ISPRS Journal of Photogrammetry and Remote Sensing 173 (2021) 95–121

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Growing status observation for oil palm trees using Unmanned Aerial Vehicle (UAV) images

Juepeng Zheng [a], Haohuan Fu [a,b], Weijia Li [a,c,*], Wenzhao Wu [a,b], Le Yu [a], Shuai Yuan [d], Wai Yuk William Tao [e], Tan Kian Pang [f], Kasturi Devi Kanniah [g]

[a] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China
[b] National Supercomputing Center in Wuxi, Wuxi 214000, China
[c] CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong, Hong Kong, China
[d] Department of Eletronic Engineering, Tsinghua University, Beijing 100084, China
[e] Insight Robotics, Hong Kong, China
[f] Refinitiv Agriculture Research, 18 Science Park Drive, 118229, Singapore
[g] Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

ARTICLE INFO

Keywords:
Individual tree detection
Growing status
Oil palm
UAV images
Deep learning

ABSTRACT

For both the positive economic benefit and the negative ecological impact of the rapid expansion of oil palm plantations in tropical developing countries, it is significant to achieve accurate detection for oil palm trees in large-scale areas. Especially, growing status observation and smart oil palm plantation management enabled by such accurate detections would improve plantation planning, oil palm yield, and reduce manpower and consumption of fertilizer. Although existing studies have already reached a high accuracy in oil palm tree detection, rare attention has been paid to automated observation of each single oil palm tree's growing status. Nowadays, with its high spatial resolution and low cost, Unmanned Aerial Vehicle (UAV) has become a promising tool for monitoring the growing status of individual oil palms. However, the accuracy is still a challenging issue because of the extreme imbalance and high similarity between different classes. In this paper, we propose a Multi-class Oil PAlm Detection approach (MOPAD) to reap both accurate detection of oil palm trees and accurate monitoring of their growing status. Based on Faster RCNN, MOPAD combines a Refined Pyramid Feature (RPF) module and a hybrid class-balanced loss module to achieve satisfying observation of the growing status for individual oil palms. The former takes advantage of multi-level features to distinguish similar classes and detect small oil palms, and the latter effectively resolves the problem of extremely imbalanced samples. Moreover, we elaborately analyze the distribution of different kinds of oil palms, and propose a practical workflow for detecting oil palm vacancy. We evaluate MOPAD using three large-scale UAV images photographed in two sites in Indonesia (denoted by Site 1 and Site 2), containing 363,877 oil palms of five categories: healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms. Our proposed MOPAD achieves an F1-score of 87.91% (Site 1) and 99.04% (Site 2) for overall oil palm tree detection, and outperforms other state-of-the-art object detection methods by a remarkable margin of 10.37–17.09% and 8.14%-21.32% with respect to the average F1-score for multi-class oil palm detection in Site 1 and Site 2, respectively. Our method demonstrates excellent potential for individual oil palm tree detection and observation of growing status from UAV images, leading to more precise and efficient management of oil palm plantations.

1. Introduction

In recent decades, oil palm plantations have expanded rapidly, particularly in the main oil palm producing countries such as Indonesia and Malaysia, which are currently two leading palm oil producers around the world (Quezada et al., 2019; Taheripour et al., 2019). Although oil palm promotes economic development and alleviates poverty for many tropical developing countries, the substitution of the

* Corresponding author at: Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China.
E-mail address: weijiali@cuhk.edu.hk (W. Li).

https://doi.org/10.1016/j.isprsjprs.2021.01.008
Received 4 August 2020; Received in revised form 31 December 2020; Accepted 5 January 2021
Available online 18 January 2021

forest to oil palm plantation has become more and more widespread in recent years, resulting in a series of negative environmental impacts. The deforestation for oil palm cultivation has drawn attention to many governmental and non-governmental institutions because of its detrimental effects on native survival species, tropical rain forests, and biodiversity, etc. (Koh, & Wilcove, 2007; Santika et al., 2017).

Oil palm tree counting and detection not only benefit oil palm industries and smallholders in yield prediction, irrigation and fertilization planning, and smart management, but also contributes to the protection of native species and ecological environment due to more reasonable oil palm plantation planning. However, only detecting the tree crown is not enough for smart oil palm plantation management (Li et al., 2017; Freudenberg et al., 2019; Santos et al.). Monitoring the growing status and disturbances (e.g., pest, disease, and mismanagement) of oil palm trees is crucial for both governments and industries (Shafri & Hamdan, 2009; Shafri et al., 2011; Mubin et al. 2019). For example, farmers can replant dead palms, weed the fields to provide a better growing environment for mismanaged palms, and duly observe the health conditions of oil palms, preventing pests and diseases from spreading the whole plantation. In addition, the precise and appropriate management of oil palm plantation is not only conducive to the improvement of the palm oil production under limited plantation area, but also benefit to achieving the production goal without extensively expanding the plantation area. To this end, the observation of oil palm's health conditions contributes to alleviating the expansion of oil palm and the threat to tropical forests and local species.

During the last two decades, satellite remote sensing technique has been adopted in many kinds of applications for oil palm plantation monitoring and management (Chong et al., 2017), such as oil palm plantation mapping (Cheng et al., 2016; Dong et al., 2019a), oil palm tree detection (Li et al., 2017; Freudenberg et al., 2019), age, carbon and yield estimation (Kanniah et al., 2012; Balasundram et al., 2013; Tan et al., 2013; Cracknell et al., 2015), change detection (Pittman et al., 2013; Cheng et al., 2019), and so on. Nowadays, with the rapid development of remote sensing techniques, it is potential to automatically and accurately detect oil palm trees from high-resolution remotely sensed images instead of time-consuming, labor-exhausting and costly manual work.

In recent years, the Unmanned Aerial Vehicle (UAV) technique has demonstrated the potentials for more fine-grained recognition and analysis, owing to its high resolution, lightweight and low cost. UAV images have been widely applied to many domains (Colomina & Molina, 2014; Wu et al., 2019; Jiang et al., 2020), including environment analysis (Bhardwaj et al., 2016; Tan et al., 2018; Zhu et al., 2019; Tu et al., 2020), precise agriculture (Zhou et al., 2017; Deng et al., 2018; Maimaitijiang et al., 2020; Bayraktar et al., 2020), forest monitoring (Liu et al., 2018; Hyyppä et al., 2020), wildlife protection (Rey et al., 2017; Kellenberger et al., 2018; Kellenberger et al., 2019), etc. Furthermore, the UAV technique is also advantageous for tree and crop-related applications, such as tree crown delineation and detection (Wallace et al., 2014; Alexander et al., 2018; Osco et al., 2020), height estimation, biomass and volume estimation (Maimaitijiang et al., 2019; Fawcett et al., 2019; Ye et al., 2019), and disease detection (Dash et al., 2017).

Nevertheless, it is still very challenging to recognize and monitor the growing status of individual trees from UAV images. The reasons leading to this are manifold. First, the number of samples for abnormal trees (e.g., dead, mismanaged, or yellowish trees) are quite scarce compared with those of healthy trees, so that we cannot acquire enough prior knowledge to recognize them. Second, the appearance and characteristics of yellowish trees or mismanaged trees are incredibly similar to those of healthy trees, resulting in significant difficulties for distinguishing these types. Third, although UAV images have very high spatial resolution, detecting smallish trees is still highly intractable. Although some recent studies roughly distinguished different health conditions (ranging from excellent to poor) of individual trees (Johansen et al., 2020), to the best of our knowledge, the multi-class monitoring of growing status for individual trees, i.e., dead, mismanaged, smallish, yellowish and healthy trees, has never been studied until now.

In this study, we focus on accurately detecting the oil palm trees and distinguishing the multi-class growing status of oil palm trees using UAV images. We propose an individual oil palm tree detection method, which achieves high accuracy for not only oil palm tree detection, but also the multi-class monitoring of growing status for each oil palm with extremely imbalanced training samples. Our main contributions include the following three respects:

(1) We collect a multi-class oil palm tree dataset from two UAV images with areas of 28.85 km^2 in total. The dataset includes nearly 300,000 oil palms with five classes of growing status: dead oil palm, healthy oil palm, mismanaged oil palm, smallish oil palm, and yellowish oil palm. The datasets and our codes are available on https://github.com/rs-dl/MOPAD.
(2) We solve the sample imbalance problem through a Class-Balanced Cross-Entropy Loss (CBCEL) and a Class-Balanced Smooth L1 Loss (CBSLL). CBCEL and CBSLL effectively re-weight the classification and regression loss modules according to the Effective Number (EN) for each class, which is able to attain notable accuracy increase on our multi-class oil palm detection dataset with extremely imbalanced samples.
(3) We propose a multi-class, high-accuracy and real-time oil palm tree detection method for UAV images, named Multi-class Oil PAlm Detection (MOPAD). On the basis of Faster RCNN, we integrate a Refined Pyramid Feature (RPF) module and a hybrid class-balanced loss module to achieve satisfying multi-class monitoring of growing status for individual oil palms. Meanwhile, we propose a practical workflow for detecting oil palm vacancy based on our detection results, which provides useful information for farmers or companies to replant the oil palms.

The remainder of this paper is organized as follows. We introduce the related work for tree crown detection in the next section, and present our study area and dataset in Section 3. Following that, we introduce our proposed MOPAD in detail in Section 4, and compare the oil palm tree detection results of our MOPAD with other object detection approaches in Section 5. In Section 6, we introduce two smart management applications for oil palm plantation, including multi-class oil palm tree distribution analysis and detecting oil palm vacancies for replanting, followed by comprehensive discussions in Section 7. At last, we conclude our paper in Section 8.

2. Related works

Existing tree crown detection algorithm can be grouped into three types, including classical image processing methods (Wulder et al., 2000; Pouliot et al., 2002; Ardila et al., 2012; Chemura et al., 2015; Gomes et al., 2018), traditional machine learning methods (Hung et al., 2012; Pu and Landry, 2012; Dalponte et al., 2014; Wang et al., 2019a, 2019b; Johansen et al., 2020), and deep learning methods (Li et al., 2017; Csillik et al., 2018; Freudenberg et al., 2019; Osco et al., 2020; Puttemans et al., 2018; Xia et al., 2019).

2.1. Classical tree crown detection

The classical tree crown detection methods comprise of local maximum filter (Wulder et al., 2000; Pouliot et al., 2002; Vastaranta et al., 2012; Khosravipour et al., 2015; Wang et al., 2016; Li et al., 2019b), template matching (Ke and Quackenbush, 2011; Murray et al., 2019), image binarization (Pitkänen, 2001; Daliakopoulos et al., 2009) and image segmentation (Ferraz et al., 2016; Weinmann et al., 2016; Qin et al., 2014; Wagner et al., 2018; Aval et al., 2018), etc. Classical image

processing methods are designed by the characteristic of morphology for tree crowns (Gomes et al., 2018; Campbell et al., 2020). In most cases, these methods can be completed using existing tools or software (e.g., eCognition and ArcGIS) without time-consuming and labor-exhausting ground truths collections (except for the template matching method) (Ardila et al., 2012; Zhang et al., 2014; Mongus & Žalik, 2015). On the other hand, the detection performance of classical image processing methods deteriorates seriously where tree crowns are crowded or overlapping. Furthermore, it generally requires manual pre-selection of the suitable tree plantation areas before conducting the automatic detection of tree crowns.

2.2. Traditional machine learning for tree crown detection

Traditional machine learning for tree crown detection commonly requires two major procedures: (1) feature extraction, (2) classifier training and prediction. Feature extraction is the first step of traditional machine learning based tree crown detection methods, which is quite crucial for classifier training and prediction. According to literature review, feature extraction in traditional machine learning methods usually adopts some handcrafted approaches, such as PCA (Solano et al., 2019), HOG (Wang et al., 2019a, 2019b) and SIFT (Malek et al., 2014). The extracted features include a variety of types, such as spectral indexes (Johansen et al., 2020), shadow information (Özcan & Ünsalan, 2020) and height information (Windrim et al., 2020), etc. Most traditional machine learning methods require annotations and supervised classifiers for detecting tree crowns, including decision trees (Pu and Landry, 2012; Yao and Wei, 2013), Random Forest (RF) (Blomley et al., 2017; Weinmann et al., 2017), Support Vector Machine (SVM) (Ardila et al., 2011; Hung et al., 2012; Dalponte et al., 2014; Dai et al., 2018), and Artificial Neural Network (ANN) (Nevalainen et al., 2017), etc. Some researchers also adopt unsupervised machine learning algorithms (i.e., K-means) for individual tree detection (Ma et al., 2020). Compared with classical tree crown detection methods, traditional machine learning based tree crown methods make a great improvement in more complex study regions, with higher recall, precision, overall accuracy and F1-score. However, some feature extraction strategies are vastly complicated, and the machine learning classifiers (i.e., RF, SVM, etc.) have particular demands for image datasets, due to their limited capacity of generalization.

2.3. Deep learning for tree crown detection

Following the success of deep learning, many tree crown detection algorithms utilized convolutional neural network (CNN) to attain promising tree crown detection results in complex scenes and large study areas. Generally, existing deep learning based tree crown detection approaches can be categorized into three types, i.e., CNN classification based method, semantic segmentation based method and object detection based method, which will be introduced as follows.

2.3.1. CNN classification based methods for tree crown detection

Most recent deep learning based tree crown detection methods adopt CNN classification based method combined with the sliding window technique (Dong et al., 2019b). They usually split the whole image into many image patches with specific window size and then classify them as the background or the tree crown through different CNN architectures, such as LeNet (Mubin et al., 2019), AlexNet (Li et al., 2017; Wu et al., 2020b, Zheng et al., 2020), VGG (Li et al., 2019a), and ResNet (Guirado et al., 2017), etc. Although CNN classification based tree crown methods generally perform much better than classical image processing methods and traditional machine learning methods in regions with overlapping and crowded trees, they almost adopted the sliding window approach to complete final detection, which is a time-consuming technique because of the relatively large amount of detected candidates of various sizes. These methods are inefficient and not flexible for detecting trees of different crown sizes because the image patch size should be predefined according to our prior knowledge (Mubin et al., 2019). Additionally, the CNN classification based methods not only require the annotation of tree crown samples, but also requires extra annotation efforts for other types such as background, other vegetation, impervious, etc.

2.3.2. Semantic segmentation based methods for tree crown detection

Semantic segmentation based tree crown detection methods are end-to-end algorithms without requiring the time-consuming sliding window technique. Unlike the CNN classification based method (i.e., generate one label for one image patch), semantic segmentation aims to provide dense labels for the whole image. Some state-of-the-art semantic segmentation models, for example, U-Net (Freudenberg et al., 2019; Wagner et al., 2020; Zhang et al., 2020), DeepLab v3+ (Morales et al., 2018; Ferreira et al., 2020) and FCN (Xiao et al., 2020; Brandt et al., 2020), etc., have been adopted into tree crown detection and achieve remarkable results. Generally, the semantic segmentation based tree crown detection method is more efficient than the CNN classification based method because it allows the detection of several trees at once. It is suitable for tree species mapping that does not require tree counting tasks (Morales et al., 2018). On the contrary, the performance of semantic segmentation methods gets worse for areas containing trees that appear to overlap with each other, accounting for detecting several overlapping or touching trees as only one tree. Apart from that, the output of semantic segmentation methods is a “probability map” or a “confidence map”, indicating the probability that a pixel belongs to the tree crown type. These methods require a post-processing step to generate the final contours of individual trees, such as the local maximum detection (Freudenberg et al., 2019; Osco et al., 2020). Furthermore, the semantic segmentation based methods focus on tree crown detection and counting, which is difficult to obtain the exact tree crown size. To this end, for identifying individual tree crowns, semantic segmentation based methods are not the best choice. It is more suitable to explore the object detection based methods as described in Section 2.3.3.

2.3.3. Object detection based methods for tree crown detection

In the past few decades, substantial object detection methods have been explored to detect various geospatial objects in the remote sensing community, including tree crown detection using high-resolution remotely sensed images (Li et al., 2020). A variety of end-to-end object detection methods have been applied to the tree crown detection field, such as YOLO v2/v3 (Puttemans et al., 2018; Itakura & Hosoi, 2020), Faster R-CNN (Zheng et al., 2019), RetinaNet (Weinstein et al., 2019, 2020; Selvaraj et al., 2020) and Mask R-CNN (Braga et al., 2020). Some researchers analyzed the accuracy and the efficiency of the above methods on tree crown detection (Santos et al., 2019; Xia et al., 2019), suggesting that two-stage object detection methods (i.e., Faster R-CNN, Mask R-CNN) often have better accuracy than one-stage object detection methods (i.e., YOLO v2/v3, RetinaNet), while one-stage algorithms greatly improve the tree crown detection speed. In general, the object detection based method is faster and more robust than other types of tree crown detection methods, efficiently alleviating the performance drop caused by complex topography, confusions with other vegetation, etc.

Nowadays, deep learning based approaches have been the tendency for tree crown detection, which achieves promising and satisfying performance. However, existing tree crown detection studies have the following limitations. On the one hand, most of them focus on one-class tree detection without subcategory. Although some researchers have separately dealt with mature and young oil palms, the method required previously defined mature oil palm regions and young oil palm regions using different sliding window sizes (Mubin et al., 2019). On the other hand, existing pest or disease detection studies are limited to diseased tree mapping (Zhang et al., 2003; Shafri et al., 2011; Da Silva et al., 2015; He et al., 2019), instead of precisely locating individual diseased tree. In addition, although some studies delve into recognizing disease

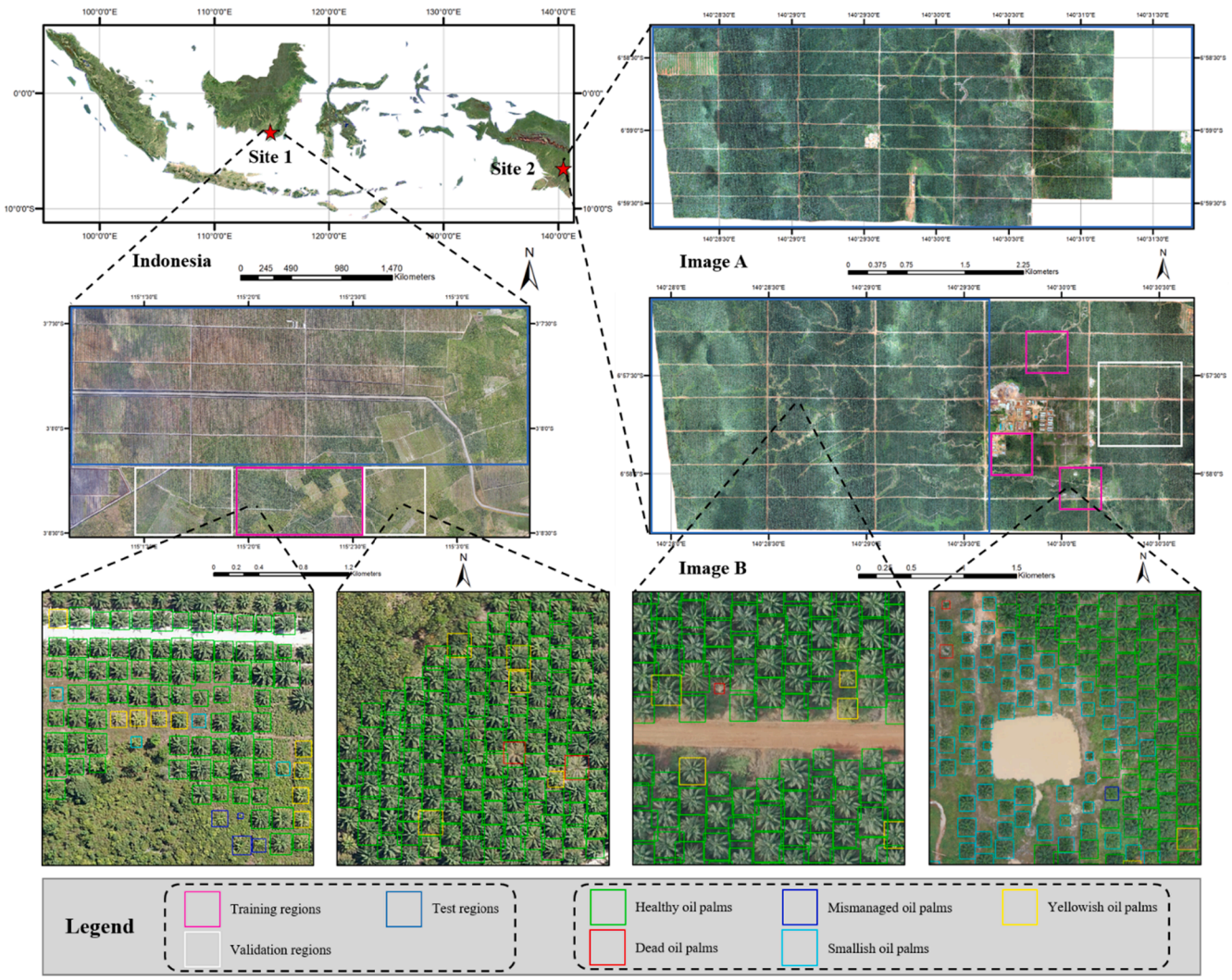

Fig. 1. The study area and UAV images used in this paper. We photogragph UAV images in two sites (i.e., Site 1 and Site 2). Site 1 has one image and Site 2 has two images (i.e., Image A and Image B). For each site, we make a clear geographical separation of training regions, validation regions and test regions, which are denoted by by pink, white, and light blue rectangles, respectively. For the growing status of oil palms, we visualize the healthy, dead, mismanaged, smallish, and yellowish oil palms by green, red, deep blue, cyan and yellow boxes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for individuals photographed by a multi-spectral sensor (Mahlein et al., 2013; Giannakis et al., 2019), they focus on disease research rather than large-scale growing status monitoring for trees or crops. As far as we know, the monitoring of multi-class growing status for individual trees has never been studied before.

3. Study area and datasets

Fig. 1 shows our study area of two sites, i.e., Site 1 and Site 2, which are located in South Kalimantan (115°2′15″E, 3°7′50″S) (Site 1) and Papua (140°29′17″E, 6°57′42″S) (Site 2) in Indonesia. South Kalimantan locates in the middle of Indonesia, where oil palm expansion contributed 3–12% of total Indonesian greenhouse gas emissions during 2000–2010 (Carlson et al., 2013). Papua locates in the southwest of Indonesia, which possesses 2% of all oil palm plantations nationwide in 2015 and is regarded as a new leading oil palm development area in the future (Austin et al., 2017). Our images are acquired by UAV with three bands (RGB) and 10 cm spatial resolution for Site 1 and 8 cm for Site 2,

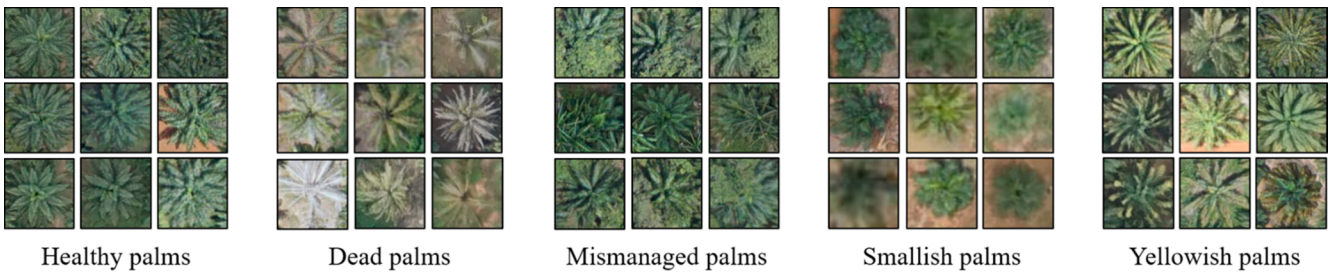

Fig. 2. Examples of different types of oil palms, including healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms.

Table 1
The number of images for training, validation and test dataset in Site 1. The training, validation and test datasets are generated from the training, validation and test regions, respectively.

Type	Training dataset	Validation dataset	Test dataset
Healthy palm	72,219	13,529	31,739
Dead palm	266	59	382
Mismanaged palm	319	81	83
Smallish palm	11,185	14,051	13,074
Yellowish palm	3005	768	2411
Total	86,994	28,488	47,689

respectively. We photograph one image in Site 1 with the size of 40,000 × 20,000 pixels (800 ha), and two images (i.e., Image A and Image B) in Site 2 with the sizes of 64, 273 × 27, 839 pixels (1145 ha) and 85,957 × 31,976 pixels (1759 ha), respectively. There are various land cover types in our study area, such as oil palm plantations, rivers, buildings, other vegetation, etc. Our aerial platform is Skywalker X8 and the type of camera is Sony a6000. Skywalker X8 (as shown in Fig. A1) is a kind of First Person View (FPV) aircraft, with strong wind resistance to guarantee the stability and efficiency of flying. The flight height is 500 m for acquiring Site 1 and 425 m for Site 2, and the cruise speed is about 15–20 m/s. Other details about data collection information and camera specifications are listed in Table A1.

In our study area, all 363,877 oil palms in two sites are manually annotated by experts. Each oil palm is annotated as one of five classes: dead palm, healthy palm, mismanaged palm, smallish palm and yellowish palm. As is shown in Fig. 2, healthy palms denote the palm that is normally cultivated. Dead palms no longer have vitality and usually have gray tree crowns. Mismanaged palms are generally surrounded by other vegetation (e.g., weeds or other trees) and have difficulty obtaining sufficient soil nutrition and sunlight. Smallish palms mean the diameter of the palm crown is relatively smaller than healthy and mature oil palms. Some of them are under a seeding stage and others are under an abnormal growing stage. Yellowish palms often suffer from pests or disease and have yellow spots on their leaves in UAV images. We show some examples of these five classes of oil palms in Fig. 2, and provide more examples photographed from side view with a closer distance in Fig. A2.

As shown in Fig. 1, we make a clear geographical separation of training regions, validation regions and test regions. Owing to different photograph acquisition conditions and environmental heterogeneity between Site 1 and Site 2, we separately train two models to validate oil palm tree detection algorithms. For Site 1, we train a model using one training region (denoted by the pink rectangle) and validate the model using two validation regions (denoted by white rectangles) during training procedures. We report the results of our model in the test region (denoted by the light blue rectangle) for Site 1. For Site 2, we train a model using three training regions (denoted by pink rectangles) in Image B and validate the model using the validation region (denoted by the white rectangle) in Image B during the training phase. We report our model's overall accuracies in one test region in Image B and the whole area in Image A in Site 2 (denoted by light blue rectangles).

We build the training datasets from the above training regions (denoted by pink rectangles in Fig. 1). First, we randomly split each training region into a certain number of images with 1024 × 1024 pixels (see in Section 5.4.1) as our training dataset. For Site 1, we split the training region into 3000 images, while for Site 2, we split each training regions in Image B into 1000 images. To this end, the training datasets of Site 1 and Site 2 both have 3000 images, all of which have at least two different oil palm classes. For the validation dataset, we generate 1000 images from the validation regions in these two sites. For the test dataset, we adopt an overlapping partition approach for the test regions in Fig. 1 (see detailed information in Section 4.4). Tables 1 and A2 list the number of oil palms in each class for training, validation and test dataset in Site 1 and Site 2, respectively. We can easily observe that healthy palm has the highest quantity and the number distribution of five classes of oil palms is extremely imbalanced, resulting in enormous challenges for accurately detecting and distinguishing different classes of oil palms. Our datasets are available on https://github.com/rs-dl/MOPAD.

4. Multi-class oil PAlm detection (MOPAD)

In this section, we introduce our proposed method, i.e., Multi-class Oil PAlm Detection (MOPAD). Our method is inherited from Faster R-CNN (Ren et al., 2015). Fig. 3 shows the framework of our MOPAD, including a Refined Pyramid Feature module (RPF), a multi-level Region proposal network (RPN) and a hybrid class-balanced loss module. We summarized the three major modules of MOPAD as follows:

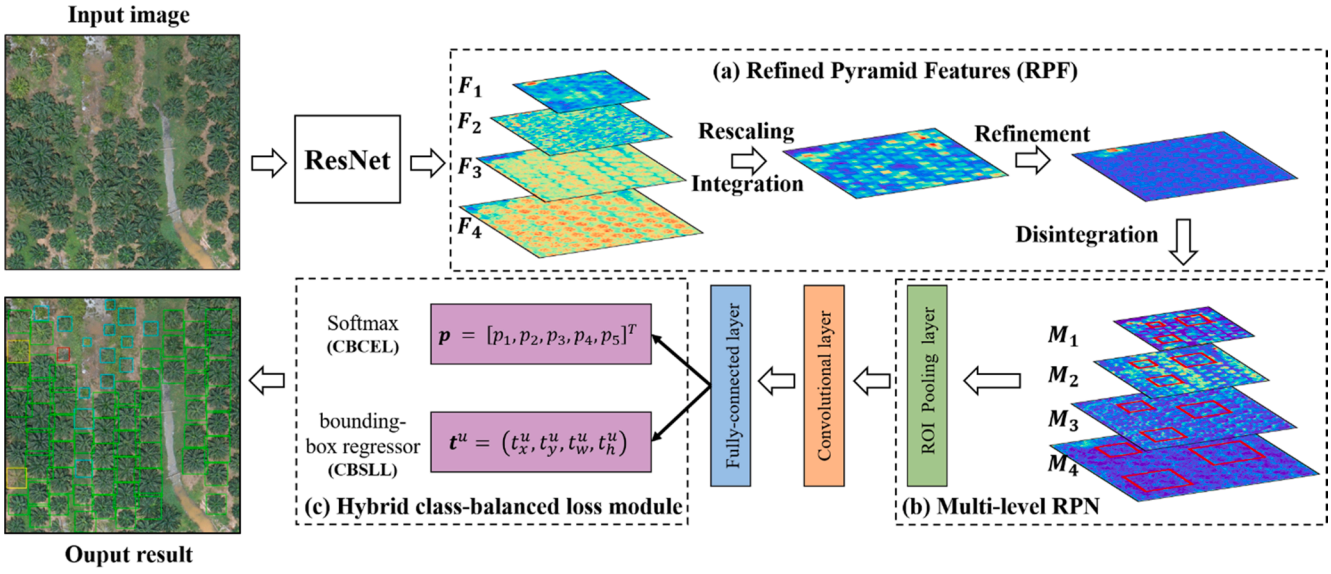

Fig. 3. The flowchart of our proposed MOPAD. Our MOPAD contains three major modules: (a) A Refined Pyramid Feature (RPF) module for feature extraction, including four steps, i.e., rescaling, integration, refinement and disintegration. (b) A multi-level Region proposal network (RPN) for generating oil palm candidates. (c) A hybrid class-balanced loss module for improving multi-class oil palm detection, including a Class-Balanced Cross-Entropy Loss (CBCEL) and a Class-Balanced Smooth L1 Loss (CBSLL).

(1) A Refined Pyramid Feature (RPF) module for feature extraction. We integrate deep-level features and shallow level features to attain more feature characteristics from different level feature maps. Specifically, integrating high level and low-level feature maps contributes to distinguish similar classes and detect small-ish oil palms. Our RPF has four steps, including rescaling, integration, refinement and disintegration.
(2) A multi-level Region proposal network (RPN) for generating oil palm candidates. We adopt an end-to-end network to generate region proposals with a wide range of scales and an aspect ratio of $\{1:1\}$ according to an oil palm tree crown's shape. Following that, we adopt a Region of Interest (RoI) pooling layer to produce the fixed size of feature vectors for all region proposals.
(3) A hybrid class-balanced loss module for improving multi-class oil palm detection. We firstly classify the region proposal as foreground (oil palm) or background and localize the oil palm using smooth L_1 loss. Because the number of samples in different classes are extremely imbalanced, we adopt a Class-Balanced Cross-Entropy Loss (CBCEL) and a Class-Balanced Smooth L_1 Loss (CBSLL) to re-weight the multi-class classification and regression loss according to the number of samples in different classes.

4.1. A Refined Pyramid Features (RPF) module

As deep level features have more semantic meanings, many networks adopt deeper CNN to extract deeper information in images. However, shallow level features are more content descriptive and have more capacity to detect small size objects. Recently, Feature Pyramid Network (FPN) (Lin et al., 2017) builds high-level semantic feature maps at different scales with lateral connections, achieving significant performance on various object detection tasks. Different from FPN, we focus on integrating deep level and shallow level features to realize complementary feature extraction.

Here, we select Residual Network (ResNet) (He et al., 2016) as our backbone. ResNet formulates the layers as learning residual modules with reference to the layer inputs, which addresses the saturated accuracy and degradation problem caused by deeper networks. In ResNet, we can acquire four different scales of feature map as the upper right of Fig. 3 shows. For FPN, it makes predictions independently at all four scales. Our approach, inspired by Pang et al. (2019), integrates four-level features and strengthens the integrated feature by refining. Following, we introduce four steps of our RPF, including integration, refinement and disintegration.

ResNet has four levels of feature maps. We denote feature maps at resolution level l as F_l and our multi-level feature maps can be denoted as $\{F_1, F_2, F_3, F_4\}$. Firstly, we rescale the multi-level feature maps into the same size. For example, if we select F_2 as the intermediate feature map resolution, F_3 and F_4 are interpolated into the same size as F_2, as well as F_1 is max-pooled. Once the multi-level feature maps are resized, the integrated feature maps (F_I) can be calculated as Eq. (1).

$$F_I = \frac{1}{N}\sum_{l_{min}}^{l_{max}} F_l \tag{1}$$

where N is the number of multi-level feature maps. l_{max} and l_{min} are represented as the indexes of the highest and lowest resolution levels, respectively. Here, we obtain the integrated multi-level feature map.

Inspired by the non-local module (Wang et al., 2018), we adopt an embedded Gaussian non-local module (denoted by Refinement) to make our integrated feature maps more discriminative. Finally, we use a reverse procedure to disintegrate and add the corresponding original feature maps to attain our refined pyramid feature maps $\{M_1, M_2, M_3, M_4\}$. Notably, each M_i has not only shallow level semantic features but also deep level semantic features.

4.2. Multi-level region proposal network (RPN)

Region proposal network (RPN) is the main contribution of Faster R-CNN (Ren et al., 2015), sharing full-image convolutional features with the detection network. In Faster R-CNN, RPN slides a dense $n \times n$ spatial window only over the last convolutional feature map. In our MOPAD, we adopt four levels of RPF (i.e., $\{M_1, M_2, M_3, M_4\}$) as the input convolutional feature maps, and attach the same spatial window to each level of our RPF. After that, we use two fully connected layers. One is the classification layer and the other is the bounding box regression layer.

In Faster R-CNN, the original default parameters of the anchor size are $\{128^2, 256^2, 512^2\}$ and the aspect ratios are $\{1:2, 1:1, 2:1\}$ for the last feature map of the Faster R-CNN. Instead, in our MOPAD, we assign an anchor size for each level of RPF. According to the size of oil palms in our image, we define the anchors to have areas of $\{32^2, 64^2, 128^2, 256^2\}$ pixels on $\{M_1, M_2, M_3, M_4\}$, respectively. As for the aspect ratio, we only assign $\{1:1\}$ at each level considering the shape of an oil palm tree crown. In addition, our oil palm candidates generation method in multi-level RPN also accelerates the training and reference process because a smaller number of anchors are generated.

4.3. A hybrid class-balanced loss module

As our training samples are extremely imbalanced among different classes, directly adopting state-of-the-art object detection approaches may have embarrassing results. Many previous researches related to imbalance problems focus on data augmentation, such as geometrical transformations on images (horizontal flipping, multi-scale strategy, adding noises, etc.) (Ren et al., 2015; Liu et al., 2016), or using random occlusion mask on images or feature maps (Zhu et al., 2015; Wang et al., 2017). However, these methods usually lead to an overfitting problem because of generating duplicated images.

On the other hand, the imbalanced problem can also be addressed by re-weighting the loss with inverse class frequency or inverse square root of class frequency. Inspired by Cui et al. (2019), we adopt a Class-Balanced Cross-Entropy Loss (CBCEL) that embed an Effective Number (EN) for loss calculation instead of directly using the sample number. The EN of samples is the expected number of samples (Cui et al., 2019) and can be calculated as:

$$EN = \frac{1-\beta^{n_u}}{1-\beta} \tag{2}$$

where n_u is the amount of samples in the ground-truth class u and β is a hyperparameter ($\beta \in [0, 1)$). It is noteworthy that $\beta = 0$ corresponds to the original loss function and $\beta \to 1$ means re-weighting through inverse class frequency. In other words, our CBCEL is reweighted by an inverse effective number. As the number of oil palm classes is 5, our predicted outputs from the classifier can be formulated as $\boldsymbol{p} = [p_1, p_2, p_3, p_4, p_5]^T$. Suppose a sample with the class label of y, the softmax cross-entropy (CE) loss for this sample can be formulated as:

$$CE_{softmax}(\boldsymbol{p}, y) = -log\left(\frac{\exp(p_u)}{\sum_{i=1}^{C}\exp(p_i)}\right) \tag{3}$$

where C is the total volume of classes. Given class u with n_u training samples, our CBCEL can be formulated as:

$$CBCEL(\boldsymbol{p}, u) = -\frac{1-\beta}{1-\beta^{n_u}} log\left(\frac{\exp(p_u)}{\sum_{i=1}^{C}\exp(p_i)}\right) \tag{4}$$

Moreover, we utilize a Class-Balanced Smooth L_1 Loss (CBSLL) for our loss function for bounding box regression. Notably, the smooth L_1 Loss is the same as the original Faster RCNN. Our CBSLL can be defined as:

Table 2
The detection results of our proposed MOPAD for the test region in Site 1.

Index	Oil palm	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm
TP	39,973	27,047	170	44	10,071	2063
FP	3280	594	63	43	2992	166
FN	7716	4692	212	39	3003	348
Precision	92.42%	97.85%	72.96%	50.57%	77.10%	92.55%
Recall	83.82%	85.22%	44.50%	53.01%	77.03%	85.57%
F1-score	87.91%	91.10%	55.28%	51.76%	77.06%	88.92%

$$CBSLL(\boldsymbol{t}^u, v) = \frac{1-\beta}{1-\beta^{n_u}} \sum_{i\in\{x,y,w,h\}} smooth_{L_1}\left(t_i^u - v_i\right) \tag{5}$$

in which

$$smooth_{L_1}(x) = \begin{cases} 0.5x^2 if|x| < 1 \\ |x| - 0.5 otherwise \end{cases} \tag{6}$$

where $\boldsymbol{t}^u = \left(t_x^u, t_y^u, t_w^u, t_h^u\right)$ is a predicted tuple; $v = \left(v_x, v_y, v_w, v_h\right)$ is a tuple of true bounding box regression targets for class u. Therefore, our multi-task loss function can be calculated as Eq. (7) and λ is a trade-off hyperparameter.

$$L(p, u, t^u, v) = CBCEL(\boldsymbol{p}, u) + \lambda^* CBSLL(\boldsymbol{t}^u, v) \tag{7}$$

4.4. Large-scale oil palm tree detection

We adopt MOPAD to predict the oil palm tree crown using three large-scale test regions in Site 1 and Site 2. As we trained our model using sub-images of 1024 × 1024 pixels (see in Section 5.4.1), we needed to crop the large-scale image into small patches to implement the large-scale image prediction (Peng et al., 2020). If we directly crop the large-scale image uniformly into small patches of 1024 × 1024 pixels without overlaps, some tree crowns will be separated into two different patches, in which case the tree crown might not be detected or the detected object is incomplete. To solve this problem, we adopt an overlapping partition approach for large-scale UAV image prediction instead of directly splitting into several 1024 × 1024 sub-images, along with every two neighbor patches in the large-scale image has an overlapping width (height) with 200 pixels to ensure corners are not overlooked by the CNN model. In this way, all the detected trees belong to at least one patch. After the oil palm tree detection of each patch, we perform an Intersection-of-Union (IoU) based merging operation to acquire final detection results. This not only solves the problem of misdetection and wrong detection in the large-scale image detection process, but also avoids duplicate detection of targets. In addition, we consider the detected healthy palms whose crown sizes are less than 50 × 50 pixels as smallish palms for Site 1 and 80 × 80 pixels for Site 2.

5. Experimental results of multi-class oil palm tree detection

In this section, we evaluate the experimental results of our proposed MOPAD for multi-class oil palm tree detection. First, we introduce the experimental setup and our evaluation metric in Section 5.1. Then we describe the multi-class oil palm tree detection results of MOPAD in Section 5.2, followed by substantial ablation studies and comparison with other state-of-the-art object detection methods in Section 5.3 and Section 5.4, respectively.

5.1. Setup and evaluation

We complement our experiments based on the MMdetection deep learning framework (Chen et al., 2019), and we set the hyperparameter λ (introduced in Eq. (7)) as 1 throughout all our experiments (Ren et al., 2015). We train our model on GeForce RTX 2080 Ti. Moreover, we choose Stochastic Gradient Descent (SGD) (Bottou, 2010) as our optimizer with a momentum strategy and the number of training epoch is 24. The backbone of MOPAD and other popular object detection methods is ResNet-101 (see Fig. C1 and Section 5.4.1) for a fair comparison. We discuss the effect of the depth of backbone in Section 5.4.1. The hyperparameter of β (introduced in Eq. (2)) is equal to 0.9999 in our experiments and we analyze the reason in Section 5.4.3. Our codes are available on https://github.com/rs-dl/MOPAD.

Our evaluation metric consists of precision, recall and F1-score, which are calculated following previous studies (Li et al., 2019a, 2019b; Zheng et al., 2020). Precision depicts the model's capability to

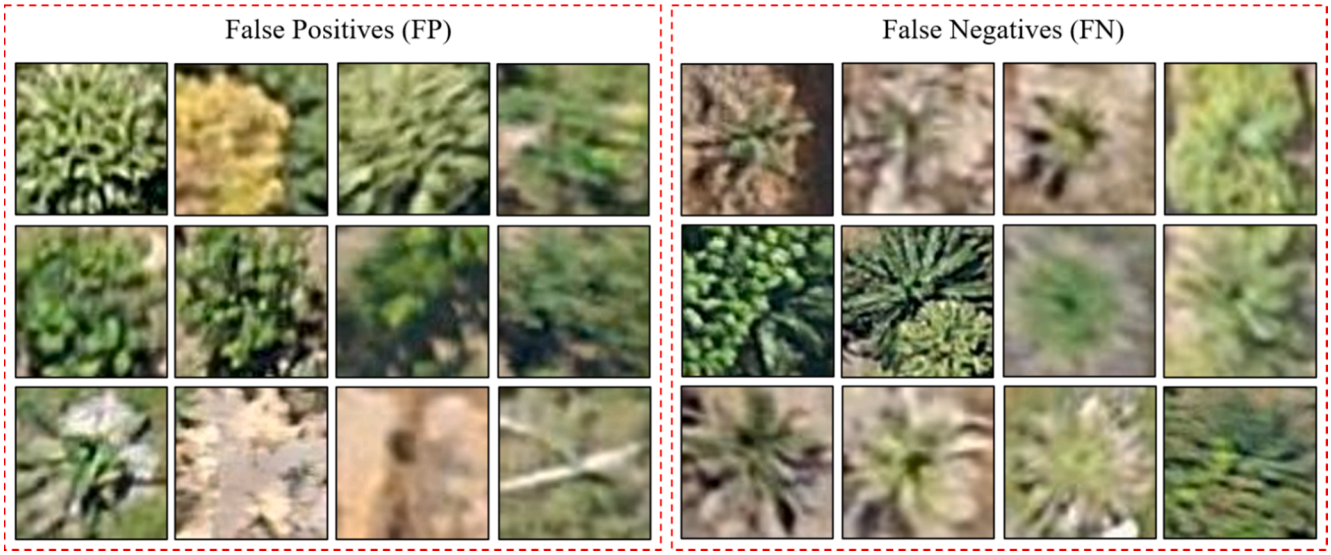

Fig. 4. Failure cases for FP and FN of oil palm detection in test regions in Site 1.

Table 3
The confusion matrix for multi-class oil palm detection of our proposed MOPAD in Site 1, in which the top row denotes the ground truth types and the left column denote the predicted types.

Index	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm
Healthy palm	27,047	3	6	364	15
Dead palm	12	170	0	15	2
Mismanaged palm	1	1	44	6	0
Smallish palm	0	16	0	10,071	10
Yellowish palm	112	1	0	14	2063

detect oil palms correctly, while recall describes the model's capability to detect ground-truth oil palms. F1-score evaluates the overall performance of the model. The above metric can be calculated from True Positives (TP), False Positives (FP) and False Negatives (FN). TP denotes the number of oil palms detected correctly; FP denotes the number of others detected as oil palms by mistake; FN denotes the number of ground-truth oil palms missing in detection results. When the IOU metric value between the detected palm and an annotated oil palm tree is equal to or greater than 0.5, the oil palm tree will be regarded as a correctly detected object.

5.2. Multi-class oil PAlm tree detection (MOPAD) results

Tables 2 and B1 list the detection results of our proposed MOPAD for the test regions in Site 1 and Site 2, with respect to TP, FP, FN, precision, recall and F1-score. The F1-scores of oil palm tree detection are 87.91% and 99.04%, with precisions of 92.42% and 98.90%, and recalls of 83.82% and 99.19%. The oil palm tree detection performance of Site 2 is much better than Site 1, with over 11% increment of F1-score. The possible reason is that the images for Site 2 have much higher quality, and the local oil palm plantation possesses better management with a unified planting pattern so that there are fewer confusions between oil palm trees and background, such as other vegetation. Figs. 4 and B1 display some examples of FP and FN for Site 1 and Site 2, respectively. We can observe that other vegetation and tree shadows can be detected as oil palm trees by mistake (in the left of Figs. 4 and B1). On the other hand, FN mainly results from sheltering by other vegetation, the small size of oil palm tree crown, and the low quality of UAV images (in the right of Figs. 4 and B1).

As for multi-class oil palm tree detection results, the average F1-scores for the five classes are 72.83% and 70.57% for the test regions in Site 1 (in Table 2) and Site 2 (in Table B1), respectively. The performance of multi-class oil palm tree detection for Site 1 is slightly better than Site 2, with a 2.16% gain of F1-score. We can observe that the healthy palm, which has the largest sample number among five classes, reaches the highest F1-score of 91% for Site 1, and over 99% for Site 2, while the mismanaged palm and the dead palm has relatively lower F1-scores, with 55.28% and 51.76% for Site 1 and 43.24% and 44.59% for Site 2. Tables 3 and B2 show the confusion matrix for multi-class oil palm tree detection of our proposed MOPAD for Site 1 and Site 2, respectively, in which the top row denotes the ground truth types and the left column denotes the predicted types. The mismanaged palm and yellowish palm have serious confusion with healthy palm because of their high similarity. We also display some examples of confusion between different classes in Fig. 5 for Site 1 (on the left) and Site 2 (on the right).

5.3. Results comparison between our method and other state-of-the-art object detection methods

We compare our method with two traditional machine learning based methods (i.e., RF and SVM), CNN classification based method (i.e., ResNet-101) and other five state-of-the-art object detection methods, including Faster R-CNN (Ren et al., 2015), Cascade R-CNN (Cai and Vasconcelos, 2018), Guided Anchoring (GA) Faster R-CNN (Wang et al., 2019a), Libra Faster R-CNN (Pang et al., 2019) and Grid R-CNN (Lu et al., 2019). RF and SVM are two prevalent machine learning methods, which have been applied in numerous tree detection studies (Dai et al., 2018; Johansen et al., 2020). For the CNN classification based method, we adopt the same backbone (i.e., ResNet-101) as our MOPAD, and the sliding window technique. Appendix C introduces the detailed workflow for the CNN classification based multi-class oil palm tree detection algorithm. Note that we employ the same post-processing procedures for both traditional machine learning methods and CNN classification based method. As for object detection based tree crown detection method, **Faster R-CNN** is a basic object detection method for MOPAD, and has

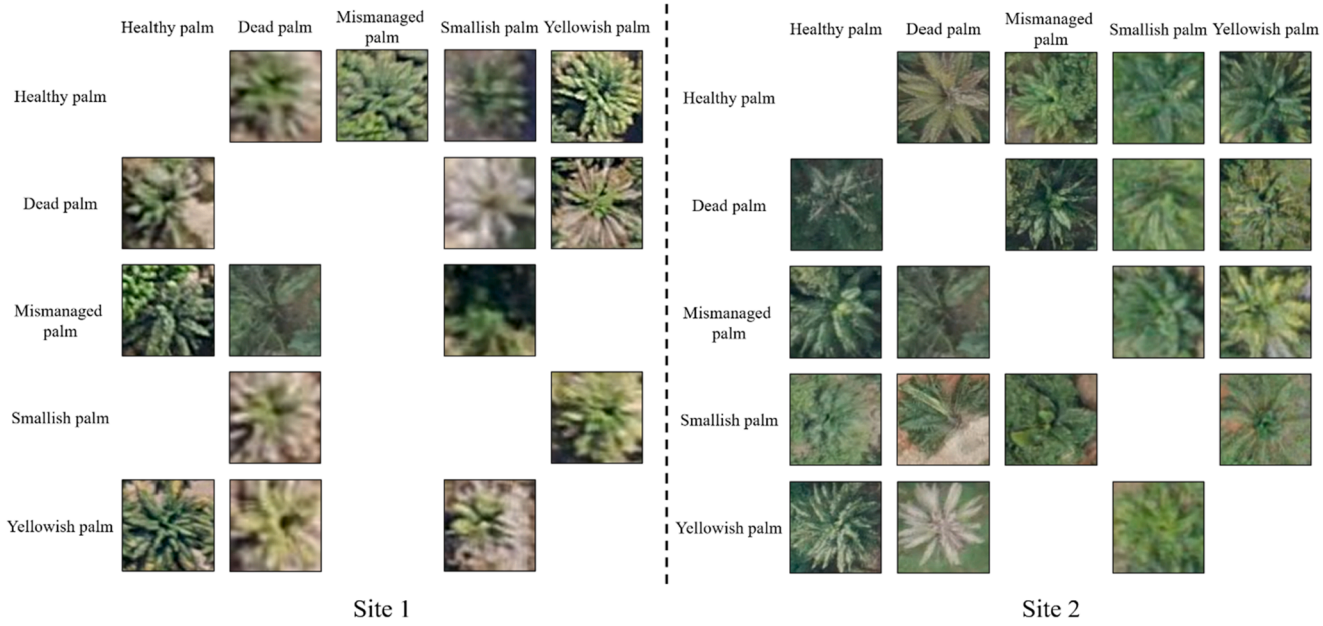

Fig. 5. Failure cases of confusions among different kinds of oil palms for Site 1 (left) and Site 2 (right), in which the top row denote the ground truth types and the left column denote the predicted types.

Table 4
The F1-score of other state-of-the-art tree crown detection methods in Site 1.

Method	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
RF	79.05%	0.46%	0.00%	37.50%	12.91%	25.98%
SVM	77.47%	0.00%	0.00%	33.45%	5.03%	23.19%
CNN (ResNet-101)	74.76%	7.33%	2.86%	35.93%	19.26%	28.03%
Faster R-CNN	90.46%	6.47%	42.48%	65.76%	73.54%	55.74%
Grid R-CNN	90.62%	13.37%	41.82%	66.14%	69.22%	56.23%
GA Faster R-CNN	88.60%	15.67%	54.70%	61.52%	71.18%	58.33%
Cascade R-CNN	**91.22%**	36.36%	40.00%	64.48%	71.46%	60.71%
Libra Faster R-CNN	91.00%	30.54%	**55.74%**	65.17%	69.87%	62.46%
MOPAD (ours)	91.10%	**55.28%**	51.76%	**77.06%**	**88.92%**	**72.83%**

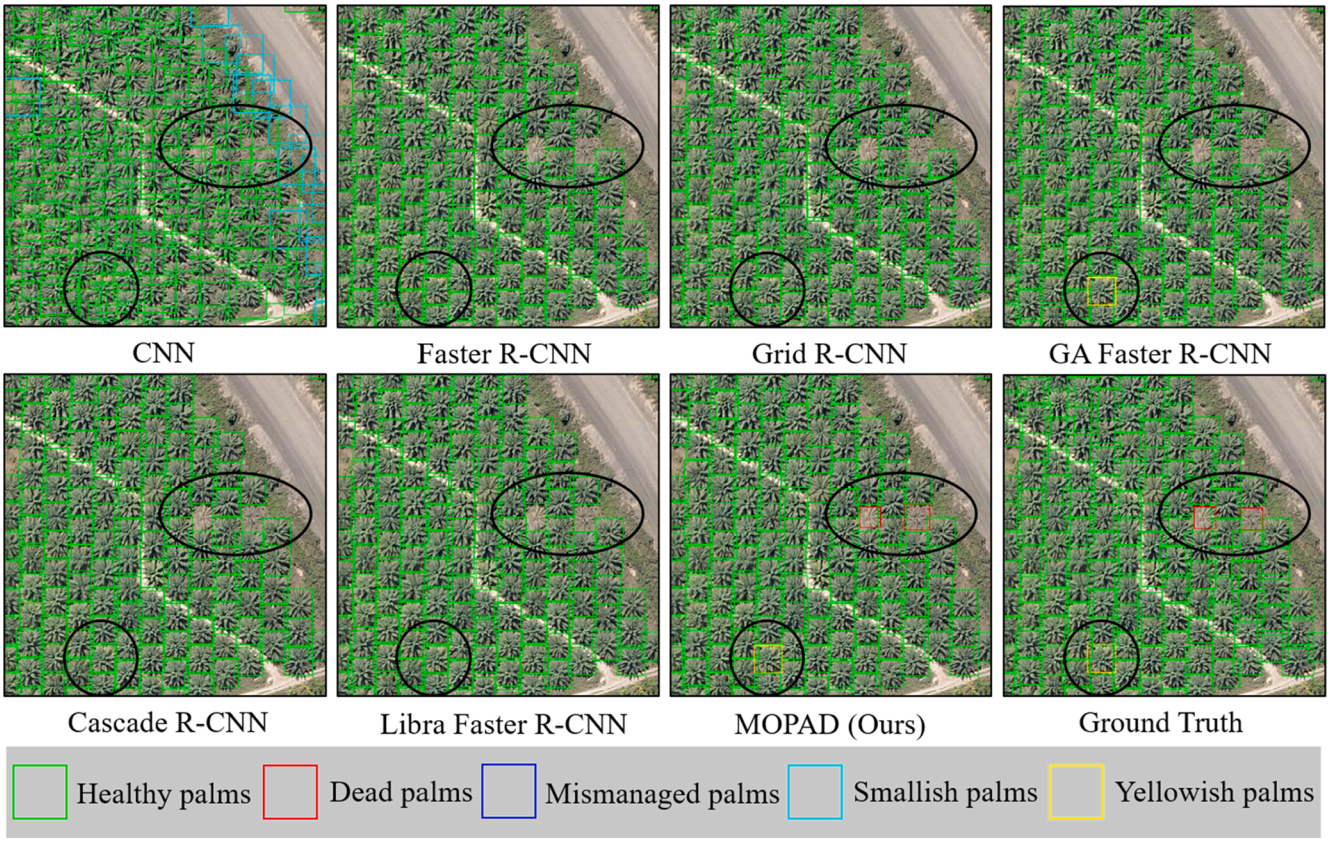

Fig. 6. Multi-class oil palm tree detection results for Region 1 in Site 1. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

been employed in several tree crown detection studies (Santos et al. 2019; Xia et al., 2019; Zheng et al., 2019). **Cascade R-CNN** consists of a sequence of detectors to avoid problems of overfitting at training procedures. **GA Faster R-CNN** leverages semantic features to guide the arbitrary anchor selection rather than using dense and predefined anchors in RPN introduced in Section 4.2. **Libra Faster R-CNN** integrates IoU-balanced sampling, balanced feature pyramid and balanced L1 loss for reducing the imbalance at different levels, which is possibly appropriate in our task with the imbalanced issue. **Grid R-CNN** captures the spatial information explicitly with a feature fusion module to achieve accurate object detection performance.

We list the test regions' results in Site 1 and Site 2 in Tables 4 and B3. Our proposed MOPAD achieves the highest F1-score for all oil palm classes except healthy oil palm and mismanaged oil palm in Site 1 (lower than Cascade R-CNN and Libra Faster R-CNN, respectively). In general, the performance of multi-class oil palm tree detection through object detection based methods are much better than traditional machine learning based methods (i.e., RF and SVM) (with at least 29.76% and 23.20% improvement for Site 1 and Site 2, respectively) and CNN classification based method (with at least 27.71% and 20.45% improvement for Site 1 and Site 2, respectively). We can observe that our method has remarkably higher multi-class oil palm detection accuracy and fewer confusions among different kinds of oil palms, with improvements of 10.37%-17.09% and 8.14%-21.32% in respect of average F1-score for Site 1 and Site 2, respectively. Faster R-CNN exhibits diminished performance with an average F1-score of 55.74% and 49.25%. Grid R-CNN, Cascade R-CNN, Libra Faster R-CNN and GA Faster R-CNN perform similarly in Site 2, achieving considerable improvements compared to Faster R-CNN with 6%-13.18% improvement. On the other hand, in Site 1, these methods attain limited accuracy gains with 0.49–6.72% improvement. As for our proposed MOPAD, it substantially ameliorates the average F1-score of multi-class growing status classification by 17.09%, compared to Faster R-CNN under extremely imbalanced samples in Site 1. Figs. 6–9 show the detection results of the CNN classification based method, the above five object detection based methods and our MOPAD, which indicate the superior results of MOPAD compared with other state-of-the-art methods. The green, red, blue, cyan and yellow rectangles represent healthy, dead, mismanaged, smallish

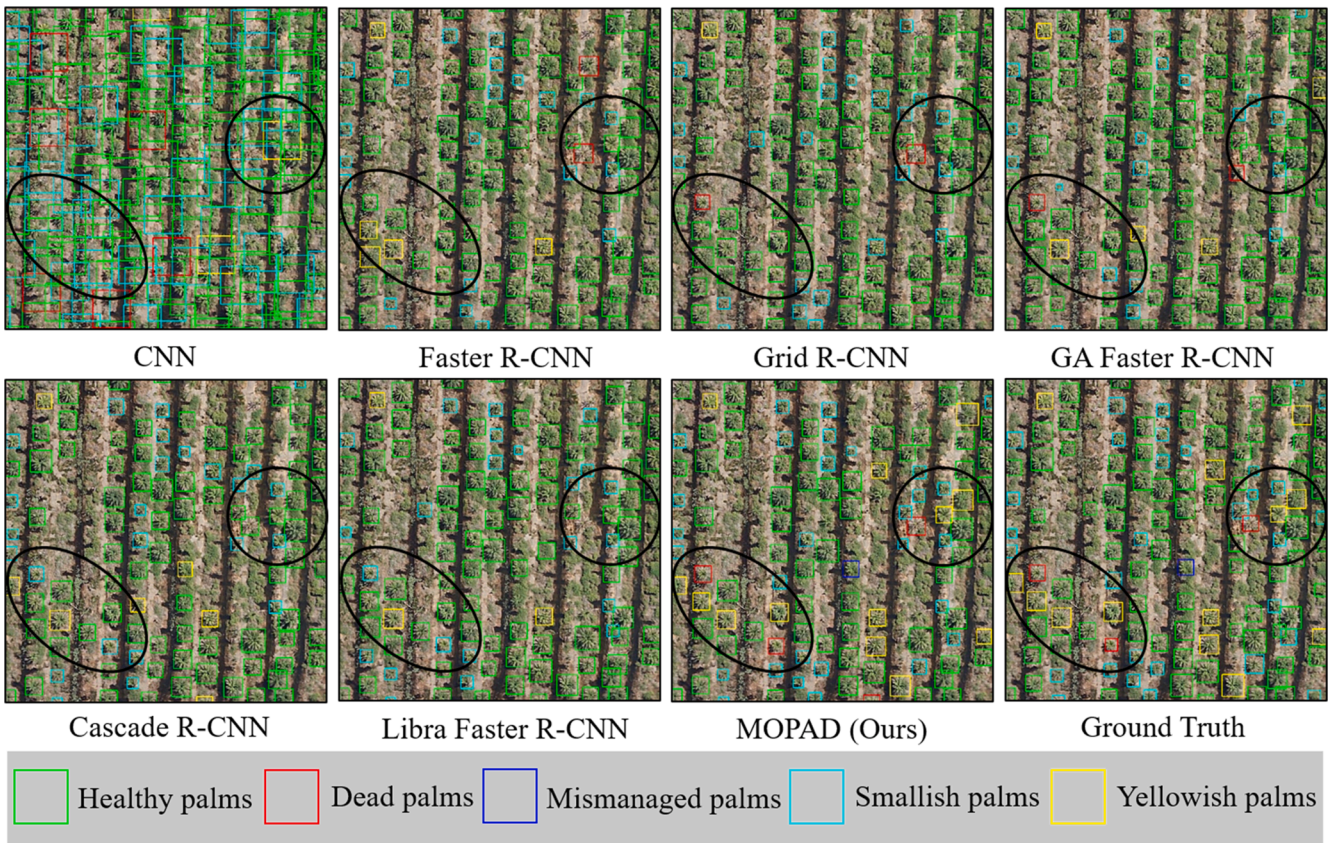

Fig. 7. Multi-class oil palm tree detection results for Region 2 in Site 1. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

and yellowish palms. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods. We can easily observe that MOPAD performs better on classes with lower samples, such as dead palms (Figs. 6 and 7), mismanaged palms (Fig. 8) and yellowish palms (Fig. 9). The improved performance for smallish palms is displayed in Fig. 10 (see in Section 5.4.2). More detection results and comparisons can be found in Figs. D1–D6 in Appendix D.

5.4. Ablation studies for our proposed MOPAD

In this part, we first assess the effect of the CNN layer's depth and the input image's size in Section 5.4.1. Then, extensive ablation studies are conducted to validate the Refined Pyramid Feature (RPF) in Section 5.4.2 and the hybrid class-balanced loss module, including the Class-Balanced Cross-Entropy Loss (CBCEL) and the Class-Balanced Smooth L_1 Loss (CBSLL) in Section 5.4.3. These experiments demonstrate the contributions of different components and provide more insights into our proposed MOPAD. Note that we take the test regions in Site 2 (the whole area in Image A and one test region in Image B, see Fig. 1) as an example in our following ablation studies.

5.4.1. Ablation study for the depth of CNN layer, the size of input images and the anchor generation

Table 5 shows the performance of different depths of ResNet, including 18, 34, 50, 101 and 152 layers. Experimental results present that ResNet-101 has the highest average F1-score of 70.57% for multi-class oil palm detection and the highest F1-score for healthy palms, smallish palms and yellowish palms. ResNet-50 reaches the highest accuracy for dead palms and mismanaged palms. In conclusion, we select ResNet-101 as the backbone of our proposed MOPAD. We further explore the optimal size of the input sub-image. The results of different input sizes are shown in Fig. B2. Not that we adopt the same input size of sub-image in both the training and test phase. When the input size is 512 $\times$ 512 pixels or 1024 $\times$ 1024 pixels, the model performs around 70% of the average F1-score, while a larger input image size of 2048 $\times$ 2048 pixels leads to a worse detection performance, especially for dead palms, which has the least amount of samples. Considering both the detection accuracy and the inference efficiency, we use 1024 $\times$ 1024 pixels as the image size in this study. As for the anchor generation, we conduct the ablation experiments about different anchor-sizes and aspect-ratios, which is displayed in Table B4. Although the ratios of $\{1:2, 1:1, 2:1\}$ (the third row in Table B4) performs slightly better than our ratios of $\{1:1\}$ (the fourth row in Table B4) under our scale setting ($\{32^2, 64^2, 128^2, 256^2\}$), we keep our ratio setting as {1:1} because it is more suitable and reasonable according to the circle shape of the oil palm tree crown. In addition, it also accelerates the training and reference process because a smaller number of anchors are generated. We can also find that our scale setting of $\{32^2, 64^2, 128^2, 256^2\}$ can further improve the average accuracies compared with using $\{128^2, 256^2, 512^2\}$.

5.4.2. Ablation study for the Refined Pyramid feature (RPF)

We show the ablation studies of RPF in Table 6. We first implement RPF only with integration. We can observe that although the naïve feature integration only obtains a 1.52% higher average F1-score than the corresponding baseline (i.e., Faster R-CNN with FPN), the F1-score of smallish palm attains 12.03% gains compared to the pure Faster R-CNN. It is noteworthy that there is no refinement or parameter added in this method. Only with this simple procedure, each resolution level obtains equal information from other level feature maps. After embedding the convolutional attention or Gaussian non-local attention (Wang et al., 2018), RPF can further improve the final results. Our RPF is able to obtain a 5.37% performance gain in terms of average F1-score for multi-

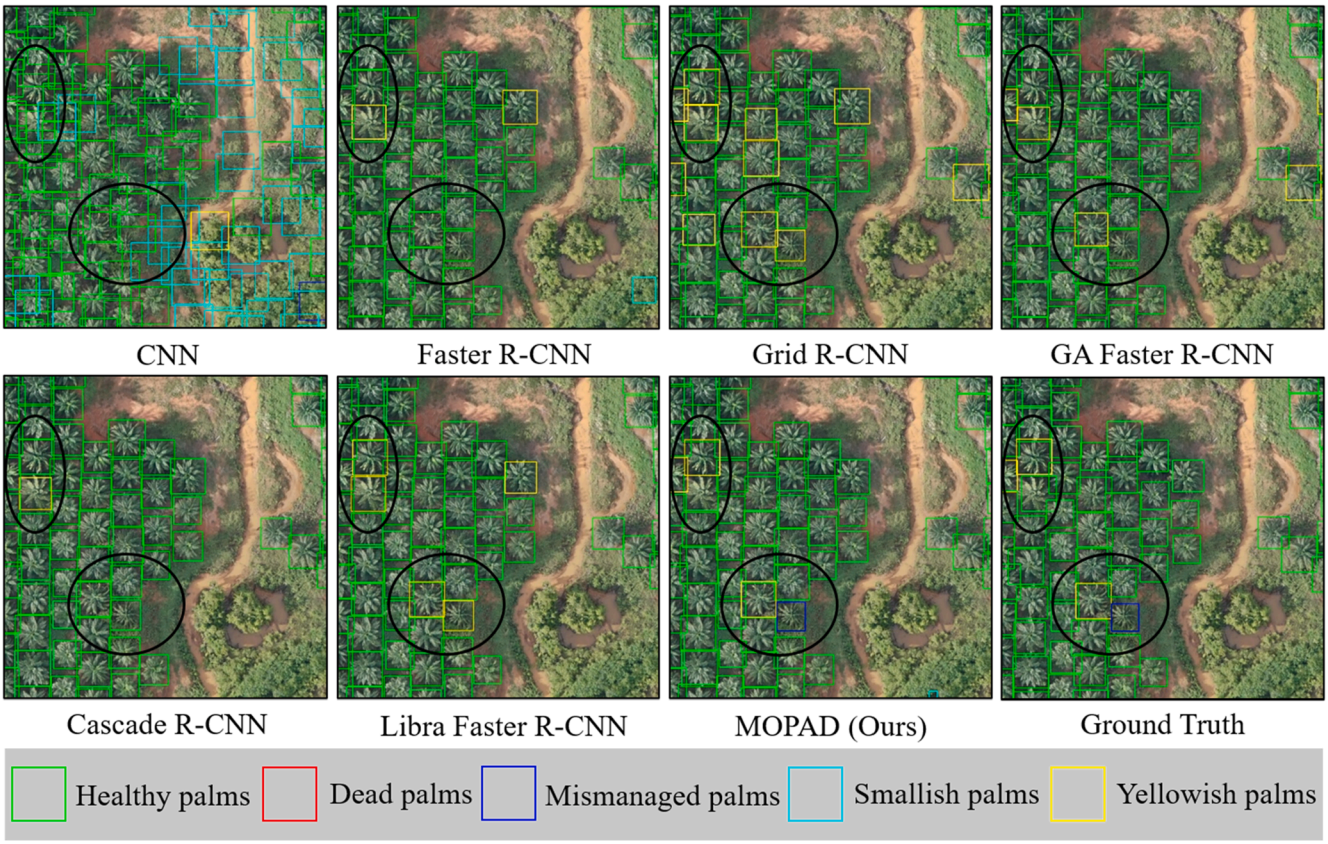

Fig. 8. Multi-class oil palm tree detection results for Region 1 in Site 2. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

class oil palm tree detection using Gaussian non-local attention. It is worth noting that our RPF achieves a considerable improvement of 13.99% for smallish palms compared with Baseline (Faster R-CNN with FPN), demonstrating that RPF efficiently integrates multi-scale features for correctly detecting smaller trees. The performance of detecting smallish oil palms of these methods are illustrated in Fig. 10. We can find that the Baseline method misses most smallish palms. Leveraging Integration and Conv-based Refinement can improve the detection results for smallish palms to some extent. Based on the non-local based refinement (our RPF), smallish palms are more accurately detected than the above methods.

5.4.3. Ablation study for the hybrid class-balanced loss module

As listed in Table 7, we verify the effectiveness of CBCEL and CBSLL separately, which gains 12.04% and 5.21% improvement for multi-class oil palm tree detection with respect to average F1-score compared to Faster R-CNN, and 13.68% and 5.99% for Faster R-CNN + RPF. Furthermore, combining CBCEL and CBSLL can make a better improvement of 13.97% for Faster R-CNN and 15.95% for Faster R-CNN + RPF. The increment of our hybrid class-balanced loss module for Faster R-CNN + RPF is higher than that for Faster R-CNN. The reason may be that RPF not only helps to detect smallish palms, but also promote to distinguish similar classes because of fusing different level features. Our hybrid class-balance loss module can effectively increase the detection accuracy by a large margin for low-number palm classes like dead palms (+29.29% and + 32.65%), mismanaged palms (+22.69% and + 21.49%) and yellowish palms (+23.84% and + 26.04%) on the basis of Faster R-CNN and Faster R-CNN + RPF, respectively. Additionally, the performance of healthy palms also has a slight increase for Faster R-CNN + RPF, with the possible reason that our hybrid class-balanced loss module alleviates the overfitting phenomenon. Although the hybrid class-balanced loss module has a relatively lower performance for smallish palms, it still helps to achieve a favorable improvement for the average F1-score of five oil palm classes.

Next, we further analyze the hyper-parameter β, which controls how fast the Effective Number (EN) grows as the number of samples increases (Cui et al., 2019). According to Eq. (2), we display the F1-scores for multi-class oil palm detection under different $\beta = \{0.9, 0.99, 0.999, 0.9999, 0.99999, 0.999999\}$. If β is closing to 1, the weights in the loss module for the class with a low number of samples become larger. Fig. 11 displays the performance of multi-class oil palm detection under different β. It is observed that as β increases, the accuracy of healthy palms consistently increases when $\beta \leq 0.999$ because of alleviating the overfitting phenomenon, while it can negatively affect the detection accuracy for smallish palms when $\beta \geq 0.9999$. For low-number oil palm classes, mismanaged palms and dead palms achieve the highest F1-score when $\beta = 0.999$ and yellowish palms reach the best performance when $\beta = 0.99999$. In conclusion, we select $\beta = 0.9999$ for our MOPAD to obtain the highest average F1-score for multi-class oil palm tree detection.

6. Experimental results of smart oil palm plantation management

In this section, we take Image A in Site 2 as an example to introduce the experimental results of two practical applications for smart oil palm plantation management based on the detection results of our proposed MOPAD. Smart oil palm plantation management benefits oil palm companies and smallholders for reasonable plantation planning, oil palm yield improvement and saving labor or fertilizer. We first introduce the distribution of oil palms in our study area in Section 6.1, and introduce how to detect the vacancies for oil palm replanting in Section 6.2.

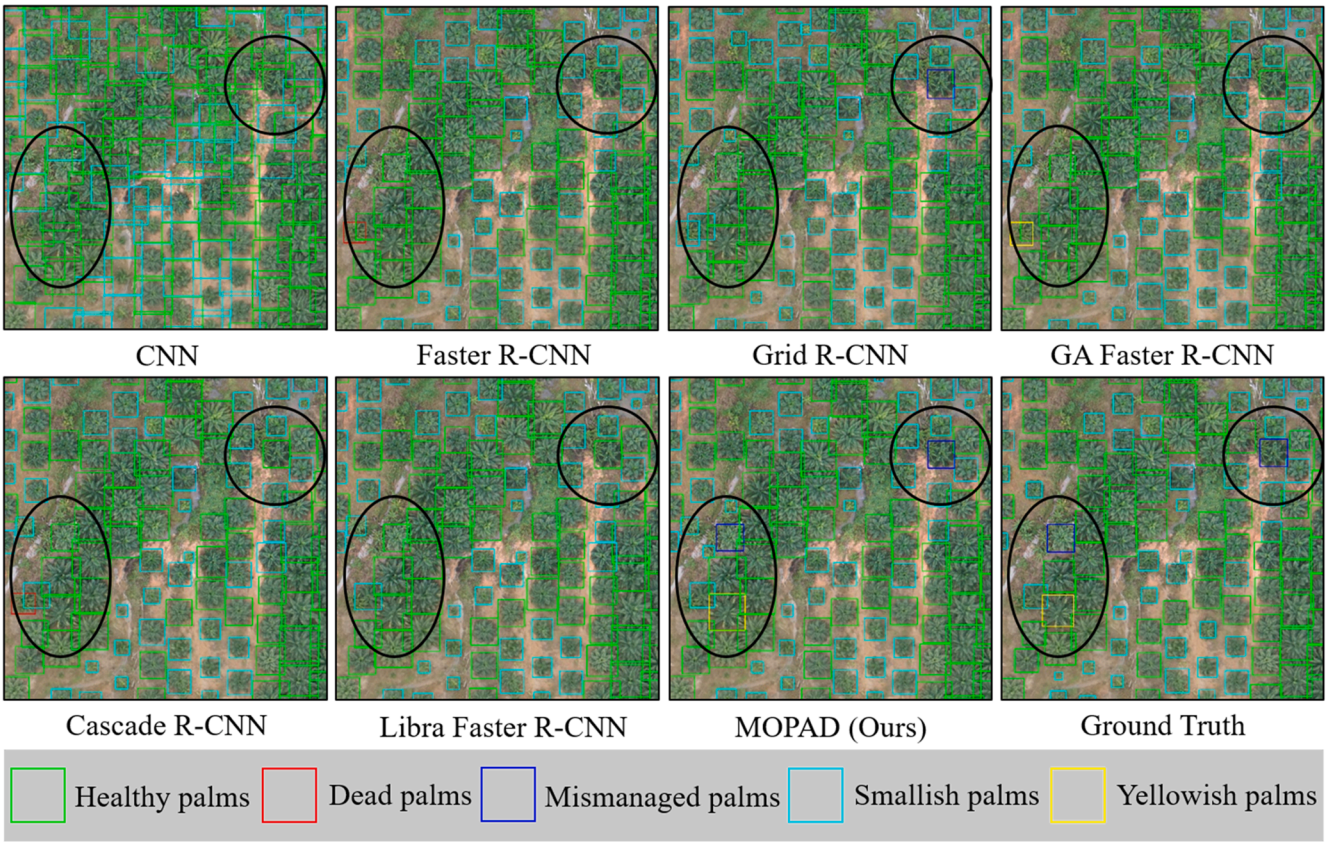

Fig. 9. Multi-class oil palm tree detection results for Region 2 in Site 2. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

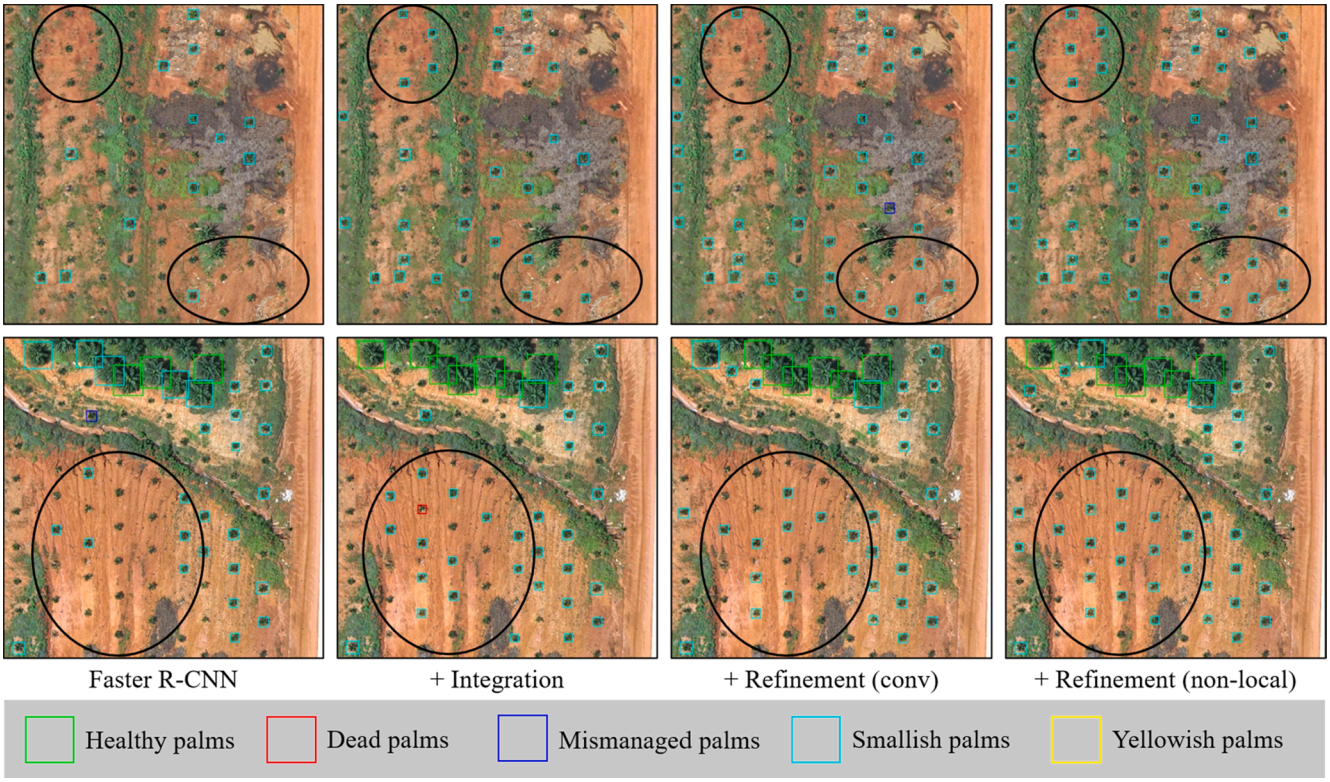

Fig. 10. Comparison of smallish oil palm detection results using different strategies. The black circles denote some examples where our Refinement (non-local) outperforms other methods.

Table 5
The F1-score of different depth of ResNet for our MOPAD.

Method	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
ResNet-18	97.14%	27.05%	35.18%	80.51%	66.69%	61.31%
ResNet-34	97.49%	44.59%	43.33%	83.64%	53.74%	64.56%
ResNet-50	98.18%	**57.38%**	**46.32%**	85.87%	48.78%	68.19%
ResNet-101	**99.43%**	43.24%	44.59%	**89.55%**	**76.04%**	**70.57%**
ResNet-152	97.22%	40.80%	43.12%	86.64%	65.92%	66.74%

Table 6
The F1-score of ablation study of RPF.

Method	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
Faster R-CNN	98.58%	7.94%	17.46%	78.25%	44.01%	49.25%
+ Integration	**98.61%**	9.33%	17.93%	90.28%	37.68%	50.77%
+ Refinement (conv)	97.99%	8.70%	21.29%	91.51%	**51.68%**	54.23%
+ Refinement (non-local) (RPF)	97.19%	**10.59%**	**23.10%**	**92.24%**	50.00%	**54.62%**

Table 7
The F1-scores of ablation study of the hybrid class-balanced loss module.

Method	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
Faster R-CNN	**98.58%**	7.94%	17.46%	**78.25%**	44.01%	49.25%
+ CBCEL	98.37%	**40.68%**	37.57%	75.39%	54.43%	61.29%
+ CBSLL	98.10%	23.94%	26.44%	75.59%	48.23%	54.46%
+CBCEL + CBSLL	98.54%	37.23%	**40.15%**	72.35%	**67.85%**	**63.22%**
Faster R-CNN + RPF	97.11%	10.59%	23.10%	**92.24%**	50.00%	54.62%
+ CBCEL	98.08%	42.22%	41.25%	88.09%	71.86%	68.30%
+ CBSLL	98.12%	20.79%	34.68%	86.16%	63.28%	60.61%
+CBCEL + CBSLL(MOPAD)	**99.43%**	**43.24%**	**44.59%**	89.55%	**76.04%**	**70.57%**

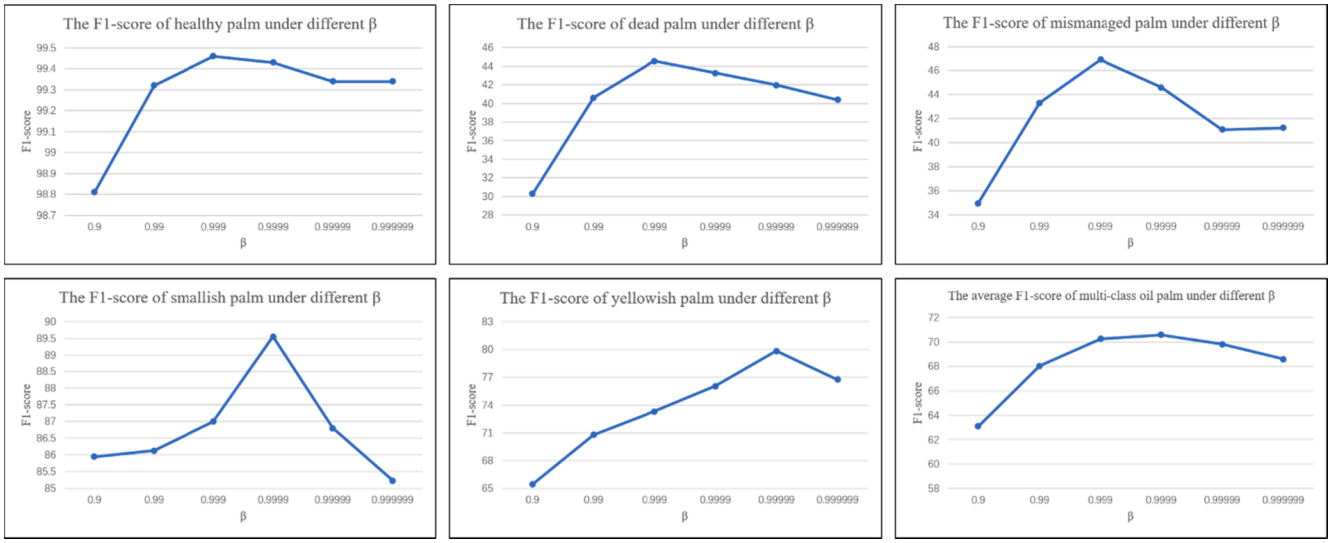

Fig. 11. The F1-score of multi-class oil palm detection under different β.

6.1. The distribution of oil palms in our research area

In our whole Image A, a number of 169,374 oil palm trees are detected using MOPAD. Our proposed MOPAD achieves the overall oil palm tree detection F1-score of 99.43%, which is a relatively satisfying result. The total inference time for oil palm detection in the whole UAV image is less than 15 min. By calculating the volume of detected oil palms per hectare for each pixel, we obtain the distribution map of our study area, as shown in Fig. 12. The map is consistent with the actual plantation distribution of the oil palms, clearly indicating the non-oil palm plantation areas, such as buildings, roads or other vegetation.

We also map the distribution of healthy palms and smallish palms in Figs. 13 and 14, respectively. As the percentage of detected healthy palms is more than 99% (see Tabel C3), the distribution of healthy palms is similar to the one of all oil palms. We can find two cases from the distribution of smallish palms. The first case is that smallish palms are aggregating together, indicating they are under a seeding stage and planted at a later time. The second case is that the discrete smallish palms may be under abnormal growth stage because their growing conditions are not as well as others that were planted at the same time, which are supposed to be taken special care and inspection.

For dead palms, mismanaged palms and yellowish palms, we locate the exact position of each tree in Fig. 15. The red and blue points show the locations of dead oil palms and mismanaged oil palms that we

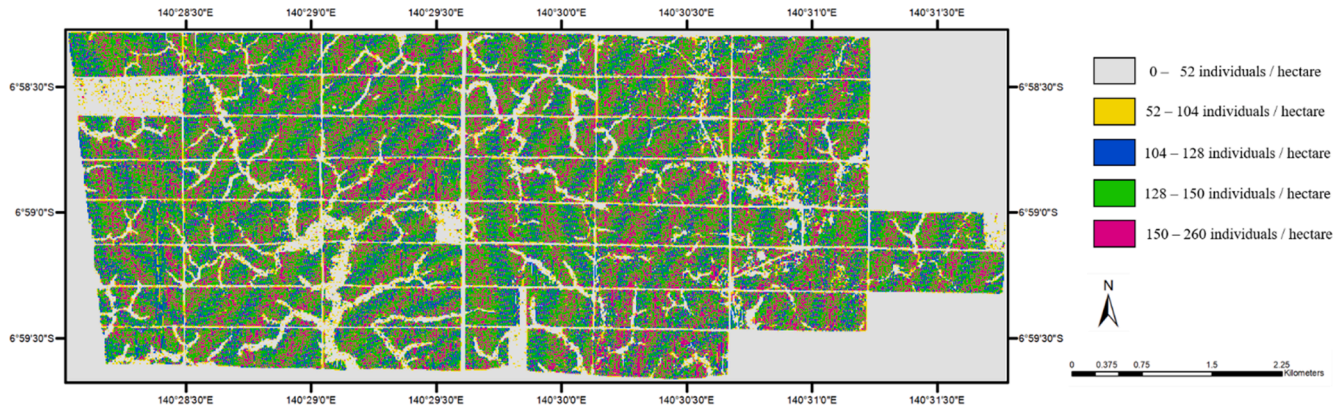

Fig. 12. The distribution of all oil palms using MOPAD.

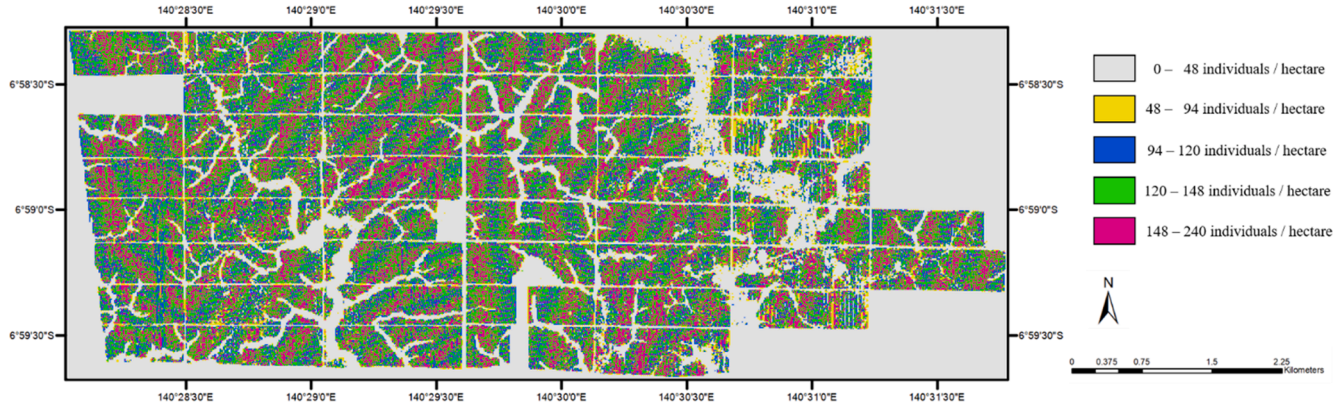

Fig. 13. The distribution of detected healthy oil palms using MOPAD.

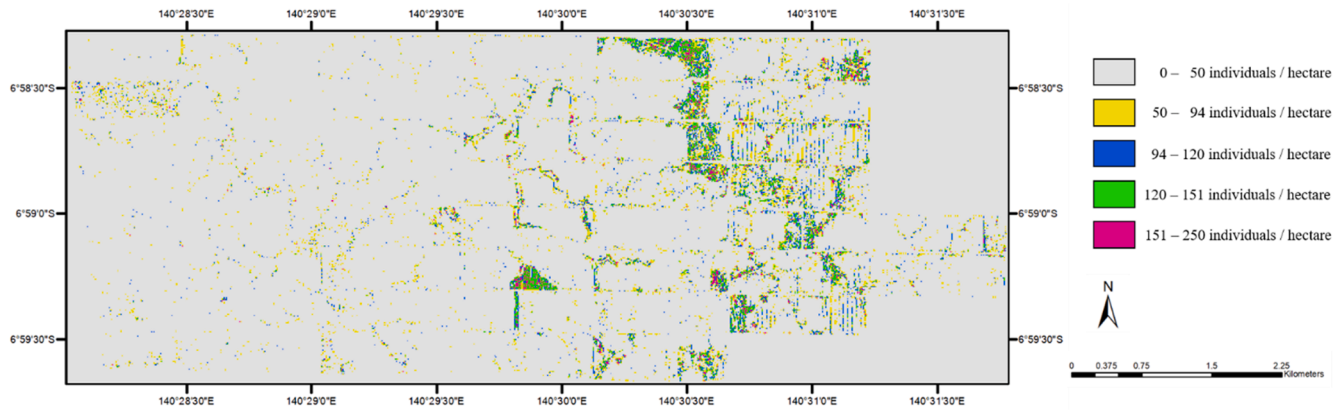

Fig. 14. The distribution of detected smallish oil palms using MOPAD.

detected, respectively. It is practical for oil palm smallholders and industries to monitor the growth of these trees and take some actions in time. For example, farmers can conveniently and efficiently find the dead palms and the mismanaged palms, so that they can replant them or weed the fields to make better management. In addition, through accurate detection of yellowish oil palm trees (denoted by yellow points in Fig. 15), oil palm industries can duly observe the health conditions of oil palms and treat pests and diseases to prevent spreading the whole plantation.

6.2. *Detecting vacancies for oil palm replanting*

Finding oil palm vacancies can effectively help farmers to replant the oil palm, increasing the harvest field and taking full advantage of the oil palm plantation area. Our proposed method for seeking vacancies of oil palms is shown in Figure 17. Firstly, we attain the detection results from the proposed MOPAD algorithm, and we split the whole UAV images by 1024×1024 pixels with a particular overlap area. Secondly, we select the split areas that have detected palms over N in order to remove regions with buildings, roads, other trees that have a low number of palms. Then we generate oil palm masks for our selected areas, which can be shown in the middle of Fig. 16. The white area denotes the oil palm area,

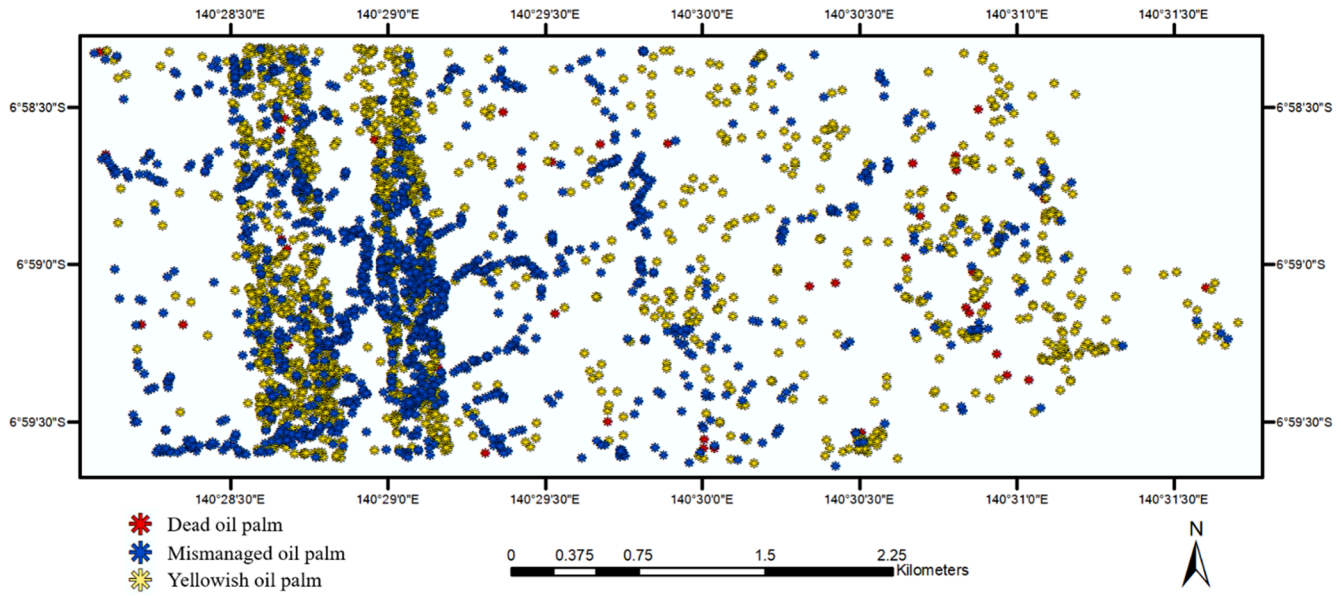

Fig. 15. The location of detected dead oil palms (denoted by red points), mismanaged oil palms (denoted by blue points) and yellowish oil palms (denoted by yellow points) using MOPAD. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

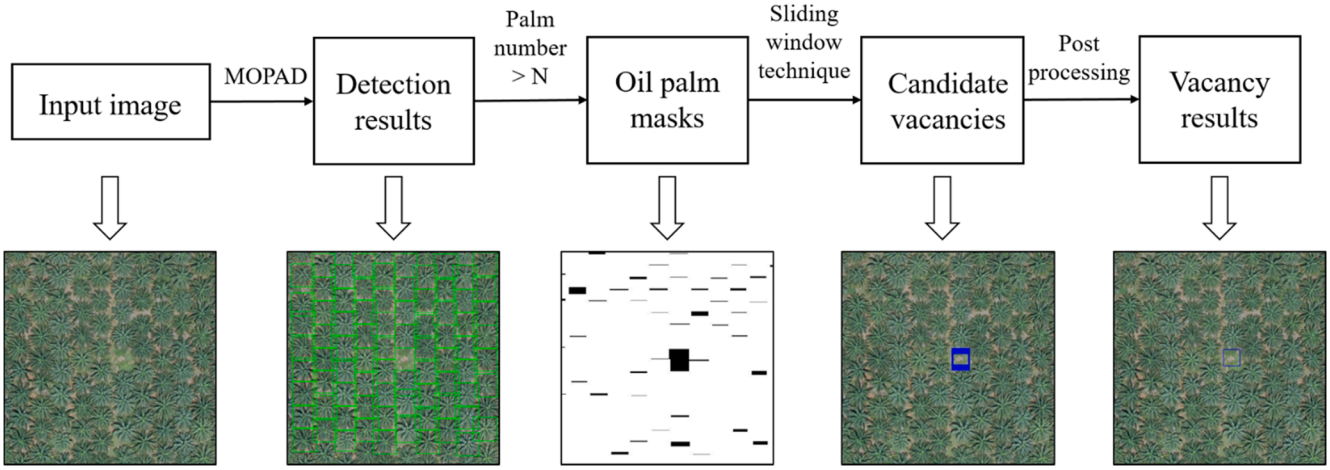

Fig. 16. Our method of finding vacancy of oil palms.

while the black area denotes the non-oil palm area. Following that, candidate vacancies are acquired by sliding window technique with a size of W × W. Finally, we merge the candidate vacancies to ensure that the distance between the center of two oil palm vacancies is more than D. In our experiments, N, W and D are set as 60, 80 and 120, respectively. Fig. 17 displays some vacancy detection results, in which the left shows examples of correctly detected vacancies and the right shows examples of vacancy detection mistakes. Because our vacancy detection method depends on the detection results of MOPAD, some false positives or false negatives of MOPAD may result in mistakes of vacancy detection.

7. Discussion

In the following, we first analyze the transferability of our proposed MOPAD for tree crown detection in Section 7.1. And then, we discuss the practical effect of growing status observation for oil palm plantation based on our MOPAD in Section 7.2 and envision the potential future works in Section 7.3.

7.1. The transferability of our proposed MOPAD for tree crown detection

In this section, we analyze the transferability of the tree crown detection model. In the experiment, we have two transfer tasks. The first one is that we train the MOPAD with training samples from Site 1 and test it on Site 2. The other one is that we train the MOPAD with training samples from Site 2 and test it on Site 1. For comparative purposes, we also present the result trained with samples from both Site 1 and Site 2, and test it on Site 1 and Site 2, respectively. Note that we resample UAV images in Site 2 to 10 cm, which is consistent with the spatial resolution of the UAV image in Site 1. Fig. 18 presents the transfer matrices of F1-score for five classes (i.e., Healthy palm, dead palm, mismanaged palm, smallish palm and yellowish palm) and their average. The site names in the left of each matric denote the source domain (training dataset), and the site names below each matric denote the target domain (test dataset). For the average F1-score, the model trained on Site 1, Site 2 and Site 1 & 2 achieves 72.83%, 59.50% and 63.15% when test on Site 1 while achieves 57.83%, 70.57% and 62.79% when test on Site 2. We can easily observe that the best detection results for both Site 1 and Site 2 are obtained from training on their own. In addition, the model trained with

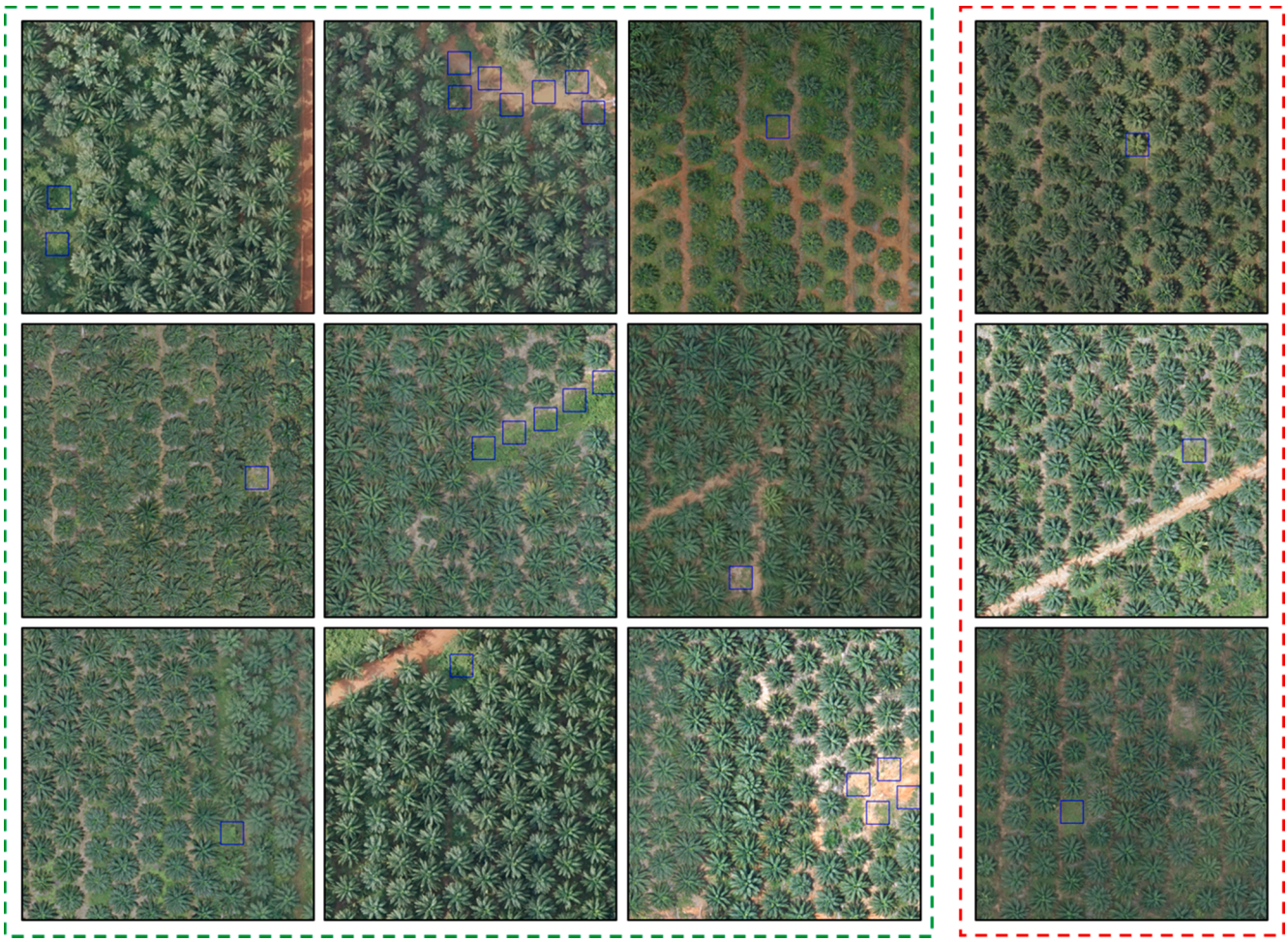

Fig. 17. Examples of our oil palm vacancy detection results. The blue squares denote the vacancies detected by our method. The left shows the correctly detected vacancies and the right shows wrong or missing vacancies because of the FP and the FN from our proposed MOPAD.

samples from Site 1 & 2 also performs worse than the model trained with samples from their own regions.

Due to the domain shift issue (Tuia et al., 2016; Liu and Shi, 2020) in which each domain has its own environment and features, it is challenging to apply the model without any transferable trick to another domain, resulting in an unavoidable deterioration of the model performance. Although the topic of domain adaptation has been explored in recent years, limited studies were conducted to tackle the domain shift problem in cross-regional or cross-temporal tree crown detection (Wu et al., 2020c; Zheng et al., 2020; Wu et al., 2020b). In the future, we will explore and develop more advanced methods for cross-instrument, cross-regional or cross-temporal tree crown detection problem.

7.2. The practical effect of growing status obsevation for oil palm plantation based on our MOPAD

The oil palm tree is a crucial economic crop in many tropical developing countries and palm oil is one of the most important agricultural export products for Malaysia and Indonesia. The yield of palm oil is highly related to the growing status of oil palm plantations. On the one hand, it is beneficial for oil palm industries and smallholders to understand the health conditions of the oil palm trees in advance, enabling them to better manage the plantation and take action in time. On the other hand, the precise and appropriate management of oil palm plantation is not only conducive to the improvement of the palm oil production under limited plantation area, but also benefit to achieving the production goal without extensively expanding the plantation area. To this end, observation of oil palm's health conditions contributes to alleviating the expansion of oil palm and the threat to tropical forests. For example, the discrete smallish palms may be under abnormal growth stage because their growing conditions are not as good as others that were planted at the same time, which are supposed to be taken special care and inspection. According to the distribution of oil palms in different health conditions, farmers can conveniently and efficiently find the dead palms and the mismanaged palms, so that they can replant them or weed the fields to make better management. Also, through accurate detection of yellowish palms, oil palm industries can duly observe the health conditions of oil palms and treat pests and diseases for plants, preventing them from spreading the whole plantation. However, existing studies mainly focus on growing status mapping for crops or trees using mid-resolution remote sensing images, such as MODIS (Huang et al., 2012) and Landsat (He et al., 2019), instead of precisely locating individual diseased tree using very-high remote sensing images. In our proposed MOPAD, we not only detect oil palm tree crowns with high accuracies, but also classify the oil palms into five growing status for each detecting oil palms: healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms. To the best of our knowledge, we are the first to detect the different growing status of oil palms and the vacancies (see Section 6.2) (Baena et al., 2017; Mubin et al., 2019; Johansen et al., 2020; Selvaraj et al., 2020). Our proposed MOPAD contains two improved strategies compared to other state-of-the-art object detection methods: i.e., the Refined Pyramid Features (RPF) and the hybrid class-balanced loss module, improving the accuracies of detecting smallish palms and the types that have quite low

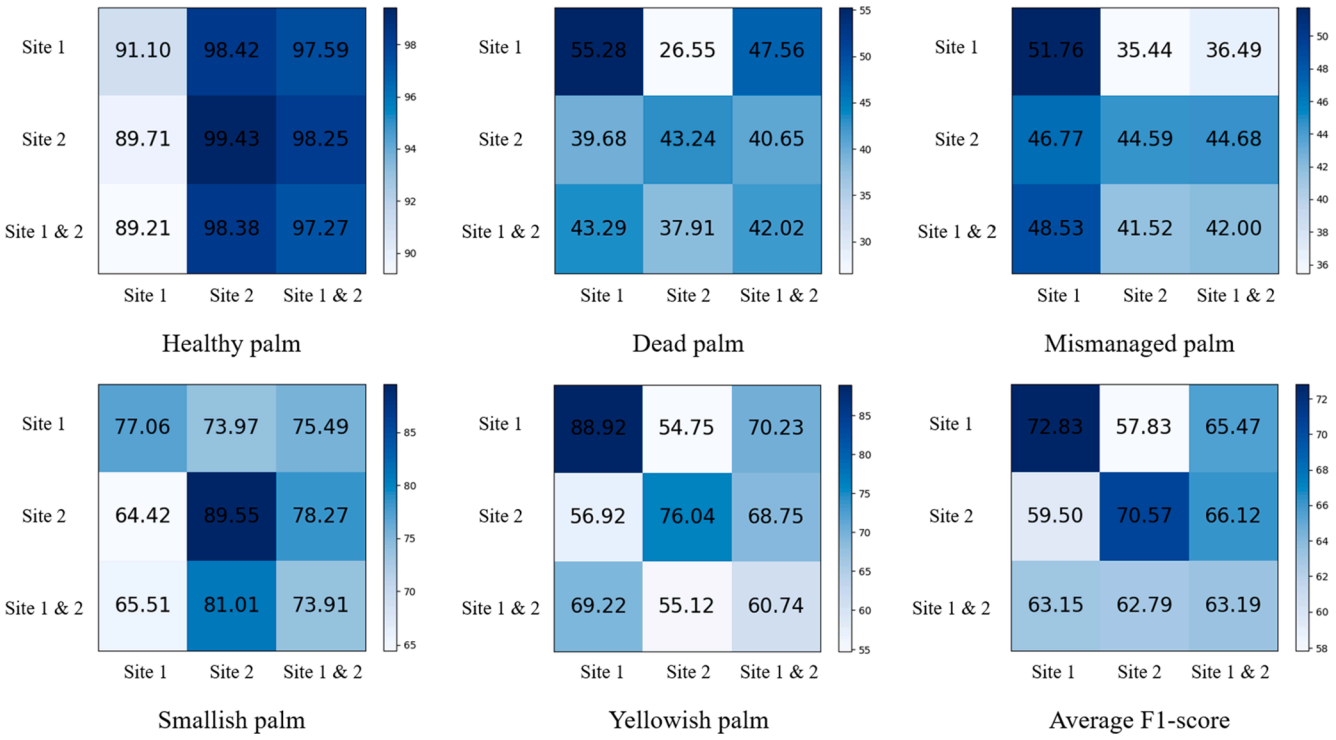

Fig. 18. The transfer matrices of F1-score for five classes (i.e., Healthy palm, dead palm, mismanaged palm, smallish palm and yellowish palm) and their average. The site names in the left of each matric denote the source domain (training dataset), and the site names below each matric denote the target domain (test dataset).

number of training samples. Furthermore, the inference time of our MOPAD is comparable with other state-of-the-art object detection methods, within 15 min to complete the whole Image A (169,374 oil palms, 17.6 km^2). As for traditional machine learning methods (i.e., RF and SVM) and CNN classification based method, the inference speed is relatively slow due to the time-consuming sliding window technique. We believe that the growing status observation for oil palm plantations based on MOPAD contributes to more efficient and smart plantation management of oil palms.

7.3. The potential future works

MOPAD is an encouraging tree crown detection method that demonstrates highly promising results for multi-class oil palm tree detection. Our method suggests that automatically monitoring the different growing status of oil palm trees is applicable for precision plantation management. However, our proposed MOPAD and its applications introduced in Section 6 still exist some limitations and could be further improved. From the data-collection perspective, given the limitation of the visible-band-only UAV images adopted in this paper rather than multi-spectral sensors, we could not exploit the inherent properties among the different growing status of oil palms. RGB-band images only capture the visible representative features, like gray leaves for dead palms and yellow spots for yellowish palms. Multi-spectral UAV images are more capable of adequately capturing intrinsic features for oil palms that are sick (Dash et al., 2017; Johansen et al., 2020), dead (Baena et al., 2017; He et al., 2019) or surrounded by other weeds or plants (Stroppiana et al., 2018). From the application perspective, although we firstly present a vacancy detection method in Section 6.2, the method is based on a brute-force area selection accounting for the number of detected palms and a time-consuming sliding-window-based technique. To our knowledge, combining with oil palm plantation mapping (Dong et al., 2019a; Xu et al., 2020) and designing an end-to-end vacancy detection algorithm might further improve the efficiency (Wu et al., 2020a; Windrim & Bryson, 2020). Furthermore, to broaden our work for smart oil palm plantation management, our work can be further applied to improve yield and biomass prediction (You et al., 2017; Zhou et al., 2017; Maimaitijiang et al., 2020) for individual oil palms, with the different growing status taken into consideration.

8. Conclusions

In this paper, we propose an accurate and multi-class oil palm tree detection algorithm using UAV images, i.e., Multi-class Oil PAlm Detection (MOPAD). MOPAD comprises three main procedures. Firstly, we design a Refined Pyramid Features (RPF) after feature extraction. RPF integrates multi-scale features to attain more accurate oil palm detection results like smallish palms. Secondly, we inherit the Region Proposal Network (RPN) from Faster R-CNN with our designed anchor sizes for oil palms. Finally, we employ a Class-Balanced Cross-Entropy Loss (CBCEL) and a Class-Balanced Smooth L_1 Loss (CBSLL) to re-weight the multi-class classification and regression loss because of the extremely imbalanced class number.

We evaluate our proposed method using three large-scale oil palm plantation images photographed by UAV in two sites located in Indonesia (Site 1 and Site 2). Our study area consists of over 360,000 oil palms with five classes of oil palms, including healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms. The healthy class has the most number of oil palm, and there exists an extremely imbalanced problem among the five classes. Our comprehensive ablation experiments show that our RPF and hybrid class-balanced loss module significantly increase the accuracies for multi-class oil palm tree detection. Our proposed MOPAD improves the average F1-score by 17.09% and 21.32% for multi-class oil palm detection compared with the Baseline (Faster R-CNN) in Site 1 and Site 2, respectively. Our method has an overall oil palm detection F1-score of 87.91% and 99.04%, and outperforms other state-of-the-art object detection algorithms (Cascade R-CNN, Libra Faster R-CNN, etc.) with an improvement of 10.37%-17.09% and 8.14%-21.32% in respect of the average F1-score for multi-class oil palm detection in Site 1 and Site 2, respectively. Furthermore, we introduce two practical applications for smart oil palm plantation management based on our detection results,

including in-depth analysis of different classes of oil palm distribution and practical and useful workflow for detecting oil palm vacancies. Our method demonstrates excellent potential for individual oil palm tree detection and multi-class monitoring of growing status using visible-band-only UAV images, leading oil palm industries and smallholders to precisely manage plantation and efficiently deal with unhealthy oil palms.

In the future, we will explore and develop a more effective and accurate multi-class oil palm tree detection method, and apply them to larger-scale and more complex areas using multi-source and multi-temporal remote sensing images. Furthermore, the strategies proposed in this paper, such as fusing and refining multi-level feature characteristics and the hybrid class-balanced loss module, are not only suitable for detecting other kinds of trees besides oil palms, but also applicative to other precise agriculture applications, such as crop mapping, disease detection, plant species classification, etc. We will explore the potential of our method for other precise agriculture applications in our future work.

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

This research was supported in part by the National Key Research and Development Plan of China (Grant No. 2017YFA0604500, 2017YFB0202204, 2017YFA0604401 and 2019YFA0606601), the National Natural Science Foundation of China (Grant No. 51761135015 and U1839206), the National Key Scientific and Technological Infrastructure project "Earth System Science Numerical Simulator Facility" (EarthLab) and by Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao).

Appendix A. Detailed information for our UAV and our datasets

Figs. A1, A2 and Tables A1 and A2.

Table A1
Data collection information and camera specifications used in this paper.

Index	Component specification
Aerial platform	Skywalker X8
Altitude	500 m for Site 1; 425 m for Site 2
Cruise speed	15–20 m/s
Camera	Sony a6000
Focal length	20 mm
Field of view [Horizontal, Vertical]	[60.87°, 42.61°]
Red [Center wavelength, Bandwidth]	[668 nm, 10 nm]
Green [Center wavelength, Bandwidth]	[560 nm, 20 nm]
Blue [Center wavelength, Bandwidth]	[475 nm, 20 nm]

Table A2
The number of images for training, validation and test dataset in Site 2. The training, validation and test datasets are generated from the training, validation and test regions, respectively.

Type	Training dataset	Validation dataset	Test dataset		
			Image A	Image B	Total
Healthy palm	216,743	17,095	153,955	788,02	232,757
Dead palm	434	71	59	30	89
Mismanaged palm	2658	51	794	242	1036
Smallish palm	47,411	12,480	12,541	1193	13,734
Yellowish palm	5724	931	2025	854	2879
Total	272,970	30,628	169,374	81,121	250,495

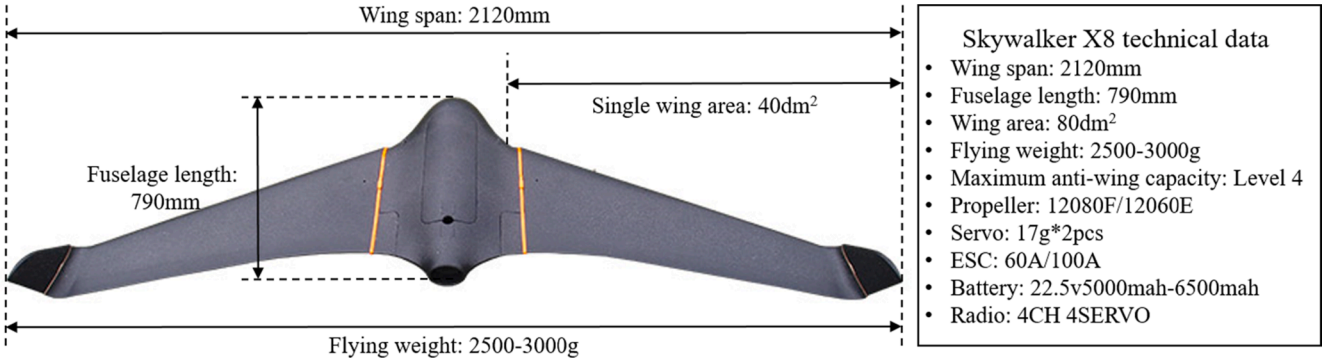

Fig. A1. The structure and technical data of Skywalker X8 used in this paper.

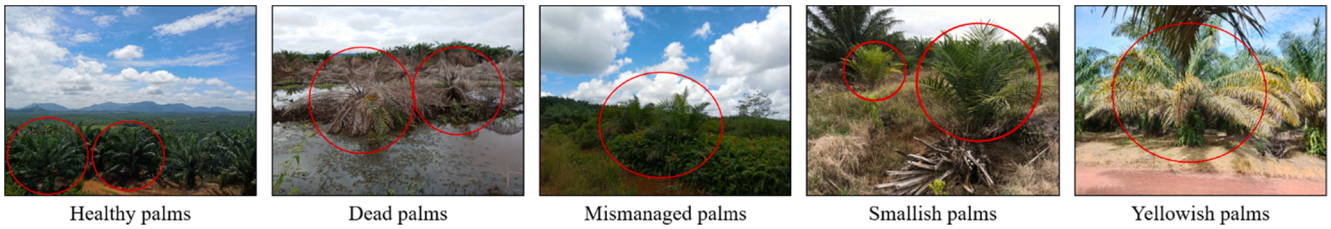

Fig. A2. Examples of different types of oil palms from side view at a closer distance, including healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms.

Appendix B. The supplemental results for Multi-class oil PAlm tree detection (MOPAD)

Figs. B1, B2 and Tables B1–B4.

Fig. B1. Failure cases for FP and FN of oil palm detection in test regions in Site 1.

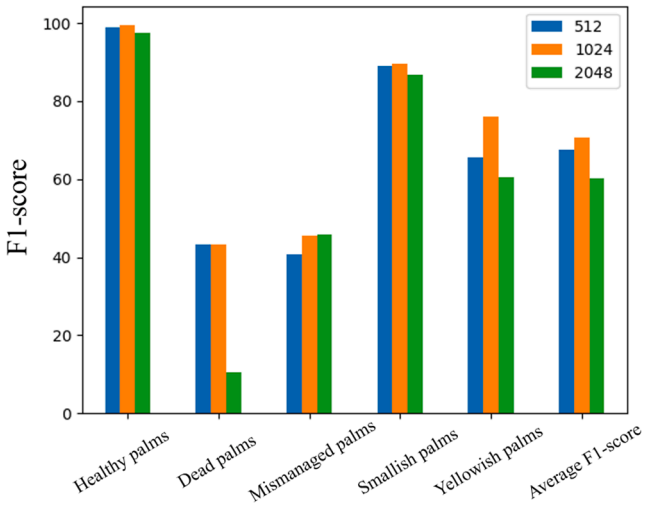

Fig. B2. The F1-scores of different size of input image for our MOPAD.

Table B1
The detection results of our proposed MOPAD for the test regions in Site 2.

Index	Oil palm	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm
TP	248,472	231,186	40	612	12,559	2183
FP	2772	1076	56	1097	1755	680
FN	2023	1571	49	424	1175	696
Precision	98.90%	99.54%	41.67%	35.81%	87.74%	76.25%
Recall	99.19%	99.33%	44.94%	59.07%	91.44%	75.82%
F1-score	99.04%	99.43%	43.24%	44.59%	89.55%	76.04%

Table B2
The confusion matrix for multi-class oil palm detection of our proposed MOPAD in Site 2, in which the top row denotse the ground truth types and the left column denotes the predicted types.

Index	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm
Healthy palm	273,283	8	137	23	639
Dead palm	17	59	3	7	26
Mismanaged palm	392	2	685	24	17
Smallish palm	103	3	93	18,236	8
Yellowish palm	632	7	0	1	2378

Appendix C. The workflow of CNN classification based method for oil palm tree detection

We firstly select six types of samples including background, healthy palms, dead palms, mismanaged palms, smallish palms and yellowish palms. Background denotes other land cover types except oil palms, such as other vegetation, buildings, roads, etc. The training workflow is displayed in Fig. C1. For fair comparison, we adopt ResNet-101 as our feature extractor. Note that the size of selected samples is 80 × 80 pixels for Site 1 and 120 × 120 pixels for Site 2, respectively. Before input the network, we resize the original image to 224 × 224 pixels for appropriate size.

For inference procedure, we adopt sliding window technique to

Table B3
The F1-scores of other state-of-the-art tree crown detection methods in Site 2.

Method	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
RF	81.06%	0.00%	0.12%	42.84%	1.60%	25.12%
SVM	86.19%	0.00%	0.73%	40.95%	2.36%	26.05%
CNN (ResNet-101)	87.11%	3.25%	2.63%	46.57%	4.44%	28.80%
Faster R-CNN	98.58%	7.94%	17.46%	78.25%	44.01%	49.25%
Grid R-CNN	98.52%	21.79%	25.17%	79.39%	51.38%	55.25%
GA Faster R-CNN	98.36%	26.57%	42.36%	79.70%	49.87%	59.37%
Cascade R-CNN	98.67%	17.96%	41.05%	80.37%	67.41%	61.09%
Libra Faster R-CNN	98.40%	22.62%	41.84%	82.63%	66.64%	62.43%
MOPAD (ours)	**99.43%**	**43.24%**	**44.59%**	**89.55%**	**76.04%**	**70.57%**

Table B4
The F1-scores of ablation study of the anchor sizes and aspect ratios.

Ratios	Scale	Healthy palm	Dead palm	Mismanaged palm	Smallish palm	Yellowish palm	Average F1-score
{1:2, 1:1, 2:1}	{128^2, 256^2, 512^2}	98.12%	6.98%	**18.83%**	68.04%	42.49%	46.89%
{1:1}	{128^2, 256^2, 512^2}	98.37%	8.63%	17.46%	65.80%	**47.41%**	47.53%
{1:2, 1:1, 2:1}	{32^2, 64^2, 128^2, 256^2}	98.40%	**8.66%**	18.65%	78.07%	43.56%	**49.47%**
{1:1}	{32^2, 64^2, 128^2, 256^2}	**98.58%**	7.94%	17.46%	**78.25%**	44.01%	49.25%

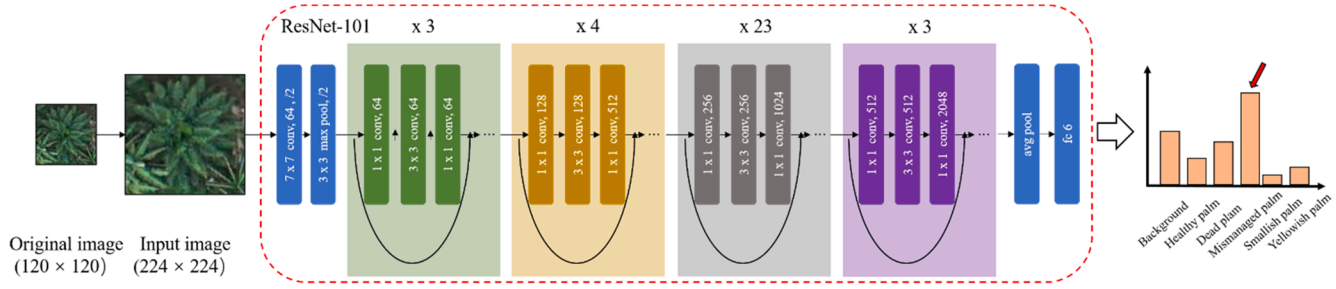

Fig. C1. The training workflow of CNN classification based method for multi-class oil palm tree detection (ResNet-101).

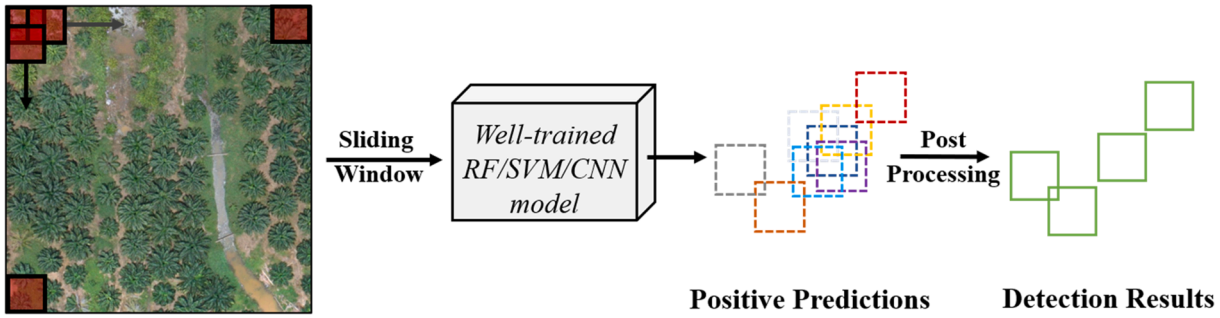

Fig. C2. The inference workflow of RF/CNN/CNN classification based method for oil palm tree detection.

predict all oil palms in the whole image, which is a time-consuming scheme. The size of the sliding window is 80 × 80 pixels for Site 1 and 120 × 120 pixels for Site 2, which is consistent with the training phase. In this study, the sliding step is set as five pixels through experimental tests. After the labels of all samples in the test regions are predicted, we merge the overlapping samples according to the Euclidean distance (Li et al., 2017). Fig. C2 illustartes the overall procedures of our inference workflow for CNN classification based method. Note that the traditional machine learning based methods, i.e., RF and SVM adopt the same inference workflow.

Appendix D. More detection results for our MOPAD and other state-of-the-art methods

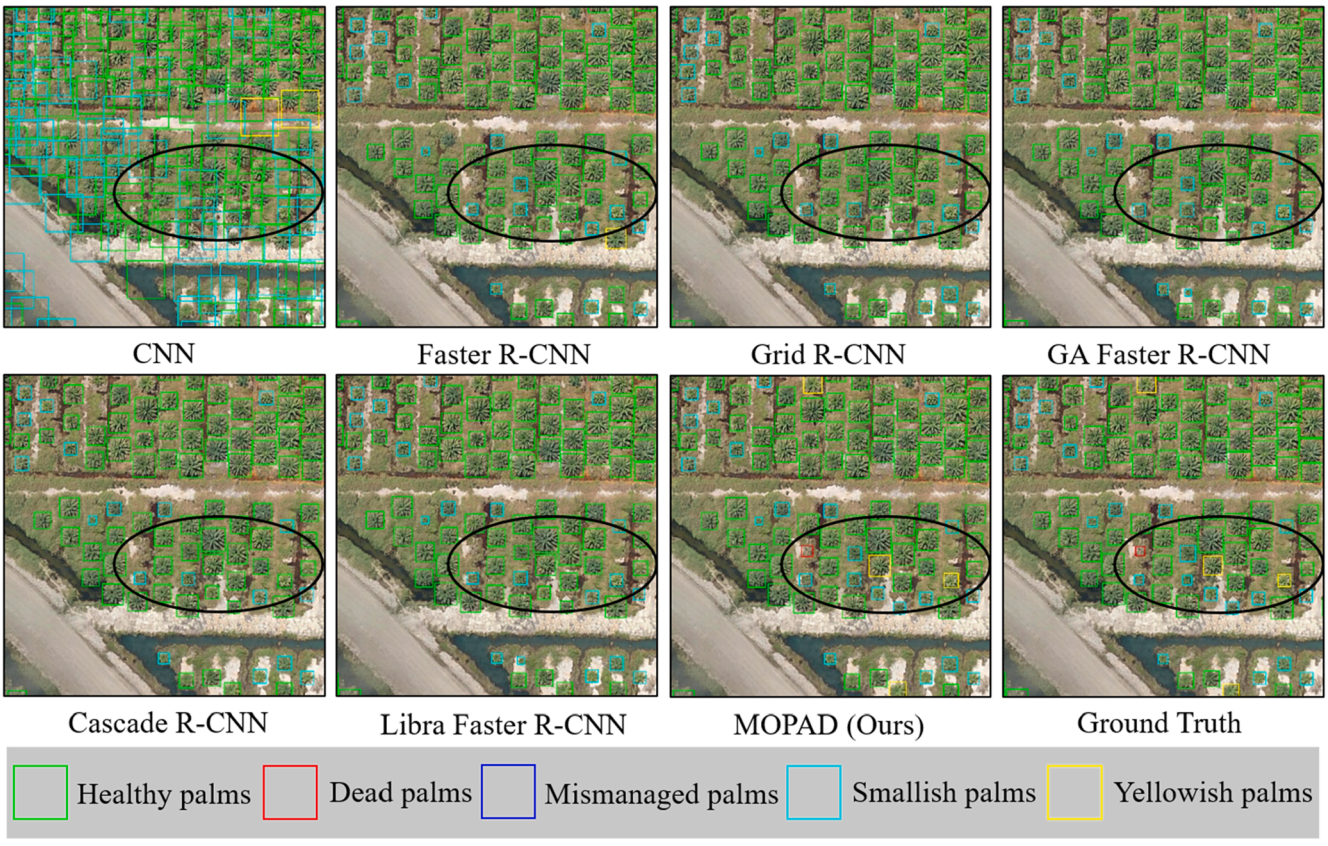

Fig. D1. Multi-class oil palm tree detection results for Region 3 in Site 1. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

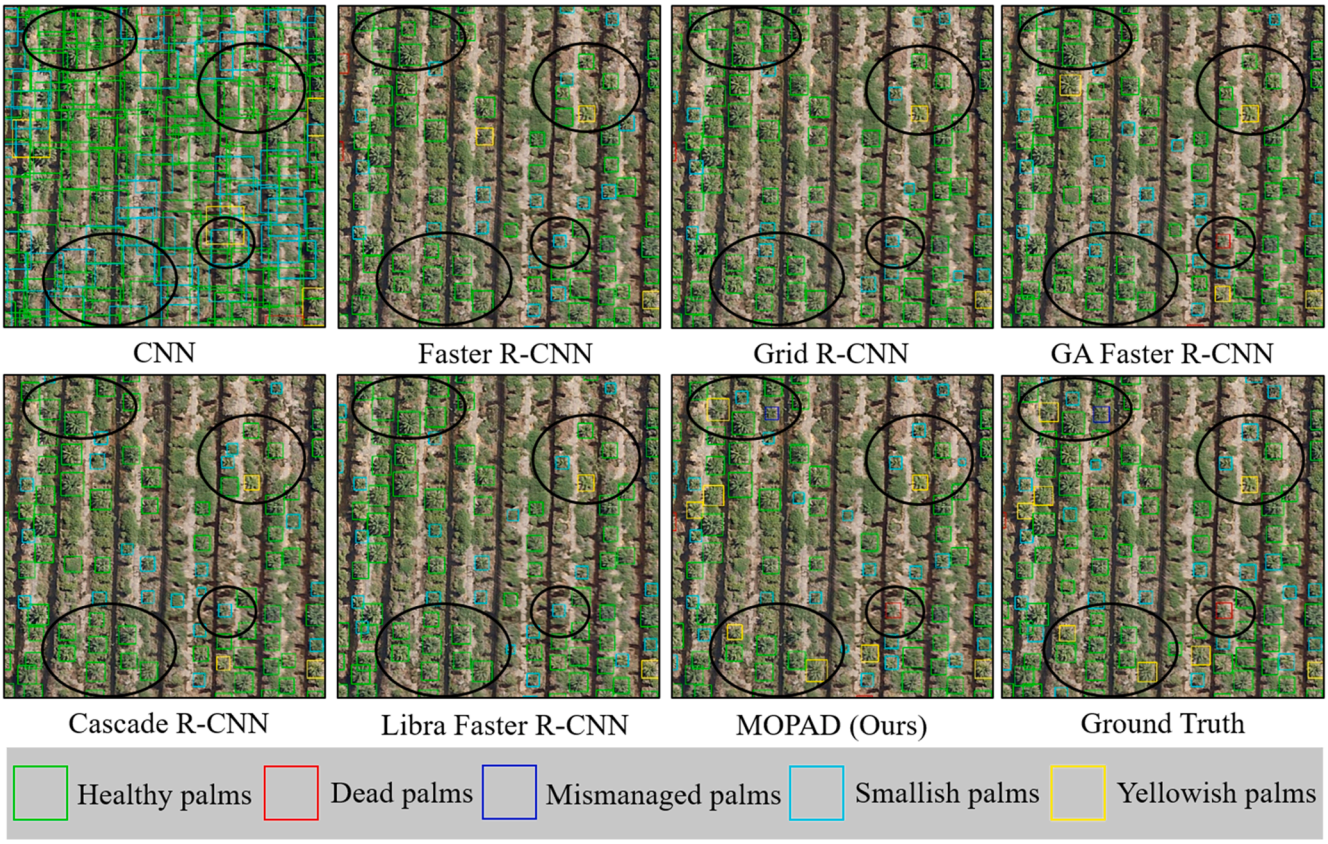

Fig. D2. Multi-class oil palm tree detection results for Region 4 in Site 1. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

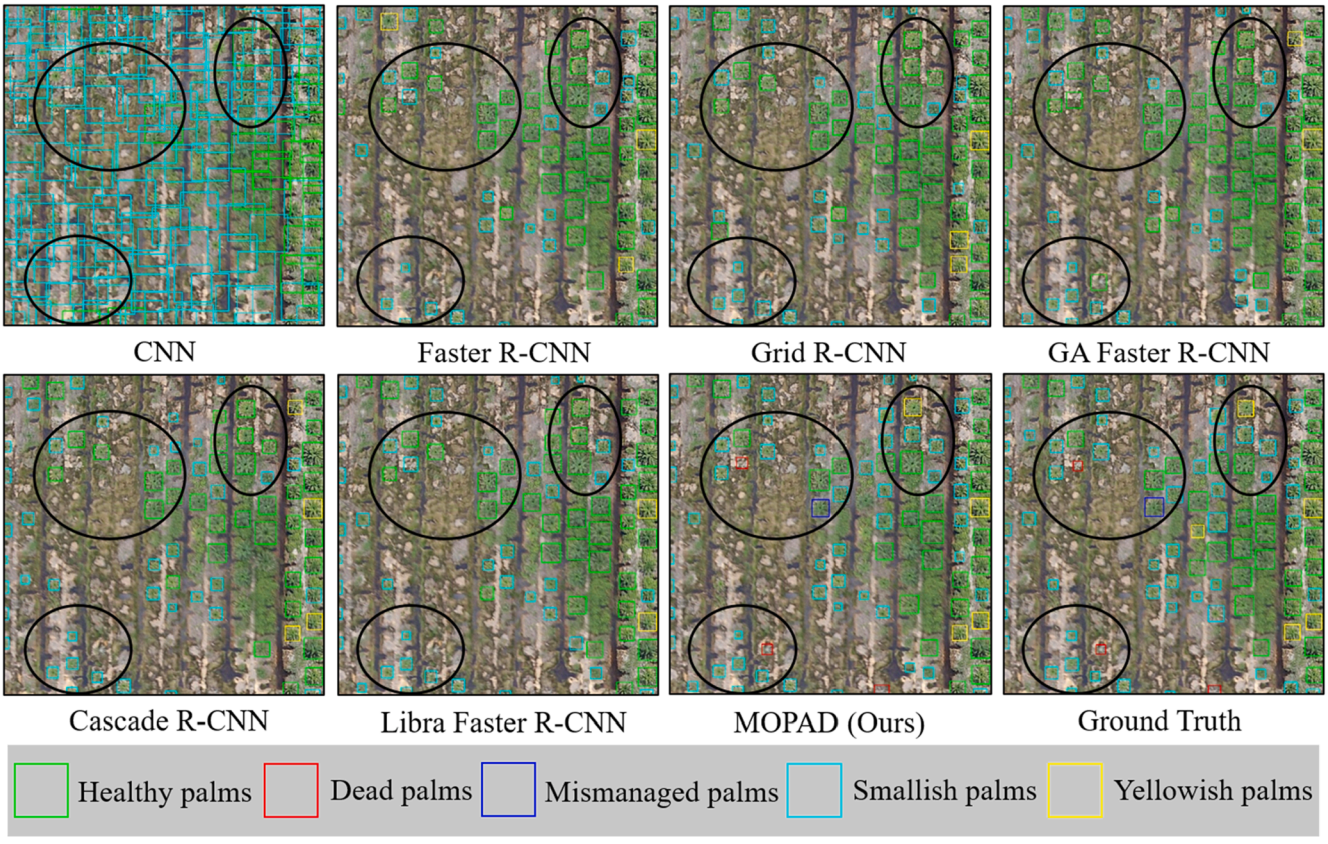

Fig. D3. Multi-class oil palm tree detection results for Region 5 in Site 1. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

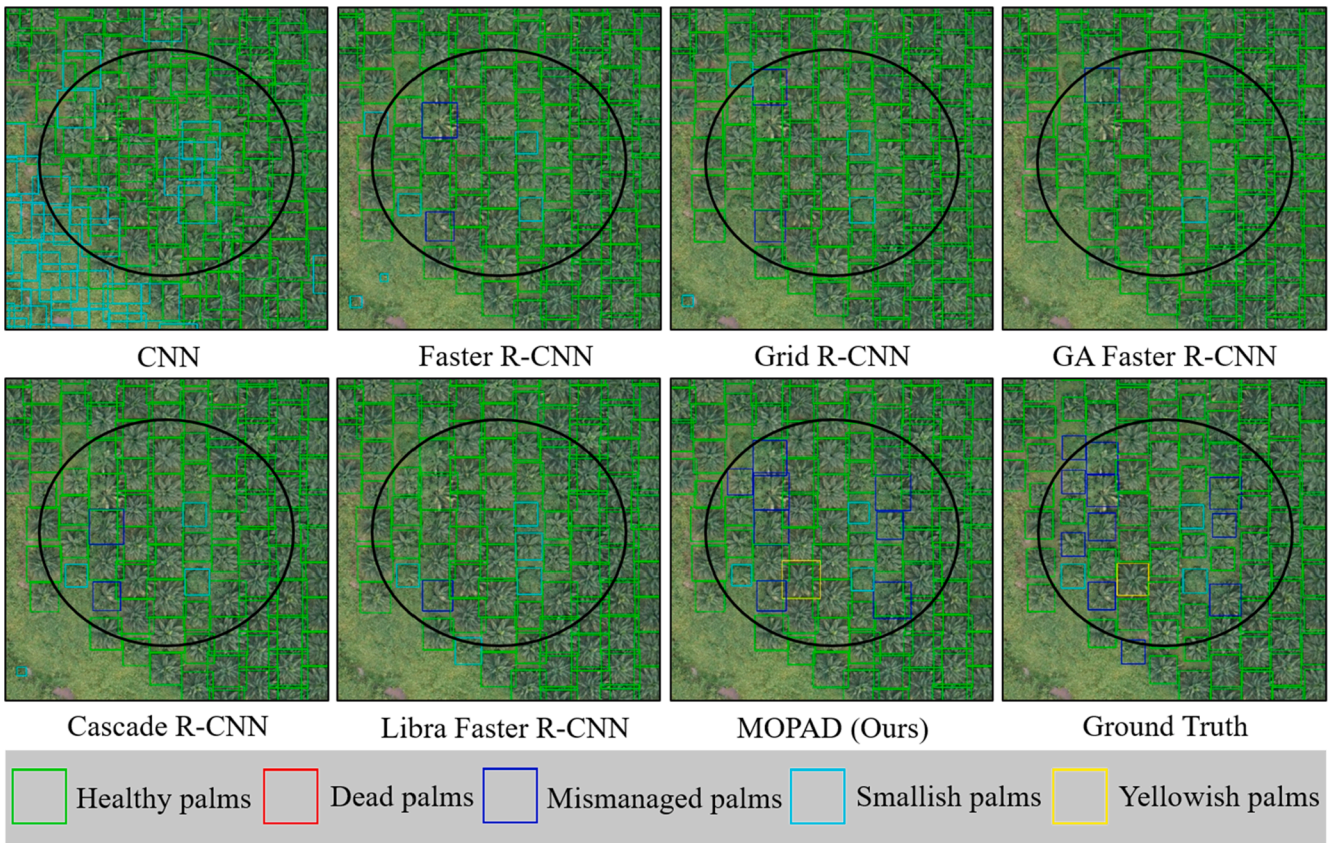

Fig. D4. Multi-class oil palm tree detection results for Region 3 in Site 2. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

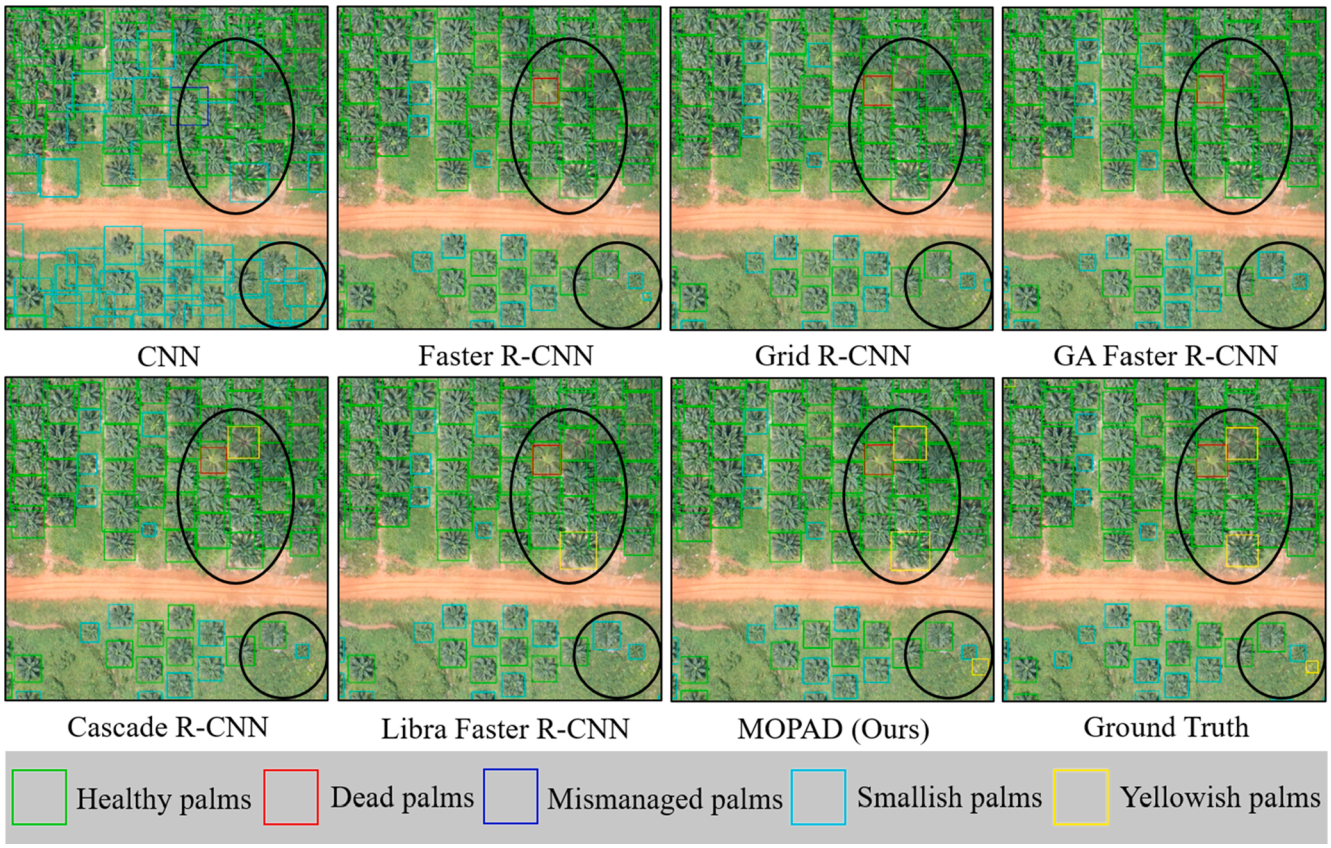

Fig. D5. Multi-class oil palm tree detection results for Region 4 in Site 2. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

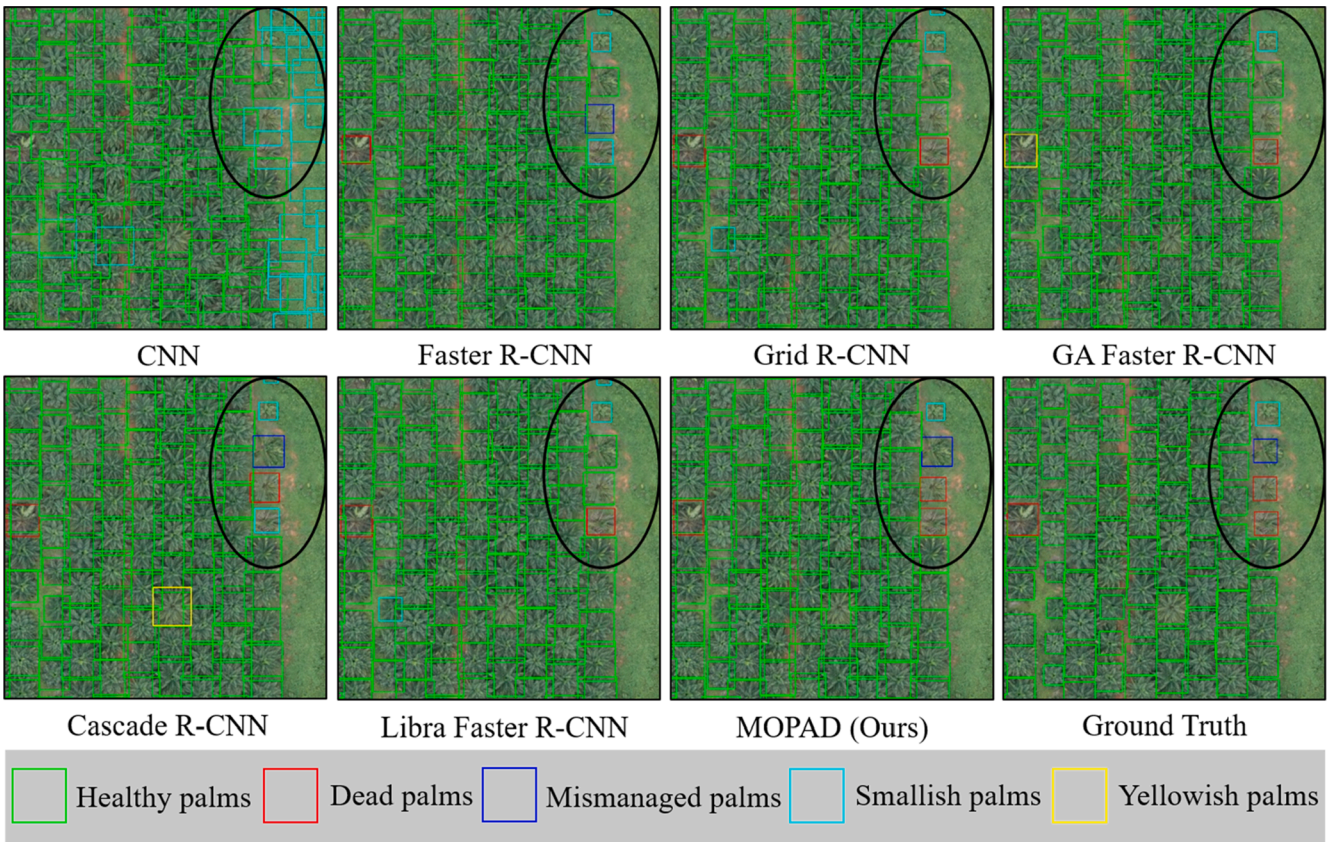

Fig. D6. Multi-class oil palm tree detection results for Region 5 in Site 2. The black circles denote some examples where our MOPAD outperforms other tree crown detection methods.

References

Alexander, C., Korstjens, A.H., Hankinson, E., Usher, G., Harrison, N., Nowak, M.G., et al., 2018. Locating emergent trees in a tropical rainforest using data from an Unmanned Aerial Vehicle (UAV). Int. J. Appl. Earth Obs. Geoinf. 72, 86–90. https://doi.org/10.1016/j.jag.2018.05.024.

Ardila, J.P., Tolpekin, V.A., Bijker, W., Stein, A., 2011. Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images. ISPRS J. Photogramm. Remote Sens. 66 (6), 762–775. https://doi.org/10.1016/j.isprsjprs.2011.08.002.

Ardila, J.P., Bijker, W., Tolpekin, V.A., Stein, A., 2012. Multitemporal change detection of urban trees using localized region-based active contours in VHR images. Remote Sens. Environ. 124, 413–426. https://doi.org/10.1016/j.rse.2012.05.027.

Austin, K.G., Mosnier, A., Pirker, J., McCallum, I., Fritz, S., Kasibhatla, P.S., 2017. Shifting patterns of oil palm driven deforestation in Indonesia and implications for zero-deforestation commitments. Land Use Policy 69, 41–48. https://doi.org/10.1016/j.landusepol.2017.08.036.

Aval, J., Demuynck, J., Zenou, E., Fabre, S., Sheeren, D., Fauvel, M., et al., 2018. Detection of individual trees in urban alignment from airborne data and contextual information: A marked point process approach. ISPRS J. Photogramm. Remote Sens. 146, 197–210. https://doi.org/10.1016/j.isprsjprs.2018.09.016.

Baena, S., Moat, J., Whaley, O., Boyd, D.S., 2017. Identifying species from the air: UAVs and the very high resolution challenge for plant conservation. PLoS ONE 12 (11), e0188714. https://doi.org/10.1371/journal.pone.0188714.

Balasundram, S.K., Memarian, H., Khosla, R., 2013. Estimating oil palm yields using vegetation indices derived from Quickbird. Life Science Journal 10 (4), 851–860.

Bayraktar, E., Basarkan, M.E., Celebi, N., 2020. A low-cost UAV framework towards ornamental plant detection and counting in the wild. ISPRS J. Photogramm. Remote Sens. 167, 1–11. https://doi.org/10.1016/j.isprsjprs.2020.06.012.

Bhardwaj, A., Sam, L., Martín-Torres, F.J., Kumar, R., 2016. UAVs as remote sensing platform in glaciology: Present applications and future prospects. Remote Sens. Environ. 175, 196–204. https://doi.org/10.1016/j.rse.2015.12.029.

Blomley, R., Hovi, A., Weinmann, M., Hinz, S., Korpela, I., Jutzi, B., 2017. Tree species classification using within crown localization of waveform LiDAR attributes. ISPRS J. Photogramm. Remote Sens. 133, 142–156. https://doi.org/10.1016/j.isprsjprs.2017.08.013.

Bottou, L., 2010. Large-Scale Machine Learning with Stochastic Gradient Descent. In: Proceedings of COMPSTAT'2010: 19th International Conference on Computational Statistics Paris France, August 22-27, 2010 Keynote, Invited and Contributed Papers. Springer Science & Business Media, p. 177.

Braga, J.R., Peripato, V., Dalagnol, R., Ferreira, M.P., Tarabalka, Y., Aragão, L.E., et al., 2020. Tree crown delineation algorithm based on a convolutional neural network. Remote Sensing 12 (8), 1288. https://doi.org/10.3390/rs12081288.

Brandt, M., Tucker, C.J., Kariryaa, A., Rasmussen, K., Abel, C., Small, J., et al., 2020. An unexpectedly large count of trees in the West African Sahara and Sahel. Nature 587 (7832), 78–82. https://doi.org/10.1038/s41586-020-2824-5.

Cai, Z., Vasconcelos, N., 2018. Cascade r-cnn: Delving into high quality object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 6154–6162.

Campbell, M.J., Dennison, P.E., Tune, J.W., Kannenberg, S.A., Kerr, K.L., Codding, B.F., Anderegg, W.R., 2020. A multi-sensor, multi-scale approach to mapping tree mortality in woodland ecosystems. Remote Sens. Environ. 245, 111853 https://doi.org/10.1016/j.rse.2020.111853.

Carlson, K.M., Curran, L.M., Asner, G.P., Pittman, A.M., Trigg, S.N., Adeney, J.M., 2013. Carbon emissions from forest conversion by Kalimantan oil palm plantations. Nat. Clim. Change 3 (3), 283–287. https://doi.org/10.1038/nclimate1702.

Chemura, A., van Duren, I., van Leeuwen, L.M., 2015. Determination of the age of oil palm from crown projection area detected from WorldView-2 multispectral remote sensing data: The case of Ejisu-Juaben district, Ghana. ISPRS J. Photogramm. Remote Sens. 100, 118–127. https://doi.org/10.1016/j.isprsjprs.2014.07.013.

Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., et al., 2019. MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155.

Cheng, Y., Yu, L., Cracknell, A.P., Gong, P., 2016. Oil palm mapping using Landsat and PALSAR: A case study in Malaysia. Int. J. Remote Sens. 37 (22), 5431–5442. https://doi.org/10.1080/01431161.2016.1241448.

Cheng, Y., Yu, L., Xu, Y., Lu, H., Cracknell, A.P., Kanniah, K., Gong, P., 2019. Mapping oil palm plantation expansion in Malaysia over the past decade (2007–2016) using ALOS-1/2 PALSAR-1/2 data. Int. J. Remote Sens. 1–20. https://doi.org/10.1080/01431161.2019.1580824.

Chong, K.L., Kanniah, K.D., Pohl, C., Tan, K.P., 2017. A review of remote sensing applications for oil palm studies. Geo-spatial Information Science 20 (2), 184–200. https://doi.org/10.1080/10095020.2017.1337317.

Colomina, I., Molina, P., 2014. Unmanned aerial systems for photogrammetry and remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 92, 79–97. https://doi.org/10.1016/j.isprsjprs.2014.02.013.

Cracknell, A.P., Kanniah, K.D., Tan, K.P., Wang, L., 2015. Towards the development of a regional version of MOD17 for the determination of gross and net primary productivity of oil palm trees. Int. J. Remote Sens. 36 (1), 262–289. https://doi.org/10.1080/01431161.2014.995278.

Csillik, O., Cherbini, J., Johnson, R., Lyons, A., Kelly, M., 2018. Identification of citrus trees from unmanned aerial vehicle imagery using convolutional neural networks. Drones 2 (4), 39. https://doi.org/10.3390/drones2040039.

Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S., 2019. Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9268–9277.
Dai, W., Yang, B., Dong, Z., Shaker, A., 2018. A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds. ISPRS J. Photogramm. Remote Sens. 144, 400–411. https://doi.org/10.1016/j.isprsjprs.2018.08.010.
Da Silva, J.M., Damásio, C.V., Sousa, A.M., Bugalho, L., Pessanha, L., Quaresma, P., 2015. Agriculture pest and disease risk maps considering MSG satellite data and land surface temperature. Int. J. Appl. Earth Obs. Geoinf. 38, 40–50. https://doi.org/10.1016/j.jag.2014.12.016.
Daliakopoulos, I.N., Grillakis, E.G., Koutroulis, A.G., Tsanis, I.K., 2009. Tree crown detection on multispectral VHR satellite imagery. Photogramm. Eng. Remote Sens. 75 (10), 1201–1211. https://doi.org/10.14358/PERS.75.10.1201.
Dalponte, M., Ørka, H.O., Ene, L.T., Gobakken, T., Næsset, E., 2014. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. Remote Sens. Environ. 140, 306–317. https://doi.org/10.1016/j.rse.2013.09.006.
Dash, J.P., Watt, M.S., Pearse, G.D., Heaphy, M., Dungey, H.S., 2017. Assessing very high resolution UAV imagery for monitoring forest health during a simulated disease outbreak. ISPRS J. Photogramm. Remote Sens. 131, 1–14. https://doi.org/10.1016/j.isprsjprs.2017.07.007.
Deng, L., Mao, Z., Li, X., Hu, Z., Duan, F., Yan, Y., 2018. UAV-based multispectral remote sensing for precision agriculture: a comparison between different cameras. ISPRS J. Photogramm. Remote Sens. 146, 124–136. https://doi.org/10.1016/j.isprsjprs.2018.09.008.
Dong, R., Li, W., Fu, H., Gan, L., Yu, L., Zheng, J., Xia, M., 2019a. Oil palm plantation mapping from high-resolution remote sensing images using deep learning. Int. J. Remote Sens. 41 (5), 2022–2046.
Dong, T., Shen, Y., Zhang, J., Ye, Y., Fan, J., 2019b. Progressive cascaded convolutional neural networks for single tree detection with google earth imagery. Remote Sensing 11 (15), 1786.
Fawcett, D., Azlan, B., Hill, T.C., Kho, L.K., Bennie, J., Anderson, K., 2019. Unmanned aerial vehicle (UAV) derived structure-from-motion photogrammetry point clouds for oil palm (Elaeis guineensis) canopy segmentation and height estimation. Int. J. Remote Sens. 40 (19), 7538–7560. https://doi.org/10.1080/01431161.2019.1591651.
Ferraz, A., Saatchi, S., Mallet, C., Meyer, V., 2016. Lidar detection of individual tree size in tropical forests. Remote Sens. Environ. 183, 318–333. https://doi.org/10.1016/j.rse.2016.05.028.
Ferreira, M.P., de Almeida, D.R.A., de Almeida Papa, D., Minervino, J.B.S., Veras, H.F.P., Formighieri, A., et al., 2020. Individual tree detection and species classification of Amazonian palms using UAV images and deep learning. For. Ecol. Manage. 475, 118397 https://doi.org/10.1016/j.foreco.2020.118397.
Freudenberg, M., Nölke, N., Agostini, A., Urban, K., Wörgötter, F., Kleinn, C., 2019. Large scale palm tree detection in high resolution satellite images using U-Net. Remote Sensing 11 (3), 312. https://doi.org/10.3390/rs11030312.
Giannakis, I., Tosti, F., Lantini, L., Alani, A.M., 2019. Health monitoring of tree trunks using ground penetrating radar. IEEE Trans. Geosci. Remote Sens. 57 (10), 8317–8326. https://doi.org/10.1109/TGRS.2019.2920224.
Gomes, M.F., Maillard, P., Deng, H., 2018. Individual tree crown detection in sub-meter satellite imagery using Marked Point Processes and a geometrical-optical model. Remote Sens. Environ. 211, 184–195. https://doi.org/10.1016/j.rse.2018.04.002.
Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., Herrera, F., 2017. Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study. Remote Sensing 9 (12), 1220. https://doi.org/10.3390/rs9121220.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.
He, Y., Chen, G., Potter, C., Meentemeyer, R.K., 2019. Integrating multi-sensor remote sensing and species distribution modeling to map the spread of emerging forest disease and tree mortality. Remote Sens. Environ. 231, 111238 https://doi.org/10.1016/j.rse.2019.111238.
Huang, Q., Li, D., Chen, Z., Liu, J., Wang, L., 2012. Monitoring of planting area and growth condition of winter wheat in China based on MODIS data. Nongye Jixie Xuebao= Trans. Chinese Soc. Agric. Mach. 43 (7), 163–167.
Hung, C., Bryson, M., Sukkarieh, S., 2012. Multi-class predictive template for tree crown detection. ISPRS J. Photogramm. Remote Sens. 68, 170–183. https://doi.org/10.1016/j.isprsjprs.2012.01.009.
Hyyppä, E., Hyyppä, J., Hakala, T., Kukko, A., Wulder, M.A., White, J.C., et al., 2020. Under-canopy UAV laser scanning for accurate forest field measurements. ISPRS J. Photogramm. Remote Sens. 164, 41–60. https://doi.org/10.1016/j.isprsjprs.2020.03.021.
Itakura, K., Hosoi, F., 2020. Automatic tree detection from three-dimensional images reconstructed from 360° spherical camera using YOLO v2. Remote Sensing 12 (6), 988. https://doi.org/10.3390/rs12060988.
Jiang, S., Jiang, C., Jiang, W., 2020. Efficient structure from motion for large-scale UAV images: a review and a comparison of SfM tools. ISPRS J. Photogramm. Remote Sens. 167, 230–251. https://doi.org/10.1016/j.isprsjprs.2020.04.016.
Johansen, K., Duan, Q., Tu, Y.H., Searle, C., Wu, D., Phinn, S., et al., 2020. Mapping the condition of macadamia tree crops using multi-spectral uav and worldview-3 imagery. ISPRS J. Photogramm. Remote Sens. 165, 28–40. https://doi.org/10.1016/j.isprsjprs.2020.04.017.
Kanniah, K.D., Tan, K.P., Cracknel, A.P., 2012. UK-DMC 2 satellite data for deriving biophysical parameters of oil palm trees in Malaysia. IEEE, pp. 6569–6572.
Ke, Y., Quackenbush, L.J., 2011. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. Int. J. Remote Sens. 32 (17), 4725–4747. https://doi.org/10.1080/01431161.2010.494184.
Kellenberger, B., Marcos, D., Tuia, D., 2018. Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. Remote Sens. Environ. 216, 139–153. https://doi.org/10.1016/j.rse.2018.06.028.
Kellenberger, B., Marcos, D., Lobry, S., Tuia, D., 2019. Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning. IEEE Trans. Geosci. Remote Sens. 57 (12), 9524–9533. https://doi.org/10.1109/TGRS.2019.2927393.
Khosravipour, A., Skidmore, A.K., Wang, T., Isenburg, M., Khoshelham, K., 2015. Effect of slope on treetop detection using a LiDAR Canopy Height Model. ISPRS J. Photogramm. Remote Sens. 104, 44–52. https://doi.org/10.1016/j.isprsjprs.2015.02.013.
Koh, L.P., Wilcove, D.S., 2007. Cashing in palm oil for conservation. Nature 448 (7157), 993. https://doi.org/10.1038/448993a.
Li, W., Fu, H., Yu, L., Cracknell, A., 2017. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. Remote Sensing 9 (1), 22. https://doi.org/10.3390/rs9010022.
Li, W., Dong, R., Fu, H., Yu, L., 2019a. Large-scale oil palm tree detection from high-resolution satellite images using two-stage convolutional neural networks. Remote Sensing 11 (1), 11.
Li, W., He, C., Fu, H., Zheng, J., Dong, R., Xia, M., et al., 2019b. A real-time tree crown detection approach for large-scale remote sensing images on FPGAs. Remote Sensing 11 (9), 1025.
Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2020. Object detection in optical remote sensing images: a survey and a new benchmark. ISPRS J. Photogramm. Remote Sens. 159, 296–307. https://doi.org/10.1016/j.isprsjprs.2019.11.023.
Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125.
Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C., 2016. Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, Cham, pp. 21–37.
Liu, K., Shen, X., Cao, L., Wang, G., Cao, F., 2018. Estimating forest structural attributes using UAV-LiDAR data in Ginkgo plantations. ISPRS J. Photogramm. Remote Sens. 146, 465–482. https://doi.org/10.1016/j.isprsjprs.2018.11.001.
Liu, S., Shi, Q., 2020. Local climate zone mapping as remote sensing scene classification using deep learning: A case study of metropolitan China. ISPRS J. Photogramm. Remote Sens. 164, 229–242. https://doi.org/10.1016/j.isprsjprs.2020.04.008.
Lu, X., Li, B., Yue, Y., Li, Q., Yan, J., 2019. Grid r-cnn. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7363–7372.
Ma, Z., Pang, Y., Wang, D., Liang, X., Chen, B., Lu, H., et al., 2020. Individual tree crown segmentation of a larch plantation using airborne laser scanning data based on region growing and canopy morphology features. Remote Sensing 12 (7), 1078. https://doi.org/10.3390/rs12071078.
Mahlein, A.K., Rumpf, T., Welke, P., Dehne, H.W., Plümer, L., Steiner, U., Oerke, E.C., 2013. Development of spectral indices for detecting and identifying plant diseases. Remote Sens. Environ. 128, 21–30. https://doi.org/10.1016/j.rse.2012.09.019.
Maimaitijiang, M., Sagan, V., Sidike, P., Maimaitiyiming, M., Hartling, S., Peterson, K.T., Fritischi, F.B., 2019. Vegetation Index Weighted Canopy Volume Model (CVMVI) for soybean biomass estimation from Unmanned Aerial System-based RGB imagery. ISPRS J. Photogramm. Remote Sens. 151, 27–41. https://doi.org/10.1016/j.isprsjprs.2019.03.003.
Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., Fritschi, F.B., 2020. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sens. Environ. 237, 111599 https://doi.org/10.1016/j.rse.2019.111599.
Malek, S., Bazi, Y., Alajlan, N., AlHichri, H., Melgani, F., 2014. Efficient framework for palm tree detection in UAV images. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (12), 4692–4703. https://doi.org/10.1109/JSTARS.2014.2331425.
Mongus, D., Žalik, B., 2015. An efficient approach to 3D single tree-crown delineation in LiDAR data. ISPRS J. Photogramm. Remote Sens. 108, 219–233. https://doi.org/10.1016/j.isprsjprs.2015.08.004.
Morales, G., Kemper, G., Sevillano, G., Arteaga, D., Ortega, I., Telles, J., 2018. Automatic segmentation of mauritia flexuosa in unmanned aerial vehicle (UAV) imagery using deep learning. Forests 9 (12), 736. https://doi.org/10.3390/f9120736.
Mubin, N.A., Nadarajoo, E., Shafri, H.Z.M., Hamedianfar, A., 2019. Young and mature oil palm tree detection and counting using convolutional neural network deep learning method. Int. J. Remote Sens. 40 (19), 7500–7515. https://doi.org/10.1080/01431161.2019.1569282.
Murray, J., Gullick, D., Blackburn, G.A., Whyatt, J.D., Edwards, C., 2019. ARBOR: a new framework for assessing the accuracy of individual tree crown delineation from remotely-sensed data. Remote Sens. Environ. 231, 111256 https://doi.org/10.1016/j.rse.2019.111256.
Nevalainen, O., Honkavaara, E., Tuominen, S., Viljanen, N., Hakala, T., Yu, X., et al., 2017. Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. Remote Sensing 9 (3), 185. https://doi.org/10.3390/rs9030185.
Osco, L.P., de Arruda, M.D.S., Junior, J.M., da Silva, N.B., Ramos, A.P.M., Moryia, É.A.S., et al., 2020. A convolutional neural network approach for counting and geolocating citrus-trees in UAV multispectral imagery. ISPRS J. Photogramm. Remote Sens. 160, 97–106. https://doi.org/10.1016/j.isprsjprs.2019.12.010.
Özcan, A.H., Ünsalan, C., 2020. Probabilistic object detection and shape extraction in remote sensing data. Computer Vision and Image Understanding, 102953.

Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., Lin, D., 2019. Libra r-cnn: towards balanced learning for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 821–830.
Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., Ye, H., 2020. Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. ISPRS J. Photogramm. Remote Sens. 169, 364–376. https://doi.org/10.1016/j.isprsjprs.2020.08.026.
Pitkänen, J., 2001. Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. Can. J. For. Res. 31 (5), 832–844. https://doi.org/10.1139/x01-013.
Pittman, A.M., Carlson, K., Curran, L.M., Ponette-Gonzalez, A., 2013. NASA satellite data used to study the impact of oil palm expansion across Indonesian Borneo. The Earth Observer 25 (5), 12–15.
Pouliot, D.A., King, D.J., Bell, F.W., Pitt, D.G., 2002. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. Remote Sens. Environ. 82 (2–3), 322–334. https://doi.org/10.1016/S0034-4257(02)00050-0.
Pu, R., Landry, S., 2012. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. Remote Sens. Environ. 124, 516–533. https://doi.org/10.1016/j.rse.2012.06.011.
Puttemans, S., Van Beeck, K., Goedemé, T., 2018. Comparing boosted cascades to deep learning architectures for fast and robust coconut tree detection in aerial images. In: Proceedings of the 13th international joint conference on computer vision, imaging and computer graphics theory and applications, vol. 5, pp. 230–241.
Qin, Y., Ferraz, A., Mallet, C., Iovan, C., 2014. Individual tree segmentation over large areas using airborne LiDAR point cloud and very high resolution optical imagery. In: 2014 IEEE Geoscience and Remote Sensing Symposium, IEEE, pp. 800–803.
Quezada, J.C., Etter, A., Ghazoul, J., Buttler, A., Guillaume, T., 2019. Carbon neutral expansion of oil palm plantations in the Neotropics. Sci. Adv. 5 (11), eaaw4418 https://doi.org/10.1126/sciadv.aaw4418.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems, pp. 91–99.
Rey, N., Volpi, M., Joost, S., Tuia, D., 2017. Detecting animals in African Savanna with UAVs and the crowds. Remote Sens. Environ. 200, 341–351. https://doi.org/10.1016/j.rse.2017.08.026.
Santika, T., Meijaard, E., Budiharta, S., Law, E.A., Kusworo, A., Hutabarat, J.A., et al., 2017. Community forest management in Indonesia: avoided deforestation in the context of anthropogenic and climate complexities. Global Environ. Change 46, 60–71. https://doi.org/10.1016/j.gloenvcha.2017.08.002.
Santos, A.A.D., Marcato Junior, J., Araújo, M.S., Di Martini, D.R., Tetila, E.C., Siqueira, H.L., et al., 2019. Assessment of CNN-based methods for individual tree detection on images captured by RGB cameras attached to UAVs. Sensors 19 (16), 3595. https://doi.org/10.3390/s19163595.
Selvaraj, M.G., Vergara, A., Montenegro, F., Ruiz, H.A., Safari, N., Raymaekers, D., et al., 2020. Detection of banana plants and their major diseases through aerial images and machine learning methods: A case study in DR Congo and Republic of Benin. ISPRS J. Photogramm. Remote Sens. 169, 110–124. https://doi.org/10.1016/j.isprsjprs.2020.08.025.
Shafri, H.Z., Hamdan, N., 2009. Hyperspectral imagery for mapping disease infection in oil palm plantationusing vegetation indices and red edge techniques. Am. J. Appl. Sci. 6 (6), 1031.
Shafri, H.Z., Anuar, M.I., Seman, I.A., Noor, N.M., 2011. Spectral discrimination of healthy and Ganoderma-infected oil palms from hyperspectral data. Int. J. Remote Sens. 32 (22), 7111–7129. https://doi.org/10.1080/01431161.2010.519003.
Solano, F., Di Fazio, S., Modica, G., 2019. A methodology based on GEOBIA and WorldView-3 imagery to derive vegetation indices at tree crown detail in olive orchards. Int. J. Appl. Earth Obs. Geoinf. 83, 101912 https://doi.org/10.1016/j.jag.2019.101912.
Stroppiana, D., Villa, P., Sona, G., Rochetti, G., Candiani, G., Pepe, M., Boschetti, M., 2018. Early season weed mapping in rice crops using multi-spectral UAV data. Int. J. Remote Sens. 39 (15–16), 5432–5452. https://doi.org/10.1080/01431161.2018.1441569.
Taheripour, F., Hertel, T.W., Ramankutty, N., 2019. Market-mediated responses confound policies to limit deforestation from oil palm expansion in Malaysia and Indonesia. Proc. Natl. Acad. Sci. 116 (38), 19193–19199. https://doi.org/10.1073/pnas.1903476116.
Tan, K.P., Kanniah, K.D., Cracknell, A.P., 2013. Use of UK-DMC 2 and ALOS PALSAR for studying the age of oil palm trees in southern peninsular Malaysia. Int. J. Remote Sens. 34 (20), 7424–7446. https://doi.org/10.1080/01431161.2013.822601.
Tan, Y., Wang, S., Xu, B., Zhang, J., 2018. An improved progressive morphological filter for UAV-based photogrammetric point clouds in river bank monitoring. ISPRS J. Photogramm. Remote Sens. 146, 421–429. https://doi.org/10.1016/j.isprsjprs.2018.10.013.
Tu, Y., Zhou, H., Lang, W., Chen, T., Li, X., Xu, B., 2020. A novel cross-sensor calibration method to generate a consistent night-time lights time series dataset. Int. J. Remote Sens. 41 (14), 5482–5502. https://doi.org/10.1080/01431161.2020.1731935.
Tuia, D., Persello, C., Bruzzone, L., 2016. Domain adaptation for the classification of remote sensing data: an overview of recent advances. IEEE Geosci. Remote Sens. Mag. 4 (2), 41–57. https://doi.org/10.1109/MGRS.2016.2548504.
Vastaranta, M., Kankare, V., Holopainen, M., Yu, X., Hyyppä, J., Hyyppä, H., 2012. Combination of individual tree detection and area-based approach in imputation of forest variables using airborne laser data. ISPRS J. Photogramm. Remote Sens. 67, 73–79. https://doi.org/10.1016/j.isprsjprs.2011.10.006.
Wagner, F.H., Ferreira, M.P., Sanchez, A., Hirye, M.C., Zortea, M., Gloor, E., et al., 2018. Individual tree crown delineation in a highly diverse tropical forest using very high resolution satellite images. ISPRS J. Photogramm. Remote Sens. 145, 362–377.
Wagner, F.H., Dalagnol, R., Tagle Casapia, X., Streher, A.S., Phillips, O.L., Gloor, E., Aragão, L.E., 2020. Regional mapping and spatial distribution analysis of canopy palms in an amazon forest using deep learning and VHR images. Remote Sensing 12 (14), 2225. https://doi.org/10.3390/rs12142225.
Wallace, L., Lucieer, A., Watson, C.S., 2014. Evaluating tree detection and segmentation routines on very high resolution UAV LiDAR data. IEEE Trans. Geosci. Remote Sens. 52 (12), 7619–7628. https://doi.org/10.1109/TGRS.2014.2315649.
Wang, Y., Hyyppä, J., Liang, X., Kaartinen, H., Yu, X., Lindberg, E., et al., 2016. International benchmarking of the individual tree detection methods for modeling 3-D canopy structure for silviculture and forest ecology using airborne laser scanning. IEEE Trans. Geosci. Remote Sens. 54 (9), 5011–5027. https://doi.org/10.1109/TGRS.2016.2543225.
Wang, X., Shrivastava, A., Gupta, A., 2017. A-fast-rcnn: Hard positive generation via adversary for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2606–2615.
Wang, X., Girshick, R., Gupta, A., He, K., 2018. Non-local neural networks. In: Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition, pp. 7794–7803.
Wang, J., Chen, K., Yang, S., Loy, C.C., Lin, D., 2019a. Region proposal by guided anchoring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2965–2974.
Wang, Y., Zhu, X., Wu, B., 2019b. Automatic detection of individual oil palm trees from UAV images using HOG features and an SVM classifier. Int. J. Remote Sens. 40 (19), 7356–7370.
Weinmann, M., Mallet, C., Brédif, M., 2016. Segmentation and localization of individual trees from MMS point cloud data acquired in urban areas. In: Tagungsband der 36. Wissenschaftlich-Technischen Jahrestagung der DGPF, Band 22, pp. 351–360.
Weinmann, M., Weinmann, M., Mallet, C., Brédif, M., 2017. A classification-segmentation framework for the detection of individual trees in dense MMS point cloud data acquired in urban areas. Remote Sensing 9 (3), 277. https://doi.org/10.3390/rs9030277.
Weinstein, B.G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. Remote Sensing 11 (11), 1309. https://doi.org/10.3390/rs11111309.
Weinstein, B.G., Marconi, S., Bohlman, S.A., Zare, A., White, E.P., 2020. Cross-site learning in deep learning RGB tree crown detection. Ecol. Inf. 56, 101061 https://doi.org/10.1016/j.ecoinf.2020.101061.
Windrim, L., Carnegie, A.J., Webster, M., Bryson, M., 2020. Tree detection and health monitoring in multi-spectral aerial imagery and photogrammetric pointclouds using machine learning. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. https://doi.org/10.1109/JSTARS.2020.2995391.
Windrim, L., Bryson, M., 2020. Detection, segmentation, and model fitting of individual tree stems from airborne laser scanning of forests using deep learning. Remote Sensing 12 (9), 1469. https://doi.org/10.3390/rs12091469.
Wu, H., Xu, Z., Wu, G., 2019. A novel method of missing road generation in city blocks based on big mobile navigation trajectory data. ISPRS Int. J. Geo-Inf. 8 (3), 142. https://doi.org/10.3390/ijgi8030142.
Wu, J., Yang, G., Yang, H., Zhu, Y., Li, Z., Lei, L., Zhao, C., 2020a. Extracting apple tree crown information from remote imagery using deep learning. Comput. Electron. Agric. 174, 105504.
Wu, W., Zheng, J., Li, W., Fu, H., Yuan, S., Yu, L., 2020b. Domain adversarial neural network-based oil palm detection using high-resolution satellite images. In: Automatic Target Recognition XXX, vol. 11394. International Society for Optics and Photonics, p. 1139406.
Wu, W., Zheng, J., Fu, H., Li, W., Yu, L., 2020c. Cross-regional oil palm tree detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 56–57.
Wulder, M., Niemann, K.O., Goodenough, D.G., 2000. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. Remote Sens. Environ. 73 (1), 103–114. https://doi.org/10.1016/S0034-4257(00)00101-2.
Xia, M., Li, W., Fu, H., Yu, L., Dong, R., Zheng, J., 2019. Fast and robust detection of oil palm trees using high-resolution remote sensing images. In: Automatic Target Recognition XXIX, vol. 10988. International Society for Optics and Photonics, p. 109880C.
Xiao, C., Qin, R., Huang, X., 2020. Treetop detection using convolutional neural networks trained through automatically generated pseudo labels. Int. J. Remote Sens. 41 (8), 3010–3030. https://doi.org/10.1080/01431161.2019.1698075.
Xu, Y., Yu, L., Li, W., Ciais, P., Cheng, Y., Gong, P., 2020. Annual oil palm plantation maps in Malaysia and Indonesia from 2001 to 2016. Earth Syst. Sci. Data 12 (2), 847–867. https://doi.org/10.5194/essd-12-847-2020.
Yao, W., Wei, Y., 2013. Detection of 3-D individual trees in urban areas by combining airborne LiDAR data and imagery. IEEE Geosci. Remote Sens. Lett. 10 (6), 1355–1359. https://doi.org/10.1109/LGRS.2013.2241390.
Ye, N., van Leeuwen, L., Nyktas, P., 2019. Analysing the potential of UAV point cloud as input in quantitative structure modelling for assessment of woody biomass of single trees. Int. J. Appl. Earth Obs. Geoinf. 81, 47–57. https://doi.org/10.1016/j.jag.2019.05.010.
You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. In: Thirty-First AAAI Conference on Artificial Intelligence.
Zhang, M., Qin, Z., Liu, X., Ustin, S.L., 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. Int. J.

Appl. Earth Obs. Geoinf. 4 (4), 295–310. https://doi.org/10.1016/S0303-2434(03)00008-4.

Zhang, J., Sohn, G., Brédif, M., 2014. A hybrid framework for single tree detection from airborne laser scanning data: A case study in temperate mature coniferous forests in Ontario, Canada. ISPRS J. Photogramm. Remote Sens. 98, 44–57. https://doi.org/10.1016/j.isprsjprs.2014.08.007.

Zhang, C., Atkinson, P.M., George, C., Wen, Z., Diazgranados, M., Gerard, F., 2020. Identifying and mapping individual plants in a highly diverse high-elevation ecosystem using UAV imagery and deep learning. ISPRS J. Photogramm. Remote Sens. 169, 280–291. https://doi.org/10.1016/j.isprsjprs.2020.09.025.

Zheng, J., Li, W., Xia, M., Dong, R., Fu, H., Yuan, S., 2019. Large-scale oil palm tree detection from high-resolution remote sensing images using faster-RCNN. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1422–1425.

Zheng, J., Fu, H., Li, W., Wu, W., Zhao, Y., Dong, R., Yu, L., 2020. Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network. ISPRS J. Photogramm. Remote Sens. 167, 154–177. https://doi.org/10.1016/j.isprsjprs.2020.07.002.

Zhou, X., Zheng, H.B., Xu, X.Q., He, J.Y., Ge, X.K., Yao, X., et al., 2017. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. ISPRS J. Photogramm. Remote Sens. 130, 246–255. https://doi.org/10.1016/j.isprsjprs.2017.05.003.

Zhu, J.Y., Krahenbuhl, P., Shechtman, E., Efros, A.A., 2015. Learning a discriminative model for the perception of realism in composite images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3943–3951.

Zhu, X., Hou, Y., Weng, Q., Chen, L., 2019. Integrating UAV optical imagery and LiDAR data for assessing the spatial relationship between mangrove and inundation across a subtropical estuarine wetland. ISPRS J. Photogramm. Remote Sens. 149, 146–156. https://doi.org/10.1016/j.isprsjprs.2019.01.021.