Cross-Regional Oil Palm Tree Detection

Wenzhao Wu[†,¶], Juepeng Zheng[†,¶], Haohuan Fu[†] (✉), Weijia Li[§], Le Yu[†]
[†]Ministry of Education Key Laboratory for Earth System Modeling,
and Department of Earth System Science, Tsinghua University, Beijing 100084, China
[§]CUHK-SenseTime Joint Lab, The Chinese University of Hong Kong, Hong Kong, China
{wumz17, zjp19}@mails.tsinghua.edu.cn, {haohuan, leyu}@tsinghua.edu.cn, weijiali@cuhk.edu.hk

Abstract

As oil palm has become one of the most rapidly expanding tropical crops in the world, detecting and counting oil palms have received considerable attention. Although deep learning has been widely applied to remote sensing image processing including tree crown detection, the large size and the variety of the data make it extremely difficult for cross-regional and large-scale scenarios. In this paper, we propose a cross-regional oil palm tree detection (CROPTD) method. CROPTD contains a local domain discriminator and a global domain discriminator, both of which are generated by adversarial learning. Additionally, since the local alignment does not take full advantages of its transferability information, we improve the local module with the local attention mechanism, taking more attention on more transferable regions. We evaluate our CROPTD on two large-scale high-resolution satellite images located in Peninsular Malaysia. CROPTD improves the detection accuracy by 8.69% in terms of average F1-score compared with the Baseline method (Faster R-CNN) and performs 4.99-2.21% better than other two state-of-the-art domain adaptive object detection approaches. Experimental results demonstrate the great potential of our CROPTD for large-scale, cross-regional oil palm tree detection, guaranteeing a high detection accuracy as well as saving the manual annotation efforts. Our training and validation dataset are available on `https://github.com/rs-dl/CROPTD`.

1. Introduction

Oil palm is a vital economic crop for many tropical developing countries, like Malaysia and Indonesia, which hold over 80% palm oil production in the world [1, 2]. Although palm oil is largely used in food products, cosmetic materials and energy source, expanding demand for oil palm plantation induce massive deforestation, wildlife habitats threat and detrimental environment effects. To this end, offering a accurate and real-time assessment of palm tree plantation in a large-scale region can bring significant impacts on both economic and ecological aspects.

However, the tremendous spatial scale and the variety of geological features across regions have made it a grand challenge with limited solutions based on manual human monitoring efforts. Although unprecedented deep learning algorithms have demonstrated potential in forming an automated approach for tree crown detection in recent years [3, 4, 5], the labelling efforts needed for covering different features in different regions hinder its effectiveness from large-scale remote sensing applications using multi-temporal and multi-sensor satellite images.

Large-scale and cross-regional oil palm tree investigation are meaningful research topics. Nowadays, the affluent remote sensing images and rapid development of deep learning algorithms bring new opportunities to large-scale and cross-regional oil palm detection. However, large-scale tree counting and detection may be confronted with remote sensing images with diverse acquisition conditions, like different sensors, illumination and environments, resulting in different distribution and domain shifts among images. One solution is to add labeled training data in new regions, which can compensate for the performance deficiency. Nevertheless, this way is fairly expensive, time-consuming and often infeasible for practical applications.

Fortunately, domain adaptation (DA) methods can help adapt the model to new data domains without adding a large quantity of new labels, which has gained lots of attention over the past decades. A variety of approaches have been developed through domain-invariantly aligning the input image or hidden feature distribution space, or both of them. Recent DA methods dive into the direction bridging the gap in feature distribution between the source and target domains using adversarial learning and achieve encouraging results [6, 7, 8, 9, 10]. Most of existing efforts are dedicated to image classification and semantic segmentation. However, object detection is a technically different

[¶]These authors contributed equally to this work

2160-7516/20/$31.00 ©2020 IEEE
DOI 10.1109/CVPRW50498.2020.00036

issue because we are supposed to focus more on the objects of interest instead of background or scene layouts.

In this paper, we propose a cross-regional oil palm tree detection (CROPTD) method using high-resolution remote sensing images located in Malaysia. Our contribution can be summarized as follows.

(1) We propose a large-scale, real-time and cross-regional oil palm tree detection via CROPTD. To the best of our knowledge, this is the first work for domain adaptive tree crown detection using end-to-end strategy.

(2) We implement our CROPTD through constructing a local domain discriminator and global domain discriminator with adversarial learning. Additionally, we propose the local attention to highlight transferable regions and alleviate negative transfer for each region.

(3) Our method achieves highest detection accuracy on two large-scale high-resolution satellite images located in Peninsular Malaysia. We release our training dataset and validation dataset, hoping that our dataset can promote the development of cross-regional tree crown detection from high-resolution satellite images.

The rest of this paper is organized as follows. We review the related works in the next section. Following that, we present our CROPTD method in Section 3. Our study area and datasets are introduced in Section 4. In Section 5, we provide detection results and compare with other state-of-the-art methods. Finally, we summarize our paper in Section 6.

2. Related Works

2.1. Tree crown detection

Existing tree crown detection algorithm can be grouped into three types, including classical image processing methods [11, 12], traditional machine learning methods [13, 14] and deep learning methods. The first two types of methods generally require sophisticated image processing procedures or very high-resolution and very high-quality images. Following the significant success of deep learning, many tree crown detection algorithms have utilized convolutional neural network (CNN) to attain eminent performance in more complex scenes and larger study area. There are three cases for deep learning tree crown detection approaches: CNN classification based methods [3, 15], semantic segmentation based methods [4, 16] and object detection based methods [5, 17].

Nowadays, object detection can be categorized as two classes: two-stage object detection [18, 19] and one-stage object detection [20, 21]. Faster R-CNN [18] is a representative two-stage object detection algorithm, where the first stage generates candidate region proposals through region proposal network (RPN) and the second stage aims at classification and bounding-box regression tasks for candidates obtained from the first stage. Our method is inherited from Faster R-CNN. With regard to tree crown detection, Zheng *et al*. firstly applied end-to-end object detection in tree crown detection using satellite images, achieving an average F1-score of over 90% in six study regions [17]. Others have utilized object detection based method for tree crown detection using higher-resolution remote sensing images photograghed by unmanned automatic vehicle [5, 22] and aerial plane [23, 24]. As traditional deep learning methods focus on grasping the texture patterns in different images, the performance of trained network in one set of satellite images would degrade significantly when moving to images that are taken in a different region or from a different source.

2.2. Domain adaptive object detection

There are a body of previous literatures in DA, most of which focus on image classification [6, 25] and semantic segmentation [26, 27] tasks. However, domain adaptive object detection is still at a relatively earlier stage. Chen *et al*. employed instance-level and image-level domain classifier with consistency regularizer to reduce the domain gap [28]. Satio *et al*. proposed strong local alignment and weak global alignment to improve the performance on the target domain [29]. The weak alignment model focuses adversarial alignment loss on images that are globally similar. The strong alignment model only puts emphasis on local receptive fields of the feature map. Furthermore, SCL [30] learns more discriminative representations by interacting different losses and training strategy. Recently, Alqasir *et al*. indicated that adapting the region proposal sub-network in Faster R-CNN is crucial [31]. Existing works strive into diminishing the domain gap in both global feature (or image) level and local feature (or instance) level. However, such efforts have not explored the transferability of local features. Especially, the difference transfer abilities in different regions have not been investigated, and not utilized.

2.3. Domain adaptation in remote sensing field

DA has been exploited in the remote sensing community to tackle with multi-temporal and multi-source satellite images, where differences in atmospheric illuminations and ground conditions can easily ruin the adaptation of a model [32]. For remote sensing image classification, Matasci *et al*. analyzed the effectiveness of TCA in multi- and hyperspectral image classification, and explored its unsupervised and semi-supervised implementation [33]. Recently, Zhu *et al*. proposed a semi-supervised adversarial learning domain adaptation framework for scene classification, reaching an overall accuracy of over 93% in different temporal aerial

images [34]. Existing DA methods applied in remote sensing field mainly concentrate on the classification tasks such as land cover and land use mapping, hyperspectral image classification and scene classification. On the contrary, the study of domain adaptive semantic segmentation and object detection in remote sensing field is relatively limited. In semantic segmentation, Benjdira *et al*. used GANs to reduce the domain shift of aerial images, improving the average segmentation accuracy from 14% to 61% [35]. Liu *et al*. reduced the discrepancy between source and target domain based on conditional generative adversarial networks. As for domain adaptive object detection, to the best of our knowledge, Koga *et al*. firstly applied CORAL and adversarial DA to a vehicle detector for satellite images [36].

For domain adaptive object detection, fully matching the whole distributions of source and target images at global level may incur performance drop. In this paper, we propose CROPTD to implement cross-regional oil palm tree detection using multi-temporal and multi-sensor satellite images. Besides global domain discriminator, we present local domain discriminator and utilize local attention value to obtain more transferable regions in an image, which is quite essential for domain adaptive object detection.

3. Methodology

In this section, we introduce our CROPTD in detail. Figure 1 illustrates the architecture of CROPTD. In this paper, we concentrate on unsupervised domain adaptive oil palm detection across two different high-resolution remote sensing images, which consists of an annotated source domain dataset $D_S = \left\{\left(x_i^S, \mathbf{y_i^S}\right)\right\}_{i=1}^{n_S}$ in the source region (R_S) and an unlabeled target domain dataset $D_T = \left\{\left(x_i^T\right)\right\}_{i=1}^{n_T}$ in the target region (R_T), where x_i is an image and $\mathbf{y_i}$ is the corresponding labels including bounding-boxes and classes.

3.1. Local transfer loss and local attention mechanism

We construct a local domain discriminator by adversarial learning and utilizing the local discriminator's output to generate an attention value for each region, distinguishing our method from Strong-Weak [29] in which the transferability of fine-grained structures in an image has not been fully utilized. The local feature map ($\mathbf{M_l}$) is extracted from the lower convolutional layers (Conv 1 in Figure 1). The $\mathbf{M_l}$ is input into the local domain classifier (G_l). D_l outputs a domain prediction map of $\mathbf{M_l}$, which has the same width (W) and height (H) as the input $\mathbf{M_l}$. Therefore, the local transfer loss can be calculated as equation (1).

$$L_l = \frac{1}{nHW}\sum_{i=1}^{n}\sum_{w=1}^{W}\sum_{h=1}^{H} D_l\left(G_l\left(M_{li}^{w,h}\right),\, d_i\right) \tag{1}$$

where $n = n_S + n_T$, is the total number of the source image (n_S) and the target image (n_T) in a batch size. d_i is the domain label of point $\mathbf{M_l}$ and D_l utilize cross entropy loss as the local transfer loss function. Notably, $d_i = 1$ represents the input image is from source domain and $d_i = 0$ represents from target domain.

Inspired by TADA [10], the transferable attention emphasizes on the distinction or similarity between two images. This concept is reasonable since we hope that the neural network could take more attention on more transferable regions. For example, in this study, we hope that more attention may be allocated to the oil palm plantation rather than buildings, roads or other vegetation area. Hence, we propose local attention mechanism to re-weight our network. We adopt local attention value to assess the transferability of a region. The local attention value for each location ($v_i^{w,h}$) of local feature map can be formulated as (2).

$$\begin{aligned} v_i^{w,h} &= 1 - E\left(\hat{d}_i^{w,h}\right) \\ E\left(p\right) &= -\sum_j p_j \cdot log\left(p_j\right) \end{aligned} \tag{2}$$

where $\hat{d}_i^{w,h} = G_l\left(M_{li}^{w,h}\right)$, is the probability of the region (w, h) in local feature map for image i belonging to the source domain, indicating that the region (w, h) belongs to the source domain when the probability approaches 1 and to the target domain when the probability approaches 0. $E\left(\cdot\right)$ is the entropy criterion, which is an uncertain measure.

Following the idea of residual mechanism [37], the local feature with local attention value ($H_{li}^{w,h}$) can be defined as (3)

$$H_{li}^{w,h} = \left(1 + v_i^{w,h}\right) \cdot M_{li}^{w,h} \tag{3}$$

3.2. Global transfer loss

The global feature map (M_g) is extracted by the whole backbone and the M_g is input into the global domain classifier (D_g). Notably, $M_g = G_{Conv2}\left(H_l\right)$ and G_{Conv2} stands for the higher convolutional layers in the backbone. D_g outputs a domain prediction result of M_g, which is the general domain classifier like other DA methods. However, here D_g utilize the focal loss [38] as the global transfer loss rather than cross entropy loss, as it can effectively ignore easy-to-classify samples yet focus on hard-to-classify samples. To this end, the global transfer loss is refined as (4).

$$L_g = \frac{1}{n}\sum_{i=1}^{n} D_g\left(G_g\left(M_g\right), d_i\right) \tag{4}$$

To stabilize the adversarial training [29], context vectors from local and global domain classifier are concatenated to

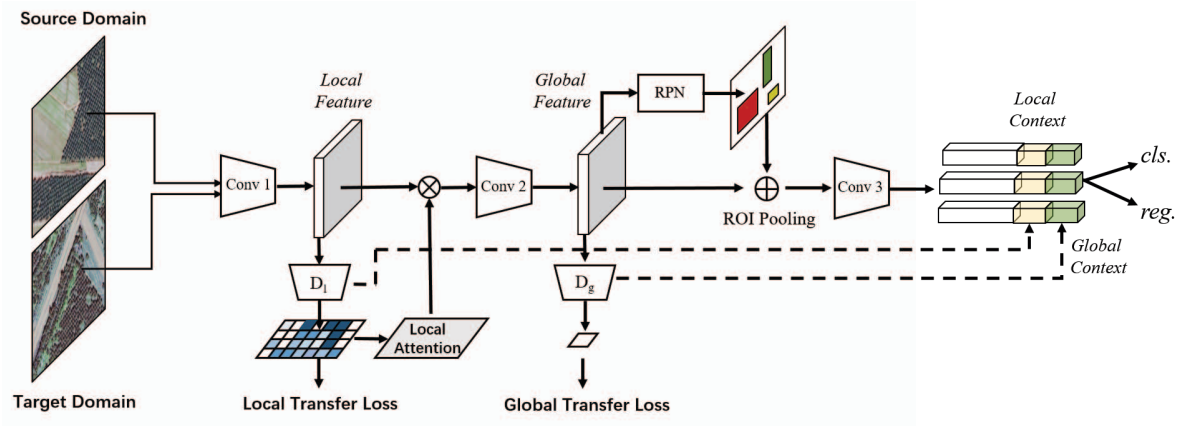

Figure 1. The flowchart of our proposed CROPTD. CROPTD contains three loss, *e.g.* local transfer loss (L_l), global transfer loss (L_g) and Faster R-CNN loss (L_{fr}). The L_l and L_g are generated by local domain discriminator (D_l) and global domain discriminator (D_g), respectively. The local attention map develops the representations of those regions with higher transferability. Conv 1, Conv 2 and Conv 3 stand for the lower convolutional layers, higher convolutional layers and other modules after RPN, respectively. Additionally, CROPTD includes context vector based regularization like Strong-Weak [29].

region-wise feature after RPN and Conv 3 as shown in Figure 1. The context vector based regularization can make the training of domain classifier more robust when minimizing the detection loss from the source images.

3.3. Overall objective loss for CROPTD.

To this end, the overall objective function can be formulated as (5).

$$C\left(\theta_F, \theta_R, \theta_{D_l}, \theta_{D_g}\right) = L_{fr} - \left(\lambda L_l + \gamma L_g\right) \tag{5}$$

where λ and γ are hyper-parameters. θ_F, θ_R, θ_{D_l} and θ_{D_g} are the parameters for backbone feature extractor, RPN, local domain discriminator and global domain discriminator. L_{fr} represents the loss components of Faster R-CNN [18] including RPN loss, classification loss and bounding-box loss. It can be summarized as (6).

$$L_{fr} = -\frac{1}{n_S}\sum_{i=1}^{n_S} D_{det}\left(R\left(F\left(x_i^S\right)\right), y_i^S\right) \tag{6}$$

where D_{det} includes all loss functions in Faster R-CNN. R and F represent the RPN and backbone extractor, respectively. The minimax optimization problem is to find the network parameters $\hat{\theta}_F$, $\hat{\theta}_R$, $\hat{\theta}_{D_l}$ and $\hat{\theta}_{D_g}$ that jointly satisfy (7).

$$\begin{aligned}\left(\hat{\theta}_F, \hat{\theta}_R\right) &= \arg\min_{\theta_F, \theta_R} C\left(\theta_F, \theta_R\right) \\ \left(\hat{\theta}_{D_g}, \hat{\theta}_{D_l}\right) &= \arg\max_{\theta_{D_g}, \theta_{D_l}} C\left(\theta_{D_g}, \theta_{D_l}\right)\end{aligned} \tag{7}$$

4. Study area and datasets

Our study area is located in the Peninsular Malaysia as is shown in Figure 2, where the oil palm plantation is expanding rapidly and threatening the local environment and native species. According to the statistics in 2016, 47% of Malaysia oil palm plantation was in the Peninsular Malaysia [39]. We have two high-resolution satellite images, Image A and Image B, in this work. Table 1 shows the elaborate information of these two satellite images. They are acquired from different sensors and locations, and the interval of photograph date is about 10 years, resulting in differences in reflectance, resolution, illumination and environmental conditions.

Table 1. The necessary information about Image A and Image B

Index	Image A	Image B
Source	QuickBird	Google Earth
Longitude	103.5991E	100.7772E
Latitude	1.5967N	4.1920N
Spectral	RGB, NIR	RGB
Acquisition	Nov 21, 2006	Dec 21, 2015
Resolution	0.6m	0.3m
Image size	12,188 × 12,576	10,496 × 10,240
Area	55.18km^2	9.67km^2
Palm number	291,827	91,357

Figure 3 and 4 display our study area and datasets. Our training dataset are collected from four regions (red rectangles) in Image A and Image B, respectively, and validation dataset are collected from one region (blue rectangles) in them. We evaluate our method by testing the whole target satellite image. Our training and validation dataset are available on `https://github.com/rs-dl/CROPTD`

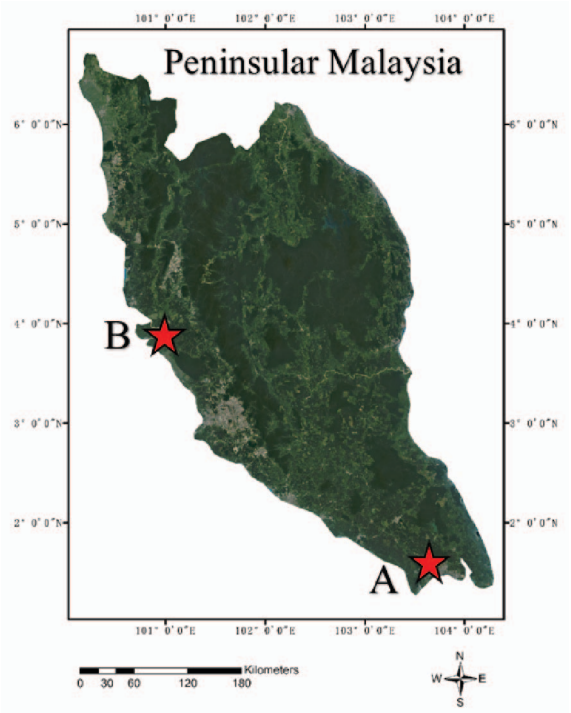

Figure 2. The location of our study area. Image A and Image B are acquired from different sensors and locations.

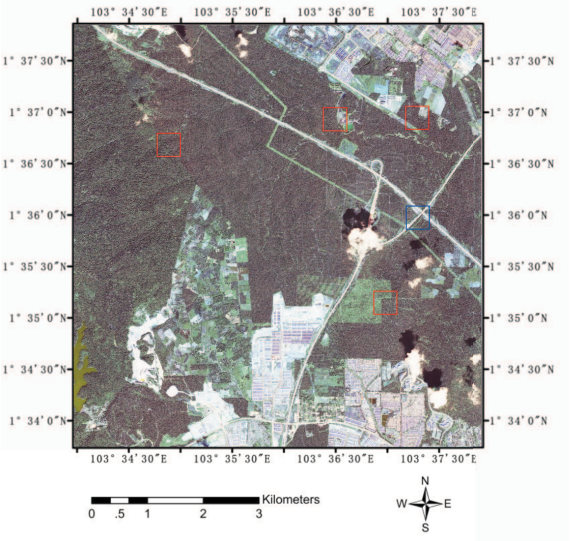

Figure 3. The dataset of Image A. Four training regions are in red rectangles and one validation region is in blue rectangle. We test Image B→ Image A on the whole Image A with 291,827 palms manually annotated.

In training and validation areas, we applied the bilinear interpolation so that the original training images are resized to 2,400 × 2,400 pixels. After that, we randomly cropped the enlarged images with 500 × 500 pixels and generated the training dataset of 4,000 samples as the input of deep

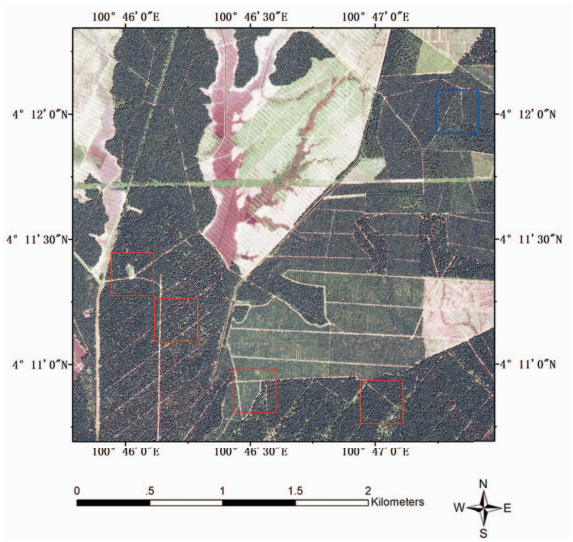

Figure 4. The dataset of Image B. Four training regions are in red rectangles and one validation region is in blue rectangle. We test Image A→ Image B on the whole Image B with 91,357 palms manually annotated.

neural network and the validation dataset of 1,000 samples. As for inference phase, we firstly resized the whole satellite image using the same scale (enlarge by 4 times both in width and height for Image A while 2 times for Image B) as above and cropped them with 66 pixels overlapped. (Given that the size of an oil palm tree in QuickBird image with 0.6m spatial resolution is about 17 × 17 pixels and 65 × 65 pixels or so after image resizing).

5. Experiments

5.1. Setup

We evaluate our CROPTD on two tasks: Image A → Image B (A → B) and Image B → Image A (B → A). We implement our method based on PyTorch [40]. Our backbone network is ResNet-101 [41] and the optimizer is SGD [42] with initial learning rate 0.001 divided by 10 every 5 epochs. We set $\lambda = 0.1$ and $\gamma = 0.1$. The modulating factor in global transfer loss is set to 5. We adopt prevalent evaluation protocal for large-scale object detection in remote sensing field, including true positives (TP), false positives (FP), false negatives (FN), precision, recall and F1-score. The precision, recall and F1-score can be calculated as (8). Additionally, the detecting palms whose probability score is higher than 0.5 and intersection-over-union (IoU) metric with ground-truth palms is higher than 0.5 will be considered as correct oil palms (TP) [24, 43].

$$precision = \frac{TP}{TP + FP}$$
$$recall = \frac{TP}{TP + FN} \quad (8)$$
$$F1 - score = \frac{2 \times precision \times recall}{precision + recall}$$

5.2. Results

We list TP, FP, FN, precision, recall and F1-score for A $\rightarrow$ B and B $\rightarrow$ A in Table 2. Our CROPTD yields an F1-score of 94.04% for A $\rightarrow$ B, and 86.67% for B $\rightarrow$ A. Figure 5 illustrates four example regions for these two transfer tasks. The green points stand for the correct detected oil palms (TP), the blue squares stand for the ground-truth oil palms that are missing (FN), and the red squares with red points stand for other types of objects like other vegetation or building corners that are detected as oil palms by mistake (FP).

Table 2. The detection results of CROPTD.

Index	A $\rightarrow$ B	B $\rightarrow$ A
TP	83,547	263,514
FP	2,771	52,718
FN	7,810	28,313
Precision	96.79%	83.33%
Recall	91.45%	90.30%
F1-score	94.04%	86.67%

Table 3 shows the comparison between our CROPTD and other state-of-the-art methods, including Faster R-CNN (Baseline) [18], DA Faster [28] and Strong-Weak [29]. We can observe that CROPTD performs better than other methods, beyond Strong-Weak by 2.21% in respect of average F1-score. It is noteworthy that not all domain adaptive object detection methods outperform Baseline. For example, the performance of DA Faster for B $\rightarrow$ A is worse than Faster R-CNN. Figure 5 describes the detection results of all of these methods. Results indicate that CROPTD outperforms other domain adaptive object detection methods in two transfer tasks, achieving less confusion between oil palms and other object instances such as other vegetation and impervious. Experimental performances demonstrate the great potential of our CROPTD for large-scale, cross-regional and real-time oil palm tree detection, guaranteeing a high detection accuracy as well as saving the manual annotation efforts.

5.3. Local attention map visualization

To evaluate whether our local attention map could focus on the desirable regions (in particular, the oil palm plantation rather than other vegetation or impervious area) in the image, we randomly sample some input images from the source domain (Image A) and target domain (Image B). As shown in Figure 6, different regions in the images have different corresponding local attention masks in our network. In each group of images, we illustrated from left to right by the original input images, the corresponding local attentions and the local attentions shown in the original input images, respectively. The red color stands for the higher local attention value while the blue color denotes the lower local attention value. Take the image on the top-right of Figure 6 as a case, where oil palm plantation surrounds around the buildings. The oil palm plantation mask is highlighted with red color while the buildings mask diminishes in blue color. These local attention map intuitively show that the local attention mechanism can produce reasonably meaningful regions for fine-grained transfer.

6. Conclusion

In this paper, we propose a large-scale, real-time and cross-regional oil palm tree detection (CROPTD) approach. Our CROPTD contains a local domain discriminator and global domain discriminator, Both of which are generated by adversarial learning. Additionally, we propose the local attention to highlight more transferable regions according to local attention value, telling our neural network where more transferable regions are. We evaluate our CROPTD on two large-scale high-resolution satellite images located in Peninsular Malaysia. Experiments demonstrate our method performs better than other state-of-the-art methods, beyond Strong-Weak by 2.21% in respect of average F1-score. In the future, we envision a more robust and efficient cross domain object detector in a larger-scale remote sensing images.

7. Acknowledgements

This research was supported in part by the National Key Research and Development Plan of China (Grant No. 2017YFA0604500, 2017YFB0202204 and No.2017YFA0604401), the National Natural Science Foundation of China (Grant No. 51761135015), and by Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao).

References

[1] L. P. Koh and D. S. Wilcove, "Cashing in palm oil for conservation," *Nature*, vol. 448, no. 7157, pp. 993–994, 2007.

[2] J. C. Quezada, A. Etter, J. Ghazoul, A. Buttler, and T. Guillaume, "Carbon neutral expansion of oil palm plantations in the neotropics," *Science Advances*, vol. 5, no. 11, p. eaaw4418, 2019.

[3] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution re-

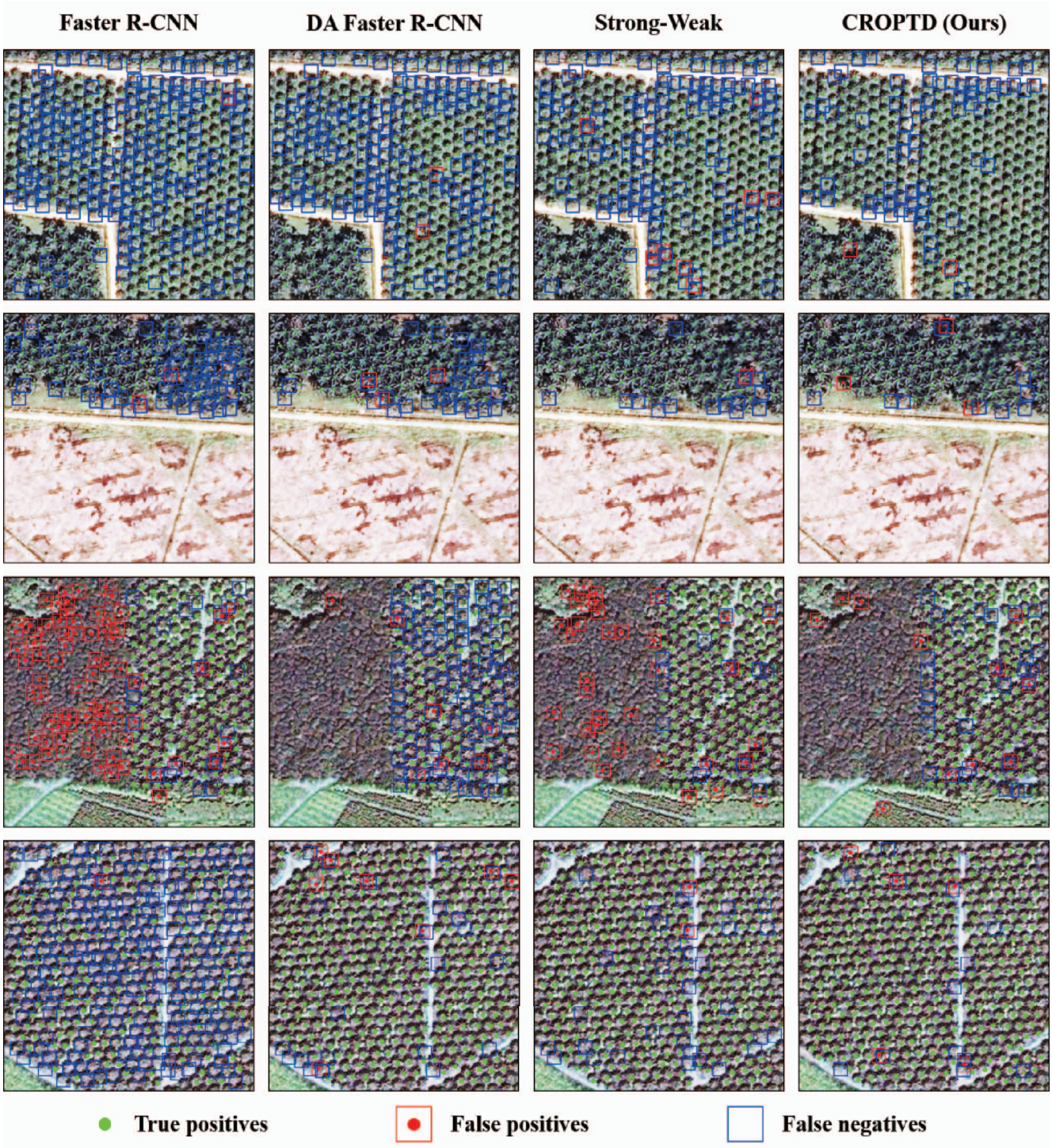

Figure 5. The results of example regions for different methods. The top two rows are the results of A→ B and the bottom two rows are the results of B→ A. From left to right, the corresponding methods are Faster R-CNN [18], DA Faster R-CNN [28], Strong-Weak [29] and CROPTD (ours), respectively. The green points stand for the correct detected oil palms (TP), the blue squares stand for the ground-truth oil palms that are missing (FN), and the red squares with red points stand for other types of objects detected as oil palms by mistake (FP).

mote sensing images," *Remote Sensing*, vol. 9, no. 1, p. 22, 2017.

[4] L. P. Osco, M. d. S. de Arruda, J. M. Junior, N. B. da Silva, A. P. M. Ramos, É. A. S. Moryia, N. N. Imai, D. R. Pereira,

Table 3. Comparison of the detection results between CROPTD and other state-of-the-art methods

Method	A→ B			B→ A			Average
	Precision	Recall	F1-score	Precision	Recall	F1-score	F1-score
Faster R-CNN [18]	**98.68%**	71.08%	82.63%	72.12%	**91.60%**	80.70%	81.67%
DA Faster R-CNN [28]	97.87%	86.76%	91.98%	80.85%	76.74%	78.75%	85.37%
Strong-Weak [29]	97.84%	87.76%	92.53%	78.98%	89.15%	83.76%	88.15%
CROPTD (ours)	96.79%	**91.45%**	**94.04%**	**83.33%**	90.30%	**86.67%**	**90.36%**

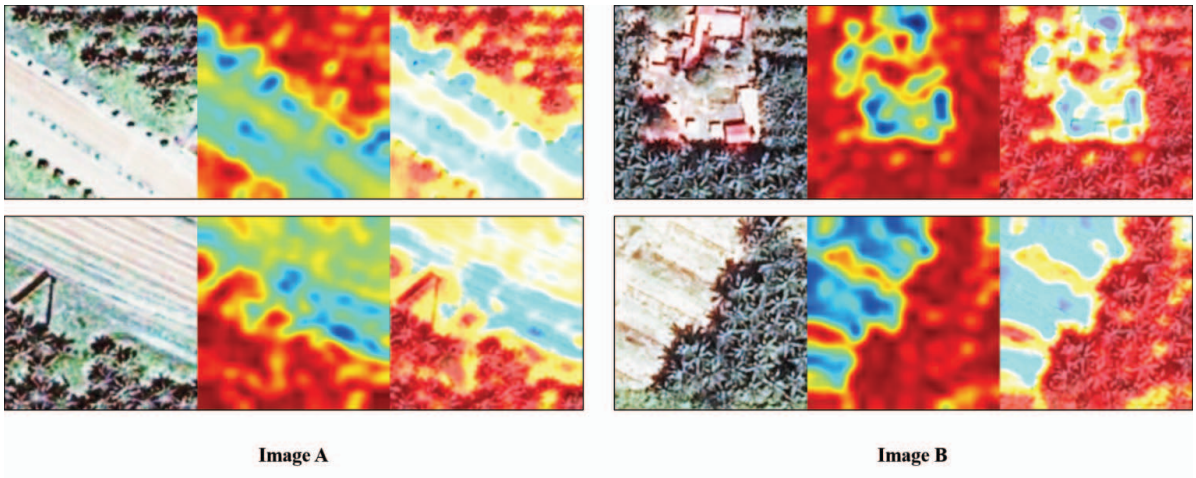

Figure 6. Local attention visualization of the last convolutional layer of Conv1 on A → B. The images on the left are randomly sampled from source domain (Image A) while the right from target domain (Image B). The red color stands for the higher local attention value while the blue color denotes the lower local attention value. In each group of images, we illustrated from left to right by the original input images, the corresponding local attentions and the local attentions shown in the original input images, respectively. These local attention map intuitively show that the local attention mechanism can produce reasonably meaningful regions for fine-grained transfer.

J. E. Creste, E. T. Matsubara, *et al.*, "A convolutional neural network approach for counting and geolocating citrus-trees in uav multispectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 160, pp. 97–106, 2020.

[5] A. A. d. Santos, J. Marcato Junior, M. S. Araújo, D. R. Di Martini, E. C. Tetila, H. L. Siqueira, C. Aoki, A. Eltner, E. T. Matsubara, H. Pistori, *et al.*, "Assessment of cnn-based methods for individual tree detection on images captured by rgb cameras attached to uavs," *Sensors*, vol. 19, no. 16, p. 3595, 2019.

[6] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," *arXiv preprint arXiv:1412.3474*, 2014.

[7] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, 2016.

[8] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7167–7176, 2017.

[9] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Advances in Neural Information Processing Systems*, pp. 1640–1650, 2018.

[10] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, "Transferable attention for domain adaptation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5345–5352, 2019.

[11] M. Wulder, K. O. Niemann, and D. G. Goodenough, "Local maximum filtering for the extraction of tree locations and basal area from high spatial resolution imagery," *Remote Sensing of environment*, vol. 73, no. 1, pp. 103–114, 2000.

[12] D. Pouliot, D. King, F. Bell, and D. Pitt, "Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration," *Remote sensing of environment*, vol. 82, no. 2-3, pp. 322–334, 2002.

[13] C. Hung, M. Bryson, and S. Sukkarieh, "Multi-class predictive template for tree crown detection," *ISPRS journal of photogrammetry and remote sensing*, vol. 68, pp. 170–183, 2012.

[14] R. Pu and S. Landry, "A comparative analysis of high spatial resolution ikonos and worldview-2 imagery for mapping urban tree species," *Remote Sensing of Environment*, vol. 124, pp. 516–533, 2012.

[15] N. A. Mubin, E. Nadarajoo, H. Z. M. Shafri, and A. Hamedianfar, "Young and mature oil palm tree detection and counting using convolutional neural network deep learning method," *International Journal of Remote Sensing*, vol. 40, no. 19, pp. 7500–7515, 2019.

[16] C. Xiao, R. Qin, and X. Huang, "Treetop detection using convolutional neural networks trained through automatically generated pseudo labels," *International Journal of Remote Sensing*, vol. 41, no. 8, pp. 3010–3030, 2020.

[17] J. Zheng, W. Li, M. Xia, R. Dong, H. Fu, and S. Yuan, "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*, pp. 1422–1425, IEEE, 2019.

[18] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, pp. 91–99, 2015.

[19] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2117–2125, 2017.

[20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*, pp. 21–37, Springer, 2016.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.

[22] M. Xia, W. Li, H. Fu, L. Yu, R. Dong, and J. Zheng, "Fast and robust detection of oil palm trees using high-resolution remote sensing images," in *Automatic Target Recognition XXIX*, vol. 10988, p. 109880C, International Society for Optics and Photonics, 2019.

[23] S. Puttemans, K. Van Beeck, and T. Goedemé, "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, 2018.

[24] B. G. Weinstein, S. Marconi, S. Bohlman, A. Zare, and E. White, "Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks," *Remote Sensing*, vol. 11, no. 11, p. 1309, 2019.

[25] X. Wang, Y. Jin, M. Long, J. Wang, and M. I. Jordan, "Transferable normalization: Towards improving transferability of deep neural networks," in *Advances in Neural Information Processing Systems*, pp. 1951–1961, 2019.

[26] Y. Zhang, P. David, and B. Gong, "Curriculum domain adaptation for semantic segmentation of urban scenes," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2020–2030, 2017.

[27] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, and T. Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International Conference on Machine Learning*, pp. 1989–1998, 2018.

[28] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, "Domain adaptive faster r-cnn for object detection in the wild," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3339–3348, 2018.

[29] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, "Strong-weak distribution alignment for adaptive object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6956–6965, 2019.

[30] Z. Shen, H. Maheshwari, W. Yao, and M. Savvides, "Scl: Towards accurate domain adaptive object detection via gradient detach based stacked complementary losses," *arXiv preprint arXiv:1911.02559*, 2019.

[31] H. Alqasir, D. Muselet, and C. Ducottet, "Region proposal oriented approach for domain adaptive object detection," in *International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 38–50, Springer, 2020.

[32] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE geoscience and remote sensing magazine*, vol. 4, no. 2, pp. 41–57, 2016.

[33] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 7, pp. 3550–3564, 2015.

[34] R. Zhu, L. Yan, N. Mo, and Y. Liu, "Semi-supervised center-based discriminative adversarial learning for cross-domain scene-level land-cover classification of aerial images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 155, pp. 72–89, 2019.

[35] B. Benjdira, Y. Bazi, A. Koubaa, and K. Ouni, "Unsupervised domain adaptation using generative adversarial networks for semantic segmentation of aerial images," *Remote Sensing*, vol. 11, no. 11, p. 1369, 2019.

[36] Y. Koga, H. Miyazaki, and R. Shibasaki, "A method for vehicle detection in high-resolution satellite images that uses a region-based object detector and unsupervised domain adaptation," *Remote Sensing*, vol. 12, no. 3, p. 575, 2020.

[37] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3156–3164, 2017.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988, 2017.

[39] K. H. D. Tang and H. M. Al Qahtani, "Sustainability of oil palm plantations in malaysia," *Environment, Development and Sustainability*, pp. 1–25, 2019.

[40] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, pp. 8024–8035, 2019.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[42] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*, pp. 177–186, Springer, 2010.

[43] A. Van Etten, D. Lindenbaum, and T. M. Bacastow, "Spacenet: A remote sensing dataset and challenge series," *arXiv preprint arXiv:1807.01232*, 2018.