Multisource-Domain Generalization-Based Oil Palm Tree Detection Using Very-High-Resolution (VHR) Satellite Images

Juepeng Zheng, Wenzhao Wu, Shuai Yuan, Haohuan Fu, *Member, IEEE*, Weijia Li, and Le Yu, *Member, IEEE*

***Abstract*— Providing accurate and timely oil palm information on a large scale is essential for both economic development and ecological significance. However, owing to different sensors, photograph acquisition conditions, and environmental heterogeneity, the large volume and the variety of the data make it extremely challenging for large-scale and cross-regional oil palm tree detection. It is computationally expensive to train a model from images covering large heterogeneous regions and all environmental conditions for continuously accumulated multisource remote sensing data. In this letter, we propose a new multisource domain generalization (DG) method, Maximum Mean Discrepancy Deep Reconstruction Classification Network (MMD-DRCN). It learns representations from multiple source domains and obtains inspiring performance in an unknown and "unseen" target domain. Besides classification loss, our MMD-DRCN distills more representative features through reconstruction loss and aligns multisource latent features by MMD loss, both of which effectively enhance the capacity of generalization. MMD-DRCN achieves an average F1-score of 82.70% in all transfer tasks, attaining a 5.83% gain compared to Baseline (a straightforward convolutional neural network (CNN) model). Experimental results demonstrate DG poses a promising potential for large-scale and cross-regional oil palm tree detection without any information of the target domain.**

***Index Terms*— Deep learning, domain generalization (DG), multiple sources, oil palm, tree crown detection.**

Manuscript received October 30, 2020; revised January 21, 2021; accepted February 18, 2021. Date of publication March 9, 2021; date of current version December 16, 2021. This work was supported in part by the National Key Research and Development Plan of China under Grant 2017YFA0604500, Grant 2017YFB0202204, and Grant 2017YFA0604401; in part by the National Natural Science Foundation of China under Grant 51761135015 and Grant U1839206; in part by the National Key Scientific and Technological Infrastructure Project "Earth System Science Numerical Simulator Facility" (EarthLab); and in part by the Center for High Performance Computing and System Simulation, Pilot National Laboratory for Marine Science and Technology (Qingdao). *(Juepeng Zheng and Wenzhao Wu contributed equally to this work.) (Corresponding author: Haohuan Fu.)*

Juepeng Zheng and Le Yu are with the Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing 100084, China, and also with the Department of Earth System Science, Tsinghua University, Beijing 100084, China (e-mail: zjp19@mails.tsinghua.edu.cn; leyu@tsinghua.edu.cn).

Wenzhao Wu is with the National Supercomputing Center in Wuxi, Wuxi 214072, China (e-mail: wumz13@tsinghua.org.cn).

Shuai Yuan is with the Department of Eletronic Engineering, Tsinghua University, Beijing 100084, China (e-mail: s-yuan16@mails.tsinghua.edu.cn).

Haohuan Fu is with the Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing 100084, China, also with the Department of Earth System Science, Tsinghua University, Beijing 100084, China, and also with the National Supercomputing Center in Wuxi, Wuxi 214072, China (e-mail: haohuan@tsinghua.edu.cn).

Weijia Li is with the CUHK-SenseTime Joint Laboratory, The Chinese University of Hong Kong, Hong Kong 999077, China (e-mail: weijiali@cuhk.edu.hk).

Digital Object Identifier 10.1109/LGRS.2021.3061726

I. Introduction

OIL PALM is a crucial economic crop in developing tropic countries, such as Indonesia and Malaysia, which hold over 80% palm oil production in the world [1]. However, oil palm is an extremely controversial topic. On one hand, palm oil has a high economic value, boosting gross domestic product and local employment. On the other hand, lavishly expanding oil palm plantation results in deforestation, environmental damages, species extinction, and so on. To this end, providing an accurate and large-scale oil palm plantation information is prominent for both economic development and ecological significance. Nowadays, with the rapid development of remote sensing techniques, it is potential to automatically and accurately detect oil palm trees using high-resolution remotely sensed images instead of time-consuming and routine manpower field surveys [2]. Nevertheless, large-scale oil palm tree detection using multitemporal or multisensor remote sensing images is still challenging under a variety of geological and temporal features across regions and images. Although various deep learning-based methods succeed in forming an automated approach for tree crown detection in recent years and exhibit encouraging performance [3]–[6], a brand new environment may exacerbate the detection results. Moreover, the labeling efforts needed for covering different features in heterogeneous regions hamper its efficiency from large-scale remote sensing applications using multitemporal and multisensor satellite images.

There are some solutions to the problem of detecting oil palms in more than one scenario. First, we may add some annotations in a new scenario (target domain) to prevent accuracy degrade, which is time-consuming and fairly expensive to some extent. Second, some studies adopt the model from the original scenario (source domain) to a new data by domain adaptation (DA) approaches without adding any costly annotations, keeping satisfying performance in the target domain [7]. DA is utilized to capture the representations and characteristics of the target domain for model adaptation with knowledge about target distribution during training. However, in the real

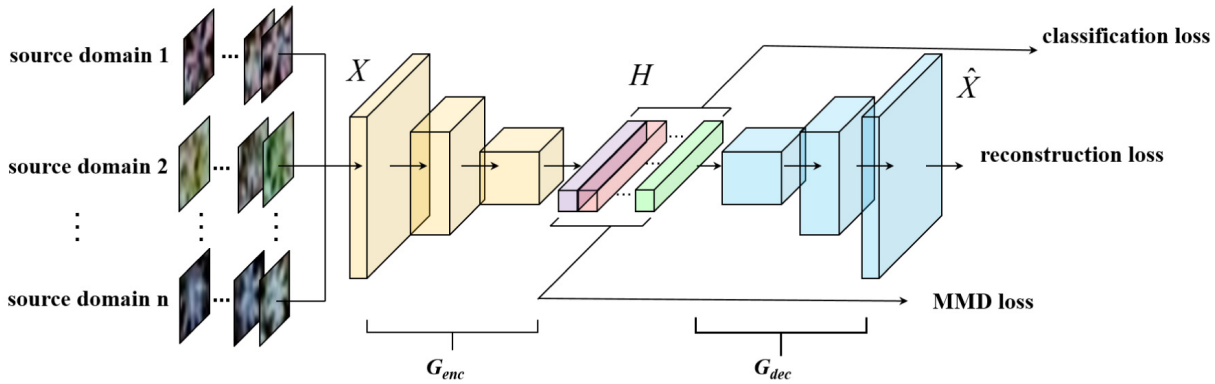

Fig. 1. Architecture of our proposed MMD-DRCN. MMD-DRCN contains four parts: 1) Feature extractor (G_{enc}) encodes multisource data ($X_1, X_2, \cdots, X_N$) as latent features, represented by ($H_1, H_2, \cdots, H_N$); 2) Classifier C learns to distinguish different class through ($H_1, H_2, \cdots, H_N$); 3) Decoder G_{dec} reconstructs the samples from multisource domains, making latent features more representative; and 4) MMD module learns invariant features among different source domains.

world, it is extremely cumbersome and infeasible to acquire new images in advance and repeatedly annotate them at once. For example, if we have to monitor the oil palm plantation in a long time series, labeling in each image or adapt model to each scenario is highly inconvenient and inefficient. Furthermore, it is difficult to collect images covering all different phases and regions during model training. Therefore, domain generalization (DG) confronts this problem by leveraging the labeled data from multiple source domains to learn a versatile and universal representation, striving to build a precise model for any "unseen" target domains [8], [9]. The common advantage for DA and DG is that we do not need to add any annotations in the target domain to prevent detection results deterioration. However, the main difference between DA and DG is that DA can utilize the information and distribution of the target domain, while the target domain is completely "unseen" to DG. In many practical applications, the target domain is "unseen," and it is intractable to train a model from images covering all geo-locations, photograph angles, and acquisition dates. DG can learn representations from one or more source domains, obtains acceptable performance in an unknown target domain. In remote sensing community, although numerous methods have been proposed for DA [10]–[13], rare attention has been paid to DG methods for remote sensing images [14]. Notably, different from the DA scenario, the target domain data set does not take part in the training process for the DG scenario.

Our contributions in this letter can be summarized as follows three aspects.

1) We propose a Maximum Mean Discrepancy Deep Reconstruction Classification Network (MMD-DRCN) to detect oil palms in a new environment from multisource high-resolution satellite images.
2) Besides classification loss, our MMD-DRCN extracts more representative features through reconstruction loss and aligns multisource latent features by MMD loss, both of which enhance the capacity of generalization.
3) Thus far, it is the first work to explore and exploit the potential of DG in remote sensing-based tree crown detection.

II. Methodology

A. Preliminary

Our goal is to start from the samples of multiple source domains and train a model that can perform well on a new target domain. In the training process, we assume to have N source domains $\mathcal{D}^S = \{\mathcal{D}_1^S, \mathcal{D}_2^S, \cdots, \mathcal{D}_N^S | N{>}1\}$. For the ith source domain, $\mathcal{D}_i^S = \{(x_i^j, y_i^j)\}$, where x^j is the jth data and y^j is corresponding label. Our target is to learn a model $f : x \rightarrow y$ to have satisfying generalization capacity on a brand new target domain $\mathcal{D}_*^T$. Note that the target domain is unavailable to the model f. In addition, the target domain covers the same categories with multisource domains ($Y_*^T = Y_1 = Y_2 = \cdots = Y_N$).

B. Network Architecture

Since the target domain is "unseen," it limits DA methods from solving this problem. As shown in Fig. 1, our proposed MMD-DRCN contains four parts.

1) Feature extractor (G_{enc}) encodes multisource data ($X_1, X_2, \cdots, X_N$) as latent features, represented by ($H_1, H_2, \cdots, H_N$).
2) Classifier C learns to distinguish different class through ($H_1, H_2, \cdots, H_N$).
3) Decoder G_{dec} reconstructs the samples from multisource domains, making latent features more representative and universal.
4) MMD module learns invariant features among different source domains.

During training, the model is supposed to minimize the reconstruction loss (L_{rec}), classification loss (L_{cls}) and MMD loss (L_{MMD}). The overall loss function can be formulated

$$L = L_{\text{cls}} + \alpha L_{\text{rec}} + \beta L_{\text{MMD}} \tag{1}$$

where α and β are tradeoff parameters for L_{rec} and L_{MMD}, respectively. We adopt cross-entropy loss as our L_{cls}

$$L_{\text{cls}} = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \sum_{k=1}^{m} y_k \log C\left(G_{\text{enc}}\left(x_i^j\right)\right) \tag{2}$$

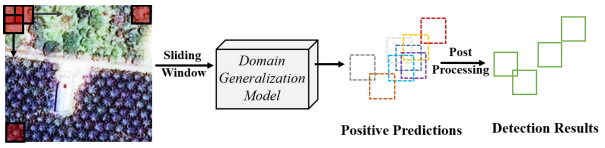

Fig. 2. Pipeline of reference phase for oil palm tree detection using MMD-DRCN.

where n_i represents the number of samples in $\mathcal{D}_i^S$ and m represents the number of class. As for L_{rec}, we utilize mean-square error (MSE)

$$L_{\text{rec}} = \sum_{i=1}^{N} \frac{1}{n_i} \sum_{j=1}^{n_i} \left\| x_i^j - G_{\text{dec}}\left(G_{\text{enc}}\left(x_i^j\right)\right)\right\|^2. \tag{3}$$

The reconstruction loss helps G_{enc} to extract more representative features, and it helps generalize our model to an unknown target domain. Inspired by Li *et al.* [9], we calculate MMD loss by minimizing the upper bound of the distribution variance among multisource domains

$$L_{\text{MMD}} = \frac{1}{N^2} \sum_{1 \le i,j \le N} MMD\left(H_i, H_j\right) \tag{4}$$

where $\text{MMD}(H_i, H_j)$ describes the MMD distance of latent features between any two source domains $(\mathcal{D}_i, \mathcal{D}_j)$. $\text{MMD}(H_i, H_j)$ can be calculated as follows:

$$\begin{aligned} \text{MMD}\left(H_i, H_j\right)^2 &= \left\| \frac{1}{n_i} \sum_{t=1}^{n_i} \phi\left(h_i^t\right) - \frac{1}{n_j} \sum_{l=1}^{n_j} \phi\left(h_i^l\right) \right\|_{\mathcal{H}}^2 \\ &= \frac{1}{n_i^2} \sum_{t=1}^{n_i} \sum_{t'=1}^{n_i} \kappa\left(h_i^t, h_i^{t'}\right) \\ &\quad + \frac{1}{n_j^2} \sum_{l=1}^{n_j} \sum_{l'=1}^{n_j} \kappa\left(h_j^l, h_j^{l'}\right) \\ &\quad - \frac{2}{n_i n_j} \sum_{t=1}^{n_i} \sum_{l=1}^{n_j} \kappa\left(h_i^t, h_j^l\right) \end{aligned} \tag{5}$$

where $\kappa(\cdot, \cdot)$ is a positive semidefinite-based kernel function. In this letter, we use the radial basis function (RBF), which is a well-established characteristic kernel

$$\kappa\left(x_i, x_j\right) = \sum_n \eta_n \exp\left\{ -\frac{1}{2\sigma_n} \left\| x_i - x_j \right\|^2 \right\} \tag{6}$$

where η_n is the weight of the nth kernel function and σ_n is the corresponding standard deviation.

C. Postprocessing

In the detection phase, we first unify these images' resolution to 0.6 m and downsample Image A, Image B, and Image C by bilinear interpolation algorithm to the same resolution of Image D. The original image is cropped to 17 × 17 pixels applying sliding window technique with 3-pixels sliding step [3]. After the prediction of the DG network, the coordinates of the positive predicted samples are merged based on the intersection-over-union (IoU) metric in order to eliminate the dense and repeated detection [13]. The final results are calculated as the average of the coordinates merged into the same group by a threshold of 0.3. Fig. 2 displays the workflow of inference steps.

TABLE I
DETAIL INFORMATION OF OUR FOUR HIGH-RESOLUTION SATELLITE IMAGES

Index	Image A	Image B	Image C	Image D
Longitude	100.78E	103.05E	102.44E	103.60E
Latitude	4.66N	5.07N	2.61N	1.60N
Resolution	0.3m	0.3m	0.3m	0.6m
Band	RGB	RGB	RGB	RGB, NIR
Source	Google Earth	Google Earth	Google Earth	QuickBird
Acquisition	Dec. 2015	Jul. 2017	Feb. 2018	Nov. 2006

III. DATA SET

Our study area is located in Peninsular Malaysia, where the oil palm plantation is expanding rapidly. We have four high-resolution remote sensing images: Image A, B, C, and D. Table I lists the detailed information of these four satellite images. They are acquired from different sensors, locations, and acquisition dates, resulting in differences in reflectance, resolution, illumination, and environmental conditions.

Fig. 3 shows our study area and the location of our four images. We collect oil palm samples in red rectangles and collect other samples (other vegetation, impervious, etc.) in blue rectangles. 80% of collected samples are training data set, while others are validation dateset. Note that we unify all images to the same resolution of 0.6 m and each sample has 17 × 17 pixels, which is consistent with our previous works [3], [13]. In each image, we select four typical regions as the test data set to evaluate the oil palm tree detection results for our proposed MMD-DRCN.

IV. EXPERIMENTS

A. Setup

We adopt Adam as optimizer with a learning rate of 0.001 and the training epoch of 150. We set tradeoff parameters α and β as 0.5 and 5. According to [9], we average RBF kernels with the bandwidth $\sigma = 1$, 5 or 10. During training, we select three images as three source domains, and the left one is the "unseen" target domain. We evaluate the capacity of the model through the performance of the target domain. To this end, we have four transfer tasks: $\{B, C, D \to A\}$, $\{A, C, D \to B\}$, $\{A, B, D \to C\}$, and $\{A, B, C \to D\}$. As for the inference phase, we adopt postprocessing procedures in II-C. We adopt prevalent evaluation protocol in our previous work [13], including true positives (TP), false positives (FP), false negatives (FN), precision, recall and F1-score. Additionally, the detecting palms whose overlapping area with ground-truth palms is higher than 0.5 will be considered as correct oil palms (TP). Our codes are available on https://github.com/rs-dl/MMD-DRCN.

B. Results

We list the detection results for four transfer tasks in Table II. Our MMD-DRCN achieves the average F1-score of 91.18%, 75.49%, 79.74%, and 84.37% when the target domain is Image A, B, C, and D, respectively. Table III

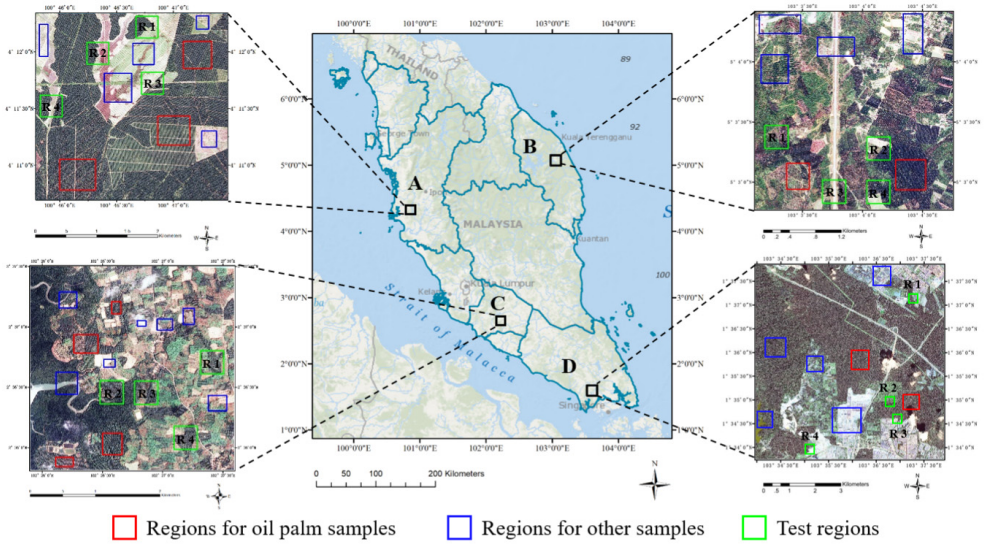

Fig. 3. Our study area is located in Peninsular Malaysia. We have four high-resolution remote sensing images: Image A, Image B, Image C, and Image D. They are acquired from different sensors, locations, and acquisition dates, resulting in differences in reflectance, resolution, and illumination.

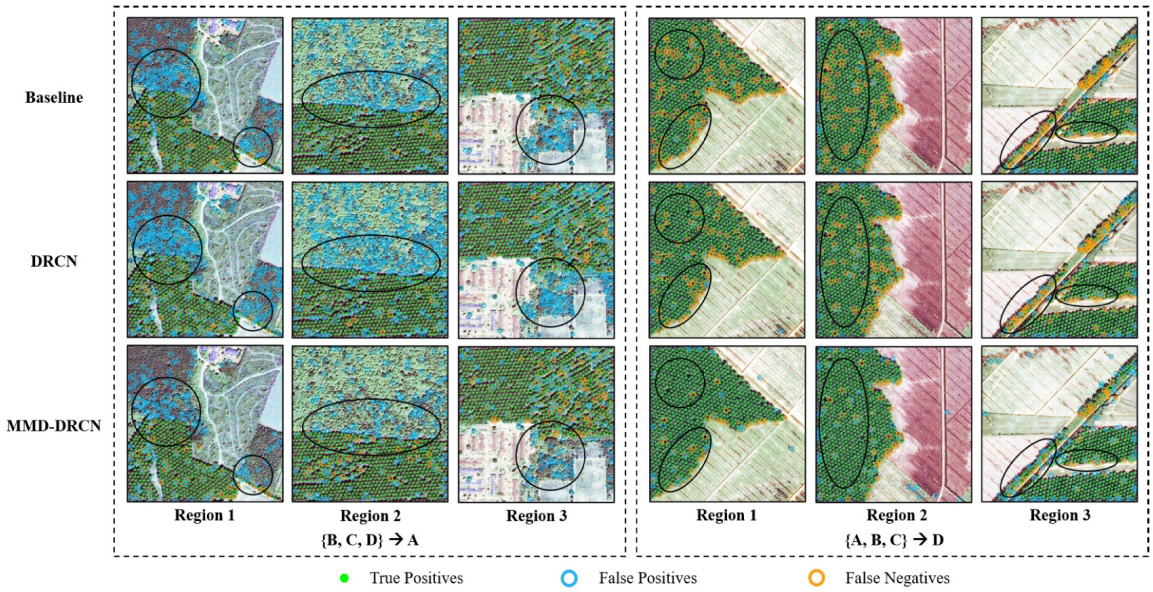

Fig. 4. Detection results for $\{B, C, D \rightarrow A\}$ (left) and $\{A, B, C \rightarrow D\}$ (right). The green points stand for the correct detected oil palms, the orange circles stand for the ground-truth oil palms that are missing, and the blue circles stand for other types of objects. The black circles denote some examples where our MMD-DRCN outperforms other methods.

also shows the precision, recall and F1-score for other methods, including Baseline, domain adversarial neural network (DANN) [15], DRCN, and conditional DANN (CDANN) [16]. Baseline means the model only learns the representations from annotations in multisource domains. DRCN means we train the model through L_{cls} and L_{rec}. DANN [15] and CDANN [16] are two state-of-the-art DG methods. It is clear to observe that our proposed MMD-DRCN achieves the highest F1-score for all circumstances with an average of 82.70%, outperforming other state-of-the-art DG methods by a margin of 2.84%–5.83%.

Fig. 4 displays the detection results for $\{B, C, D \rightarrow A\}$ (left) and $\{A, B, C \rightarrow D\}$ (right). The green points stand for the correct detected oil palms, the orange circles stand for the ground-truth oil palms that are missing, and the blue circles stand for other types of objects. The black circles denote some examples where our MMD-DRCN outperforms other methods. We can find that MMD-DRCN has less FP (blue circles) than Baseline and DRCN. However, DRCN only has slight improvement for FP, even some negative transfer effect happening.

C. Sensitive Analysis

As MMD-DRCN contains three loss functions, we explore different tradeoff parameters in this part. We set $\alpha \in [0.1, 10]$

TABLE II

DETECTION RESULTS FOR FOUR TRANSFER TASKS. "R A1" DENOTES THE REGION 1 IN IMAGE A

Index	$\{B, C, D \rightarrow A\}$				$\{A, C, D \rightarrow B\}$				$\{A, B, D \rightarrow C\}$				$\{A, B, C \rightarrow D\}$			
	R A1	R A2	R A3	R A4	R B1	R B2	R B3	R B4	R C1	R C2	R C3	R C4	R D1	R D2	R D3	R D4
TP	976	767	790	462	457	572	1,013	640	701	711	920	624	795	624	742	419
FP	88	44	26	94	286	619	250	184	462	322	145	210	104	182	157	161
FN	129	52	44	71	81	89	129	49	54	114	55	119	101	47	113	51
Precision	91.73%	94.57%	96.81%	83.09%	61.51%	48.03%	80.21%	77.67%	60.28%	68.83%	86.39%	74.82%	88.43%	77.42%	82.54%	72.24%
Recall	88.33%	93.65%	94.72%	86.68%	84.94%	86.54%	88.70%	92.89%	92.85%	86.18%	94.36%	83.98%	88.73%	93.00%	86.78%	89.15%
F1-score	90.00%	94.11%	95.76%	84.85%	71.35%	61.77%	84.24%	84.60%	73.10%	76.53%	90.20%	79.14%	88.58%	84.50%	84.61%	79.81%
Average F1-score	91.18%					75.49%				79.74%				84.37%		

TABLE III

PRECISION, RECALL AND F1-SCORE OF DIFFERENT METHODS FOR FOUR TRANSFER TASKS (%)

Method	$\{B, C, D \rightarrow A\}$			$\{A, C, D \rightarrow B\}$			$\{A, B, D \rightarrow C\}$			$\{A, B, C \rightarrow D\}$			**Average**		
	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score	precision	recall	F1-score
Baseline	**95.42**	79.37	86.62	62.68	81.49	70.04	57.86	**90.97**	70.37	73.29	90.06	80.45	72.31	85.47	76.87
DANN	94.27	82.96	88.24	59.42	**89.69**	71.38	71.39	78.40	74.07	**85.60**	77.95	80.99	74.84	85.08	78.67
DRCN	94.70	82.93	88.41	63.38	88.61	73.04	73.02	85.23	78.42	67.42	**90.55**	76.69	74.62	86.83	79.14
CDANN	94.54	84.61	89.28	65.35	83.54	73.05	**74.21**	77.19	74.96	84.46	80.91	82.15	77.16	84.04	79.86
MMD-DRCN	91.55	**90.85**	**91.18**	**66.83**	88.27	**75.49**	72.58	89.34	**79.74**	80.16	89.41	**84.37**	**77.78**	**89.47**	**82.70**

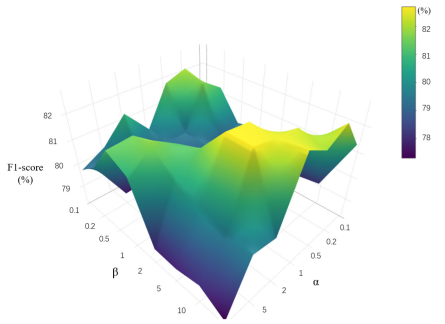

Fig. 5. F1-score under different tradeoff parameters, with $\alpha \in [0.1, 10]$ and $\beta \in [0.1, 50]$. We set α and β as 0.5 and 5 in our MMD-DRCN to obtain the optimal pairs of tradeoff parameters.

and $\beta \in [0.1, 50]$. Fig. 5 illustrates the F1-score under different α and β. To this end, we choose α and β as 0.5 and 5 in our MMD-DRCN to obtain the optimal pairs of tradeoff parameters.

V. CONCLUSION

In this letter, we propose a new multisource DG scheme, MMD-DRCN. It learns representations from multiple source domains and obtains an acceptable performance in an unknown and "unseen" target domain. Besides classification loss, our MMD-DRCN extracts more representative features through reconstruction loss and aligns multisource latent features by MMD loss, both of which enhance the capacity of generalization. Our target domain data set does not take part in the training process. MMD-DRCN achieves an average F1-score of 82.70% in all transfer tasks, attaining a 5.83% gain compared to Baseline (a straightforward convolutional neural network (CNN) model). Experimental results demonstrate DG poses a promising potential for large-scale and cross-regional oil palm tree detection without any information of the target domain. In the future, we will explore more advanced DG methods for more remote sensing applications.

REFERENCES

[1] J. C. Quezada, A. Etter, J. Ghazoul, A. Buttler, and T. Guillaume, "Carbon neutral expansion of oil palm plantations in the neotropics," *Sci. Adv.*, vol. 5, no. 11, Nov. 2019, Art. no. eaaw4418.

[2] D. Marinelli, C. Paris, and L. Bruzzone, "An approach to tree detection based on the fusion of multitemporal LiDAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1771–1775, Nov. 2019.

[3] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sens.*, vol. 9, no. 1, p. 22, Dec. 2016.

[4] I. E. D. Souza and A. X. Falcao, "Learning CNN filters from user-drawn image markers for coconut-tree image classification," *IEEE Geosci. Remote Sens. Lett.*, early access, Sep. 11, 2020, doi: 10.1109/LGRS.2020.3020098.

[5] J. Zheng *et al.*, "Growing status observation for oil palm trees using unmanned aerial vehicle (UAV) images," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 95–121, Mar. 2021.

[6] K. Johansen *et al.*, "Mapping the condition of macadamia tree crops using multi-spectral UAV and WorldView-3 imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 165, pp. 28–40, Jul. 2020.

[7] W. M. Kouw and M. Loog, "A review of domain adaptation without target labels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 766–785, Mar. 2021.

[8] M. Ghifary, W. B. Kleijn, M. Zhang, and D. Balduzzi, "Domain generalization for object recognition with multi-task autoencoders," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2551–2559.

[9] H. Li, S. J. Pan, S. Wang, and A. C. Kot, "Domain generalization with adversarial feature learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5400–5409.

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

[11] W. Wu, J. Zheng, W. Li, H. Fu, S. Yuan, and L. Yu, "Domain adversarial neural network-based oil palm detection using high-resolution satellite images," *Proc. SPIE*, vol. 11394, Apr. 2020, Art. no. 1139406.

[12] W. Wu, J. Zheng, H. Fu, W. Li, and L. Yu, "Cross-regional oil palm tree detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 56–57.

[13] J. Zheng *et al.*, "Cross-regional oil palm tree counting and detection via a multi-level attention domain adaptation network," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 154–177, Sep. 2020.

[14] C. Persello and L. Bruzzone, "Relevant and invariant feature selection of hyperspectral images for domain generalization," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 3562–3565.

[15] Y. Ganin *et al.*, "Domain-adversarial training of neural networks," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, May 2015.

[16] Y. Li *et al.*, "Deep domain generalization via conditional invariant adversarial networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 624–639.