Automatic Detection of Oil Palm Tree from UAV Images Based on the Deep Learning Method

Xinni Liu, Kamarul Hawari Ghazali, Fengrong Han & Izzeldin Ibrahim Mohamed

Published online: 22 Oct 2020.

Submit your article to this journal ⍈

View related articles ⍈

View Crossmark data ⍈

Taylor & Francis
Taylor & Francis Group

Check for updates

Automatic Detection of Oil Palm Tree from UAV Images Based on the Deep Learning Method

Xinni Liu, Kamarul Hawari Ghazali, Fengrong Han, and Izzeldin Ibrahim Mohamed

Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pekan, Malaysia

ABSTRACT

Palm oil is a major contributor to Malaysia's GDP in the agriculture sector. The sheer vastness of oil palm plantations requires a huge effort to administer. An oil palm plantation in regards to the irrigation process, fertilization, and planning for planting new trees require an audit process to correctly count the oil palm trees. Currently, the audit is done manually using aerial view images. Therefore, an effective and efficient method is imperative. This paper proposes a new automatic end-to-end method based on deep learning (DL) for detection and counting oil palm trees from images obtained from unmanned aerial vehicle (UAV) drone. The acquired images were first cropped and sampled into small size of sub-images, which were divided into a training set, a validation set, and a testing set. A DL algorithm based on Faster-RCNN was employed to build the model, extracts features from the images and identifies the oil palm trees, and gives information on the respective locations. The model was then trained and used to detect individual oil palm tree based on data from the testing set. The overall accuracy of oil palm tree detection was measured from three different sites with 97.06%, 96.58%, and 97.79% correct oil palm detection. The results show that the proposed method is more effective, accurate detection, and correctly counts the number of oil palm trees from the UAV images.

Introduction

The most important economic crop in Malaysia and other tropical countries such as Indonesia, Nigeria, and Thailand are the oil palm trees. It is one of the main contributors to Malaysia's Gross Domestic Product (GDP) and foreign exchange incomes. Palm oil has also become the main source of vegetable oil due to its high production yield compared to rapeseed, sunflower, and soybean (Hansen et al. 2015). Malaysia has a total of 4.49 million hectares of oil palm plantations, which produces approximately 17.73 million tons of palm oil each year (Nambiappan et al. 2018). Hence, one of the key tasks in plantation management is to count the number of oil palm trees, also known as the oil palm tree audit. The process allows the plantation manager to estimate the

CONTACT Xinni Liu ✉ kamarul@ump.edu.my 📄 Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Pekan, Malaysia

yield of plantation and plan the weeding and fertilization strategy (Huth et al. 2014). The audit also facilitates the irrigation improvement process, planting of new trees, and managing old and dead trees (Kattenborn et al. 2014).

Recent studies show that there are three main methods of classification and detection of oil palm trees which are image processing, machine learning, and deep learning (DL). Implementation of the image processing method is more direct in which the algorithm is used to extract the features of an object such as shape, edge, and color, then perform direct manipulation or transformation of the pixel values of the images. The challenging part of image processing in object detection is when the object is crowded in one place, overlapped with other objects or the background is complex (Sheng et al. 2012). In reference (Nam et al. 2016), Scale Invariant Feature Transform (SIFT) based on image processing methods was used to detect oil palm trees. SIFT key points were extracted together with local binary patterns to analyze the texture of the overall region from aerial images. The method achieved 91.11% accuracy of oil palm tree detection. More progressive image processing method has been used to mask areas of palm oil plantations (Korom et al. 2014). The features of the oil palm trees were derived using the vegetation indices and kappa statistics. Optimal thresholding was used to distinguish between a tree and non-tree region. The method yielded a 77% accuracy of oil palm tree detection. In reference (Manandhar, Hoegner, and Stilla 2016), the oil palm tree shape feature was represented by the result of the circular autocorrelation of the polar shape matrix. The shape feature was extracted using the sliding window method. An algorithm to detect the local maximum was used to detect the palm trees in different tough scenarios. The average detection accuracy obtained was 84%.

For the machine learning methods, hand-crafted features are still needed by the simple trainable classifier to segment the oil palm tree image as individual tree crowns and extract the feature before applying a predefined classier for detection and classification. In reference (Rizeei et al. 2018), object-based image analysis (OBIA) and support vector machine (SVM) method were used for oil palm tree counting in Worldview-3 satellite image and range (LiDAR) airborne imagery. The oil palm tree detection result was 98% on the testing site, but only satellite and LiDAR airborne imageries were used in this method. Machine learning methods are highly dependent on hand-crafted features. When dealing with new data, new hand-crafted features based on the new data are required, which can be costly. The extraction method is then repeated with the new features; however, the process proves to be tedious (Fassnacht et al. 2016).

Presently, DL methods have been widely used in many applications, particularly in image detection and classification (Alliez et al. 2017; Kemker, Salvaggio, and Kanan 2018; Rezaee et al. 2018). DL is essentially a subset of machine learning which belongs to the broader family of artificial intelligence.

DL is based on artificial neural network (ANN) that has a network of numerous hidden layers. Unlike machine learning algorithms that almost always require structured data, deep learning has networks that are capable of supervised or unsupervised learning from labeled or unlabeled data. Li et al. (Li et al. 2016) used DL to classify oil palm trees from high-resolution remote sensing images collected by the QuickBird satellite. A sliding window and post-processing approaches were used to detect the oil palm trees. This method achieved a detection accuracy of 96%. Maciel et al. (Zortea et al. 2018) trained two independent convolutional neural networks with two input sizes 32×32 and 64×64 for capturing both the coarse and fine details of the oil palm tree image. The estimated probabilities for each oil palm tree were combined by averaging the results from two detection networks, and the overall accuracies for the method were in the range between 91.2% and 98.8%. Both of the two methods used DL image classification methods.

From the previous studies in (Korom et al. 2014; Manandhar, Hoegner, and Stilla 2016; Nam et al. 2016), the drawback is essentially on the accuracy of the detection. DL methods have proven to increase the accuracy percentage (Li et al. 2016; Zortea et al. 2018). In this paper, an oil palm tree image detection algorithm using a DL Faster RCNN approach is described. The input images are aerial images from the UAV drone, which is presently widely used in modern precision agriculture.

Deep Learning Method for Oil Palm Tree Counting

Deep Learning Method Overview

The method proposed in this paper employed state-of-the-art DL based on the Faster RCNN approach to build and train the oil palm tree detection model (Ren et al. 2017). High-resolution drone images were cropped into multiple sub-images and were arranged into training, validation, and testing data. Figure 1 shows the flowchart of the proposed DL method. Figure 2 shows the sample of the original high-resolution oil palm tree image in a plantation. The images in the training dataset go through a random process of brightness enhancement for augmentation purposes. The augmentation process increases the total number of the training dataset, which reduces the overfitting problem (Guo et al. 2016). The training dataset and the validation dataset were made into labeled data. Next, the convolutional neural network (CNN) was implemented based on the Faster RCNN framework. The labeled images in the training dataset were used to train the network and measure the detection accuracy based on the testing dataset. The network parameters of the Faster RCNN were tuned continuously until the best combinations which contribute to a high percentage of detection were obtained. These combinations were also used for accuracy analysis.
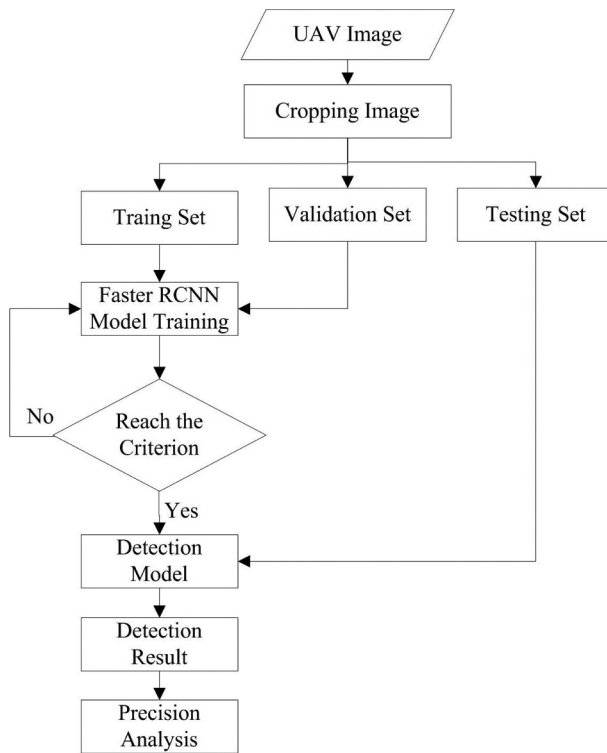
Figure 1. Flowchart of the proposed DL method.

Figure 2. High-resolution of a palm tree image used.

Faster RCNN

The Faster RCNN model consisted of the convolution network, region proposal network (RPN), region of interest (ROI) pooling, softmax classification, and bounding box regression. Figure 3 shows the training process of the method using the datasets described in Section 2.1. Once the network was trained, the performance of the network model was tested using the validation and testing dataset.

For the network training, the labeled training dataset was introduced to the convolution layer, which will extract the deep features of the image to produce the feature map of the input image. The feature map was then fed into the RPN layer. The RPN was trained to yield the region with probable oil palm trees. The probable region is marked by an orange box which contains the probability score of oil palm trees. The obtained feature map and the proposed region were fed into the ROI layer to generate proposed regions based on the previously obtained feature map. The information is then fed to the fully connected layer and by using the softmax classification and bounding box regression, the final tree detection was obtained. The final network was validated and tested.

The loss function during the training contains the classification loss, L_{cls} and the region loss, L_{reg} is shown by Equation (1), where i indicates the bounding box index, p_i is the foreground softmax probability, and p_i^* is the probability for ground truth box, t_i indicates the predicted bounding box, t_i^* is the corresponding ground truth box for the foreground bounding box. The training aims to minimize the loss. N_{cls} indicates the number of class and N_{reg} indicates the number of the proposal regions, parameter μ is used to balance the difference between N_{cls} and N_{reg}. The classification loss L_{cls} was calculated as the loss after softmax classification, the regression loss L_{reg} for bounding box regression was calculated by Smooth L1 loss, shown as Equation (2). Parameter t_i^* is the ground truth box while $t_i = \{t_x, t_y, t_w, t_h\}$ indicates the parameter of the predicted bounding box. The non-maximum suppression was used to further obtain the proposed region containing the oil palm trees.
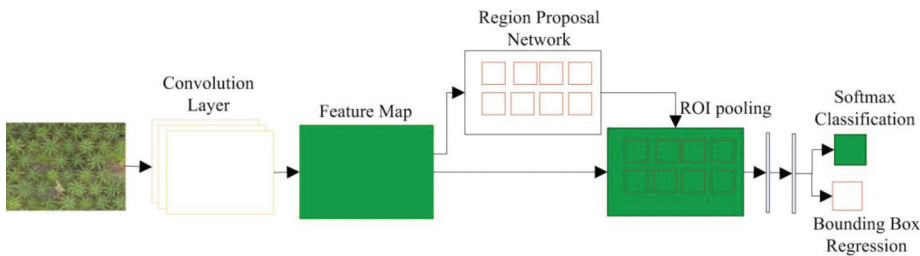

Figure 3. The training process of faster RCNN.

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}\left(p_i, p_i^*\right) + \mu \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}\left(t_i, t_i^*\right) \tag{1}$$

$$L_{reg}\left(t_i, t_i^*\right) = \sum_{i \in \{x, y, w, h\}} \text{Smooth}L1\left(t_i - t_i^*\right) \tag{2}$$

$$\text{Smooth}L1(X) = \begin{cases} 0.5X^2 \text{if} |X| < 1 \\ |X| - 0.5 \text{otherwise} \end{cases} \tag{3}$$

Experimental Analysis

Dataset

In this research, the aerial view image data were acquired by a drone services company. The images were of a 22 hectares oil palm tree plantation. The drone has a band order of red, blue, and green (RGB) and its ground sample distance (GSD) is 5 cm per pixel. The plantation contains different age groups of trees between 2 and 16 years old. The oil palm trees are crowded in some areas, overlapped with some other vegetation, or sparsely distributed in other areas, as shown in Figure 4. The original image is 7837 × 23608, which was cropped into 700 sub-images with 600 × 500 pixels; 500 of these images were set as the training data, 100 images were chosen as the validation dataset and another 100 images as the testing dataset. For the DL network to achieve optimum accuracy, the training dataset should consist of all possible conditions of the oil

Figure 4. Oil palm tree plantation.

(a)Crowded palm trees   (b)Overlapped with other vegetation

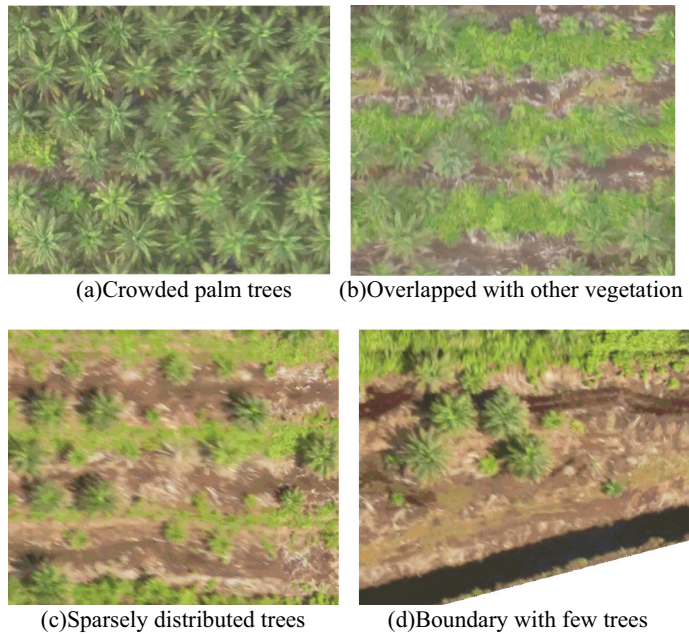(c)Sparsely distributed trees   (d)Boundary with few trees

Figure 5. The training samples.

palm trees in terms of age, surrounding vegetation, and background. Shown in Figure 5 are the samples of training dataset with different ages and distribution of trees; crowded mature palm trees, overlapped with other vegetation, sparsely distributed and boundaries with a few palm trees.

Faster RCNN Parameter Configuration

The popular classical VGG16 network trained from ImageNet was used as the backbone of the convolution network (Karen Simonyan 2014). The weights of first 1–5 convolutional layer remain unchanged while the weights of the 6th-7th convolutional layer were converted to the fully connected layers. The weights were changed accordingly, where 6th full connected layer has 512 kernels of 7×7 and 7th full connected layer has 4096 kernels of 1×1, the ROI pooling size is 7, all the dropout layers and the 8th full connected layer were removed. The parameter configuration during Faster RCNN network training is shown in Table 1. The RPN network yielded 300 proposal boxes using the initial parameter values shown in Table 1. The softmax function was used to classify the proposal feature maps as an object or backgrounds. The bounding box regression method was used to obtain a more accurate proposal box. There are two classes in this study, namely the object and the background. The object is the oil palm tree while the background is the vegetation, empty land, and boundary.

Table 1. Parameter set of faster RCNN.

Parameters	Value
Learning rate	0.001
Momentum	0.9
Gamma	0.1
Weight decay	0.0005
Proposal boxes	300

Experiment Result and Accuracy Assessment

Experiment Results

The results of the presented model based on the Faster RCNN model are shown in Figure 6. The detected oil palm trees were marked with red rectangle boxes together with the confidence scores which indicates the score of the confidence that the object is oil palm tree. The oil palm trees not detected were manually marked with blue rectangle boxes. Figure 6a shows an image of mature and crowded oil palm trees. All of the oil palm trees were successfully detected, and the confidence score for oil palm trees is more than 0.9. Figure 6b shows similar mature oil palm trees but they were overlapped with other vegetation, only one tree was not detected, and the confidence score obtained for this image is more than 0.97. In Figure 6c, only three small oil palm trees

(a) Crowded palm tree　　　(b) Overlapped with vegetations

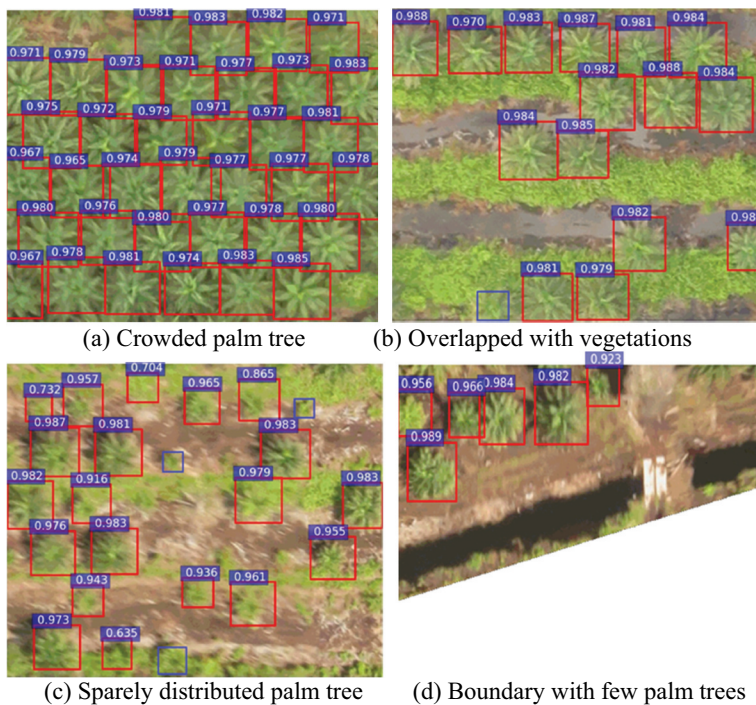(c) Sparely distributed palm tree　　(d) Boundary with few palm trees

Figure 6. Detection results on four samples. (a) Crowded palm tree, (b) overlapped with vegetations, (c) sparely distributed palm tree, (d) boundary with few palm trees.

were not detected, and the confident score four small detected oil palm trees are 0.70, 0.73, 0.86, and 0.63, respectively, while for the other detected oil palm trees it is more than 0.9. Figure 6d shows young oil palm trees on the boundary which were all successfully detected, and the confidence score is more than 0.92. It can be seen that Figure 6b, d have few palm trees, but the background of Figure 6d is clearer and simpler than 6(b) and 6(c), so all the oil palm trees are successfully detected in 6(c). Based on the characteristic of the detection image and the detection result, it is observed that oil palm trees that appear small in the images and oil palm trees that overlapped with other vegetation are more difficult to detect.

Accuracy Assessment

Accuracy assessment was done by comparing the detected and the non-detected tree to the number of actual trees in the image to measure the precision, recall, and overall accuracy. The process was conducted manually. The precision, recall, and overall accuracy (OA) are described by Equations (4)-(6), where TP is the true positive detection result, indicating the total number of correctly detected oil palm trees, FP is the false positive detection result, which gives the number of objects incorrectly detected as oil palm trees, and FN is the false negative detection result, which expresses the number of oil palm trees that were not detected.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

$$OA = \frac{\text{Precision} + \text{Recall}}{2} \tag{6}$$

The result of 100 testing dataset is given in Table 2. A total of 2326 oil palm trees were correctly detected, 56 other vegetation were detected as oil palm trees, and 60 oil palm trees were not detected. The precision, recall and OA are 97.65%, 97.49%, and 97.57%, respectively.

The performance of the proposed method was tested against the traditional machine learning methods, i.e. ANN-based (Fadilah et al. 2012) and SVM-based methods (Dalponte et al. 2015). Three other sites were selected for the validation purposes; site A where most of the palm trees overlap with other

Table 2. Detection result of the testing set.

Data	TP	FP	FN	Precision	Recall	OA
Testing set	2326	56	60	97.65%	97.49%	97.57%

vegetations, site B where the palm tree in the plantation has small size, and site C with most of the mature oil palm trees crowded. For the ANN- and SVM-based methods, the images were cropped into individual oil palm trees with background and then labeled as the training set which was used to train the ANN and SVM network to obtain parameters to develop the ANN and SVM models. The models were then implemented to detect the oil palm trees in the test images using a sliding window method (Pibre et al. 2017). For the proposed method, the input images of each site were cropped to 600×500 pixels or less than 600×500 pixels sub-images. These sub-images were then fed into the aforementioned pretrained Faster RCNN model. The detection results for site A, site B, and site C are shown in Tables 3, 4 and 5, respectively.

The detection results of all three sites show a similar pattern; the proposed method has the highest TP, and the lowest FP and FN compared to the results of the ANN- and SVM-based methods. The stark difference between the FP results of the proposed method compared to the two other methods is the most interesting result. The FP produced by the traditional methods is tenfold more compared to the proposed method. The huge number of FP could have easily contributed to the low percentage of precision and OA of the ANN- and SVM-based method. The detection results from all three sites show that the proposed method has the highest precision number and OA percentage with a difference of approximately 10% to 17%.

Table 3. Detection result of site A.

Approach	TP	FP	FN	Precision	Recall	OA
ANN	705	386	25	64.62%	96.58%	80.60%
SVM	697	325	33	68.20%	95.48%	81.84%
Proposed method	710	23	20	96.86%	97.26%	97.06%

Table 4. Detection result of site B.

Approach	TP	FP	FN	Precision	Recall	OA
ANN	834	365	36	69.56%	95.86%	82.71%
SVM	832	263	38	75.94%	95.63%	85.79%
Proposed method	845	25	25	97.13%	96.02%	96.58%

Table 5. Detection result of site C.

Approach	TP	FP	FN	Precision	Recall	OA
ANN	785	182	31	81.18%	96.20%	88.69%
SVM	788	225	28	77.79%	96.57%	87.18%
Proposed method	796	16	20	98.03%	97.55%	97.79%

Conclusions

This study proposed a novel approach for detecting oil palm trees from UAV images. It employed state-of-the-art DL based on the Faster RCNN and image pre-processing method to detect oil palm trees from drone images of the plantation. The performance of the proposed method was compared to traditional machine learning-based ANN and SVM methods on three different sites. The overall accuracies obtained for the proposed method were more than 96%. The proposed method outperforms the other two machine learning methods in terms of precision and overall accuracy. The results also showed that the proposed method was effective and robust in the detection of oil palm trees in various plantation conditions. The object detection was unsuccessful when the size of the oil palm trees in the images was too small, there are overlapping background and low resolution of the object. Nonetheless, based on the model pretrained in this paper, it takes 1.5 hour to detect and count the 22 hectares oil palm tree plantation in this study, the results were promising which suggests that it has the potential to be used in practical applications.

Funding

References

Dalponte, M., L. T. Ene, M. Marconcini, T. Gobakken, and E. Næsset. 2015. Semi-supervised SVM for individual tree crown species classification. *ISPRS Journal of Photogrammetry and Remote Sensing* 110:77–87. doi:10.1016/j.isprsjprs.2015.10.010.

Fadilah, N., J. M. Saleh, H. Ibrahim, and Z. A. Halim. 2012. Oil palm fresh fruit bunch ripeness classification using artificial neural network. Paper presented at the 2012 4th International Conference on Intelligent and Advanced Systems (ICIAS2012), Kuala Lumpur, Malaysia, June 12–14.

Fassnacht, F. E., H. Latifi, K. Stereńczak, A. Modzelewska, M. Lefsky, L. T. Waser, and A. Ghosh. 2016. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment* 186:64–87. doi:10.1016/j.rse.2016.08.013.

Guo, Y., Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew. 2016. Deep learning for visual understanding: A review. *Neurocomputing* 187:27–48. doi:10.1016/j.neucom.2015.09.116.

Hansen, S. B., R. Padfield, K. Syayuti, S. Evers, Z. Zakariah, and S. Mastura. 2015. Trends in global palm oil sustainability research. *Journal of Cleaner Production* 100:140–49. doi:10.1016/j.jclepro.2015.03.051.

Huth, N. I., M. Banabas, P. N. Nelson, and M. Webb. 2014. Development of an oil palm cropping systems model: Lessons learned and future directions. *Environmental Modelling and Software* 62:411–19. doi:10.1016/j.envsoft.2014.06.021.

Kattenborn, T., M. Sperlich, K. Bataua, and B. Koch. 2014. Automatic single tree detection in plantations using UAV-based photogrammetric point clouds. *The International Archives of*

Photogrammetry, Remote Sensing and Spatial Information Sciences 40 (3):139. doi:10.5194/isprsarchives-XL-3-139-2014.

Kemker, R., C. Salvaggio, and C. Kanan. 2018. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. *ISPRS Journal of Photogrammetry Remote Sensing Letters* 145:60–77. doi:10.1016/j.isprsjprs.2018.04.014.

Korom, A., M. Phua, Y. Hirata, and T. Matsuura. 2014. Extracting oil palm crown from WorldView-2 satellite image. Paper presented at the IOP Conference Series: Earth and Environmental Science, Volume 18, 8th International Symposium of the Digital Earth (ISDE8) August 2013, Kuching, Sarawak, Malaysia 26–29.

Li, W., H. Fu, L. Yu, and A. Cracknell. 2016. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sensing* 9 (1):22–35. doi:10.3390/rs9010022.

Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P., 2017. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):645–657. doi:10.1109/TGRS.2016.2612821

Manandhar, A., L. Hoegner, and U. Stilla. 2016. Palm tree detection using circular autocorrelation of polar shape matrix. *ISPRS Annals of Photogrammetry, Remote Sensing Spatial Information Sciences* 3:465–72.

Nam, L., L. Huang, X. J. Li, and J. Xu. 2016. An approach for coverage path planning for UAVs. Paper presented at the 2016 IEEE 14th international workshop on advanced motion control (AMC), Auckland, New Zealand 411–416.

Nambiappan, B., A. Ismail, N. Hashim, N. Ismail, D. Nazrima, N. Abdullah, . . . N. Ain. 2018. Malaysia: 100 years of resilient palm oil economic performance. *Journal of Oil Palm Research* 30 (1):13–25. doi:10.21894/jopr.2018.0014.

Pibre, L., M. Chaumon, G. Subsol, D. Lenco, and M. Derras. 2017. *How to deal with multi-source data for tree detection based on deep learning.* Paper presented at the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Montreal, QC, Canada, November 14–16.

Ren, S., K. He, R. Girshick, and J. Sun. 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6):1137–49. doi:10.1109/TPAMI.2016.2577031.

Rezaee, M., Y. Zhang, R. Mishra, F. Tong, and H. Tong. 2018. Using a VGG-16 network for individual tree species detection with an object-based approach. Paper presented at the 2018 10th IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), Beijing, China.

Rizeei, H. M., H. Z. Shafri, M. A. Mohamoud, B. Pradhan, and B. Kalantar. 2018. Oil palm counting and age estimation from WorldView-3 imagery and LiDAR data using an integrated OBIA height model and regression analysis. *Journal of Sensors* 2018.

Sheng, G., W. Yang, T. Xu, and H. Sun. 2012. High-resolution satellite scene classification using a sparse coding based multiple feature combination. *International Journal of Remote Sensing* 33 (8):2395–412. doi:10.1080/01431161.2011.608740.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556

Zortea, M., M. Nery, B. Ruga, L. B. Carvalho, and A. C. Bastos. 2018. Oil-palm tree detection in aerial images combining deep learning classifiers. Paper presented at the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, July 22–27.