Article

Large-Scale Oil Palm Tree Detection from High-Resolution Satellite Images Using Two-Stage Convolutional Neural Networks

Weijia Li ^{1,2} , Runmin Dong ^{1,2}, Haohuan Fu ^{1,2,*} and Le Yu ^{1,2,*}

¹ Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China; liwj14@mails.tsinghua.edu.cn (W.L.); drml7@mails.tsinghua.edu.cn (R.D.)

² Joint Center for Global Change Studies (JCGCS), Beijing 100084, China

* Correspondence: haohuan@tsinghua.edu.cn (H.F.); leyu@tsinghua.edu.cn (L.Y.); Tel.: +86-158-1084-7817 (H.F.); +86-137-1894-4875 (L.Y.)

Received: 2 November 2018; Accepted: 19 December 2018; Published: 20 December 2018

Abstract: Being an important economic crop that contributes 35% of the total consumption of vegetable oil, remote sensing-based quantitative detection of oil palm trees has long been a key research direction for both agriculture and environmental purposes. While existing methods already demonstrate satisfactory effectiveness for small regions, performing the detection for a large region with satisfactory accuracy is still challenging. In this study, we proposed a two-stage convolutional neural network (TS-CNN)-based oil palm detection method using high-resolution satellite images (i.e. Quickbird) in a large-scale study area of Malaysia. The TS-CNN consists of one CNN for land cover classification and one CNN for object classification. The two CNNs were trained and optimized independently based on 20,000 samples collected through human interpretation. For the large-scale oil palm detection for an area of 55 km², we proposed an effective workflow that consists of an overlapping partitioning method for large-scale image division, a multi-scale sliding window method for oil palm coordinate prediction, and a minimum distance filter method for post-processing. Our proposed approach achieves a much higher average F1-score of 94.99% in our study area compared with existing oil palm detection methods (87.95%, 81.80%, 80.61%, and 78.35% for single-stage CNN, Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN), respectively), and much fewer confusions with other vegetation and buildings in the whole image detection results.

Keywords: oil palm tree; object detection; convolutional neural networks; deep learning; high-resolution satellite imagery

1. Introduction

Oil palm is one of the most rapidly expanding crops in tropical regions due to its high economic value [1], especially in Malaysia and Indonesia, the two leading oil palm-producing countries. The palm oil produced from the oil palm trees can be widely used for many purposes, e.g., producing cooking oil, food additive, cosmetics, industrial lubricants, and biofuels, etc. [2]. In recent years, palm oil has become the world's most consumed vegetable oil, making up 35% of the total consumption of vegetable oil [3]. As a result of the increasing demand of palm oil, a considerable amount of land (e.g., existing arable land and forests) has been replaced by oil palm plantation areas [4,5]. The expansion of oil palm plantation areas has caused serious environmental problems, e.g., deforestation, the reduction of biodiversity, and the loss of ecosystem functioning, etc. [6–8]. The economic benefits of the oil palm

expansion and its negative impact on the natural environment has become a controversial topic with great concern [9,10].

The remote sensing technique is a powerful tool for various oil palm plantation monitoring studies, including oil palm mapping, land cover change detection, oil palm detection and counting, age estimation, pest or disease detection, and yield estimation, etc. [11–14]. High-resolution remote sensing image-based oil palm tree detection and counting is a significant research issue in many aspects. First, it can provide precise information for related ecology models such as the palm oil yield estimation, the biomass estimation and carbon stock calculation, etc. [15,16]. Second, it provides essential information for monitoring the growing and distribution condition of oil palms, which leads to better replantation, fertilization, irrigation plans, and more efficient utilization of resources [17]. Moreover, the appropriate management of oil palm plantation is beneficial for improving the production per unit area and achieving the production goal without expanding the plantation area, thus alleviating the expansion of oil palm and its threat to tropical forests to some extent, and contributing to the protection of the biodiversity and ecosystem [8].

Over the past few decades, existing methods for detecting oil palms or tree crowns can basically be classified into three categories: (1) the traditional image processing-based methods; (2) the traditional machine learning-based methods and (3) the deep learning-based methods. The traditional image processing methods for tree crown detection include local maximum filter [18–20], image binarization [21,22], template matching [23,24], and image segmentation [25,26], etc. For instance, Pouliot et al. [27] proposed an algorithm for tree crown detection and delineation through local maximum filter-based potential tree apex identification, transects sampling, and minimum distance filtering, etc., obtaining the highest detection accuracy of 91% in a study area of about 0.017 km². Shafri et al. [28] proposed an oil palm extraction method based on spectral and texture analysis, edge enhancement and image segmentation, etc., obtaining the ex of 95% in 10 regions with a total area of about 0.78 km². Srestasathiern and Rakwatin [29] proposed an oil palm detection approach using feature selection, semi-variogram computation and local maximum filtering, etc., obtaining the detection accuracy of 90–99% in several regions with a total area of about 1 km². Dos Santos et al. [24] proposed a shadow extraction and template matching-based method for babassu palm tree detection, in which only 75.45% of the trees were correctly detected in a study area of about 95 km². In general, traditional image processing-based methods have the following advantages. First, these methods are relatively simple for implementation and can be completed using existing tools. Second, the tree crowns are not only detected but also delineated in most studies [30,31]. Third, the manually labeled samples are not required in traditional image processing-based methods (except for the template matching method). However, these methods have some major limitations. For most methods, the performance of tree crown detection deteriorates seriously where tree crowns are very crowded or overlapping. The local maximum filtering-based approach cannot detect young oil palms grown in green areas effectively, as the pixel value of the tree apex is not always the local maximum in these areas. Moreover, it often requires manual selection of the certain tree plantation areas before the automatic detection of tree crowns.

The traditional machine learning-based tree crown detection methods usually require feature extraction, image segmentation, classifier training, and prediction, etc. [32–34]. The commonly used classifiers in tree crown detection studies include the support vector machine (SVM), the random forest (RF), and the artificial neural network (ANN), etc. For instance, Malek et al. [35] proposed a palm tree detection method using SIFT descriptor-based feature extraction and extreme learning machine classifier, obtaining the detection accuracy of 89–96% in a study area of about 0.057 km². López-López et al. [36] proposed an image segmentation and SVM classifier based method for detecting unhealthy trees, obtaining the classification accuracy of 65–90% in a study area of about 900 km². Nevalainen et al. [37] proposed a method for tree crown detection and tree species classification, including local maximum filter-based tree crown detection, feature extraction, and RF and ANN-based tree species classification, obtaining the classification accuracy of 95% in a study area

of about 0.08 km². The traditional machine learning-based methods do not require manual selection of the tree plantation area. Many studies not only detect tree crowns but also identify tree species. The machine learning-based methods often outperform traditional image processing-based methods in complex regions. However, some feature extraction and tree classification methods have certain demands for image datasets. For instance, the methods proposed in References [35–37] require very high-resolution UAV images, hyperspectral and thermal imagery, and point cloud and hyperspectral images, respectively. These requirements bring limitations to large-scale oil palm detection issues to some extent.

Since 2014, deep learning has been applied to the remote sensing field and has achieved excellent performance in many studies [38–41], including multi-spectral and hyperspectral image-based land cover mapping [42–46], high resolution image-based scene classification [47–49], object detection [50–54], and semantic segmentation [55–57], etc. In recent years, several studies applied the deep learning-based methods to the oil palm or tree crown detection studies, all of which are based on the sliding window technique combined with the CNN. In 2016, we proposed the first deep learning-based method for oil palm detection in our previous study and achieved the detection accuracy of 96% in a pure oil palm plantation area in Malaysia [17]. Guirado et al. [58] proposed a deep learning-based method for scattered shrub detection, obtaining the detection accuracy of 93.38% and 96.50% in two study areas of 0.053 km². The paper concluded that deep learning-based methods outperforms object-based methods in regions where the images for tree detection are relatively different from the images for training sample collection. Pibre et al. [59] proposed a deep CNN-based approach for detecting tree crowns in different sizes using a multi-scale sliding window technique and a multi-source data fusion method, obtaining the highest tree crown size detection accuracy of 75% in a study area of smaller than 0.1 km². In summary, the deep learning-based tree crown detection methods proposed in existing studies are all based on the features of the individual tree crown, without taking the features of the plantation region into consideration. Although achieving relatively good detection results in small study areas (<5 km²), existing deep learning-based tree crown detection methods are not suitable for oil palm tree detection in the large-scale study area, in which there exists great similarity between the oil palm trees and other vegetation types in many regions.

In this research, we proposed a two-stage convolutional neural network-based method (TS-CNN) for oil palm detection in large-scale study area (around 55 km²) located in Malaysia, using high-resolution satellite images (i.e., Quickbird). We collected 20,000 samples in four types (i.e., oil palm, background, other vegetation/bareland, and impervious/cloud) from the study region through human interpretation. Two convolutional neural networks were trained and optimized based on the 20,000 samples, one for land cover classification and the other for object classification. We designed a large-scale image division method and a multi-scale sliding-window-based method for oil palm coordinate prediction, and proposed a minimum distance filter-based method for post-processing. Experiment results show that there exist much fewer misclassifications between oil palms and other vegetation or buildings in the image detection results of our proposed TS-CNN compared with the other 10 oil palm detection methods. Our proposed TS-CNN achieves the average F1-score of 95% in our study area compared with the manually labeled ground truth, which is 7–16% greater than the existing single-stage classifier-based methods. Moreover, the link of the 5000 manually-labeled oil palm sample coordinates and the whole oil palm detection results of our proposed method can be found at the end of this paper. We hope this dataset can be beneficial to other oil palm-related studies.

2. Study Area and Datasets

Figure 1 shows the location of our study area (103°35′56.82″E, 1°35′47.96″N, in the south of Malaysia) and the QuickBird satellite image (acquired on 21 November 2006) used for the oil palm detection in this research. There are various land cover types in this area, including oil palm plantation areas, forests, grassland, bare land, impervious, and cloud, etc. The size of the QuickBird satellite image used in this study is 12,188 × 12,576 pixels. The spatial resolution of its one panchromatic

band and four multi-spectral bands are 0.6 m and 2.4 m respectively. Moreover, we employ a spectral sharpening method (i.e., Gram-Schmidt) and remove the NIR band, resulting in an image with three spectral bands (Red, Green and Blue) in 0.6-m spatial resolution for further processes.

Twenty-thousand sample coordinates were collected from multiple regions selected within the whole study area, including 5000 oil palm sample coordinates, 5000 background sample coordinates, 5000 other vegetation/bare land sample coordinates, and 5000 impervious/cloud sample coordinates. Figure 1 shows several examples of sample coordinates collected from different regions. The oil palm sample coordinates and the background sample coordinates were collected from the two black rectangular areas through manual interpretation. Only the sample coordinates located in the center of an oil palm tree crown will be interpreted as the oil palm tree sample coordinates (denoted by red points), while other coordinates will be interpreted as background sample coordinates (denoted by yellow points). The sample coordinates in other vegetation/bare land type (denoted by green points) were randomly generated from the five green square areas. Similarly, the sample coordinates in impervious/cloud type (denoted by blue points) were randomly generated from the five blue square areas and the road area denoted by blue polygon. All 20,000 sample coordinates will be randomly divided into a training dataset with 17,000 coordinates and a validation dataset with 3000 coordinates. In addition, we select six typical areas (denoted by red squares) for method evaluation, which will be described in Section 4.

Figure 1. The location of the study area; the QuickBird satellite image used for the oil palm detection; several examples of sample coordinates collected from different regions; and six selected regions for method evaluation. The red, yellow, green, and blue points denote the oil palm tree, background, other vegetation/bare land and impervious/cloud sample coordinates, respectively.

3. Methods

3.1. Motivations and Overview of Our Proposed Method

In our previous study, we proposed a single CNN-based approach for oil palm detection in a pure oil palm plantation area [17]. The detection results show many misclassifications between oil palms and other vegetation if we directly apply the single CNN-based method to the large-scale study area with various land cover types, even if we collect thousands of samples of other vegetation types. The main reason is that in many regions it is hard to differentiate oil palm trees from other vegetation when only using the input image that has a similar size to a single oil palm tree. However, due to the

fact that the neighboring oil palm trees often have similar spatial distances (i.e., around 7.5–9 m in our study area) while other types of vegetation are often distributed randomly, it will be much easier to identify the oil palms from other types of vegetation if we take the special pattern or texture of the oil palm plantation areas into consideration.

In this research, we proposed a two-stage CNN-based method for oil palm detection from a large-scale QuickBird satellite image, which solves the above-mentioned problems of our previous work and improves the detection results to a great extent. Figure 2 shows the overall flowchart of this method, including the training and optimization of TS-CNN (Figure 2a–h) and the large-scale oil palm detection (Figure 2a–l). First, we built a sample dataset in three classes (i.e., oil palm plantation area, other vegetation/bareland, and impervious/cloud) with a size of 65×65 pixels from the sample coordinates (Figure 2c), and we train and optimize the first CNN for land cover classification (Figure 2d). The sample image size for land cover classification can influence the detection accuracy to some extent, which will be analyzed in Section 4.1. Similarly, we built a sample dataset in four classes (i.e., oil palm, background, other vegetation/bareland, and impervious/cloud) with a size of 17×17 pixels (similar to the largest tree crown size of the oil palms in our study area) from the sample coordinates (Figure 2f), and we train and optimize the second CNN for object classification (Figure 2g). For the oil palm tree detection in the large-scale remote sensing image, first we proposed an overlapping partitioning method to divide the large-scale image into multiple sub-images (Figure 2i). Second, we proposed a multi-scale sliding-window technique to collect image datasets (Figure 2j), and utilized the two optimal CNN models (Figure 2e,h) obtained in training phases to predict labels for the image datasets and to obtain the oil palm tree coordinates. Third, we applied a minimum distance filter-based post-processing method to the predicted oil palm tree coordinates (Figure 2k) and obtained the final oil palm detection results (Figure 2l). The details of each step will be described in Sections 3.2 and 3.3.

Figure 2. The overall flowchart of the proposed TS-CNN-based method.

3.2. Training and Optimization of the TS-CNN

3.2.1. Training and Optimization of the CNN for Land Cover Classification

For training and optimization of the TS-CNN, the 20,000 sample coordinates collected in Section 2 are randomly divided into 17,000 training sample coordinates and 3000 validation sample coordinates. To build the sample dataset for land cover classification, we extract the training and validation image samples in a size of 65×65 pixels from the whole image based on their corresponding sample coordinates, and combine the samples of oil palm class and background class into the samples of oil palm plantation class. The size of the image sample for land-cover classification can influence the detection accuracy to some extent, which will be analyzed in Section 4.1. Then we use the 17,000 training samples and the 3000 validation samples (in three classes, with a size of 65×65 pixels) to train the first CNN for land-cover classification and calculate the classification accuracies.

Figure 3 shows the framework of this step. The CNN used for land-cover classification is based on the AlexNet architecture [60]. It consists of five convolutional layers (the first, second and fifth are followed with a max-pooling layer) and three fully-connected layers. We utilize ReLU [60] as the activation function and adopt the dropout technique to prevent the substantial overfitting. Moreover, we optimize the main hyper-parameters of CNN (e.g., the number of convolution kernels in each

convolutional layer, the number of hidden units in each fully-connected layer, and the batch size) to obtain the optimal hyper-parameter setting for land-cover classification (achieving the highest accuracy for 3000 validation samples in 65×65 pixels).

Figure 3. The framework of the CNN for land-cover classification.

3.2.2. Training and Optimization of the CNN for Object Classification

To build the sample dataset for object classification, we extract the training and validation image samples in a size of 17×17 pixels (similar to the largest tree crown size of the oil palms in our study area) from the whole image based on their corresponding sample coordinates. Then we use the 17,000 training samples and the 3000 validation samples (in four classes, with a size of 17×17 pixels) to train the second CNN for object classification and to calculate the classification accuracy. Figure 4 shows the framework of this step. The basic architecture of the second CNN is the same as the one of the first CNN. The main hyper-parameters of the second CNN for object classification are optimized independently from the first CNN described in Section 3.2.1. Similarly, we will obtain the optimal hyper-parameter setting for object classification (achieving the highest accuracy for 3000 validation samples in 17×17 pixels).

Figure 4. The framework of the CNN for object classification.

3.3. Oil Palm Tree Detection in the Large-Scale Study Area

3.3.1. Overlapping Partitioning Based Large-Scale Image Division

As mentioned in Section 2, the image size of our study area is $12,188 \times 12,576$ pixels. To improve the efficiency of the large-scale oil palm tree detection, we divide the whole image into multiple sub-images in 600×600 pixels so that each sub-image can be processed in parallel. If we directly divide the large-scale image into multiple sub-images without overlaps, some tree crowns will be divided into two sub-images and might not be detected correctly. To avoid this situation, we design an overlapping partitioning method for the large-scale image division. In our proposed method, every two neighbor sub-images in the large-scale image have an overlapping region in a user-defined width (20 pixels in this study). In this way, all the oil palm trees will belong to at least one sub-image. Correspondingly, we will perform a minimum distance filter-based post-processing after the palm tree detection of each sub-image to avoid repeated detection of the same oil palm tree resulting from the overlapping partitioning, which will be described in Section 3.3.3.

3.3.2. Multi-Scale Sliding Window Based Image Dataset Collection and Label Prediction

After dividing the large-scale image into multiple sub-images, we designed a multi-scale sliding-window technique to collect two image datasets for each sub-image and to predict labels for all samples in the image datasets, as shown in Figure 5. Each sample of the first image dataset corresponds to a window of 65×65 pixels. Each sample of the second image dataset corresponds to a window of 17×17 pixels. Taking both the processing efficiency and the detection accuracy into consideration, the optimal sliding distance of a window in each sliding step is three pixels in our study area (obtained through experiments). After the sample collection of the two image datasets, we used the first optimal CNN model to predict land-cover types for the first image dataset, and used the second optimal CNN model to predict the object types for the second image dataset. A sample coordinate will be predicted as the coordinate of an oil palm tree if both its corresponding sample in the first image dataset is predicted as the oil palm plantation area type and its corresponding sample in the second image dataset is predicted as the oil palm type.

Figure 5. Multi-scale sliding-window-based image dataset collection and label prediction.

3.3.3. Minimum Distance Filter Based Post-Processing

After completing the overlapping partitioning-based large-scale image division and the multi-scale sliding-window-based label prediction, we will obtain all samples predicted as the oil palm types and their corresponding spatial coordinates in the whole image. At this point, multiple predicted oil palm coordinates are often located within the same oil palm tree crown, as we partition the large-scale image into multiple sub-images with overlap and the sliding distance of the window in each step is smaller than the actual distance between two oil palm trees. To avoid the repeated detection of the same oil palm tree, we apply a minimum distance filter-based post-processing method to all predicted oil palm coordinates, as shown in Figure 5c,d. According to the actual situation of the oil palm tree plantation, the Euclidean spatial distance between two neighbor oil palms should not be smaller than 10 pixels (i.e., 6 m) in our study area. Thus, we iteratively merge each group of oil palm coordinates into one coordinate if their Euclidean distance is smaller than a given threshold. In each iteration, we increased the threshold by one pixel (from 3 to 10 pixels), and the merging results remain unchanged after the threshold increased to 8 pixels. After the minimum distance filter-based post-processing step, we will obtain the final results of the oil palm detection in the whole study area.

4. Experimental Results Analysis

4.1. Hyper-Parameter Setting and Classification Accuracies

In this study, the main hyper-parameters of the CNN for land-cover classification and the CNN for object classification are optimized independently. After the hyper-parameter optimization, we will obtain the optimal hyper-parameter setting of two CNNs with the highest classification accuracy of the 3000 validation samples. For both CNNs, the sizes of the convolution kernel, max-pooling kernel, and mini-batch are 3, 2 and 10, respectively. The maximum number of iterations for CNN training is set as 100,000. For the first CNN for land-cover classification, we obtain the highest classification accuracy of 95.47% of the 3000 validation samples in three classes when the number of convolution kernels in five convolutional layers and the number of neurons in three fully-connected layers are 24-64-96-96-64-800-800-3. For the second CNN for object classification, we obtain the highest classification accuracy of 91.53% on the 3000 validation samples in four classes when the number of convolution kernels in five convolutional layers and the number of neurons in three fully-connected layers are 25-60-45-45-65-800-800-4. Moreover, the size of the image sample for land-cover classification can influence the detection results to some extent. Experiment results show that the classification

accuracies on 3000 validation samples are 92%, 95% and 91% when the sizes of the image sample are 45×45 pixels, 65×65 pixels, and 85×85 pixels, respectively. Thus, we chose 65×65 pixels as the sample image size for land-cover classification.

4.2. Evaluation of the Oil Palm Tree Detection Results

To evaluate the oil palm detection results of our proposed method quantitatively, we selected six typical regions in the large-scale study area and compared the detection results of these regions with the manually labeled ground truth. The precision, the recall and the F1-score of the six selected regions are calculated according to Equations (1)–(3), where the True Positive (TP) indicates the number of oil palms that are detected correctly, the False Positive (FP) indicates the number of other objects that are detected as oil palms by mistake, and the False Negative (FN) indicates the number of oil palms not detected. An oil palm will be considered as correctly detected only if the Euclidean spatial distance between its coordinate and an oil palm coordinate in the ground truth dataset is smaller than or equal to 5 pixels [17].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (3)$$

Table 1 displays the performance of the TS-CNN-based method in six typical regions selected from the large-scale study area, in respect to TP, FP, FN, Precision, Recall and F1-score. The locations of the six selected regions for method evaluation can be found in Figure 1 (denoted by the red squares and numbered from top to bottom). Each selected region includes various land cover types, e.g., oil palm plantation areas, other vegetation, bare land, buildings, and roads, etc. Figure 6 shows the image detection results in six selected regions, where the red points indicate the detected oil palms of our proposed method, the green squares indicate the oil palms that are not detected, and the blue squares indicate other types of objects that are mistakenly detected as oil palms. We can find that our proposed method achieves excellent oil palm detection results with very few confusions with other vegetation or buildings, and the F1-scores of the oil palm detection results are 91–96% in six regions.

Table 1. The detection results of our proposed TS-CNN-based method.

Index	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
TP	445	1034	956	1114	385	1028
FP	12	47	35	38	35	29
FN	33	44	35	78	40	49
Precision	97.37%	95.65%	96.47%	96.70%	91.67%	97.26%
Recall	93.10%	95.92%	96.47%	93.46%	90.59%	95.45%
F1-score	95.19%	95.79%	96.47%	95.05%	91.12%	96.34%

Figure 6. The image detection results of TS-CNN-based method (a–f) in six selected regions.

4.3. Oil palm Tree Detection Results in the Whole Study Area

In the whole study area, a total number of 264,960 oil palm trees are detected based on our proposed method. The link of the predicted oil palm coordinates of the whole study area can be found at the end of this paper. The total processing time for the oil palm detection in the whole study area (including the data pre-processing, oil palm detection and post-processing) is about 6 hours (Library: Tensorflow, Hardware: single NVIDIA Tesla K40 GPU). We obtain the density map of our study area through calculating the number of detected oil palms per hectare for each pixel, as shown in Figure 7. For most regions in the oil palm plantation area, the densities of oil palms are between 108 to 180 per hectare, which is consistent with the actual plantation situation of the oil palms (the spatial distance between two neighbor oil palm trees is often 7.5–9 m).

Figure 7. The density map of detected oil palms in our study area.

5. Discussion

5.1. Detection Results of Different CNN Architectures

In this section, we will analyze the influence of various CNN architectures on the oil palm tree detection results in the six selected regions. For both the first CNN for land cover classification (CNN 1) and the second CNN for object classification (CNN 2), we evaluate the detection accuracies of three widely-used CNN architectures: (1) LeNet, including 2 convolutional layers, 2 max-pooling layers, and 1 fully-connected layers; (2) AlexNet, including 5 convolutional layers, 3 max-pooling layers, and 3 fully-connected layers; (3) VGG-19, including 16 convolutional layers, 5 max-pooling layers, and 3 fully-connected layers. Thus, there are nine combinations of the CNN architectures for our proposed TS-CNN-based method. Table 2 shows the F1-score of six selected regions obtained from nine combinations of the CNN architectures. The numbers in bold type indicate the best case (the highest F1-score) for each region. We find that the combination of AlexNet + AlexNet architecture used in this

study achieves the highest F1-score in most of the regions (slightly lower than LeNet + AlexNet for Region 3 and Region 6) and achieves the highest average F1-score in six regions amongst different combinations of CNN architectures.

Table 2. The F1-score of six selected regions obtained from different CNN architectures.

CNN 1	CNN 2	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
LeNet	LeNet	90.79%	93.69%	94.22%	90.78%	79.73%	94.86%
LeNet	AlexNet	93.86%	94.42%	96.83%	94.52%	90.11%	96.53%
LeNet	VGG-19	91.12%	90.73%	92.57%	90.06%	76.72%	93.92%
AlexNet	LeNet	91.92%	93.39%	94.14%	91.69%	80.18%	95.09%
AlexNet	AlexNet	95.19%	95.79%	96.47%	95.05%	91.12%	96.34%
AlexNet	VGG-19	89.53%	91.07%	93.45%	91.70%	77.30%	94.14%
VGG-19	LeNet	88.96%	86.77%	92.37%	89.18%	72.48%	93.51%
VGG-19	AlexNet	91.35%	87.53%	95.37%	93.16%	84.28%	95.08%
VGG-19	VGG-19	86.22%	85.81%	92.58%	86.91%	67.20%	91.63%

5.2. The Detection Result Comparison of Different Methods

To further evaluate the proposed TS-CNN-based method, we implemented four existing oil palm detection methods for comparison with the proposed method, including the single-stage CNN, SVM, Random Forest, and Artificial Neural Network-based method (denoted by CNN, SVM, RF and ANN, respectively). The CNN-based method outperforms other oil palm detection methods (the template matching-based method, the ANN-based method, and the local maximum filter-based method) in all selected regions in our previous studies [17]. The SVM, RF and ANN are broadly used machine learning methods in the remote sensing domain due to their excellent classification accuracy [61,62] and good performance in many tree crown detection or tree species classification studies [33,36,37,63]. Thus, we chose these methods for comparison with the TS-CNN-based method. We also implemented six other two-stage methods using the different combinations of the above classifiers (denoted by SVM+SVM, RF+RF, ANN+ANN, CNN+SVM, CNN+RF, and CNN+ANN, respectively) for further evaluating the performance of the two-stage strategy for the oil palm detection.

We also tried applying the Faster-RCNN [64] method to the oil palm detection based on the same sample dataset used for TS-CNN. Two-thousand sample images in 500×500 pixels were randomly generated from the selected regions (described in Figure 1) for the training of Faster-RCNN. However, the detection results are very bad in all six regions, which might be due to the limited number of labeled sample images, the small size of oil palm objects, and the relatively low resolution of the satellite images used in this study. The Faster-RCNN and other end-to-end object detection methods will be further explored and improved for the oil palm detection in our future research.

All of these methods include the same processes of overlapping partitioning-based large-scale image division and minimum distance filter-based post-processing, and different processes of classifier training and optimization, image dataset collection, and label prediction. For the single classifier-based method (i.e., CNN, SVM, RF, and ANN), we trained and optimized the single classifier for object classification, and utilized the single-scale sliding-window-based method for image dataset collection and label prediction [17]. For the two-stage classifier-based method (i.e., CNN+CNN, SVM+SVM, RF+RF, ANN+ANN, CNN+SVM, CNN+RF, and CNN+ANN), we trained and optimized one classifier for land cover classification and one classifier for object classification, and utilized the multi-scale sliding window-based method for image dataset collection and label prediction.

Tables 3–8 shows the detection results of all methods in each selected regions in terms of TP, FP, FN, Precision, Recall, and F1-score. The detection results of all six regions (in terms of F1-score) are summarized in Table 9. The numbers in bold type indicate the best case (the highest value for TP, Precision, Recall and F1-score, and the lowest value for FP and FN) among different methods. Figures 8 and 9 show the image detection results of all methods in Region 2 and Region 4. We find that our proposed TS-CNN-based method outperforms other methods in many aspects. First, the

average F1-score in six regions of TS-CNN is 94.99%, which is 7.04–16.64% higher than those obtained from the existing one-stage methods (CNN, SVM, RF, and ANN), and 2.51–11.45% higher than those obtained from other classifiers-based two-stage methods (SVM+SVM, RF+RF, ANN+ANN, CNN+SVM, CNN+RF, and CNN+ANN). Second, from both the detection accuracy and the detection image results, we can conclude that the two-stage classifier-based methods outperform the single classifier-based methods for all classifiers (i.e., CNN, SVM, RF, and ANN). The CNN classifier outperforms SVM, RF and ANN classifiers for both two-stage and one-stage classifier-based methods in our study area. Moreover, there exists much fewer misclassifications between oil palms and other vegetation or buildings in the detection results of our proposed method, compared with those obtained from the other 10 methods (See the top areas in Figure 8 and the bottom left areas in Figure 9). Nevertheless, the oil palm detection results of our proposed method still have some weaknesses in the following aspects. On one hand, the recalls of the two-stage classifier-based methods are often slightly lower than the one-stage classifier-based methods. On the other hand, there are still some misclassifications between oil palm trees and other vegetation types, especially around the boundary of two land-cover types. These issues should be solved in our future study.

Table 3. The detection results of each method in Region 1.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	445	12	33	97.37%	93.10%	95.19%
Existing methods using single classifier	CNN	458	147	20	75.70%	95.82%	84.58%
	SVM	448	221	30	66.97%	93.72%	78.12%
	RF	457	253	21	64.37%	95.61%	76.94%
	ANN	438	296	40	59.67%	91.63%	72.28%
Two-stage methods using different combinations of classifiers	SVM+SVM	403	76	75	84.13%	84.31%	84.22%
	RF+RF	448	37	30	92.37%	93.72%	93.04%
	ANN+ANN	428	40	50	91.45%	89.54%	90.49%
	CNN+SVM	434	24	44	94.76%	90.79%	92.74%
	CNN+RF	440	25	38	94.62%	92.05%	93.32%
	CNN+ANN	427	41	51	91.24%	89.33%	90.27%

Table 4. The detection result of each method in Region 2.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	1034	47	44	95.65%	95.92%	95.79%
Existing methods using single classifier	CNN	1047	121	31	89.64%	97.12%	93.23%
	SVM	1041	153	37	87.19%	96.57%	91.64%
	RF	1061	166	17	86.47%	98.42%	92.06%
	ANN	1036	223	42	82.29%	96.10%	88.66%
Two-stage methods using different combinations of classifiers	SVM+SVM	1030	116	48	89.88%	95.55%	92.63%
	RF+RF	1011	74	67	93.18%	93.78%	93.48%
	ANN+ANN	984	104	94	90.44%	91.28%	90.86%
	CNN+SVM	990	59	88	94.38%	91.84%	93.09%
	CNN+RF	1008	65	70	93.94%	93.51%	93.72%
	CNN+ANN	986	109	92	90.05%	91.47%	90.75%

Table 5. The detection result of each method in Region 3.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	956	35	35	96.47%	96.47%	96.47%
Existing methods using single classifier	CNN	976	169	15	85.24%	98.49%	91.39%
	SVM	960	218	31	81.49%	96.87%	88.52%
	RF	970	264	21	78.61%	97.88%	87.19%
	ANN	944	277	47	77.31%	95.26%	85.35%
Two-stage methods using different combinations of classifiers	SVM+SVM	938	151	53	86.13%	94.65%	90.19%
	RF+RF	957	136	34	87.56%	96.57%	91.84%
	ANN+ANN	930	147	61	86.35%	93.84%	89.94%
	CNN+SVM	944	72	47	92.91%	95.26%	94.07%
	CNN+RF	953	71	38	93.07%	96.17%	94.59%
	CNN+ANN	929	124	62	88.22%	93.74%	90.90%

Table 6. The detection results of each method in Region 4.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	1114	38	78	96.70%	93.46%	95.05%
Existing methods using single classifier	CNN	1137	139	55	89.11%	95.39%	92.14%
	SVM	1113	218	79	83.62%	93.37%	88.23%
	RF	1136	275	56	80.51%	95.30%	87.28%
	ANN	1095	287	97	79.23%	91.86%	85.08%
Two-stage methods using different combinations of classifiers	SVM+SVM	1006	117	186	89.58%	84.40%	86.91%
	RF+RF	1125	200	67	84.91%	94.38%	89.39%
	ANN+ANN	1078	213	114	83.50%	90.44%	86.83%
	CNN+SVM	1104	94	88	92.15%	92.62%	92.38%
	CNN+RF	1124	85	68	92.97%	94.30%	93.63%
	CNN+ANN	1083	173	109	86.23%	90.86%	88.48%

Table 7. The detection result of each method in Region 5.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	385	35	40	91.67%	90.59%	91.12%
Existing methods using single classifier	CNN	388	298	37	56.56%	91.29%	69.85%
	SVM	379	737	46	33.96%	89.18%	49.19%
	RF	392	932	33	29.61%	92.24%	44.83%
	ANN	365	742	60	32.97%	85.88%	47.65%
Two-stage methods using different combinations of classifiers	SVM+SVM	377	499	48	43.04%	88.71%	57.96%
	RF+RF	382	830	43	31.52%	89.88%	46.67%
	ANN+ANN	356	614	69	36.70%	83.76%	51.04%
	CNN+SVM	376	125	49	75.05%	88.47%	81.21%
	CNN+RF	389	113	36	77.49%	91.53%	83.93%
	CNN+ANN	360	153	65	70.18%	84.71%	76.76%

Table 8. The detection result of each method in Region 6.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Our proposed method	CNN+CNN	1028	29	49	97.26%	95.45%	96.34%
Existing methods using single classifier	CNN	1056	56	21	94.96%	98.05%	96.48%
	SVM	1055	87	22	92.38%	97.96%	95.09%
	RF	1059	85	18	92.57%	98.33%	95.36%
	ANN	1032	157	45	86.80%	95.82%	91.09%

Table 8. Cont.

Method Type	Method Name	TP	FP	FN	Precision	Recall	F1-Score
Two-stage methods using different combinations of classifiers	SVM+SVM	1024	56	53	94.81%	95.08%	94.95%
	RF+RF	997	45	80	95.68%	92.57%	94.10%
	ANN+ANN	970	60	107	94.17%	90.06%	92.07%
	CNN+SVM	1034	46	43	95.74%	96.01%	95.87%
	CNN+RF	1036	52	41	95.22%	96.19%	95.70%
	CNN+ANN	1009	70	68	93.51%	93.69%	93.60%

Table 9. Summary of the detection results of all methods in terms of F1-score.

Method	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6	Average
CNN+CNN	95.19%	95.79%	96.47%	95.05%	91.12%	96.34%	94.99%
CNN	84.58%	93.23%	91.39%	92.14%	69.85%	96.48%	87.95%
SVM	78.12%	91.64%	88.52%	88.23%	49.19%	95.09%	81.80%
RF	76.94%	92.06%	87.19%	87.28%	44.83%	95.36%	80.61%
ANN	72.28%	88.66%	85.35%	85.08%	47.65%	91.09%	78.35%
SVM+SVM	84.22%	92.63%	90.19%	86.91%	57.96%	94.95%	84.48%
RF+RF	93.04%	93.48%	91.84%	89.39%	46.67%	94.10%	84.75%
ANN+ANN	90.49%	90.86%	89.94%	86.83%	51.04%	92.07%	83.54%
CNN+SVM	92.74%	93.09%	94.07%	92.38%	81.21%	95.87%	91.56%
CNN+RF	93.32%	93.72%	94.59%	93.63%	83.93%	95.70%	92.48%
CNN+ANN	90.27%	90.75%	90.90%	88.48%	76.76%	93.60%	88.46%

Figure 8. The image detection results of all methods (a–k) in Region 2.

Figure 9. The image detection results of all methods (a–k) in Region 4.

6. Conclusions

In this research, we proposed a two-stage convolutional neural network-based oil palm detection method, using a large-scale QuickBird satellite image located in the south of Malaysia. Twenty-thousand sample coordinates in four classes were collected from the study area through human interpretation. For the training and optimization of the TS-CNN, we built a sample dataset in three classes with a size of 65×65 pixels to train and optimize the first CNN for land-cover classification, and built a sample dataset in four classes with a size of 17×17 pixels to train and optimize the second CNN for object classification. For the oil palm tree detection in large-scale remote sensing images, we proposed an overlapping partitioning method to divide the large-scale image into multiple sub-images. We proposed a multi-scale sliding-window technique to collect image datasets and predict the label for each sample of the image datasets to obtain the predicted oil palm coordinates. Finally, we applied a minimum distance filter-based post-processing method and obtained the final oil palm detection results.

In the experiment results analysis, we calculated the detection accuracies of six selected regions through comparing the detection results with the manually-labeled ground truth. The TS-CNN-based method achieves the highest average F1-score of 95% among the 11 oil palm detection methods in our study area. There exist much fewer confusions between oil palms and other vegetation or buildings in the detection image results of our proposed TS-CNN compared with other methods. Moreover, we obtained the oil palm density map from the detection results in the whole study area. We also analyzed the influence of different combinations of CNN architectures on the oil palm detection results. In our future research, we will explore the potential of more state-of-the-art object detection algorithms in oil palm tree detection studies. We will also design more accurate and efficient oil palm tree detection methods for various large-scale study areas.

Supplementary Materials: The 5000 manually-labeled oil palm sample coordinates and the whole oil palm detection results of our proposed method can be found in <https://github.com/liweijia/oil-palm-detection>.

Author Contributions: Conceptualization, W.L., H.F. and L.Y.; Data curation, L.Y.; Formal analysis, W.L.; Funding acquisition, H.F.; Investigation, W.L.; Methodology, W.L.; Project administration, H.F.; Resources, L.Y.; Software, W.L. and R.D.; Supervision, H.F. and L.Y.; Validation, W.L. and R.D.; Visualization, W.L. and R.D.; Writing—original draft, W.L.; Writing—review & editing, H.F. and L.Y.

Funding: This work was supported in part by the National Key R&D Program of China (Grant No. 2017YFA0604500 and 2017YFA0604401), by the National Natural Science Foundation of China (Grant No. 5171101179), and by Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao).

Acknowledgments: The authors would like to thank the editors and reviewers for their valuable comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Cheng, Y.; Yu, L.; Xu, Y.; Lu, H.; Cracknell, A.P.; Kanniah, K.; Gong, P. Mapping oil palm extent in Malaysia using ALOS-2 PALSAR-2 data. *Int. J. Remote Sens.* **2018**, *39*, 432–452. [[CrossRef](#)]
- Lian, P.K.; Wilcove, D.S. Cashing in palm oil for conservation. *Nature* **2007**, *448*, 993–994.
- Chong, K.L.; Kanniah, K.D.; Pohl, C.; Tan, K.P. A review of remote sensing applications for oil palm studies. *Geo-Spat. Inf. Sci.* **2017**, *20*, 184–200. [[CrossRef](#)]
- Abram, N.K.; Meijaard, E.; Wilson, K.A.; Davis, J.T.; Wells, J.A.; Ancrenaz, M.; Budiharta, S.; Durrant, A.; Fakhruzz, A.; Runting, R.K.; et al. Oil palm-community conflict mapping in Indonesia: A case for better community liaison in planning for development initiatives. *Appl. Geogr.* **2017**, *78*, 33–44. [[CrossRef](#)]
- Cheng, Y.; Yu, L.; Zhao, Y.; Xu, Y.; Hackman, K.; Cracknell, A.P.; Gong, P. Towards a global oil palm sample database: Design and implications. *Int. J. Remote Sens.* **2017**, *38*, 4022–4032. [[CrossRef](#)]
- Barnes, A.D.; Jochum, M.; Mumme, S.; Haneda, N.F.; Farajallah, A.; Widarto, T.H.; Brose, U. Consequences of tropical land use for multitrophic biodiversity and ecosystem functioning. *Nat. Commun.* **2014**, *5*, 5351. [[CrossRef](#)] [[PubMed](#)]
- Busch, J.; Ferretti-Gallon, K.; Engelmann, J.; Wright, M.; Austin, K.G.; Stolle, F.; Turubanova, S.; Potapov, P.V.; Margono, B.; Hansen, M.C.; et al. Reductions in emissions from deforestation from Indonesia's moratorium on new oil palm, timber, and logging concessions. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1328–1333. [[CrossRef](#)] [[PubMed](#)]
- Cheng, Y.; Yu, L.; Xu, Y.; Liu, X.; Lu, H.; Cracknell, A.P.; Kanniah, K.; Gong, P. Towards global oil palm plantation mapping using remote-sensing data. *Int. J. Remote Sens.* **2018**, *39*, 5891–5916. [[CrossRef](#)]
- Koh, L.P.; Miettinen, J.; Liew, S.C.; Ghazoul, J. Remotely sensed evidence of tropical peatland conversion to oil palm. *Proc. Natl. Acad. Sci. USA* **2011**, *1*, 201018776. [[CrossRef](#)]
- Carlson, K.M.; Heilmayr, R.; Gibbs, H.K.; Noojipady, P.; Burns, D.N.; Morton, D.C.; Walker, N.F.; Paoli, G.D.; Kremen, C. Effect of oil palm sustainability certification on deforestation and fire in Indonesia. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 121–126. [[CrossRef](#)]
- Cracknell, A.P.; Kanniah, K.D.; Tan, K.P.; Wang, L. Evaluation of MODIS gross primary productivity and land cover products for the humid tropics using oil palm trees in Peninsular Malaysia and Google Earth imagery. *Int. J. Remote Sens.* **2013**, *34*, 7400–7423. [[CrossRef](#)]
- Tan, K.P.; Kanniah, K.D.; Cracknell, A.P. Use of UK-DMC 2 and ALOS PALSAR for studying the age of oil palm trees in southern peninsular Malaysia. *Int. J. Remote Sens.* **2013**, *34*, 7424–7446. [[CrossRef](#)]
- Gutiérrez-Vélez, V.H.; DeFries, R. Annual multi-resolution detection of land cover conversion to oil palm in the Peruvian Amazon. *Remote Sens. Environ.* **2013**, *129*, 154–167. [[CrossRef](#)]
- Cheng, Y.; Yu, L.; Cracknell, A.P.; Gong, P. Oil palm mapping using Landsat and PALSAR: A case study in Malaysia. *Int. J. Remote Sens.* **2016**, *37*, 5431–5442. [[CrossRef](#)]
- Balasundram, S.K.; Memarian, H.; Khosla, R. Estimating oil palm yields using vegetation indices derived from Quickbird. *Life Sci. J.* **2013**, *10*, 851–860.
- Thenkabail, P.S.; Stucky, N.; Griscom, B.W.; Ashton, M.S.; Diels, J.; Van Der Meer, B.; Enclona, E. Biomass estimations and carbon stock calculations in the oil palm plantations of African derived savannas using IKONOS data. *Int. J. Remote Sens.* **2004**, *25*, 5447–5472. [[CrossRef](#)]

17. Li, W.; Fu, H.; Yu, L.; Cracknell, A. Deep learning based oil palm tree detection and counting for high-resolution remote sensing images. *Remote Sens.* **2016**, *9*, 22. [[CrossRef](#)]
18. Wulder, M.; Niemann, K.O.; Goodenough, D.G. Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery. *Remote Sens. Environ.* **2000**, *73*, 103–114. [[CrossRef](#)]
19. Wang, L.; Gong, P.; Biging, G.S. Individual tree-crown delineation and treetop detection in high-spatial-resolution aerial imagery. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 351–357. [[CrossRef](#)]
20. Jiang, H.; Chen, S.; Li, D.; Wang, C.; Yang, J. Papaya Tree Detection with UAV Images Using a GPU-Accelerated Scale-Space Filtering Method. *Remote Sens.* **2017**, *9*, 721. [[CrossRef](#)]
21. Pitkänen, J. Individual tree detection in digital aerial images by combining locally adaptive binarization and local maxima methods. *Can. J. For. Res.* **2001**, *31*, 832–844. [[CrossRef](#)]
22. Daliakopoulos, I.N.; Grillakis, E.G.; Koutroulis, A.G.; Tsanis, I.K. Tree crown detection on multispectral VHR satellite imagery. *Photogramm. Eng. Remote Sens.* **2009**, *75*, 1201–1211. [[CrossRef](#)]
23. Ke, Y.; Quackenbush, L.J. A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *Int. J. Remote Sens.* **2011**, *32*, 4725–4747. [[CrossRef](#)]
24. Dos Santos, A.M.; Mitja, D.; Delaître, E.; Demagistri, L.; de Souza Miranda, I.; Libourel, T.; Petit, M. Estimating babassu palm density using automatic palm tree detection with very high spatial resolution satellite images. *J. Environ. Manag.* **2017**, *193*, 40–51. [[CrossRef](#)] [[PubMed](#)]
25. Yang, J.; He, Y.; Caspersen, J.P.; Jones, T.A. Delineating Individual Tree Crowns in an Uneven-Aged, Mixed Broadleaf Forest Using Multispectral Watershed Segmentation and Multiscale Fitting. *IEEE J. Sel. Top. Appl. Earth Observ.* **2017**, *10*, 1390–1401. [[CrossRef](#)]
26. Chemura, A.; van Duren, I.; van Leeuwen, L.M. 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.* **2015**, *100*, 118–127. [[CrossRef](#)]
27. Pouliot, D.A.; King, D.J.; Bell, F.W.; Pitt, D.G. Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sens. Environ.* **2002**, *82*, 322–334. [[CrossRef](#)]
28. Shafri, H.Z.; Hamdan, N.; Saripan, M.I. Semi-automatic detection and counting of oil palm trees from high spatial resolution airborne imagery. *Int. J. Remote Sens.* **2011**, *32*, 2095–2115. [[CrossRef](#)]
29. Srestasathiern, P.; Rakwatin, P. Oil palm tree detection with high resolution multi-spectral satellite imagery. *Remote Sens.* **2014**, *6*, 9749–9774. [[CrossRef](#)]
30. Gomes, M.F.; Maillard, P.; Deng, H. Individual tree crown detection in sub-meter satellite imagery using Marked Point Processes and a geometrical-optical model. *Remote Sens. Environ.* **2018**, *211*, 184–195. [[CrossRef](#)]
31. Ardila, J.P.; Bijker, W.; Tolpekin, V.A.; Stein, A. Multitemporal change detection of urban trees using localized region-based active contours in VHR images. *Remote Sens. Environ.* **2012**, *124*, 413–426. [[CrossRef](#)]
32. Pu, R.; Landry, S. A comparative analysis of high spatial resolution IKONOS and WorldView-2 imagery for mapping urban tree species. *Remote Sens. Environ.* **2012**, *124*, 516–533. [[CrossRef](#)]
33. Hung, C.; Bryson, M.; Sukkarieh, S. Multi-class predictive template for tree crown detection. *ISPRS J. Photogramm.* **2012**, *68*, 170–183. [[CrossRef](#)]
34. Dalponte, M.; Ørka, H.O.; Ene, L.T.; Gobakken, T.; Næsset, E. Tree crown delineation and tree species classification in boreal forests using hyperspectral and ALS data. *Remote Sens. Environ.* **2014**, *140*, 306–317. [[CrossRef](#)]
35. Malek, S.; Bazi, Y.; Alajlan, N.; AlHichri, H.; Melgani, F. Efficient framework for palm tree detection in UAV images. *IEEE J. Sel. Top. Appl. Earth Observ.* **2014**, *7*, 4692–4703. [[CrossRef](#)]
36. López-López, M.; Calderón, R.; González-Dugo, V.; Zarco-Tejada, P.J.; Fereres, E. Early detection and quantification of almond red leaf blotch using high-resolution hyperspectral and thermal imagery. *Remote Sens.* **2016**, *8*, 276. [[CrossRef](#)]
37. Nevalainen, O.; Honkavaara, E.; Tuominen, S.; Viljanen, N.; Hakala, T.; Yu, X.; Hyyppä, J.; Saari, H.; Pölönen, I.; Imai, N.N.; et al. Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sens.* **2017**, *9*, 185. [[CrossRef](#)]
38. Chen, X.; Xiang, S.; Liu, C.L.; Pan, C.H. Vehicle detection in satellite images by hybrid deep convolutional neural networks. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 1797–1801. [[CrossRef](#)]

39. Han, W.; Feng, R.; Wang, L.; Cheng, Y. A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification. *ISPRS J. Photogramm.* **2017**, *145*, 23–43. [[CrossRef](#)]
40. Long, Y.; Gong, Y.; Xiao, Z.; Liu, Q. Accurate object localization in remote sensing images based on convolutional neural networks. *IEEE T. Geosci. Remote.* **2017**, *55*, 2486–2498. [[CrossRef](#)]
41. Li, W.; Fu, H.; Yu, L.; Gong, P.; Feng, D.; Li, C.; Clinton, N. Stacked autoencoder-based deep learning for remote-sensing image classification: A case study of African land-cover mapping. *Int. J. Remote Sens.* **2016**, *37*, 5632–5646. [[CrossRef](#)]
42. Zhang, C.; Sargent, I.; Pan, X.; Li, H.; Gardiner, A.; Hare, J.; Atkinson, P.M. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sens. Environ.* **2018**, *216*, 57–70. [[CrossRef](#)]
43. Huang, B.; Zhao, B.; Song, Y. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. *Remote Sens. Environ.* **2018**, *214*, 73–86. [[CrossRef](#)]
44. Mou, L.; Ghamisi, P.; Zhu, X.X. Unsupervised spectral-spatial feature learning via deep residual Conv-Deconv network for hyperspectral image classification. *IEEE T. Geosci. Remote.* **2018**, *56*, 391–406. [[CrossRef](#)]
45. Chen, Y.; Lin, Z.; Zhao, X.; Wang, G.; Gu, Y. Deep learning-based classification of hyperspectral data. *IEEE J. Sel. Top. Appl. Earth Observ.* **2014**, *7*, 2094–2107. [[CrossRef](#)]
46. Makantasis, K.; Doulamis, A.D.; Doulamis, N.D.; Nikitakis, A. Tensor-based classification models for hyperspectral data analysis. *IEEE T. Geosci. Remote. Sens.* **2018**, *9*, 1–15. [[CrossRef](#)]
47. Zou, Q.; Ni, L.; Zhang, T.; Wang, Q. Deep Learning Based Feature Selection for Remote Sensing Scene Classification. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 2321–2325. [[CrossRef](#)]
48. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. *Remote Sens.* **2015**, *7*, 14680–14707. [[CrossRef](#)]
49. Liu, Y.; Zhong, Y.; Fei, F.; Zhu, Q.; Qin, Q. Scene Classification Based on a Deep Random-Scale Stretched Convolutional Neural Network. *Remote Sens.* **2018**, *10*, 444. [[CrossRef](#)]
50. Rey, N.; Volpi, M.; Joost, S.; Tuia, D. Detecting animals in African Savanna with UAVs and the crowds. *Remote Sens. Environ.* **2017**, *200*, 341–351. [[CrossRef](#)]
51. Cheng, G.; Zhou, P.; Han, J. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 7405–7415. [[CrossRef](#)]
52. Ding, P.; Zhang, Y.; Deng, W.J.; Jia, P.; Kuijper, A. A light and faster regional convolutional neural network for object detection in optical remote sensing images. *ISPRS J. Photogramm.* **2018**, *141*, 208–218. [[CrossRef](#)]
53. Tang, T.; Zhou, S.; Deng, Z.; Lei, L.; Zou, H. Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks. *Remote Sens.* **2017**, *9*, 1170. [[CrossRef](#)]
54. Han, X.; Zhong, Y.; Zhang, L. An efficient and robust integrated geospatial object detection framework for high spatial resolution remote sensing imagery. *Remote Sens.* **2017**, *9*, 666. [[CrossRef](#)]
55. Li, W.; He, C.; Fang, J.; Fu, H. Semantic Segmentation based Building Extraction Method using Multi-source GIS Map Datasets and Satellite Imagery. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 238–241.
56. Fu, G.; Liu, C.; Zhou, R.; Sun, T.; Zhang, Q. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sens.* **2017**, *9*, 498. [[CrossRef](#)]
57. Wang, H.; Wang, Y.; Zhang, Q.; Xiang, S.; Pan, C. Gated convolutional neural network for semantic segmentation in high-resolution images. *Remote Sens.* **2017**, *9*, 446. [[CrossRef](#)]
58. Guirado, E.; Tabik, S.; Alcaraz-Segura, D.; Cabello, J.; Herrera, F. Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study. *Remote Sens.* **2017**, *9*, 1220. [[CrossRef](#)]
59. Pibre, L.; Chaumont, M.; Subsol, G.; Ienco, D.; Derras, M. How to deal with multi-source data for tree detection based on deep learning. In Proceedings of the GlobalSIP: Global Conference on Signal and Information Processing, Montreal, QC, Canada, 14–16 November 2017.
60. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–8 December 2012; pp. 1097–1105.

61. Li, W.; Fu, H.; You, Y.; Yu, L.; Fang, J. Parallel Multiclass Support Vector Machine for Remote Sensing Data Classification on Multicore and Many-Core Architectures. *IEEE J. Sel. Top. Appl. Earth Observ.* **2017**, *10*, 4387–4398. [[CrossRef](#)]
62. Fassnacht, F.E.; Latifi, H.; Stereńczak, K.; Modzelewska, A.; Lefsky, M.; Waser, L.T.; Straub, C.; Ghosh, A. Review of studies on tree species classification from remotely sensed data. *Remote Sens. Environ.* **2016**, *186*, 64–87. [[CrossRef](#)]
63. Dalponte, M.; Ene, L.T.; Marconcini, M.; Gobakken, T.; Næsset, E. Semi-supervised SVM for individual tree crown species classification. *ISPRS J. Photogramm.* **2015**, *110*, 77–87. [[CrossRef](#)]
64. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 7–12 December 2015; pp. 91–99.

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).