A Comparison of Deep Learning Methodology for Predicting Palm Oil Trees

Raudlah Hawin AYANI, Astrid CALISTA, and Mulyanto, Indonesia

Key words: Palm Oil, Tree Counting, Faster R-CNN, YOLOv3, Single Shoot Detector, Retina-Net, Mask R-CNN

SUMMARY

Indonesia is the largest country of palm oil production in the world. According to the Central Statistics Agency 2020, Indonesia's CPO (Crude Palm Oil) production reached 44.8 million tons. Many Indonesian CPO is supported by the distribution of CPO in various provinces in Indonesia, such as Riau, North Sumatera, Central Kalimantan, East Kalimantan, and West Kalimantan. From these backgrounds, the appropriate method for palm oil tree counting is needed to find high accuracy and applicability for business actors to estimate the yield figures and predict market needs. In optimizing palm oil tree counting vastly and efficiently, it is necessary to develop artificial intelligence for spatial analysis support. Artificial Intelligence or GeoAI is the technology to support palm oil tree counting by learning from data training in machines and automatically detecting the object. Furthermore, it will compare the object detection deep learning efficiency to view the suitable method. Several object detection methods were compared, such as faster R-CNN, YOLOv3 (You Only Look Once), Single Shoot Detector (SSD), and Retina-Net. Using a DJI Phantom 4 multispectral drone with a spatial resolution of 5 cm, the researcher collected the data from drone acquisition generated by PT. Aria Agri Indonesia. The research location focused on the Central Kalimantan area with 15 hectares, and the data was processed by ArcGIS Pro software. Based on model testing results, it is successfully tried the faster R-CNN method detected 3666 palm oil trees with a confidence threshold of 0,5 is equal to a 50% confidence threshold, YOLOv3 detected 7739 palm oil trees with a confidence threshold of 0,5 is equal to a 50% confidence threshold, Single Shoot Detector detected 1650 palm oil trees with a confidence threshold of 0,73 is equal to a 73% confidence threshold, and Retina-Net detected 1848 palm oil trees with a confidence threshold of 0,74 is equal to a 74% confidence threshold. The manual digitization used for the data validation and the total number of palm oil trees is 1667. A single Shoot Detector is the closest result to manual digitization for the deep learning object detection method.

A Comparison of Deep Learning Methodology for Predicting Palm Oil Trees

Raudlah Hawin AYANI[1,2], Astrid CALISTA[1,2], and Mulyanto[2], Indonesia

1 Ikatan Surveyor Indonesia
2 Aria Agri Indonesia

1. INTRODUCTION

Indonesia is the largest oil palm plantation, reaching 15.08 million hectares (ha). The palm oil plantation area increased from the previous year is 1.5%. Most palm oil plantations are owned by large private plantations 55.8%, people's plantations 40.34%, and large state plantations 3.84% (Junaedi, 2022). Based on the Central Statistics Agency, 2020, Indonesia's CPO (Crude Palm Oil) production reached 44.8 million tons. To monitor a large area of palm oil plantations, it is necessary to use fast and precise technology to calculate the plantation inventory. One technology implementation for plantation inventory is the tree-counting process. Tree counting automatically can help the plantation owner to reduce the consuming tree counting manually. Performing the tree counting manually will consume more time, energy, and cost than combined technology. Using remote sensing and artificial intelligence can accelerate counting palm oil trees faster than manually.

Remote sensing and machine learning are an excellent technology for forest inventory management and play an essential role in terms of economic sustainability (Yao et al., 2021). These technologies can help the plantation officer to inventory. One remote sensing technology is UAV (Unmanned Aerial Vehicle), which can apply to counting the palm oil tree. The remote sensing system includes data acquisition, transmission, processing, and storage (Yin, 2019). The spatial analysis imagery data can be generated from the UAV data acquisition. The drafter can digitize or annotate the object as a palm oil tree. This model is used for deep artificial intelligence modelling. The artificial intelligence part of deep learning for building an inventory of individual palm tree oil automatically counts and geo-location (Ammar & Kouba, 2020). Deep learning uses palm oil tree training as the modelling data to generate the machine training data automatically. This study will compare several deep learning methods, such as R-CNN, YOLOv3, SSD, and Retina-Net.

2. DEEP LEARNING MODELING

The effectiveness of deep learning uses pattern classification techniques in machine learning methods. The optical image must train from a deep learning model using instance segmentation, labelling, location prediction, and pixel-based semantic segmentation mask. Segmentation is a field of computer vision that combines object detection and semantic segmentation (Kristal et al., 2020). After performing the segmentation stage, next is executing

the object detection methods often used for deep learning, including Faster-CNN, YoloV3, Single Shoot Detect, and Retina-Net.

R-CNN, or Regions with CNN Features, is an object detection model that uses high-capacity CNNs to bottom-up region proposals in order to localize and segment objects. It uses selective search to identify some bounding-box object region candidates ("regions of interest") and then extracts features from each region for classification. Faster R-CNN designed one backbone convolutional neural network (the region proposal network (RPN)) for both missions to reduce the processing cost during the inference. Therefore, the RPN network is trained independently for every task using the multi-task loss defined below (Ammar A, 2021):

$$L(P_i, t_i) = \frac{1}{NCls} \sum_i Lcls(P_i, t_i^*) + \lambda \frac{i}{Nreg} \sum_i p_i^* Lreg(t_i, t_i^*) \quad 1$$

Where:

- pi is the probability that the ith anchor inside a mini-batch corresponds to an object; the network generates this probability.
- $p * i$ is a binary value that equals one if the anchor is positive and 0 otherwise. The anchor is positive if it has the highest IoU overlap with one ground-truth box or if the IoU overlaps with the ground-truth box is superior to 0.7. The anchor is negative if, for all the ground-truth boxes, the IoU overlap is inferior to 0.3.
- ti corresponds to the coordinates of the bounding box predicted by the network.
- $t * i$ corresponds to the ground-truth bounding box's coordinates for which the anchor is positive. • Lcls corresponds to the classification loss.
- Lreg corresponds to the regression loss.
- Ncls and Nreg are the normalization factors.
- λ corresponds to the weight used to balance the two losses.

The architecture of YOLO uses a Convolutional Neural Network (CNN) to divide the images into grids, predict coordinates, class, and probabilities, and apply a single neural network (Wibowo et al., 2022). The YOLOv3 CNN network will learn to generate two pieces of information for every grid cell: the class probability vector and the list of bounding box coordinates. Every bounding box is associated with a confidence score. All grids' information is assembled to generate the final list of detected objects inside the input image. For every grid cell, YOLOv3 associates only one anchor. Then, it estimates for every anchor a list of five values:

$$([t_x, t_y, t_w, t_h], \text{Conf}) \quad 2$$

where [tx, ty, tw, th] are four parameters used to generate the bounding box coordinates of the object, and Con f is the confidence score of the bounding box. Hence, YOLOv3 generates from the input image one 3D matrix with the following dimension:

$$N \times N\ (3 * (5+C)) \quad 3$$

where (N × N) is the number of grid cells: (13 ∗ 13) in the case of the image size being (416 ∗ 416). C is the number of classes the system is trained in (two in our case). (3 ∗ (5 + C) corresponds to the five parameters detected in Equation but at three different scales. This is to ensure the robustness of the model against scale variance.

SSD is a single-stage object detection method that discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. At prediction time, the network generates scores for the presence of each object category in each default box and adjusts the box to match the object shape better. Additionally, the network combines predictions from multiple feature maps with different resolutions to handle objects of various sizes naturally (Liu et al., 2015). The multiBox objective is where the SSD training objective gets its inspiration, but it handles more object classifications. Let x p ij = 1, 0 serve as a match indicator for category p's i-th default box and jth ground truth box. Under the matching as mentioned above scheme, can have P I x p ij 1. Localization loss (loc) and confidence loss (conf) are weighted sums that make up the overall objective loss function (Liu, W., Rabinovich, A., Berg, 2016):

$$\mathrm{L(x,c,l,g)} = \frac{1}{N}\sum_i \mathrm{Lconf(x,c)} + \alpha Lloc(x,l,g)) \qquad 4$$

where N represents how many default boxes were found to match. Wet set the loss to 0 if N = 0. A Smooth L1 loss between the ground truth box (g) parameters and the anticipated box (l) parameters represents the localization loss (Girshick, R, 2015).

Retina-Net is a one-stage object detection model that utilizes a focal loss function to address class imbalance during training. Focal loss applies a modulating term to the cross-entropy loss to focus learning on complex negative examples. Retina-Net is a single, unified network comprising a *backbone* network and two task-specific *subnetworks*. The backbone is responsible for computing a convolutional feature map over an entire input image and is an off-the-self convolutional network. (Lin et al., 2017). Retina-Net architecture is used throughout the object-detecting procedure. As shown in figure 1, the Retina-Net design consists of three parts: a feature extraction backbone network, two classifications, and box regression subnetworks. (Lin, Goyal, Girshick, He, & Doll'ar, 2018)

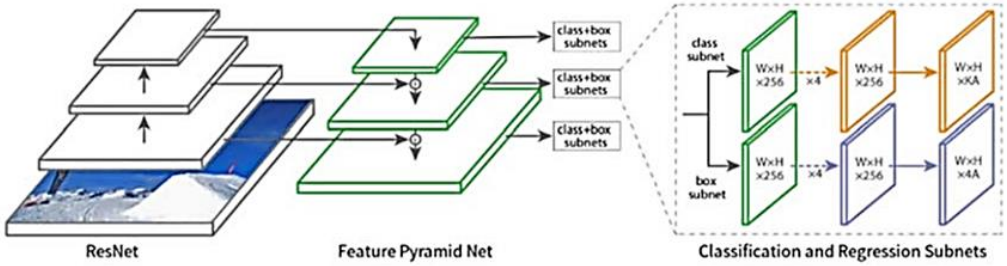

Figure 1. Retina-Net architecture using ResNet Backbone (Lin, et al., 2018)

3. DATA AND METHODOLOGY

In this study, the researcher collected the data drone DJI Phantom 4 Multispectral and was collected from PT. Aria Agri Indonesia, to be processed into the raster image and deep learning analysis.

In addition, this research has a research location which is located in Central Kalimantan. Figure 2 describes the study location of this research

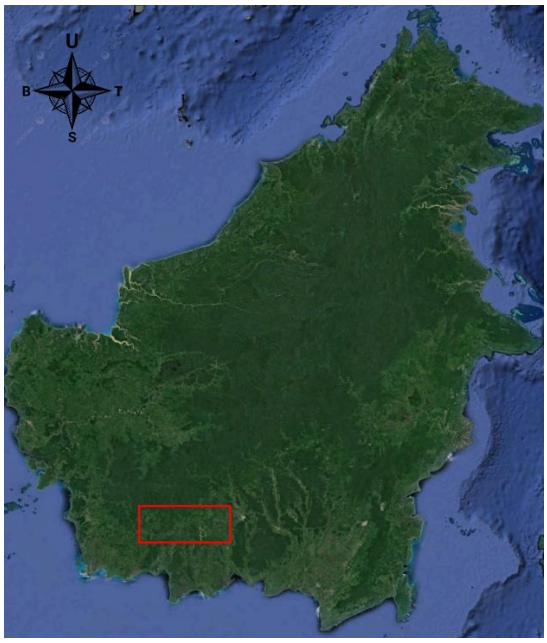

Figure 2. Location of the research

The raw image, with a spatial resolution of 5 cm, has image property information, such as the dimension is 1600 x 1300, width is 1600 pixels, height is 1300 pixels, and horizontal and vertical resolution is 96 dpi. The camera property of the drone camera includes using the camera maker DJI with FC6363 model, F-stop is f/2.2, exposure time is 1/670 sec, the focal length is 6 mm, max aperture is 2.27, and metering mode is center-weighted average. The GPS properties of the image location are latitude and longitude 2° 1’ and 113° 2’, with an altitude of 147.648 m.

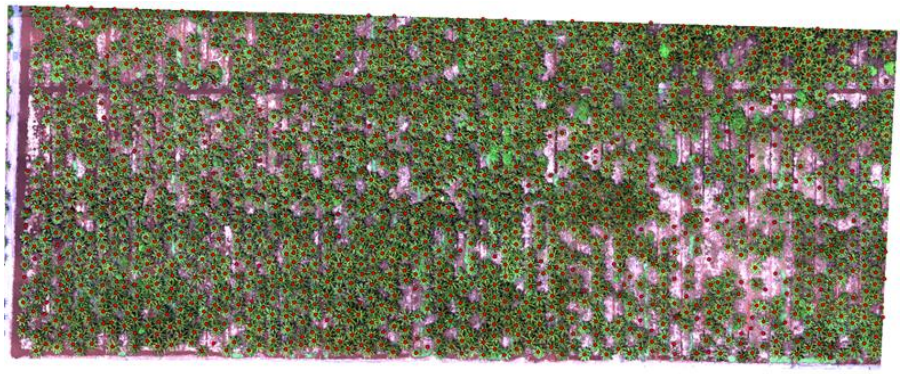

Figure 3. The raster image for tree counting modeling

The raw data was built into orthophoto using Agisoft metashape software to create from raw data. The process builds an orthophoto from the alignment, builds dense clouds to combine the point-point from the alignment photo, and interpolates the point-point into a 3D object. Furthermore, build mesh was processed to interpolate and reconstruct from the tie point or dense could to form a covering area. Build texture and ortho mosaic in the following process to generate the raster image. The raster image (Figure 3) needs spatial analysis for tree counting palm oil. The researcher calculated the palm oil tree based on the raster image and counted 1667 palm oil trees. This data is used to validate deep learning data to check the confidence level of tree counting manually and automatically. This research adopts the deep learning method approach to derive palm oil tree counting modelling. The general method is presented in Figure 4

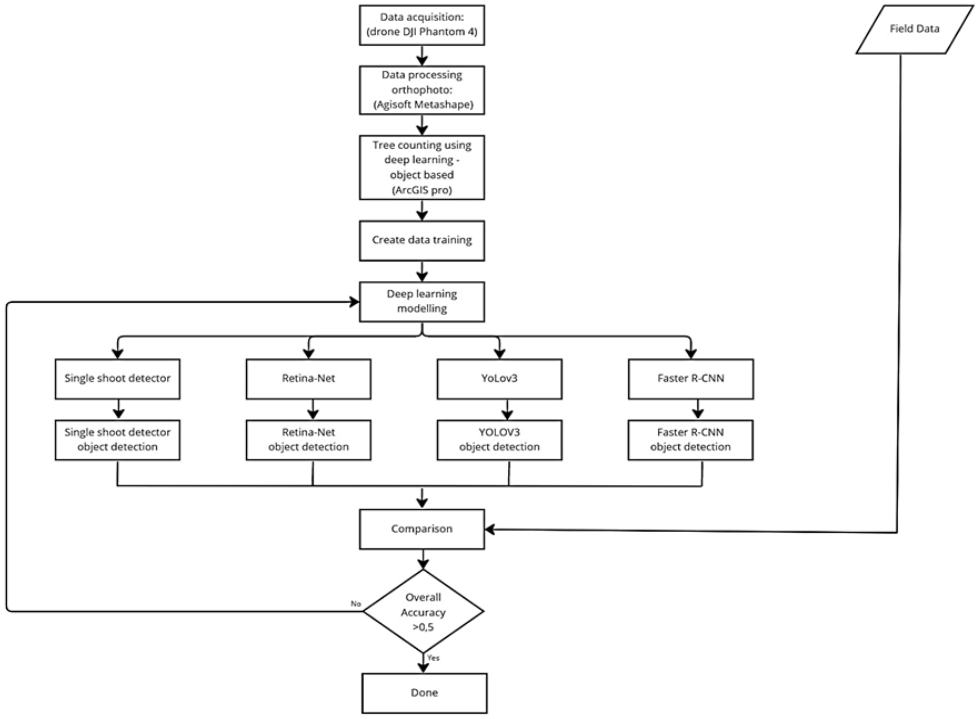

Figure 4. The flowchart of deep learning modelling

Figure 4 shows several stages to obtain a model representing the actual field reality state. Therefore, the author describes in detail the stages with the following points:

a. **Data acquisition**
   This study was taken using the Drone DJI 4 Phantom, and the output is an aerial photo of the area study
b. **Data processing orthophoto**
   This stage of processing drone raw data into ortho mosaic images that overlap with the other images
c. **Tree counting**
   This tree calculation stage uses deep learning technology with several methods, such as Faster-CNN, YoloV3, Single Shot Detector, and Retina-Net. Several settings in this stage were made for padding, threshold, batch size, and nms overlap, excluded in each method to improve data accuracy
d. **Create data training**
   Data training is an important parameter to generate the process of tree counting using deep learning. If the correct data training is available, the systems can increase the accuracy of feature extraction, pattern recognition, and complex problem-solving.
e. **Deep learning model**
   This step is to run a training data deep learning model using a raster as an input to produce the feature class that contains the objects found.
f. **Comparison**
   This stage is for analysing deep learning and field data results. The analysis includes the results of tree counting from several methods and the closeness of values between deep learning results with field data.
g. **Quality Control of Tree Counting using Deep Learning:**
   This stage is for controlling the quality of tree counting using deep learning and setting the data of validity threshold or accuracy. If the results exceed or are equal to 0.5, it has a confidence level of 50%. This method model is considered close to the actual field reality model.

4. RESULTS AND DISCUSSION

Based on the drone DJI Phantom 4 multispectral data was processed to orthophoto images. Those images as the material for tree counting by deep learning models. Deep learning has an architecture for each method to process a system for training, and the training can continuously improve the models. The researchers tested several deep learning methods, including Faster-CNN, YoloV3, Single Shoot Detect, and Retina-Net. The results of this study are shown in figure 5.

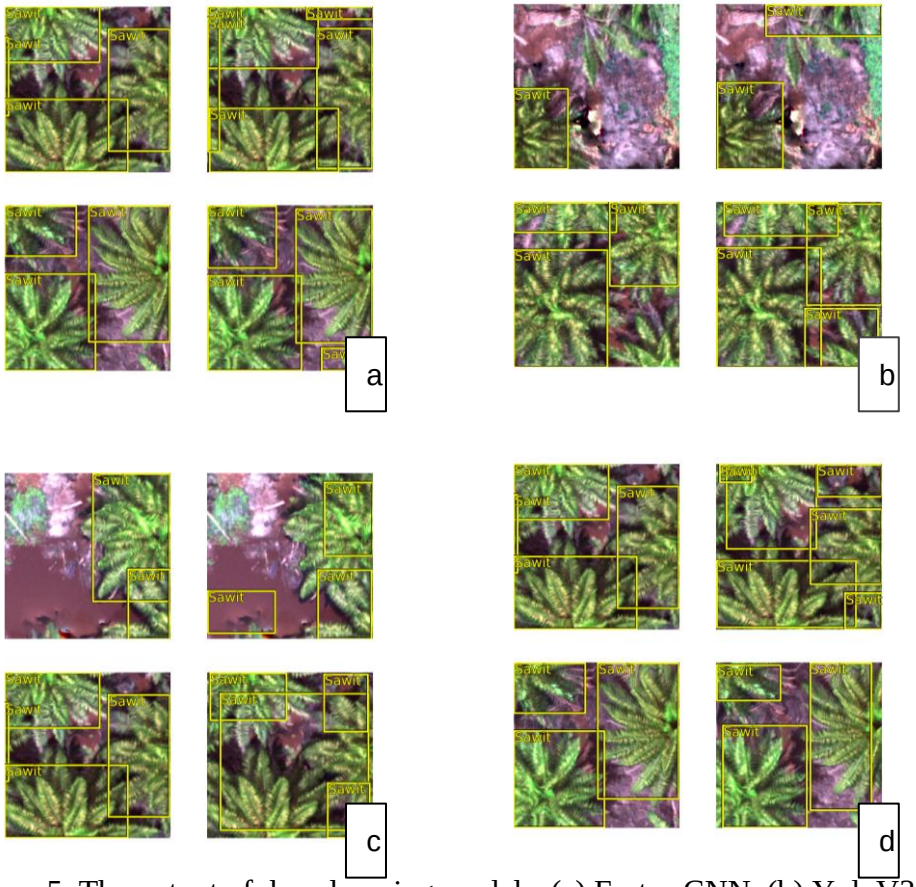

Figure 5. The output of deep learning models. (a) Faster-CNN, (b) YoloV3, (c) Single Shoot Detect, (d) Retina-Net

a. Tree counting using deep learning results

Tree counting with several methods in ArcGIS pro has a result shown in table 1:

Table 1. The results of tree counting using deep learning

No	Methods	Count of Trees	Confidence Threshold
1	Faster-CNN	3666	0,5
2	YoloV3	7739	0,5
3	Single Shoot Detect	1650	0,73
4	Retina-Net	1848	0,74

b. Comparison between deep learning results and field data

Based on the results of data processing using deep learning compared with field data, we found the difference between the two, which is displayed in table 2:

Table 2. Comparison between deep learning results and field data

No	Methods	Count of Trees Deep Learning	Field Data	Deviation
1	Faster CNN	3666	1667	1999
2	YoloV3	7739	1667	6072
3	Single Shoot Detect	1650	1667	17
4	Retina-Net	1848	1667	181

c. Train and Validation Graph

This graph represents the training and field validation data, where the relationship between the batch process and data loss is presented as shown in figure 6:

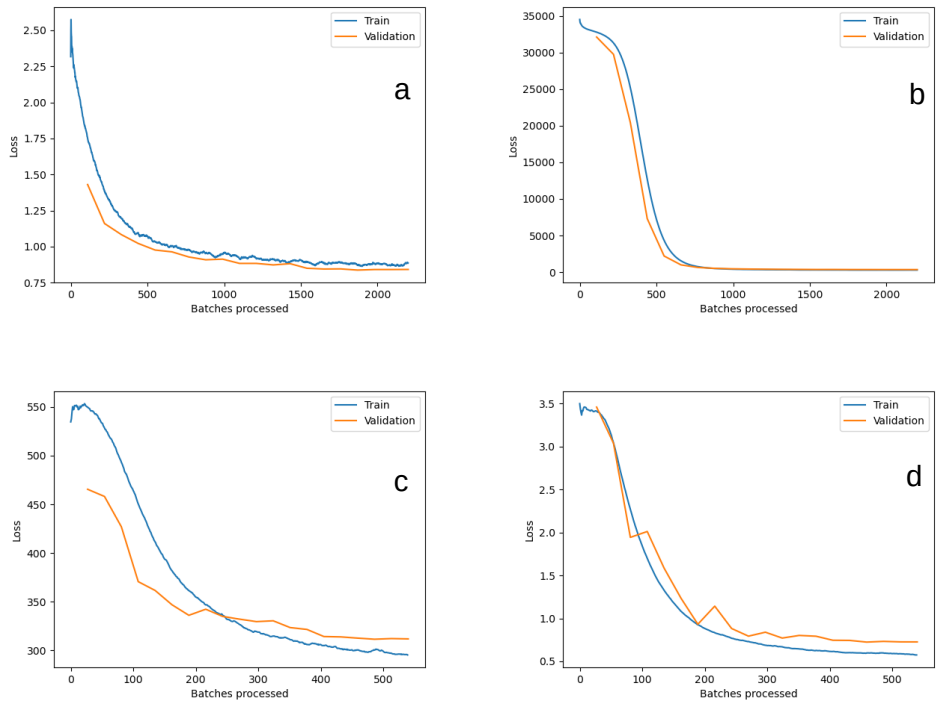

Figure 6. Statistics of tree counting using deep learning. (a) Faster-CNN, (b) YoloV3, (c) Single Shoot Detect, (d) Retina-Net

According to the tree counting deep learning training, Faster-CNN, YoloV3, and Retina-Net, the results of batch value show are close to the loss value. The training data do not represent manual counting because the backbone model of the YoloV3 is Darknet-53, while the other backbone models are Resnet. However, the YoloV3 method is more suitable for object detection through gaps between plants. The other backbone model in the form of Resnet

is SSD, and the cloud is seen that the value of the train data is higher than the validation value. Then, at some point found, the values overlap. So that this model represents the close results between training data to data validation, the SSD models can facilitate palm oil tree modelling with high-accuracy results.

5. CONCLUSION

This study investigates Single Shoot Detect (SSD) using the Resnet backbone method that approaches field data. The tree counting results generated by this method are 1650, while the field data, which is considered validation data, has 1667 tree counting results. The conclusion is that there are 17 more palm trees in the field data than in the Single Shoot Detect method. The Single Shot Detect method approaches the reliability of field data with a confidence threshold of 0.73; statistically, it can be stated to have a confidence level of 73%.

ACKNOWLEDGMENTS

The corresponding author is funded by Ikatan Surveyor Indonesia (Indonesia Surveyors Association). The author thanks Aria Agri Indonesia for providing the data to process tree counting modelling and ArcGIS Pro for providing the tree counting modelling system.

REFERENCE

Junaedi, J. (2022). INDONESIA IS THE BIGGEST GRANT OF OIL PALM CRUDE PALM OIL (CPO) IN THE WORLD BUT FACING THE PROBLEM OF OIL SCARCITY SURPRISE COOKING OIL PRICES. *International Journal of Social Science, 2*(4), 1779–1790. https://doi.org/10.53625/ijss.v2i4.4137

Yao, L., Liu, T., Qin, J., Lu, N., & Zhou, C. (2021). Tree counting with high spatial-resolution satellite imagery based on deep neural networks. *Ecological Indicators*, *125*. https://doi.org/10.1016/j.ecolind.2021.107591

Yin, N., Liu, R., Zeng, B., & Liu, N. (2019). A review: UAV-based Remote Sensing. *IOP Conference Series: Materials Science and Engineering*, *490*(6). https://doi.org/10.1088/1757-899X/490/6/062014

Ammar, A., & Koubaa, A. (2020). *Deep-Learning-based Automated Palm Tree Counting and Geolocation in Large Farms from Aerial Geotagged Images*. http://arxiv.org/abs/2005.05269

Wibowo, H., Sitanggang, I. S., Mushthofa, M., & Adrianto, H. A. (2022). Large-Scale Oil Palm Trees Detection from High-Resolution Remote Sensing Images Using Deep Learning. *Big Data and Cognitive Computing*, *6*(3). https://doi.org/10.3390/bdcc6030089

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2015). *SSD: Single Shot MultiBox Detector*. https://doi.org/10.1007/978-3-319-46448-0_2

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). *Focal Loss for Dense Object Detection*. http://arxiv.org/abs/1708.02002

Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: NIPS. (2015)

Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR. (2016)

Girshick, R.: Fast R-CNN. In: ICCV. (2015)

Liu, W., Rabinovich, A., Berg, A.C.: ParseNet: Looking wider to see better. In: ILCR. (2016)

kristal, A. (2022). *Analysis of the Reliability of Building Footprint Extraction Using a Deep Learning Approach Based on Mask R-CNN from Aerial Photos*. *17*(2), 273.

BIOGRAPHICAL NOTES

M. Sc. RAUDLAH HAWIN AYANI

2014 - 2018	B.Eng in Geomatics Engineering, Institut Teknologi Sepuluh Nopember, Indonesia
2018-2019	M.Eng in Geomatics Engineering, Institut Teknologi Sepuluh Nopember, Indonesia
2019-2021	M.Sc in Geomatics, National Cheng Kung, Taiwan
2021	Author of Crustal Deformation of the Kendeng Fault Branches Area from GNSS and InSAR Data in Surabaya City, Indonesia in Geoicon Conference
2022	Countryside GIS Expert in Ministry of National Development Planning/Bappenas Republic of Indonesia
2022	Head of GIS in Aria Agri Indonesia
2022	Young Surveyor of Indonesian Surveyors Associate

B. Eng ASTRID CALISTA

2017 - 2021	B.Eng in Geomatics Engineering, Institut Teknologi Sepuluh Nopember, Indonesia
2019 - 2020	GIS Analyst in Mapfan Olrait I-Sense Technology
2021 - 2022	Head of Dept. Geospatial in Surabaya Survey Solutions
2021	Author of Study of Sentinel-1A Satellite Image Data Utilization to Observe Deformation of Post Earthquakes using DINSAR Method (*Differential Interferometry Synthetic Apreture Radar*) (Case Study: Pesisir Barat Regency, Lampung) in POMITS Institut Teknologi Sepuluh Nopember, Publication
2022	GIS Data Processing in PT. Aria Agri Indonesia

2022 Young Surveyor of Indonesian Surveyors Associate

MULYANTO
2016 - 2021 Data engineer in PT. AMZ Geoinfo Solution
2022 GIS Data Processing in PT. Aria Agri Indonesia

CONTACTS
M.Sc Raudlah Hawin Ayani
Indonesian Surveyors Association
Jakarta, Indonesia
raudlahhawin@gmail.com