Smart Agricultural Technology 10 (2025) 100748

Contents lists available at ScienceDirect

Smart Agricultural Technology

journal homepage: www.journals.elsevier.com/smart-agricultural-technology

An efficientnet-based model for classification of oil palm, coconut and banana trees in drone images

Kwee Kim Teo [a], Nurul Fazmidar Binti Mohd Noor [a,*], Shivakumara Palaiahnakote [b], Mohamad Nizam Bin Ayub [a]

[a] Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia
[b] School of Science, Engineering and Environment (SEE), University of Salford, Manchester, UK

ARTICLE INFO

Keywords:
Image segmentation
Trees detection
Classification
Palm
Coconut
Banana
Drone images

ABSTRACT

Oil palm tree detection and classification from coconut and banana trees is vital for increasing the production of oil palm businesses globally, particularly in Malaysia. Since oil palm, coconut, and banana trees share common characteristics such as tree shape and structure, classification is challenging. Further, this work considers images captured by drones, which adds complexity to the classification problem. Unlike most existing methods that primarily detect oil palm trees, the proposed work aims to detect and classify multiple tree types. Inspired by the success of the Segment Anything Model (SAM), a generalized model for object segmentation, we adapted SAM for detecting and localizing oil palm, coconut, and banana trees in drone images. Similarly, motivated by the efficiency and effective feature extraction of EfficientNetB3, we integrated it for the classification task. The proposed model combines SAM for detection and EfficientNetB3 for classification in an end-to-end architecture. To evaluate its performance, we conducted experiments on a dataset collected from a Malaysian drone services company, featuring frames captured across diverse locations. Results demonstrate that the proposed method significantly outperforms state-of-the-art approaches. For detection, the proposed SAM achieves F1-scores of 97 %, 89 %, and 91 % for oil palm, coconut, and banana trees, respectively. For classification, the proposed model achieves F1 scores of 92 %, 88 %, and 91 % for oil palm, coconut, and banana trees, respectively. The results show that the proposed method is superior to the existing methods.

1. Introduction

According to the Malaysian Palm Oil Council [1], in 2020, Malaysia held a significant position in the global palm oil industry, contributing substantially to both production and exports, with shares of 25.8 % and 34.3 %, respectively. After Indonesia, Malaysia has become the second-largest producer of oil palm in the world. In 2021, Malaysia's oil palm plantations covered a total area of 5.74 million hectares (Mha), yielding an approximate production of 18.12 million tonnes (Mt) of palm oil [2]. Ranked as Malaysia's fourth-largest industrial crop, coconut follows oil palm, rubber, and rice. Under favorable conditions, including a suitable climate, sufficient rainfall, and a conducive environment, coconut trees can readily produce fruit [3]. According to the Ministry of Agriculture and Food Industries [4], as of 2020, the area of coconut plantations in Peninsular Malaysia was 55,573.1 hectares (ha), producing 451,693.1 metric tonnes (t) of coconuts. Meanwhile, in Sabah, Sarawak, and Labuan, coconut plantations cover 29,368.9 hectares (ha), yielding a total production of 109,291.3 metric tonnes (t).

Similarly, bananas have been recognized as a key fruit for export markets [5]. According to statistics from the Ministry of Agriculture and Food Industries [6] for 2018, the planted area of banana trees in Johor was 8304 hectares (ha), yielding 116,966 metric tonnes (t). Pahang ranked second with 4975 hectares (ha), producing 62,421 metric tonnes (t). Another state, Sabah, has banana plantations covering 3602 hectares (ha), producing 43,835 metric tonnes (t) of bananas. It is noted that Malaysia's banana exports amounted to USD 12.6 million in 2021, positioning the country as the 51st largest exporter of bananas globally. In the same year, bananas ranked as the 582nd most exported product from Malaysia [7]. For visualization, the statistics of different tree types in terms of area and production are represented in Fig. 1. In summary, the above discussion suggests that there is high demand for improving the production of the three mentioned crops. However, meeting the current demand is challenging due to unforeseen causes, including multiple diseases and unpredictable changes in nature. To alleviate this

* Corresponding author.
E-mail address: fazmidar@um.edu.my (N.F.B.M. Noor).

https://doi.org/10.1016/j.atech.2024.100748
Received 26 August 2024; Received in revised form 21 December 2024; Accepted 22 December 2024
Available online 27 December 2024

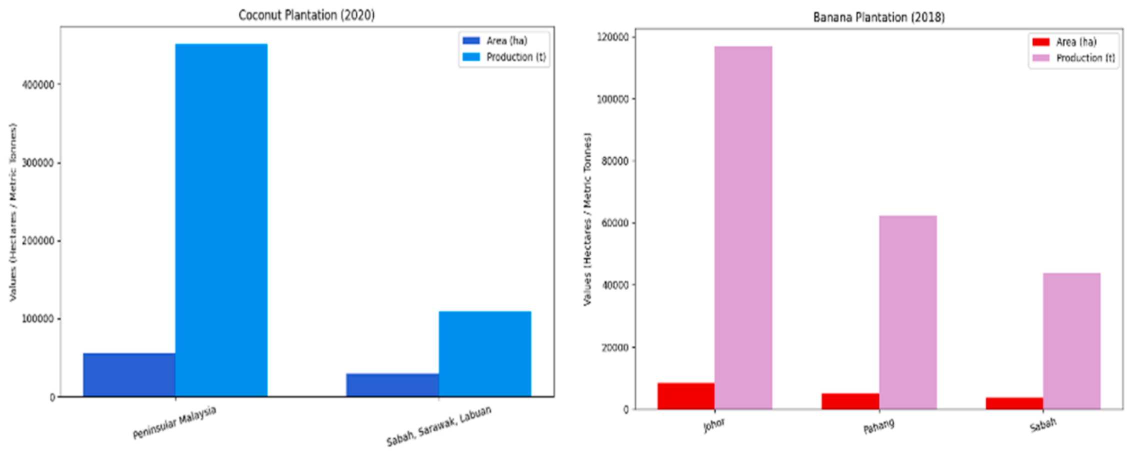


Fig. 1. Statistics for coconut and banana plantations in terms of total area and production in 2020 and 2018 respectively.

(a) Multiple palm trees (b) Multiple coconut trees (c) Mixed banana and palm trees

Fig. 2. Detecting trees by the existing method [8].

(a) Multiple palm trees (b) Multiple coconut trees (c) Mixed banana and palm trees

Fig. 3. Detecting trees irrespective of different types by the proposed method.

problem, it is necessary to identify healthy and unhealthy trees, which in turn requires the detection and classification of oil palm, coconut, and banana trees to determine the reasons for declining production.

Models have been developed in the past for detecting oil palm trees and counting them from drone images [8–10]. However, these models are limited to detection and counting and do not include classification. The complexity of detection and classification increases, especially when trees of similar species coexist within the same plantation area. Additionally, the inherent visual similarities shared by these trees, when observed from an aerial view, make the problem of detection and classification even more complex and challenging. This is evident from the sample images presented in Fig. 2 and Fig. 3, where one can see crowded oil palm trees, coconut trees, and mixed tree species, and where the state-of-the-art method [8] fails to detect all the trees in all three images properly. Although the existing method [8] was developed for palm tree counting in remotely sensed images, it does not perform well with the images shown in Fig. 2. On the other hand, the proposed model performs better than the existing method as shown in Fig. 3. When detection fails, it automatically affects classification performance. Thus, one can conclude that the detection and classification of oil palm, coconut, and banana trees remain an open challenge.

Hence, this work proposes a new unified model that integrates a model called the Segment Anything Model (SAM) for detection and EfficientNetB3 for classification. The motivation for exploring SAM is that this model has strong generalization ability and was developed using an advanced language model for segmenting different objects, animals, and people [11]. This work adapts the existing SAM to detect trees regardless of type. Similarly, EfficientNet is an efficient model for effective feature extraction [12], and we adapt EfficientNetB3 for classification. Overall, the proposed work intelligently fuses both models to achieve the best results for detection and classification. The contributions are threefold: (i) Exploring the well-known segmentation model (SAM) for detecting oil palm, coconut, and banana trees; (ii) Deploying EfficientNetB3 in a novel way for the classification of oil palm, coconut, and banana trees; (iii) The way the proposed work fuses these models into an end-to-end solution for successful classification.

2. Related work

As mentioned in the previous section, there are not many methods for classification of trees. Therefore, we review the related methods of tree counting and detection and classification.

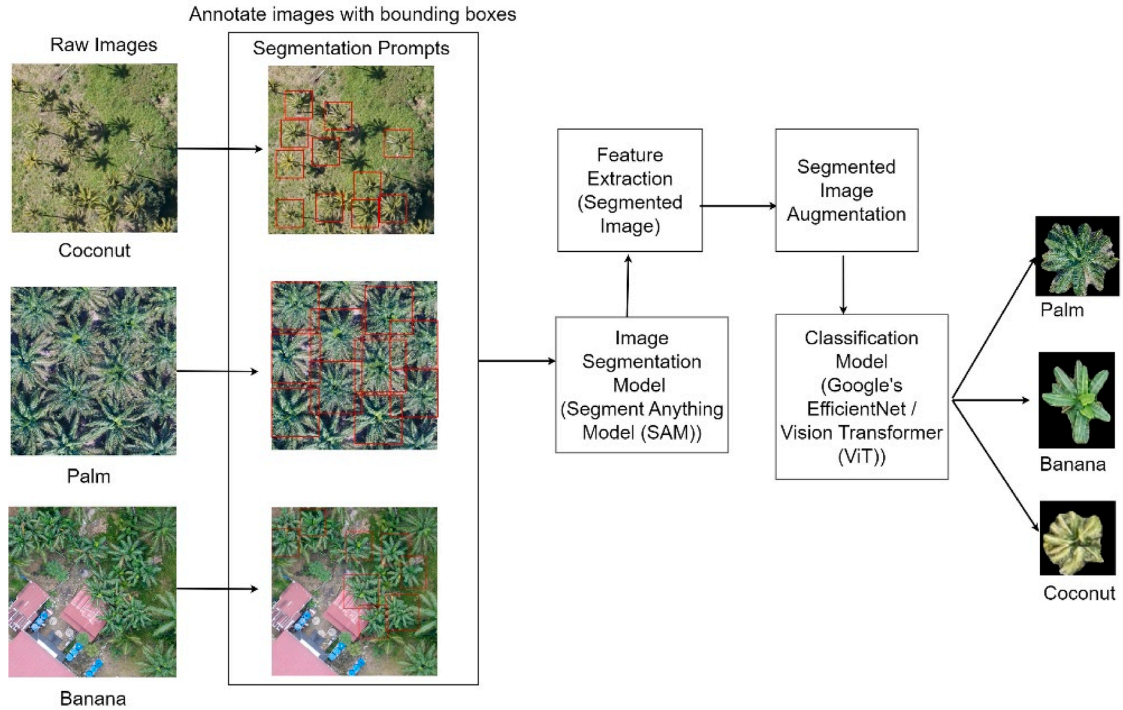


Fig. 4. End-to-End Proposed model for classification of palm, coconut and banana trees.

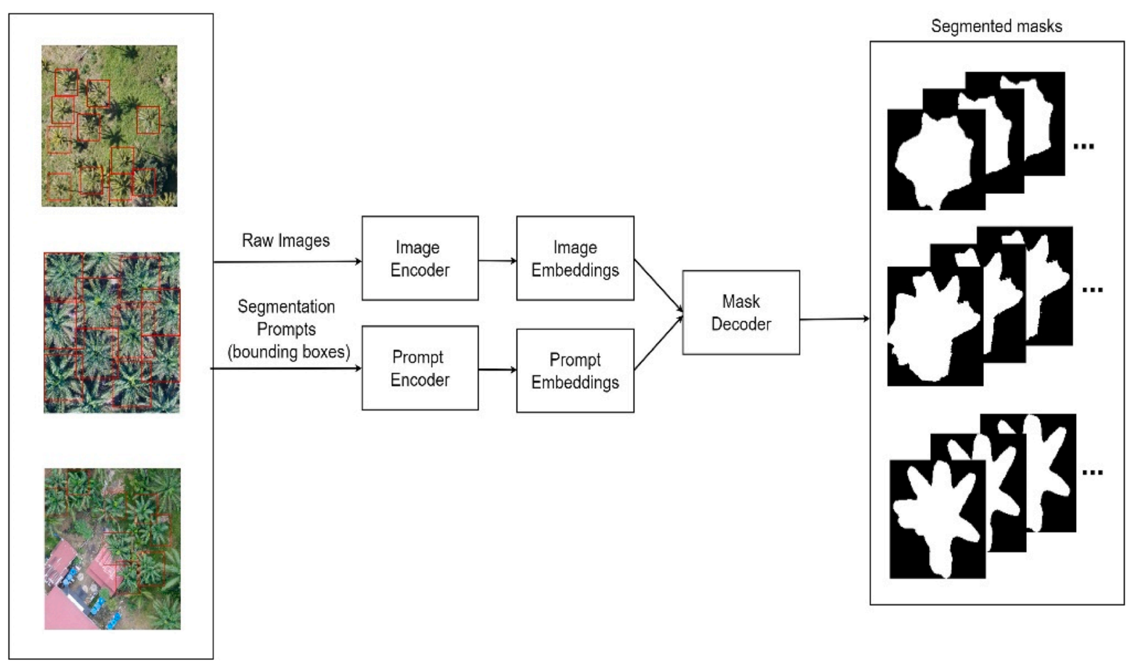


Fig. 5. The proposed SAM architeture for detecting trees.

In 2018, the research [13] focused on the creation and evaluation of an automated algorithm utilizing deep learning principles to construct an inventory of individual oil-palm trees. The methodology involves the integration of outcomes from two separate convolutional neural networks (CNNs) for achieving the best results. However, the key limitation is its reliance on orthomosaics derived from RGB imagery collected by UAVs with a pixel spacing of 10–20 cm/pixel. This limitation may affect the generalizability. The method [14] introduced an interesting application of deep learning in individual palm tree inventory using high-resolution remote sensing imagery. The research advocates for the implementation of the RetinaNet, a cutting-edge deep learning detection method to establish a model capable of classifying and pinpointing palm trees within the study area through RGB imagery. However, the method is fine tuned on particular dataset to improve the performance. The authors [8] focus on the automatic detection and counting of oil palm trees using Microsoft Bing Maps Very High Resolution (VHR) satellite imagery and Unmanned Aerial Vehicle (UAV) image data. This data is processed through a combination of semi-automatic image processing and deep learning techniques. However, the primary limitation of this method is the need for extensive manual data labeling.

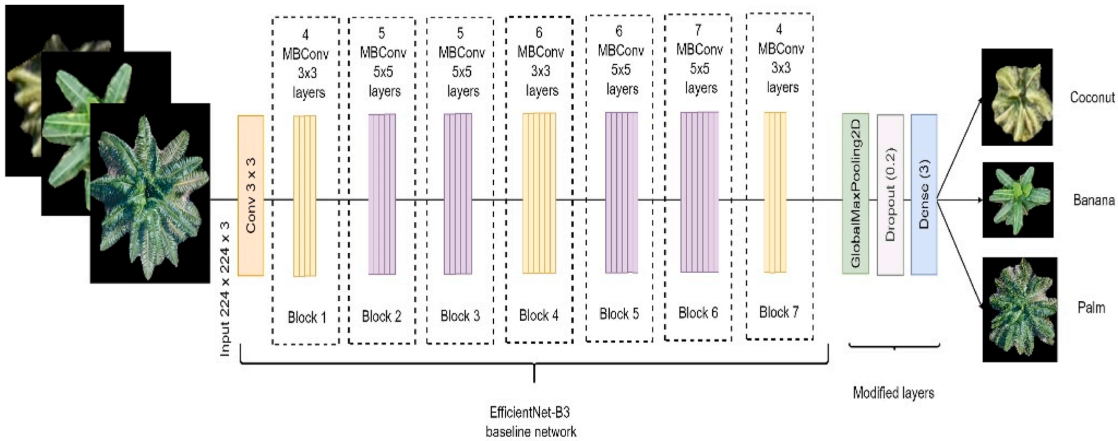


Fig. 6. The general architecture of EfficientNetB3 consists of 7 blocks, each containing a variable number of modules.

The authors [15] conducted a study utilizing high-resolution satellite images to identify oil palm trees across a substantial area in Malaysia. The authors gathered 20,000 sample coordinates, dividing them randomly into 17,000 for training and 3000 for validation. Employing a two-stage convolutional neural network (TS-CNN), the method incorporated a region proposal network and a land-cover classification network based on the AlexNet architecture. The method works well for simple images but not images containing crowded trees. In 2020, the researchers [16] developed deep learning techniques, specifically the Faster RCNN approach, for detecting oil palm trees from UAV images. It is noted that the peformance of the method depends on a large amount of labelled data. The authors [17] proposed a combination of image processing techniques and deep learning algorithms to detect and count banana plants in a farm. However, the limitation of this method is that it only focuses on detecting and counting banana plants. According to the work [18], used a deep learning-based system for automated palm tree detection and geolocation from UAV images. The models used Faster R-CNN, YOLOv3, YOLOv4, and EfficientDet for tree detection. The study [19] explores the use of deep learning techniques, specifically employing Masked Region-based Convolutional Neural Networks (Mask R-CNN) with ResNet50 and ResNet101 architectures, for the identification and segmentation of coconut trees in aerial imagery. The approach requires high-resolution aerial imagery, which may not be available in all areas.

In summary, it is noted from the review that most methods focused detection but not classification. In addition, none of the methods aimed at developing single model for detection and classification of three different types of trees, namely, oil palm, coconut and banana trees. Therefore, this work aims to develop a model for detection and classification.

3. Proposed methodology

As discussed in the previous section, the classification of oil palm, coconut, and banana trees is challenging. To reduce the complexity of classification, the proposed work detects tree regions irrespective of tree types by excluding the background information. The detection step is based on the observation that oil palm, coconut, and banana trees share common characteristics, namely shape and structure. Therefore, to detect the trees irrespective of type, we adapt the Segment Anything Model (SAM) in this work, which outputs the tree regions. Although the structure of oil palm, coconut, and banana trees is almost the same, the leaves and branches of the crowns of each tree type differ. This observation motivated us to explore EfficientNetB3 for classification by feeding the output of detection as input. The steps of the proposed work are shown in Fig. 4, where one can see an end-to-end model for detection and classification.

3.1. *Segment Anything Model (SAM) for localizing trees regions*

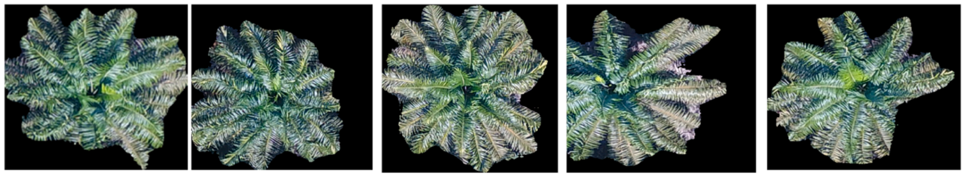
SAM is a promptable segmentation model developed by the FAIR team at Meta AI [11], as shown in Fig. 5. It demonstrates the steps of SAM for detecting trees. The localization prompts (e.g., bounding boxes) are encoded and projected into the embedding space as visual features. The bounding boxes help specify the regions within the image where trees are located, enabling the model to focus on these specific areas for further processing. Given the bounding boxes, SAM localizes the tree regions from the background. This localization is crucial as it simplifies the image by removing irrelevant details and focusing on the tree crown structures. The mask decoder is responsible for generating localization masks based on the encoded features and localization prompts. It consists of several lightweight convolutional layers and a mask decoder head. The final output is a set of localization masks for the objects of interest, such as coconut, banana, and palm trees, based on the provided prompts. This localization process forms the foundation for subsequent classification, enabling the extraction of meaningful features unique to each tree species.

3.2. *Classification of palm, coconut and banana trees*

The localized tree regions provided by the step presented in the previous section serve as input for classification. As mentioned in the proposed methodology section, the arrangement of leaves and branches in the crowns of oil palm, coconut, and banana trees is unique. This observation leads to the proposal of EfficientNetB3 for classification. In the process of developing a classification model, the proposed deep learning algorithm is Google's EfficientNet as shown in Fig. 6.

Google EfficientNet is a noteworthy family of convolutional neural network (CNN) models specifically designed to excel in image classification tasks by achieving a harmonious balance between accuracy and efficiency. EfficientNets introduce a novel scaling approach through a compound coefficient, which uniformly scales all dimensions of depth, width, and resolution [12]. EfficientNets are built upon a baseline network, which is a mobile-sized CNN model called MobileNet. This baseline network serves as the starting point for scaling. EfficientNets use MBConv [12] as the building block for their convolutional layers. This work involves transfer learning using a pre-trained EfficientNetB3 model, originally trained on the ImageNet benchmark dataset.

To use EfficientNetB3 for tree classification with segmented images obtained from the Segment Anything Model (SAM), it is necessary to preprocess the localized tree images to match the input size of the chosen EfficientNetB3 model. To reduce the spatial dimensions of the feature maps to a single vector, a GlobalMaxPooling2D layer is added on top of the EfficientNetB3 backbone, which can then be fed into the subsequent fully connected layers. After the GlobalMaxPooling2D layer,



(a) Sample successfully classified palm trees

(b) Sample successfully classified coconut trees

(c) Sample successfully classified banana trees

Fig. 7. Classification of palm, coconut, and banana trees of the proposed method.

(a) Palm trees (b) Coconut trees (c) Banana trees (d) Mixed trees

Fig. 8. Sample images of different trees of our dataset. (a) crowded palm trees, (b) Sparsely distributed coconut treees, (c) Overlapped banana trees and (d) Mixed of palm, coconut and banana trees.

Table 1
Implementation Details of ViT.

Hyperparameters	Values
Input image size	(224, 224, 3)
Patch size	16
Number of channels	3
Embedding dimension	768
Number of encoder layers	12
Number of attention heads	12
Hidden size (MLP dimension)	3072
Dropout rate	0.1
Output layer activation function	Softmax
Epochs	30
Batch size	5
Optimizer	Adam
Initial learning rate	0.001
Loss function	Cross-entropy

a Dropout layer with a probability of 0.2 is added to the network to enhance the model's generalization ability. The final Dense layer, with 3 units (for palm, coconut, and banana), produces logits for each class. The Softmax activation function converts these logits into probabilities, allowing the model to classify the input localized tree image into one of the three tree classes. The effectiveness of the proposed EfficientNetB3 is illustrated in Fig. 7, where the model classifies the tree correctly.

Table 2
Implementation details of the Proposed EfficientNetB3.

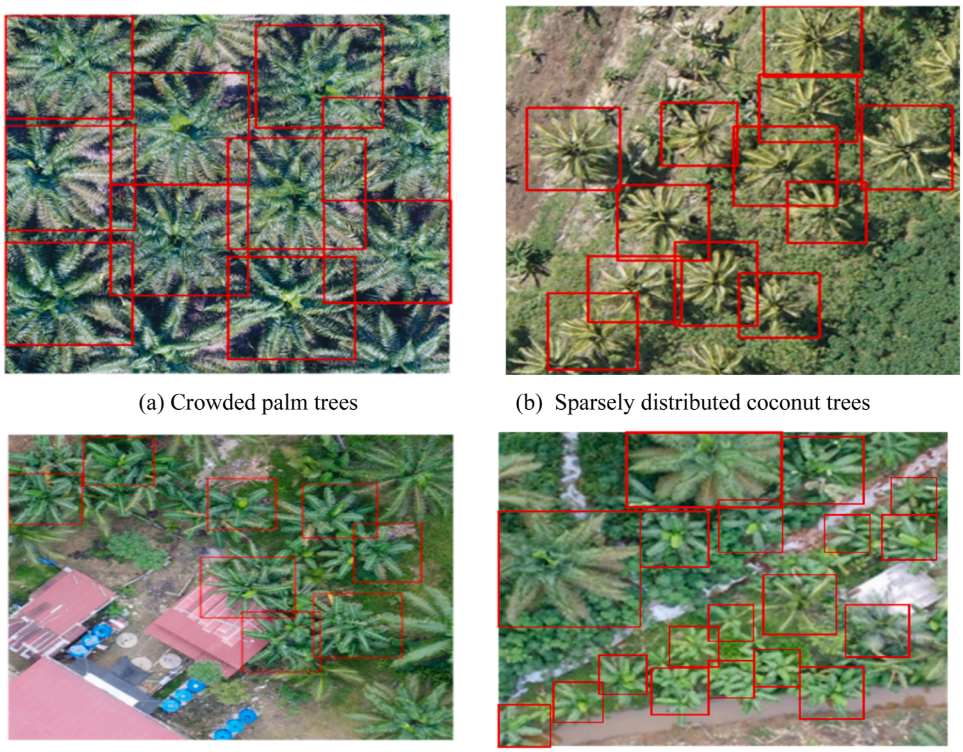
Hyperparameters	Values
Input shape	(224, 224, 3)
Dropout rate	0.2
Output layer activation function	Softmax
Epochs	30
Batch size	4
Optimizer	RMSprop
Initial learning rate	2×10^{-5}
Learning rate decay factor	0.9
Loss function	Categorical cross-entropy

4. Experimental results

For experimentation, we constructed our own dataset with the assistance of a drone services company. To evaluate the performance of the proposed method, we conducted an ablation study, along with detection and classification experiments. To demonstrate the effectiveness of the proposed method and its superiority over state-of-the-art techniques, we implemented ViT and MobileNet V2 for classification and YOLOv3 for detection in a comparative study. ViT utilizes a transformer-based architecture, treating image patches as tokens to effectively capture long-range dependencies and global features. In contrast, MobileNet V2 [21] is a lightweight convolutional neural

Table 3
Validating the performance of different variants of the Google EfficientNet in terms of classification.

Architectures	Palm tree				Coconut tree				Banana tree			
	P	R	F	Accuracy	P	R	F	Accuracy	P	R	F	Accuracy
B0	0.84	0.80	0.84	0.88	0.76	0.741	0.76	0.84	0.83	0.80	0.81	0.88
B1	0.84	0.86	0.85	0.90	0.79	0.81	0.81	0.86	0.84	0.82	0.83	0.89
B2	0.88	0.92	0.90	0.93	0.85	0.84	0.85	0.89	0.87	0.84	0.85	0.90
B3	**0.889**	**0.96**	**0.92**	**0.94**	**0.87**	**0.89**	**0.88**	**0.92**	**0.95**	**0.88**	**0.91**	**0.94**



(a) Crowded palm trees

(b) Sparsely distributed coconut trees

(c) Overlapped banana trees and

(d) Mixed of palm, coconut and banana trees.

Fig. 9. Detecting trees irrespective of types and background complexities by the proposed method.

network (CNN) designed for mobile and embedded devices. It employs depthwise separable convolutions and an inverted residual structure, striking a balance between efficiency and performance. For the detection task, we employed an object-based deep learning method with YOLOv3 [8], which is known for its real-time object detection capabilities. YOLOv3 uses a single neural network to predict bounding boxes and class probabilities, making it efficient for detecting and localizing objects in images.

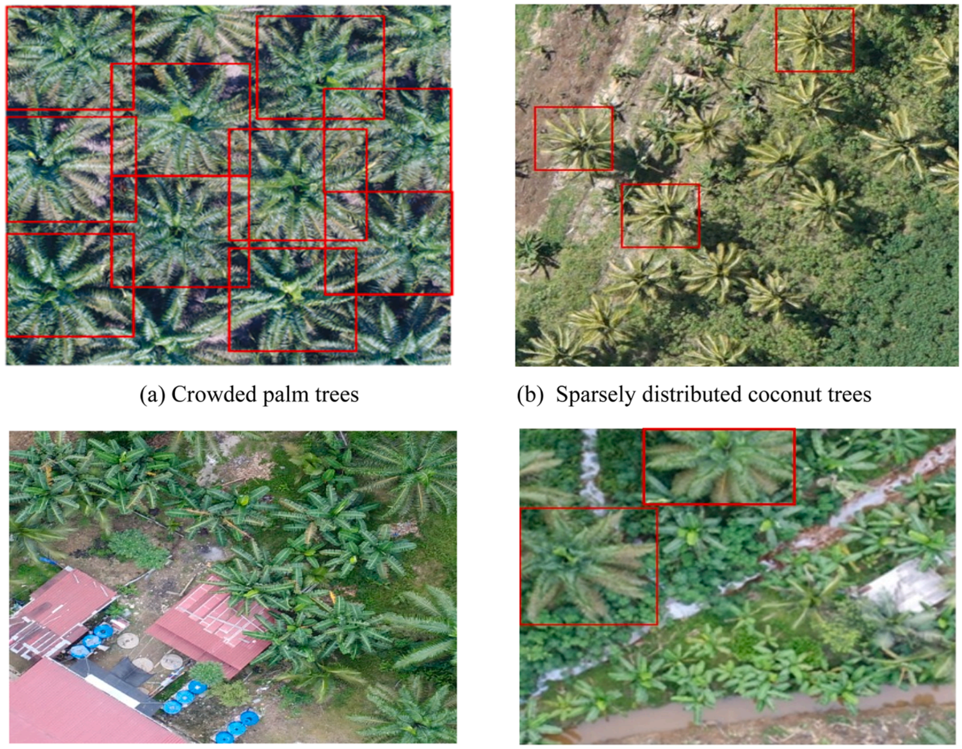
4.1. Dataset and evaluation

Sample images captured by drones are shown in Fig. 8, where it can be seen that each sample has its complexities in detecting trees and classification. The first image shown in Fig.8(a) represents crowded oil palm trees, which are dense so it is not so easy to separate trees. The second image shown in Fig. 8(b) contains coconut trees which are all scattered with other plants as a background. The third image in Fig. 8(c) shows the banana tree with the house and other objects as background. The fourth image in Fig. 8(d) includes oil palm, coconut, and banana trees with complex backgrounds. Overall, the collected data represent all possible cases of real situations in Malaysia. In addition, external factors such as open environment, day and evening make the dataset more complex.

Since the data is collected from the company for experimentation, the details of the drone and the payload are not available to the public. This is part of the agreement and copyright between our work and the company. In general, as per our knowledge, the standard drone with a standard configuration is used for capturing videos. The videos are captured at different times of the day.

The data collection occurred in the afternoon, within the timeframe of 10AM to 6PM. The dataset is carefully curated to include a diverse representation of banana, coconut, and palm trees in different lighting conditions and environmental settings. It encompasses both natural habitats and agricultural plantations to reflect real-world scenarios accurately. Recognizing the importance of a balanced dataset for effective classification, a deliberate data augmentation strategy was implemented to augment the available images and achieve a more equitable distribution among the tree species. The augmented dataset for each species was expanded to 450 images. The training set for each species comprised 315 images, while the validation set included 135 images, and 25 original images were reserved as testing set, ensuring sufficient data for training, validation, and testing for the classification model. The dataset sourced from the drone services company forms a critical foundation for the successful implementation and validation of our localization and classification methodologies for banana, coconut, and palm trees classification.

For measuring the performance of the proposed method, we use standard metrics as defined in Eqs. (1)-(4). For trees localization and



(a) Crowded palm trees

(b) Sparsely distributed coconut trees

(c) Overlapped banana trees and

(d) Mixed of palm, coconut and banana trees.

Fig. 10. Detecting trees irrespective of types and background complexities trees by the existing method.

Table 4
Detection performance of the proposed and existing methods.

Methods	Palm tree			Coconut tree			Banana tree		
	P	R	F	P	R	F	P	R	F
Proposed SAM	**0.98**	**0.97**	**0.97**	**0.89**	**0.88**	**0.89**	**0.92**	**0.92**	**0.91**
Putra et al. [8]	0.90	0.95	0.93	0.75	0.77	0.76	0.78	0.82	0.80

Table 5
Classification performance of the proposed and existing models.

Methods	Palm tree				Coconut tree				Banana tree			
	P	R	F	Acc	P	R	F	Acc	P	R	F	Acc
Proposed Method	0.88	**0.96**	**0.92**	**0.94**	**0.87**	**0.89**	**0.88**	**0.92**	**0.95**	**0.88**	**0.91**	**0.94**
ViT	**0.91**	0.92	**0.92**	**0.94**	0.83	0.84	0.83	0.89	0.87	0.88	0.87	0.92
Carvalho et al. [21]	0.85	0.90	0.87	0.92	0.83	0.84	0.83	0.82	0.82	0.85	0.83	0.88

classification, Accuracy (1), Recall (2), Precision (3) and F-Score (4) are used.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 - Score = 2 * \frac{(Precision\ *\ Recall)}{(Precision\ +\ Recalll)} \tag{4}$$

To demonstrate the superiority of the proposed method over state-of-the-art approaches, we implemented the method from [8], which was designed for counting oil palm trees in remote sensing images, to compare its detection performance with the proposed method. Since the method [8] uses deep learning for oil palm tree counting as the proposed work, it is relevant to compare it with the proposed method to show that the scope is limited to oil palm tree counting. For classification, we implemented the ViT [20] as it is popular like Efficient for visual feature extraction and the method [21] which was developed for tree species classification. The approach [21] uses the MobileNetwork for classification and it can classify different trees. This is the reason to choose the method for comparing with the proposed method. The implementation details of the existing ViT and the proposed EfficientNetB3 are presented in Table 1 and Table 2, respectively. These details can be used for reprodudcibiltiy of the propsoed method for classfication.

4.2. Ablation study

In the proposed methodology section, the work uses the

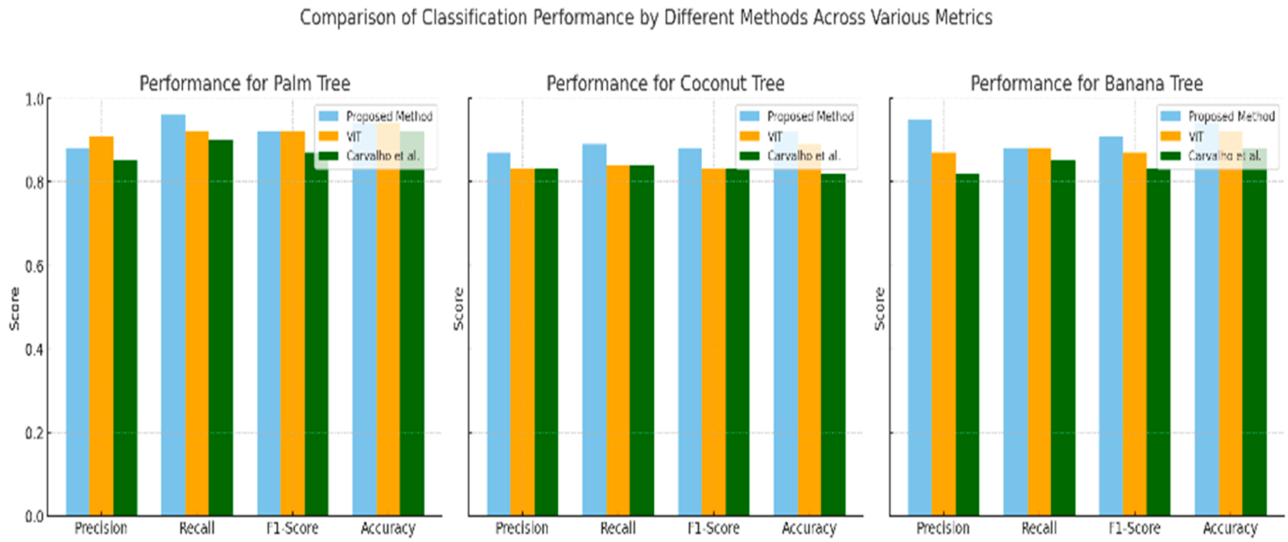


Fig. 11. Comparison of classification performance by different methods across various metrics.

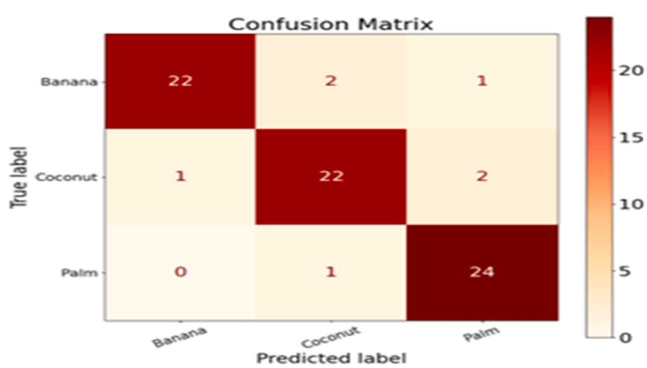


Fig. 12. Confusion matrix for the proposed method.

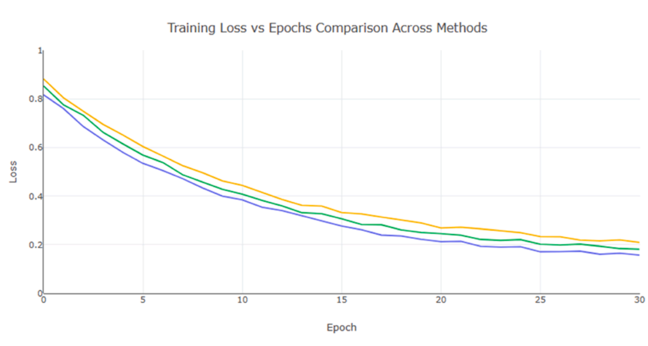


Fig. 13. Training loss vs epochs comparison accros methods.

EfficientNetB3 variant to classify oil palm, coconut, and banana trees. However, it is noted that there are other variants of an EfficientNet, namely, B0, B1, and B2. To show the effectiveness of B3, we conducted experiments on our dataset for B0, B1, and B2 and calculated the measures for all three categories of trees in Table 3. It is observed from Table 3 that the B3 is the best for all three categories in terms of all the measures. This shows that the B3 is better than B0, B1, and B2. It is also true that B2 is better than B1 while B1 is better than B0. Therefore, as the EfficientNet model advances, the model's performance improves. Hence, we can conclude that B3 is effective compared to B0, B1, and B2 for the classification of oil palm, coconut, and banana trees.

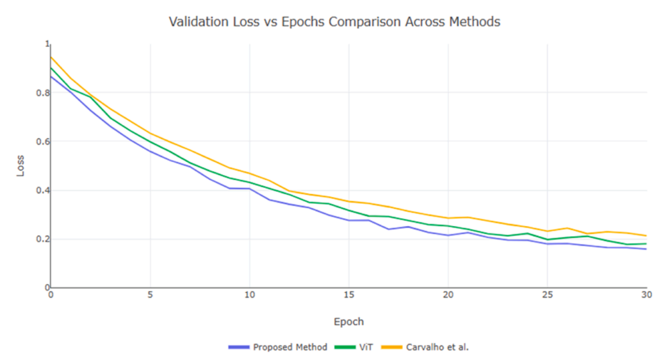


Fig. 14. Validation loss vs epochs comparison accross methods.

4.3. Experiments on detection and localization

Qualitative results for detection of tree irrespective of types and categories are shown in Fig. 9(a)-(d), where it is noted that the proposed SAM works well for all the adverse situations. This shows that the adapting SAM is capable of achieving the best detection performance. In the same way, when we look at the results of the existing method [8] for the same images as shown in Fig. 10(a)-(d), the existing method does not perform well compared to the proposed model as it misses trees in all the images.

The key reason is that since the scope of the method is limited to oil palm tree detection and counting, it detects oil palm and coconut trees, but it fails completely to detect banana trees. This indicates that the existing method [8] is not robust to different types of trees and it lacks generalization ability compared to the proposed method. On the other hand, the way the proposed work adapts the SAM makes a huge difference for multiple types of trees in different situations.

The quantitative results of both the proposed and existing methods, as presented in Table 4, lead to the same conclusions. In Table 4, the detection results exhibit outstanding performance across all three tree types compared to the existing method [8]. The model shows high precision, recall, and F1-scores, indicating its robust capability in accurately detecting these tree types. These results underscore the model's overall effectiveness, particularly excelling with palm trees while maintaining high accuracy for banana and coconut trees, making it a highly reliable for detection tasks. When we analyse the performance of the proposed method on all the three types of trees, the results are

(a) Blurred nature of the image (b) Shadows can create misleading shapes

Fig. 15. Coconut trees appear indistinct causing them to resemble palm tree.

almost consistent. This shows that the proposed model is stable and reliable. In the case of existing method [8], the results are not consistent and hence the method is not reliable.

4.4. Experimental on classification

The performance of the proposed method, ViT, and the existing methods across all three categories is presented in Table 5 and Fig. 11. It shows that the proposed method outperforms all other methods across all three categories in terms of all measures, except for the classification of oil palm trees, where the ViT method performs better. Therefore, one can argue that the proposed model is superior to the state-of-the-art method for classification of oil palm, coconut, and banana trees. The same conclusions can be drawn from Fig. 11. In addition, the confusion matrix shown in Fig. 12 shows the proposed method is consistent and stable for classification of each class. Since the Vision Transformer is also good at extracting visual features like the proposed EfficientNetB3, it performs better for oil palm tree classification, but it does not perform well for the other two categories. The reason is that trees do not provide many visual features; rather, they provide more texture features. Additionally, the oil palm tree provides a rich texture pattern compared to coconut and banana trees. So, the ViT scores the best results for oil palm trees compared to the other two trees.

However, when we consider the overall performance of the proposed method on all three categories, our method is the best. On the other hand, the existing method [21], based on MobileNetV2, is primarily designed for computational efficiency rather than extracting nuanced features from diverse tree species. While it excels in lightweight applications, its reliance on depthwise separable convolutions and linear bottlenecks makes it less capable of capturing the distinct texture and visual features necessary for the successful classification of oil palm, coconut, and banana trees. This limitation highlights the superiority of the proposed method, which demonstrates a more robust feature extraction capability tailored for this specific task. When we look at the experiments on training and validation loss vs the number of epochs as shown in Fig. 13 and Fig. 14, the proposed method is efficient in terms of training and validation compared to the existing models. The reason is the proposed EfficientNetB3 uses the lightweight MobileNet for feature extraction and classification.

Sometimes, when the images are affected by shadow, lighting and too high distance, the method may miss to extract vital information from the images as shown in Fig. 15(a) and Fig. 15(b), where we can see the crown shape visibility is obscured. In these situations, the proposed model does not perform well and hence it fails to detect correct region, which leads to misclassification. This is challenging but it is beyond the scope of the present study. However, this problem can be fixed exploring Transformer for extracting local information such that it can detect trees regions properly. If the detection steps outputs tree regions correctly, the method can classify successfully, which will be our future work.

5. Conclusion and future work

In this study, we have proposed a new end-to-end model for detection and classification of oil palm, coconut and banana trees. The model called Segment Anything Model is explored for tree region detection irrespective of types. The detected and localized regions are supplied to the proposed EfficientNetB3 for feature extraction and classification. We believe this is the first work made an attempt to detect and classify the three types of trees in drone images. The way the proposed work fuses detection and classification model is the key contribution of the proposed work compared to the existing works. The experiments on complex and diverse datasets demonstrate the effectiveness of the proposed method over state-of-the-art methods. However, when the images captured by drone lost vital information especially shapes like crown and trunk textures, the proposed model does not perform well. In this situation, the detection steps miss an important region in the images. This leads to misclassification. This can be solved by investigating Transformer based models with large computer vision models, which will be our future work.

Ethics statement

Not applicable: This manuscript does not include human or animal research.

CRediT authorship contribution statement

Kwee Kim Teo: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Nurul Fazmidar Binti Mohd Noor:** Writing – review & editing, Supervision. **Shivakumara Palaiahnakote:** Writing – review & editing, Supervision. **Mohamad Nizam Bin Ayub:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The study was significantly enhanced through the invaluable contribution of Aonic Sdn. Bhd. for sharing the dataset. We extend sincere appreciation to the company for their generosity, as this research would not have been feasible without access to their dataset. Special thanks to Lee Sheng Shiang (Patrick) at Aonic Sdn. Bhd. for his important role in facilitating data access. The dataset played a crucial role in shaping the findings of this study, fostering a deeper understanding in

classification of banana, coconut and palm trees.

Data availability

The authors do not have permission to share data.

References

[1] MPOC, Malaysian Palm Oil Industry Has A Long-Term Plan To Address EU Regulation – MPOC, Malaysian Palm Oil Council (2023). https://www.mpoc.org.my/malaysian-palm-oil-industry-has-a-long-term-plan-to-address-eu-regulation/ (accessed Sep. 29).

[2] A.P. Hj. Ghulam Kadir, "Overview of the malaysian oil palm industry 2021 foreword," MPOB. Accessed: Sep. 29, 2023. [Online]. Available: https://bepi.mpob.gov.my/images/overview/Overview2021.pdf.

[3] I.O.I. Team, "Coconut," IOI Group. https://www.ioigroup.com/our-businesses/coconut (accessed Sep. 29, 2023).

[4] Bernama, MAFI AIMS TO BOOST COCONUT PRODUCTION BY 50 PCT IN 2023, BERNAMA.com (2022). Nov. 11,, https://bernama.com/en/business/news.php?id=2137608 (accessed Sep. 29, 2023).

[5] B.C. Tan, "Volume 9 | Issue 1 | Can Banana be a Success Story for Malaysia? - FAMA," www.fama.gov.my, 2022. https://www.fama.gov.my/volume-9-issue1-can-banana-be-a-success-story-for-malaysia- (accessed Sep. 29, 2023).

[6] DOSM, "Department of statistics malaysia," ministry of economy department of statistics malaysia official portal, 2022. https://www.dosm.gov.my/portal-main/landingv2 (accessed Sep. 29, 2023).

[7] OEC, "Bananas in Malaysia," OEC - The Observatory of Economic Complexity. https://oec.world/en/profile/bilateral-product/bananas/reporter/mys (accessed Sep. 29, 2023).

[8] Y.C. Putra, A.W. Wijayanto, Automatic detection and counting of oil palm trees using remote sensing and object-based deep learning, Remote Sens. Appl. Soc. Environ. 29 (2023) 100914, https://doi.org/10.1016/j.rsase.2022.100914.

[9] P.N. Chowdhury, P. Shivakumara, L. Nandanwar, F. Samiron, U. Pal, T. Lu, Oil palm tree counting in drone images, Pattern Recognit. Lett. 153 (2022) 1–9, https://doi.org/10.1016/j.patrec.2021.11.016.

[10] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, S. Yuan, Large-scale oil palm tree detection from high-resolution remote sensing images using faster-RCNN, IEEE Xplore (2019). https://ieeexplore.ieee.org/document/8898360.

[11] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W. Lo, P. Dollár and R. Girshick, "Segment Anything," arXiv:2304.02643 [cs], 2023,: https://arxiv.org/abs/2304.02643.

[12] M. Tan and Q.V. Le, "EfficientNet: rethinking Model Scaling for Convolutional Neural Networks," arXiv:1905.11946 [cs.LG], May 28, 2019. https://arxiv.org/abs/1905.11946.

[13] M. Zortea, M. Nery, B. Ruga, L.B. Carvalho, and A.C. Bastos, "Oil-palm tree detection in aerial images combining deep learning classifiers," IEEE Xplore, 2018. https://ieeexplore.ieee.org/document/8519239.

[14] M. Culman, S. Delalieux, K.V. Tricht, Palm tree inventory from aerial images using retinanet, IEEE Xplore (2020). https://ieeexplore.ieee.org/document/9105246.

[15] W. Li, R. Dong, H. Fu, L. Yu, Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks, Remote Sens. 11 (1) (2018) 11, https://doi.org/10.3390/rs11010011.

[16] X. Liu, K.H. Ghazali, F. Han, I.I. Mohamed, Automatic detection of oil palm tree from UAV images based on the deep learning method, Appl. Artif. Intell. 35 (1) (2020) 13–24, https://doi.org/10.1080/08839514.2020.1831226.

[17] B. Neupane, T. Horanont, N.D. Hung, Deep learning based banana plant detection and counting using high-resolution red-green-blue (RGB) images collected from unmanned aerial vehicle (UAV), PLoS. One 14 (10) (2019) e0223906, https://doi.org/10.1371/journal.pone.0223906.

[18] A. Ammar, A. Koubaa, B. Benjdira, Deep-Learning-based automated palm tree counting and geolocation in large farms from aerial geotagged images, Agronomy 11 (8) (2021) 1458, https://doi.org/10.3390/agronomy11081458.

[19] M.S. Iqbal, H. Ali, S.N. Tran, T. Iqbal, Coconut trees detection and segmentation in aerial imagery using mask region-based convolution neural network, IET Comput. Vis. 15 (6) (2021), https://doi.org/10.1049/cvi2.12028.

[20] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit and N. Houlsby, "An Image is Worth 16×16 Words: transformers for image recognition at scale," arXiv:2010.11929 [cs], 2020 https://arxiv.org/abs/2010.11929.

[21] M. de A. Carvalho, J. Marcato, J.A.C. Martins, P. Zamboni, C.S. Costa, H. L. Siqueira, M.S. Araújo, D.N. Gonçalves, D.E.G. Furuya, L.P. Osco, A.P.M. Ramos, J. Li, A.A. de Castro, W.N. Gonçalves, A deep learning-based mobile application for tree species mapping in RGB images, Int. J. Appl. Earth Observat. Geoinf. 114 (2022) 103045, https://doi.org/10.1016/j.jag.2022.103045.