-131
[4] Pahan I. 2008. Panduan Lengkap Kelapa Savit (Jakarta: Penebar Swadaya) p 22
[5] Whelan B, J T 2014 Precision Agriculture for Grain Production Systems, F. Crop. Res. 9 133-155
[6] Amirruddin AD, Muharam FM, Ismail MH, Ismail MF, Tan NP, Karam DS 2020 Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest. Comput Electron Agric. 1 169
[7] Harahap M, Elfrida J, Agusman P, Rafael M, Abram R, Andrianto . 2019 Sistem Cerdas Pemantauan Arus Lalu Lintas Dengan YOLO (You Only Look Once v3) Semin. Nas. APTIKOM 367–376
[8] Bochkovskiy A, Wang C-Y, Liao H-YM 2020 YOLOv4: Optimal Speed and Accuracy of Object Detection arXiv
[9] Huang R, Pedoeem J, Chen C 2019 YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers Proc 2018 IEEE International Conference on Big Data, Big Data 2018 Institute of Electrical and Electronics Engineers p 2503–2510
[10] Wang C, Wang Y, Han Y, Song L, Quan Z 2017 CNN-based object detection solutions for embedded heterogeneous multicore SoCs Proc. Asia South Pacific Des. Autom. Conf. ASP-p 105–110
[11] Lorentius, Christopher Albert Adipranata R, Tjondrowiguno A 2020 Pengenalan Aksara Jawa dengan Menggunakan Metode Convolutional Neural Network. e-Proc Eng. 7(1) p 2558–2567
[12] Santos A, Ariyanto G. 2018 Implementasi deep learning berbasis keras untuk pengenalan wajah Emit. J. Tek. Elektro. 18(01) p 15–21
[13] Redmon J, Divvala S, Girshick R, Farhadi A 2019 You only look once: unified, real-time object detection. arXiv
[14] Widystuti W, Darmawan JBB. 2018. Pengaruh jumlah data set terhadap akurasi pengenalan dalam deep convolutional network. Konf. Nas. Sist. Inf. p 8–9
[15] Natan O, Gunawan AI, Dewantara 2019 Grid SVM: Aplikasi Machine Learning dalam Pengolahan Data Akuakultur. J. Rekayasa Elektr. 15(1)