A Case Study on Detecting and Mapping Individual Coconut Trees using YOLOv3 in Conjunction with UAV Remote Sensing for Smart Plantation Management

RMDM Karunarathna[1#], KHN Kulapathi[1] and KLPI Liyanage[1]

[1]*Centre for Defence Research and Development, Sri Lanka*

[#]r.m.d.karunarathna@gmail.com

Abstract:*Location and number data of individual coconut trees are important for surveying planting areas, predicting coconut yield, and managing and planning coconut plantations. This data had usually obtained through manual investigation and statistics, which is time-consuming and tedious. Deep learning object recognition models, widely used in computer vision, can provide an opportunity to accurately identify individual coconut trees, which is essential for rapid data acquisition and the reduction of human error. This study proposes an approach to identify individual coconut trees and map their spatial distribution by combining deep learning with unmanned aerial vehicle (UAV) remote sensing. UAV remote sensing collected high-resolution true-colour images of coconut trees at the Mahayaya Coconut Model Plantation in Sri Lanka. An image dataset of deep learning models of individual coconut trees (ICTs) had constructed by visual description and field survey based on coconut tree images captured by UAV remote sensing. YOLOv3 was selected to train, validate and test the image dataset of coconut trees. The results show that the average accuracy of the YOLOv3 model for validation reaches 91.7%. The number of ICTs in the study area was calculated using YOLOv3, and their spatial distribution map was created using the non-maximum suppression method and ArcGIS software. This study will provide basic data and technical support for smart coconut plantation management in Mahayaya coconut model plantation and other coconut-producing areas.*

Keywords: *Individual Coconut Tree (ICT) detection; deep learning; YOLOv3; remote sensing; Unmanned Aerial Vehicle (UAV); spatial distribution*

1. Introduction

Smart plantation management requires accurate planting area survey, disease and pest prevention and control, and data on the location and characteristics of individual coconut trees (ICT) in a plantation for coconut yield prediction. Traditionally, field surveys and statistics were used to collect this data, such as locations, spatial distribution, number of ICTs, etc. These surveys are time-consuming, labour-intensive and expensive but fail to meet the requirements of smart plantation management. There is a need to develop a fast, cheap and accurate methodology for ICT investigation to obtain this data.

Remote sensing images of coconut trees in relatively large plantations can be captured by satellite or aerial photography. When using satellite remote sensing, cloudy weather is initially a major challenge, and due to the poor quality of images, coconut trees are difficult to detect. Limiting the spatial resolution of satellite imagery is another major challenge for accurately identifying ICTs. Aerial photography is taking pictures using manned or unmanned aircraft. Manned aerial vehicles are not suitable for ICTs detection due to high costs and difficult operations. Unmanned aerial vehicles (UAV) remote sensing is the best option to accomplish this task. Drones are a subset of UAVs that are generally very small, light and inexpensive. A drone usually has one or more high-resolution cameras that can capture medium to high-quality images depending on the height it flies (Jintasuttisak, Edirisinghe & Elbattay 2022). UAV remote sensing has automation, intelligence, and specialization advantages to quickly obtain space remote sensing information such as land, resources, environment, and events, and conduct real-time processing, modelling, and analysis of advanced emerging aerial remote sensing technology solutions (Li & Li 2014). Recently, it has been widely used in many practical fields, such as photometry, precision agriculture (Khanal, Fulton & Shearer 2017), geohazard assessment (Li & Li 2014), forest fire detection (Ghali, Akhloufi & Mseddi 2022), and environmental monitoring (Wu, Shan, Lai & Zhou 2022; Immerzeel et al. 2014). UAV remote sensing has great potential to quickly and economically acquire image data of coconut trees in plantations.

In the recent decade, with the development of computer hardware devices and the rapid development of artificial intelligence (AI) technology, deep learning convolutional neural network (CNN), the core technology of AI, has pioneered new object recognition methods and feature extraction. In remote sensing images (Osco et al. 2021; Zhang, Zhang & Du 2016). Many CNN architectures have been proposed for object recognition in computer vision and image analysis, and they are divided into two categories, two-stage and one-stage models. (Girshick, Donahue, Darrell & Malik 2014) proposed an R-CNN (Region-based Convolutional Neural Network) two-object detection model based on classification problems. Based on R-CNN, fast RCNN and fast R-CNN are then proposed to improve performance and accuracy. (Redmon, Divvala, Girshick & Farhadi 2016) single-stage based object recognition model YOLO (You Look Only Once). The YOLO model not only simplifies the neural network size but also improves the recognition speed while improving the recognition accuracy. (K et al. 2022) proposed a pipeline based on YOLOv2 to perform fast multiscale object detection in large-scale satellite imagery. (2021) (Osco et al. 2021) present a comprehensive review of the fundamentals of deep learning related to UAV-based imagery, providing a key reference for integrating deep learning with UAS remote sensing for ICT detection.

More recently, (dos Santos et al. 2019) proposed and evaluated the use of CNN-based methods combined with high spatial resolution UAV imagery in red-green-blue (RGB) to identify legally protected tree species. Three state-of-the-art object detection methods were evaluated: fast R-CNN, YOLOv3 and RetinaNet. RetinaNet gave the most accurate results, with an average accuracy of 92.64%. Satellite imagery analysis by (Brandt et al. 2020) found isolated tree canopies over a large area of West Africa. Their results show that mapping the location and size of each tree worldwide can be done quickly with some limitations (Brandt et al. 2020; Hanan & Anchang 2020). (Safonova, Guirado, Maglinets, Alcaraz-Segura & Tabik 2021) used masked R-CNN and UAV imagery for olive tree canopy and shadow segmentation to further estimate the biomass of individual trees. (Sun et al. 2022) applied an end-to-end tree count deep learning framework (CMask R-CNN) to regional tree recognition by calculating the tree population in the subtropical metropolis Guangzhou and representing the crown of each tree. (Hu et al. 2022) presented a pipeline for tracking and clustering 259 peach tree crowns based on UAV images of a peach orchard in Southeast China and constructed conditional generative adversarial networks (cGANs) to extract the crown area. The results of (Yu et al. 2022) showed that the mask-R-CNN model achieved the highest accuracy (F1 score = 94.68%) for identifying a single tree compared to the local maxima algorithm and marker-limited watershed segmentation.

Motivated by the great progress in single tree detection by deep learning and UAV remote sensing, we proposed an approach to build an accurate Individual Coconut Tree (ICT) detection model by combining deep learning with UAV remote sensing images to fill the gap in the above studies. With this model, the location and spatial distribution of ICTs can be quickly and accurately mapped, and the number of ICTs can also be quickly calculated. We envision that state-of-the-art deep learning methods can detect ICTs in high-resolution, true-colour images with low cost, high accuracy, and high performance. Coconut trees are selected as a case study and will be empirically determined as a deep-transfer learning model for training and validation to test the YOLOv3 hypothesis. It aims to provide reliable and timely baseline data and technical support for intelligent plantation management and precision farming development.

More specifically, three main contributions were reported in this paper. First, high-resolution images gagged by the UAV sensor were set for a sample set of tree images of individual coconut trees. Second, using the new data, the yolov3 model was trained and evaluated for realizing accurate and fast detection of ICTs. Third, a thematic map showing the location, spatial distribution, and the number of the ICTs in the large-scale coconut plantations of the study area. It can provide important reference information for precision plantation management.

2. Materials and Methods

A. Study area

A large coconut plantation is selected as the experimental study area. It is located in Makandura, Gonawila. (Figure 1). The coconut tree (Cocos nucifera) is a member of the palm tree family (Arecaceae) and the only living species of the genus Cocos. The coconut tree provides food, fuel, cosmetics, folk medicine and building materials, among many other uses. A mature coconut tree's height can reach 20-22 meters on average, and its crown diameter generally goes to 8-9 meters. Thus, it can be recognized in high-resolution UAV-based images by visual interpretation.

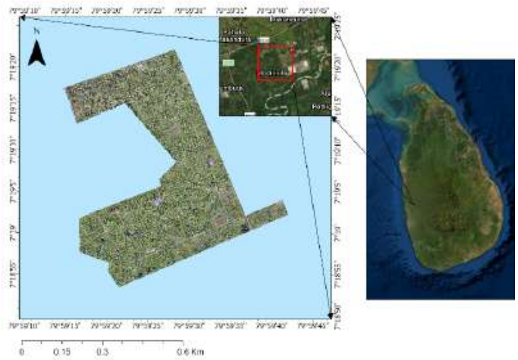

Figure 1. Overview of the experimental study area. The location map shows the study area in a red rectangle

Therefore, detecting ICTs, mapping their location and spatial distribution, and counting their planting area and the number of trees are important. It is desired to provide reference information for cultivation area investigation, yield prediction, and smart plantation management and plan in Sri Lanka.

B. Our proposed approach for individual coconut tree detection and mapping.

This study proposed an approach for detecting ICTs, mapping their spatial distribution, and counting their planting area and number by integrating deep transfer learning of YOLOv3 with high-resolution low-altitude UAV remote sensing images. The workflow of the proposed approach is illustrated in Figure 2, containing six steps shown as follows.

i. Capturing and processing UAV remote sensing images;
ii. Creating a dataset of Individual Coconut Tree Image Samples (ICTIS);

iii. Training, validating, and testing the YOLOv3 model;
iv. Evaluating the accuracy and performance, CocoNet, for the detection of ICTs will be obtained;
v. Mapping the location and spatial distribution of ICTs using the predicted results of CocoNet;
vi. Counting the planting area and the number of ICTs

To test and validate our proposed approach, the coconut trees were selected as the example targets to carry on the study on ICT detection and their spatial distribution mapping. The main methods and critical steps of the workflow are explained in detail in Figure 2.

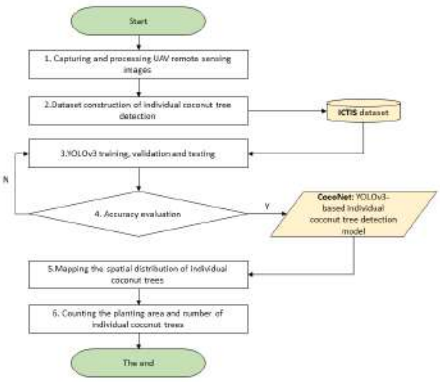

Figure 2. Workflow chart of our proposed approach to detecting and mapping individual coconut trees integrated YOLOv3 with UAV remote sensing.

1) Capturing and processing UAV remote sensing images: The DJI M300 RTK multispectral drone was used as a UAV system to capture low-altitude remote sensing images, equipped with 1/2.3" CMOS, 12 MP, including one RGB sensor for visible light imaging and five monochrome sensors for multispectral imaging (Blue, Green, Red, Red-Edge and Near-Infrared bands). It integrates RTK-enabled GNSS, including GPS and Galileo. So, it can capture high-quality multi-band remote sensing images without ground control points required in the traditional aerial survey. Furthermore, it can provide efficient tools for farmers in precision agriculture, significantly improving the efficiency of environmental data acquisition. To obtain high-quality UAV raw data, aerial photography tasks need to be planned before take-off.

A flight altitude of 150 m was set to capture high-quality UAV raw data with a spatial resolution of 5 cm, with 60% heading and lateral overlaps. Figure 3 shows some examples of the raw true-colour images collected by UAV aerial photography, which were used later to construct a dataset of RGB true-colour image samples of ICTs.

Figure 3. Examples of the raw true-colour images captured by UAV remote sensing.

The original images obtained by UAV remote sensing on date was pre-processed to generate a digital orthographic mosaic image model of the study area. The pre-processing steps mainly include:

i. Confirming the integrity of original image data, including camera parameters in the segment and segment attributes and GNSS information;
ii. Establishing engineering files and importing original image data, creating engineering, adding image data, setting image attributes, and camera model parameters in the Pix 4D Mapper software;
iii. Automatic processing of the UAV images, including initialization, point cloud encryption, regional 3D reconstruction, and digital orthographic image model generation

The digital orthographic mosaic image model of the study area generated through the above processes is shown in Figure 4. It spent about 60 minutes, two flights of UAV aerial photography, completing the task of capturing remote sensing images in the entire study area. During this task, the battery onboard needed to be replaced multiple times, causing the flight to start from a different place and thus, some images of different extent areas would be captured and the study area is consist with 3600 coconut trees and the area covered was 0.3 km^2.

Figure 4. Digital mosaic orthographic image of the study area.

2) Construction of a dataset of individual coconut tree image samples: The processed mosaic orthographic images of coconut trees were imported into ArcGIS Pro 2.8, and its deep learning module was used to label individual coconut tree samples. After labelling, cropping, and exporting, an individual coconut tree detection dataset based on a UAV remote sensing image was generated and named the Individual Coconut Tree Image Samples (ICTIS) dataset. The steps are as follows. Firstly, we created a shapefile of the surface element class vectors in ArcGIS Pro, drew circle elements for the coconut sample annotations manually according to the records of field investigation, added a class file in the properties table of surface element class vector, and identified the individual coconut tree sample's category. The annotation example is shown in Figure 5a. Secondly, the polygon feature-class file was used to export the images and their corresponding annotated sample data, suitable for the subsequent research requirements. The digital orthographic image of the study area was cropped into clip images with the size of 640 × 640 and zero overlaps. The images without coconut tree annotations were excluded when exporting in ArcGIS Pro. Lastly, a dataset of ICTIS was created according to the PASCAL VOC (Everingham, van Gool, Williams, Winn & Zisserman 2010) data format by combining all exported clip images, with a total of 570 images. The label example of the clip images of the dataset obtained after cropping and exporting is shown in Figure 5b. The actual label of an individual coconut tree is the minimum bounding rectangle of the drawn circle, which will be the ground truth for model training and validation in deep learning.

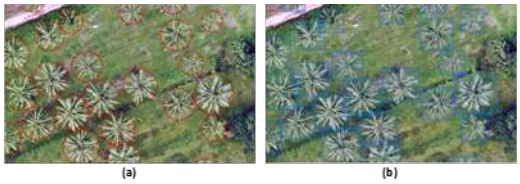

Figure 5. Annotation samples (a) in ArcMap and label samples (b) of the dataset of individual coconut tree image samples in the study area.

3) YOLOv3 deep learning object detection model: Based on the dataset ICTIS, the single-stage object detection algorithm of YOLOv3 (You Only Look Once) was empirically selected to train, validate, and test the model for individual coconut tree detection. YOLOv3 is the third version of the YOLO model family and has been widely used in object detection tasks such as pedestrians, vehicles, and ships. The YOLO model is divided into three parts: backbone network (Backbone), neck network (Neck), and head network (Head) (Figure 6). A backbone network is used to extract features from the input data; the neck network collects and distributes features of different scales; the head network is used to judge the positioning and category of the target box. In YOLOv3, the backbone network adopts a cross-stage local network (Wang et al. 2020) to solve the problem of gradient information duplication and gradient disappearance of network optimization; it adopts a path aggregation network (Liu, Qi, Qin, Shi & Jia 2018) and spatial pyramid pooling network (He, Zhang, Ren & Sun 2015). As a neck network, the model enhances the detection of objects with different scaling scales to identify the same object of different scales; the head network uses the same detection layer as YOLOv1 and YOLOv2, applies the best anchor box to the feature map, and generates the final output vector with category probability, object score, and prediction bounding box.

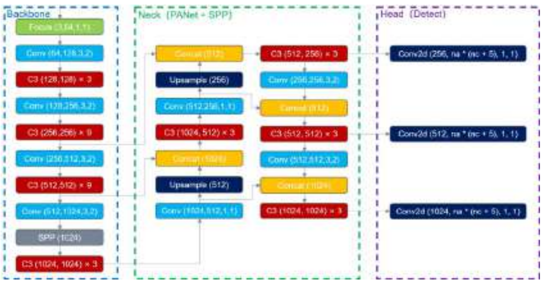

Figure 6. The YOLOv3 model structure.

The YOLOv3 model was used for training and validation. This model was named CocoNet for short. Finally, CocoNet was used to detect ICTs in the experimental area, and its spatial distribution map was made.

C. Evaluation metrics.

Model evaluation is very important in deep learning. Only by choosing an appropriate evaluation method can we quickly discover potential problems with the model in the training process and find suitable ways to optimize the model. The confusion matrix is not only a standard format for evaluating the accuracy but also a visualization tool capable of using special matrices to present the effect of model performance. The confusion matrix consists only of positive and negative examples. Table 1 shows the confusion matrix for a classic example of binary classification. Each column represents a predicted value, and each row represents an actual category.

Table 1. Confusion matrix of binary classification of artificial intelligence.

Confusion matrix		**Predicted label**	
		true	**false**
Actual label	positive	TP*	FP
	negative	TN	FN

* TP (True Positive) means that the actual category of the sample is positive, and the result predicted by the model is also positive. TN (True Negative) means that the actual category of the sample is negative, and the model predicts it to be negative. FP (False Positive) means that the actual category of the sample is negative, but the model predicts it to be positive. FN (False Negative) means that the actual category of the sample is positive, but the model predicts it as negative.

The present study is an example of a binary classification. We evaluated the accuracy of the trained YOLOv3 model using the precision, recall, F_1 score, and average precision(AP).

1) Precision and recall: According to Table 1, the precision (P) and recall (R) metrics are defined as Equations 1 and 2, respectively. Precision indicates the percentage of samples that were actually positive out of all results that were predicted to be positive samples. The recall indicates the ratio of samples predicted positive by the classifier to the actual number of positive samples. Also called sensitivity, it represents the classifier's sensitivity to the category of positive examples.

$$P = \frac{TP}{(TP+FP)} \tag{1}$$

$$R = \frac{TP}{(TP + FN)} \tag{2}$$

Where, P and R denote the precision and recall, respectively. TP, FP, and FN indicate the same meanings as in Table 1.

2) F1-score: The F1-score is the harmonic mean of precision and recall, taking both metrics into account in Equation 3.

$$F_1 = \frac{2 \times \mathrm{P} \times \mathrm{R}}{(\mathrm{P+R})} \tag{3}$$

Where, F_1 denotes the F1-score; P and R denote the precision and recall, respectively.

3) Average precision: In the domain of deep learning object recognition, Average Precision (AP) measures how well a model recognizes a particular category and is represented by a Precision-Recall Curve (PRC) plot (Figure 7). The PRC chart is a horizontal recall and vertical precision, and is a monotonically decreasing curve. The area under the PR curve for a particular category is defined as AP as defined in Equation 4.

$$AP = \int_0^1 f(c)d(c) \tag{4}$$

Where, AP is the average precision and f(c) is the precision recall curve for category c. The closer the curve is to the upper right corner in the PR plot, the more accurate the model is. In addition to using the model to estimate the area under the curve, we can also draw a line with a slope of 1 on the PRC plot and the intersection of this line with the PR curve is the equilibrium point F1. This score is known as F1 score.

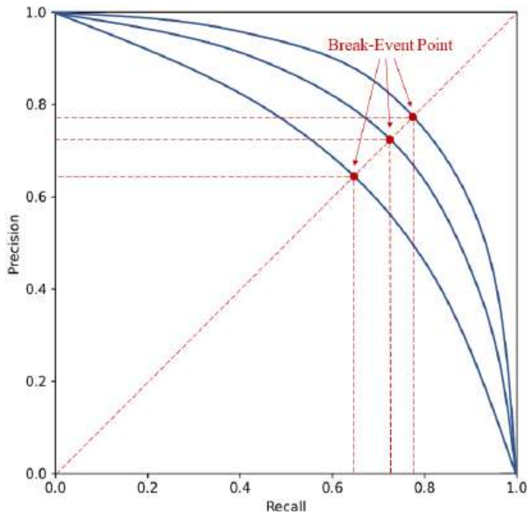

Figure 7. Precision-Recall Curves

4) Average precision: In object detection, the strength of a representation model is not only the prediction probability of categories, but also the accuracy of the positioning of prediction boxes. The Intersection over Union (IoU) ratio is commonly used as the matching degree evaluation metric for predicted bounding boxes and ground truth boxes in a data set (Figure 8), and their area intersection and intersection ratio are calculated according to the Equation 5. The higher the ratio value, the better the match. The ideal result is a perfect overlap between the prediction box and the ground truth box that achieves a ratio of 1.

$$IoU = \frac{\mathrm{Area(B)} \cap \mathrm{Area(G)}}{\mathrm{Area(B)} \cup \mathrm{Area(G)}} \tag{5}$$

Where, The area of the prediction bounding box is shown by Area(B) , and The area of the ground truth box is shown by Area(G).

Figure 8. Intersection over the union of ground truth and prediction bounding box.

The threshold criterion for positive is IoU > 0.5, otherwise negative. Therefore, AP@0.5 used below represents the average precision when IoU > 0.5 and AP@0.5:0.95 used below represents the average accuracy when IoU is between 0.5 and 0.95. Furthermore, inference time is also an important metric for evaluating the model's ability in object detection. Frames per second (FPS) is commonly used to measure model inference speed.

5) Experimental environment and setup: The experimental platform was configured with the following:

- AMD RYZEN 7 5800X CPU with 3.8GHz processor
- MSI RTX 3050 VENTUS 8GB GPU independent graphics card
- 64-bit Windows 10
- Python 3.7,
- PyTorch 1.8.1

By cross-validation, 513 (90%) images were randomly selected as the training set and 28 (5%) images were selected as the validation set. The remaining 28 (5%) images were used as a test set to test the final model.

3. Result and Discussion

A. Model Characteristics

1) Accuracy of the YOLOv3 model: Figure 9 shows the average precision (AP@0.5) variation of the YOLOv3 model during the training process from 20 to 120 epochs. With the help of the pre-trained model weights, the model can achieve high accuracy quickly. After 75 epochs of training, the AP of YOLOv3 reaches the fitting state and remains stable.

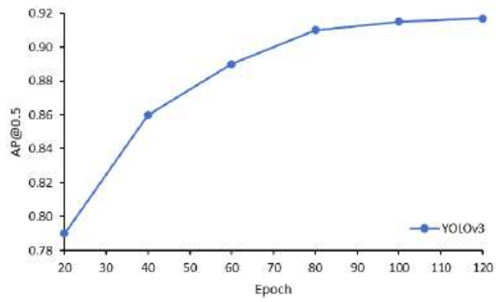

Figure 9. Average precision (AP) variation of each YOLOv3 model (20~120 epochs).

By comparing the accuracy evaluation metrics of the model in Table 2, it can be found that YOLOv3 accuracy performance is suitable for ICT detection.

Table 2. Results of the evaluation metrics of the YOLOv3 model.

A. **odel**	*B.* **recision**	*C.* **ecall**	*D.* **1 Score**	*E.* **P@0.5**
F. OLOv3	*G.* .871	*H.* .859	*I.* .867	*J.* .917

*AP@0.5 means the average precision when the Intersection over Union > 0.5

2) Training and validation loss of the YOLOv3 model: Based on the training and test loss curves in Figure 10, the model has performed well, and there is no overfitting. If the training loss value is close to the value of the validation loss, the model is not overfitting. The lower the loss, the better the accuracy of the model.

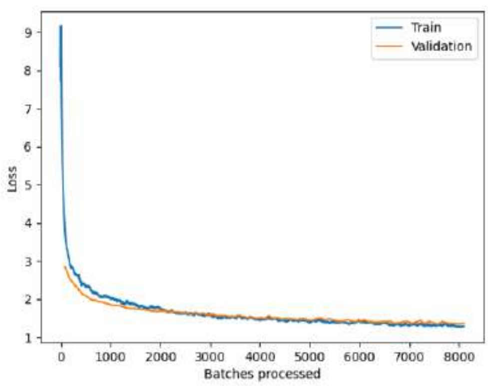

Figure 10. Training and validation losses curve of the YOLOv3 model.

B. Statistics and mapping of the individual coconut trees

A thematic map (Figure 11) was made to show the spatial distribution of the detected ICTs using the retrained YOLOv3 model. As shown in the two inserted square boxes of Figures 11b and 11c, it can be found that different sizes of ICTs are almost detected accurately, indicating that CocoNet could have a high enough accuracy to complete the task of individual coconut tree detection. It is necessary to test further and validate the feasibility of the application of CocoNet in other coconut plantations.

The original square labels detected by using CocoNet were converted into the corresponding circles to reduce the overlapping effect on the map. The planting area and the number of the detected ICTs in the experimental study area were counted with the ArcGIS Pro software and shown in the thematic map. The results show that the cultivated area is 297,156.83 m^2, and the total number of coconut trees is 3,306.

C. Limitations and future work

Despite a lot of hard work, there are some limitations in dataset creation, deep learning model selection and design, and hyperparameter optimization. First, although we acquired both RGB and multispectral images using UAV remote sensing, only the UAV-based RGB images were used to construct the ICTIS dataset in the present study. The UAV-based multispectral images will be used in future studies to improve the model's accuracy. Other UAV-based high-resolution images such as hyperspectral or LiDAR imagery would be better options for the detection of ICTs because their more spectral information or highly effective point cloud data (Hu et al. 2022; Jaskierniak et al. 2021) could reveal more detailed features and improve the performance of CNNs that helps distinguish ICTs from the images. Second, more CNN models such as Faster R-CNN, U-Net, SDD, and Mask RCNN (Yu et al. 2022; Safonova et al. 2021; dos Santos et al. 2019) should be trained and tested to obtain a better model to fulfil the task. The structure of the model could even be modified to improve accuracy and performance for better precise applications of smart plantation management. Third, data augmentation and hyperparameter optimization need to be further carried out to obtain a more robust performance model. These all deserve further research.

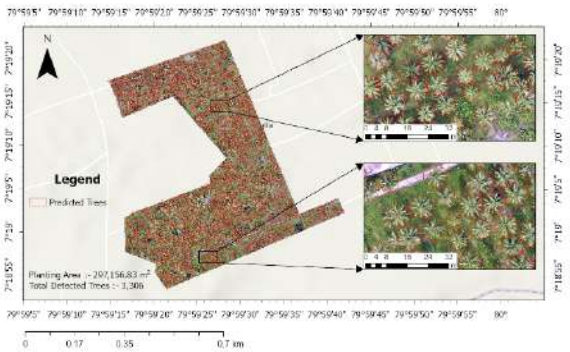

Figure 11. A thematic map shows the spatial distribution of the detected individual Coconut trees using the YOLOv3 and their planting area and number information with two enlarged square regions inserted.

In the future, our proposed approach can be used to obtain these kinds of spatial and attribute data about the individual coconut trees in a plantation. These data could be easily integrated into a smart plantation management system that could provide fast growth monitoring of individual coconut trees, accurate yield estimation of the coconut, real-time disease prevention and control, and precision cultivation and management. Town-level, county-level and city-level thematic maps of ICTs will be made through our proposed approach in the coming study. The coconut yield estimation based on the thematic map of ICTs will be an important topic in our future research.

4. Conclusion

In the present study, we proposed a deep learning approach to detecting and mapping individual coconut trees in UAV remote sensing imagery, taking the coconut trees in Mahayaya Coconut Model Garden, Sri Lanka, as an example. UAV remote sensing technology was applied to acquire high spatial-resolution images of the study area. These images were pre-processed in the Pix 4D Mapper software. A dataset of individual coconut tree image samples (ICTIS) was constructed through visual interpretation and the deep learning tools in the ArcGIS software combined with fieldwork investigation. YOLOv3 object detection model was used to train and validate the dataset. The evaluation results show that the model achieves relatively high detection accuracy. The trained YOLOv3 model, namely CocoNet, was thus selected to detect and post-process ICTs in the whole mosaic orthographic image of the study area. Finally, a spatial distribution thematic map of ICTs was made according to the detection results. This study provides reference information for related research and smart plantation management.

References

Brandt, M, Tucker, CJ, Kariryaa, A, Rasmussen, K, Abel, C, Small, J, Chave, J, Rasmussen, LV, et al. 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. *Nature*. 587(7832). doi.org/10.1038/s41586-020-2824-5.

Everingham, M, van Gool, L, Williams, CKI, Winn, J & Zisserman, A. 2010. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*. 88(2). doi.org/10.1007/s11263-009-0275-4.

Ghali, R, Akhloufi, MA & Mseddi, WS. 2022. Deep Learning and Transformer Approaches for UAV-Based Wildfire Detection and Segmentation. *Sensors*. 22(5). doi.org/10.3390/s22051977.

Girshick, R, Donahue, J, Darrell, T & Malik, J. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi.org/10.1109/CVPR.2014.81.

Hanan, NP & Anchang, JY. 2020. doi.org/10.1038/d41586-020-02830-3.

He, K, Zhang, X, Ren, S & Sun, J. 2015. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 37(9). doi.org/10.1109/TPAMI.2015.2389824.

Hu, J, Zhang, Y, Zhao, D, Yang, G, Chen, F, Zhou, C & Chen, W. 2022. A Robust Deep Learning Approach for the Quantitative Characterization and Clustering of Peach Tree Crowns Based on UAV Images. *IEEE Transactions on Geoscience and Remote Sensing*. 60. doi.org/10.1109/TGRS.2022.3142288.

Immerzeel, WW, Kraaijenbrink, PDA, Shea, JM, Shrestha, AB, Pellicciotti, F, Bierkens, MFP & de Jong, SM. 2014. High-resolution monitoring of Himalayan glacier dynamics using unmanned aerial vehicles. *Remote Sensing of Environment*. 150. doi.org/10.1016/j.rse.2014.04.025.

Jaskierniak, D, Lucieer, A, Kuczera, G, Turner, D, Lane, PNJ, Benyon, RG & Haydon, S. 2021. Individual tree detection and crown delineation from Unmanned Aircraft System (UAS) LiDAR in structurally complex mixed species eucalypt forests. *ISPRS Journal of Photogrammetry and Remote Sensing*. 171. doi.org/10.1016/j.isprsjprs.2020.10.016.

Jintasuttisak, T, Edirisinghe, E & Elbattay, A. 2022. Deep neural network based date palm tree detection in drone imagery. *Computers and Electronics in Agriculture*. 192. doi.org/10.1016/j.compag.2021.106560.

K, L, Karnick, S, Ghalib, MR, Shankar, A, Khapre, S & Tayubi, IA. 2022. A novel method for vehicle detection in high-resolution aerial remote sensing images using YOLT approach. *Multimedia Tools and Applications*. doi.org/10.1007/s11042-022-12613-9.

Khanal, S, Fulton, J & Shearer, S. 2017. doi.org/10.1016/j.compag.2017.05.001.

Li, D & Li, M. 2014. Research advance and application prospect of unmanned aerial vehicle remote sensing system. *Wuhan Daxue Xuebao (Xinxi Kexue Ban)/Geomatics and Information Science of Wuhan University*. 39(5). doi.org/10.13203/j.whugis20140045.

Liu, S, Qi, L, Qin, H, Shi, J & Jia, J. 2018. Path Aggregation Network for Instance Segmentation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. doi.org/10.1109/CVPR.2018.00913.

Osco, LP, Marcato Junior, J, Marques Ramos, AP, de Castro Jorge, LA, Fatholahi, SN, de Andrade Silva, J, Matsubara, ET, Pistori, H, et al. 2021. doi.org/10.1016/j.jag.2021.102456.

Redmon, J, Divvala, S, Girshick, R & Farhadi, A. 2016. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. V. 2016-December. doi.org/10.1109/CVPR.2016.91.

Safonova, A, Guirado, E, Maglinets, Y, Alcaraz-Segura, D & Tabik, S. 2021. Olive tree biovolume from uav multi-resolution image segmentation with mask r-cnn. *Sensors*. 21(5). doi.org/10.3390/s21051617.

dos Santos, AA, Marcato Junior, J, Araújo, MS, di Martini, DR, Tetila, EC, Siqueira, HL, Aoki, C, Eltner, A, et al. 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVS. *Sensors (Switzerland)*. 19(16). doi.org/10.3390/s19163595.

Sun, Y, Li, Z, He, H, Guo, L, Zhang, X & Xin, Q. 2022. Counting trees in a subtropical mega city using the instance segmentation method. *International Journal of Applied Earth Observation and Geoinformation*. 106. doi.org/10.1016/j.jag.2021.102662.

Wang, CY, Mark Liao, HY, Wu, YH, Chen, PY, Hsieh, JW & Yeh, IH. 2020. CSPNet: A new backbone that can enhance learning capability of CNN. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. V. 2020-June. doi.org/10.1109/CVPRW50498.2020.00203.

Wu, Y, Shan, Y, Lai, Y & Zhou, S. 2022. Method of calculating land surface temperatures based on the low-altitude UAV thermal infrared remote sensing data and the near-ground meteorological data. *Sustainable Cities and Society*. 78. doi.org/10.1016/j.scs.2021.103615.

Yu, K, Hao, Z, Post, CJ, Mikhailova, EA, Lin, L, Zhao, G, Tian, S & Liu, J. 2022. Comparison of Classical Methods and Mask R-CNN for Automatic Tree Detection and Mapping Using UAV Imagery. *Remote Sensing*. 14(2). doi.org/10.3390/rs14020295.

Zhang, L, Zhang, L & Du, B. 2016. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*. 4(2). doi.org/10.1109/MGRS.2016.2540798.