

# Partial Domain Adaptation for Scene Classification From Remote Sensing Imagery

Juepeng Zheng<sup>ID</sup>, *Graduate Student Member, IEEE*, Yi Zhao<sup>ID</sup>, Wenzhao Wu, Mengxuan Chen,  
Weijia Li, and Haohuan Fu<sup>ID</sup>, *Member, IEEE*

**Abstract**—Although domain adaptation approaches have been proposed to tackle cross-regional, multitemporal, and multisensor remote sensing applications since they do not require any human interpretation in the target domain, most current works assume identical label space across the source and the target domains. However, in real-world applications, we often transfer knowledge from a large-scale dataset with rich annotations to a small-scale target dataset with scarcity of labels. In most cases, the label space of the source domain is usually large enough to subsume that of the target domain, which is termed partial domain adaptation. In this article, we propose a new partial domain adaptation algorithm for remote sensing scene classification and our proposed method contains three major parts. First, we employ a progressive auxiliary domain module to alleviate the negative transfer effect caused by outlier classes. Second, we adopt an improved domain adversarial neural network (DANN) with multiweights to better encourage domain confusion. Last but not least, we design an attentive complement entropy regularization to improve the prediction confidence for samples and avoid untransferable samples (such as the samples belonging to outlier classes in the source domain) being mistakenly classified. We collect three common remote sensing datasets to evaluate our proposed method. Our method achieves an average accuracy of 79.36%, which considerably outperforms other state-of-the-art partial domain adaptation methods with an average accuracy improvement of 1.90%–12.45% and attaining a 13.67% gain compared to the straightforward deep learning model (ResNet-50). The experiment results indicate that our approach shows promising prospects for solving more general

Manuscript received 30 August 2022; revised 31 October 2022; accepted 10 December 2022. Date of publication 14 December 2022; date of current version 20 January 2023. This work was supported in part by the National Key Research and Development Plan of China under Grant 2020YFB0204800; in part by the National Natural Science Foundation of China under Grant T2125006, Grant U1839206, and Grant 42201358; and in part by the Jiangsu Innovation Capacity Building Program under Project BM2022028.

(Corresponding authors: Weijia Li; Haohuan Fu.)

Juepeng Zheng, Yi Zhao, and Mengxuan Chen are with the Ministry of Education Key Laboratory for Earth System Modeling, the Department of Earth System Science, and the Department of Earth System Science, Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Tsinghua University, Beijing 100084, China (e-mail: zjp19@mails.tsinghua.edu.cn; y-zhao19@mails.tsinghua.edu.cn; chenmx21@mails.tsinghua.edu.cn).

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

Weijia Li is with the School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 519082, China (e-mail: liweij29@mail.sysu.edu.cn).

Haohuan Fu is with the Ministry of Education Key Laboratory for Earth System Modeling, the Department of Earth System Science, and the Department of Earth System Science, Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Tsinghua University, Beijing 100084, China, and also with the National Supercomputing Center, Wuxi 214072, China (e-mail: haohuan@tsinghua.edu.cn).

Digital Object Identifier 10.1109/TGRS.2022.3229039

and practical domain adaptation problems where the label space of the source domain subsumes that of the target domain.

**Index Terms**—Adversarial learning, deep learning, negative transfer effect, partial domain adaptation, remote sensing, scene classification.

## I. INTRODUCTION

ALTHOUGH deep learning algorithms have already been proved to be highly successful over a wide variety of tasks in the remote sensing field [2], [3], [4], [5], [6], it requires manual annotations in the training phase for a particular distribution (source domain) to learn effective representations. However, when directly applying models obtained from source domains to target domains, models will lose their accuracy due to different data distributions and variances caused by different sensors, surface environment, and so on. For example, as shown in Fig. 1, remote sensing images with the same annotations but derived from different datasets (i.e., AID [7] and NWPU-RESISC45 [8]) show significant differences in spectral distributions. We can easily observe the notable discrepancy in colors, textures, and other characteristics between these two different remote sensing datasets. Furthermore, we also display the spatial texture information differences between these two datasets with respect to the same specific category. The light blue, orange, and light green histograms, respectively, represent the mean histogram of max, contrast, and homogeneity values through gray-level co-occurrence matrix (GLCM) [1] in one specific category. Nevertheless, one common solution for this dilemma is to fine-tune pretrained networks on task-specific datasets, and it may be impractical because of prohibitively expensive label collection.

Fortunately, the domain adaptation approach provides us a way to deal with the abovementioned problem by seeking to minimize the discrepancy in the absence of accessing the target label information [9], [10], [11], [12]. Existing domain adaptation approaches generally focus on the standard domain adaptation, where the source label space is equal to the target label space. However, with the development of big data, large-scale datasets with abundant labels become available, and in most cases, the label space of the source domain is usually large enough to subsume that of the target domain, which is termed partial domain adaptation. For example, we often need to transfer knowledge of deep neural networks from a large-scale dataset with rich annotations (e.g., ImageNet [13]) to a small-scale target dataset with

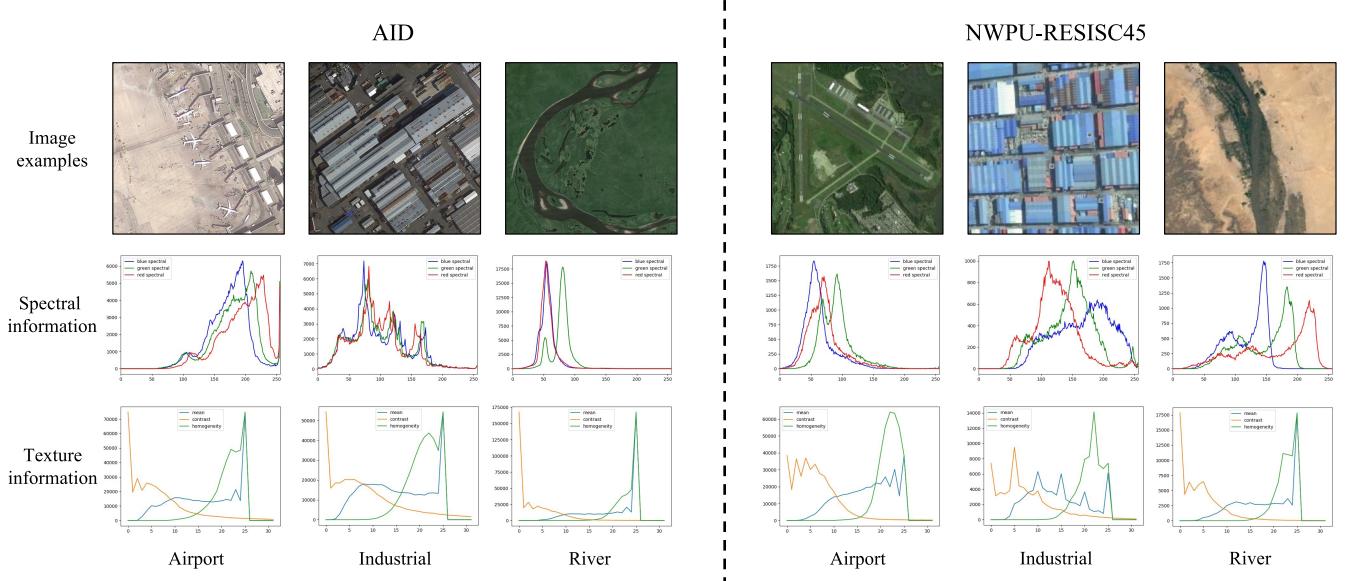


Fig. 1. Spectral difference and texture difference between two different remote sensing image datasets [i.e., AID (Left) and NWPU-RESISC45 (Right)] in the same category, for instance, airport, industrial, and river. The first line displays image examples for these categories. The middle line displays spectral information. The red, green, and blue histograms, respectively, represent the mean histogram of red, green, and blue bands in one specific category. The  $x$ -axis denotes the pixel values ranging from 0 to 255 and the  $y$ -axis represents the statistics of times for each pixel value. The last line displays texture information. The light blue, orange, and light green histograms, respectively, represent the mean histogram of max, contrast, and homogeneity values through GLCM [1] in one specific category.

scarcity of annotations (e.g. AID [7] or NWPU-RESISC45 [8] for remote sensing community). As shown in Fig. 2, partial domain adaptation (bottom) is more general and challenging than standard domain adaptation (top) since the outlier source classes (e.g., “Island” and “Snowberg”) will produce unpleasant impact when discriminating the classes of the target domain. Therefore, there are two major difficulties in partial domain adaptation scenarios. The one is that the outlier source classes will make the well-known negative transfer bottleneck more prominent. The other is that it is nontrivial to identify which classes are outlier source classes since the target classes are unknown during training.

In practical applications, partial domain adaptation methods have a great potential for remote sensing applications than previously standard domain adaptation methods. For example, the label space of the existing remote sensing dataset is not usually identical to our particular tasks. Until now, we have 30-m land cover annotations for 11 land cover types in 2013 [14]. When we directly adopt domain adaptation approaches to transfer the knowledge from the previous samples to map a particular country, city, or region (such as New York) in different years, the samples of some classes (such as snow/ice and tundra) do not belong to the target domain, which may inevitably cause negative transfer effect. Some researchers may manually remove the samples of unnecessary classes before applying domain adaptation [15], but it is labor-consuming and depends on a piece of prior knowledge. In many cases, we even do not know which part of the source label space is shared with the target label space since their annotations are unavailable during the training progress. For example, if we want to make a forest inventory (tree species classification) where few people have been there by using remote sensing

images, it is obviously impossible for us to know the exact tree species in advance. The best technique is that we utilize the rich tree species data bank collected from other regions and adopt partial domain adaptation algorithms to map the tree species in the depopulated forest, which not only saves human interpretation and extra annotations but also eliminates the negative transfer effect caused by standard domain adaptation approaches.

In this article, we emphasize the partial domain adaptation problem, that is, the label space of the source domain is a subspace of that of the target domain. According to the necessary of partial domain adaptation scenarios in the remote sensing domain and the characteristics of remote sensing images, we make corresponding optimizations from them. To sum up, our contributions in this context can be highlighted as the following three aspects.

- 1) We propose a partial domain adaptation algorithm for remote sensing scene classification to address the partial domain adaptation scenarios. To the best of our knowledge, this work is among the first attempts on the partial domain adaptation issue in the remote sensing community.
- 2) We design a progressive auxiliary domain module (PADM) simply borrowed from the source domain dataset, alleviating the imbalanced class problem in partial domain adaptation scenarios. We also improve the domain adversarial neural network (DANN) by two kinds of weighting schemes and design the attentive complement entropy regularization (ACER).
- 3) We conduct extensive experiments on three public remote sensing datasets (i.e., AID, NWPU-RESISC45, and UC Merced). Our method attains a 13.67%

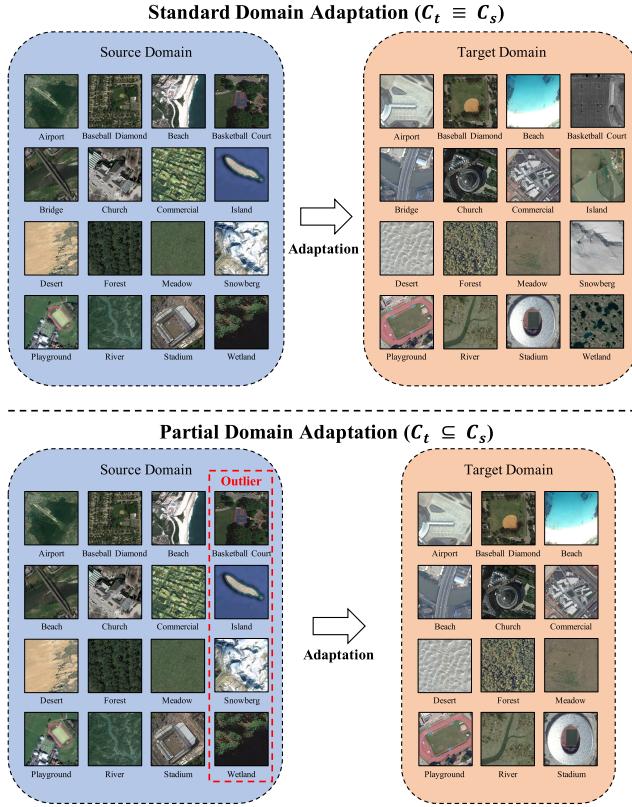


Fig. 2. Standard domain adaptation scenario (Top) and the partial domain adaptation scenario (Bottom). In the standard domain adaptation scenario, the source label space is equal to the target label space, while in the partial domain adaptation scenario, the target label space is a subset of the source label space. The partial domain adaptation scenario is difficult since the source domain may have some outlier classes not appearing in the target domain, e.g., “Island” and “Snowberg.” These outlier source classes will make the well-known negative transfer bottleneck more prominent. Another technical difficulty is that it is nontrivial to identify which classes are outlier source classes since the target classes are unknown during training.

gain compared to the straightforward CNN model (ResNet-50) and yields 1.90%–12.45% improvements compared to other state-of-the-art partial domain adaptation methods.

In the remainder of this context, we first present the related works in Section II. Then, we elaborate our proposed method in Section III, and introduce our collected dataset in Section IV. In Section V, we analyze the performance of our proposed method and compare it with other state-of-the-art partial domain adaptation approaches and other standard domain adaptation methods, followed by disclosing the negative transfer effect and further discussions in Section VI. Finally, we summarize our work in Section VII.

## II. RELATED WORK

### A. Domain Adaptation

Domain adaptation approaches emphasize aligning the model to new data distributions without utilizing a large number of labor-consuming annotations and recently has been paid much attention in the machine learning domain [9], [16], [17]. A rich line of domain adaptation methods can help to diminish the discrepancy between the source domain

and the target domain through two major approaches, either based on moment matching or adversarial learning. Moment matching-based domain adaptation approaches aim at minimizing the distribution discrepancy in feature space [18], [19], [20], [21], [22], [23]. Adversarial learning-based domain adaptation approaches focus on incorporating a classifier in the source domain with gradient reversal to confuse a domain discriminator. The discriminator (i.e., a binary classifier) can identify whether the input image comes from the source or the target domain. After applying adversarial learning, it cannot recognize different domains well [24], [25], [26], [27], [28], [29]. At present, the adversarial learning-based domain adaptation approaches have become a mainstream field of DA issues.

However, the aforementioned domain adaptation approaches generally focus on the standard domain adaptation, where the source domain and the target domain share the identical label space. In real applications, especially in the big data era, the label space of the source domain is usually larger than that of the target domain in most cases. The standard domain adaptation no longer meets the requirement of mostly cross-domain tasks and more attention needs to be paid to the partial domain adaptation scenario.

### B. Partial Domain Adaptation

Different from standard domain adaptation, in partial domain adaptation scenarios, the label space of the target domain is a subset of that of the source domain. Several methods have been proposed to solve the partial domain adaptation problem in the computer vision community. Existing partial domain adaptation methods concentrate on reducing the training weights for the classes or samples that are not considered in the label space of the target domain. For example, selective adversarial network (SAN) [30] employs a multidiscriminator weighted by the percentages of predicted results to the target domain. Importance weighted adversarial net (IWAN) [31] only utilizes a domain discriminator weighted by the probability of being a target sample. Partial domain adaptation network (PADA) [32] adds the average predictions of the target domain simultaneously to the source classifier and the domain discriminator. Different from PADA, example transfer network (ETN) [33] and balanced adversarial alignment and adaptive uncertainty suppression (BA<sup>3</sup>US) [34] adopted an entropy minimization principle to reweight the source classifier and the domain discriminator. Deep residual correction network (DRCN) [35] plugs one residual block into the source network along with the task-specific feature layer, which effectively enhances the adaptation from source to target and explicitly weakens the influence from the irrelevant source classes. Other works employ reinforcement learning [36], [37], generative adversarial network (GAN) [38], or graph convolutional network (GCN) [39] to deal with partial domain adaptation scenarios.

To date, most of the existing efforts tackle partial domain adaptation by reducing the weights for the outlier classes in the source domain. They are admittedly quite effective but constitute only one side of the coin. These methods heavily

rely on high-accuracy “pseudo” target prediction results to acquire suitable weights, while rare attention has been paid to adding the outlier classes during the adversarial learning module, which pursues the balance between label distributions across the source and the target domains.

### C. Domain Adaptation in Remote Sensing

Although researchers are committed to finding solutions for remote sensing image classification, two major difficulties prevent their efforts from achieving a wide range of applications [40]. On one hand, labeled data are not always adequate for each scenario. On the other hand, data acquired with different sensors under different environments put high demand for the generalization ability of the models. Therefore, domain adaptation has been introduced into the remote sensing field to deal with large-scale and long-time-series applications using multisource and multitemporal remote sensing images, in which differences in ground environment and photographed instrument may readily impact the model’s transferable capacity [41]. Nowadays, domain adaptation effectively minimizes the distribution gaps between images due to different sensors and conditions and emerges in all kinds of remote sensing applications ranging from classification [42], [43], [44], [45], [46], [47], [48], [49], semantic segmentation [50], [51], [52], [53], [54], [55], object detection [56], [57], [58], [59], and regression tasks [60], [61]. Domain adaptation algorithms have been proven to be effective and vital in cross-regional, large-scale, and multitemporal remote sensing applications.

In the remote sensing community, the off-the-shelf domain adaptation methods mainly concentrate on the standard domain adaptation issue. Although works have strived into some advanced transfer learning scenarios in the remote sensing field, such as multisource domain adaptation [62], [63], multitarget domain adaptation [64], [65], open-set domain adaptation [66], [67], and domain generalization [68]. As for partial domain adaptation, only coordinate partial adversarial domain adaptation (CPADA) [69] has explored the potential in the satellite images classification. However, the reweighting strategy of CPADA is quite similar to PADA [32], and CPADA only evaluates one transfer task (i.e., from NWPU-RESISC45 [8] to UC Merced [70]) for different remote sensing datasets. In this context, we propose a new partial domain adaptation algorithm for remote sensing scene classification to eliminate the negative transfer effect brought by outlier classes in the source domain, which is a more intractable but practical task in real-world applications. To the best of our knowledge, this work is one of the first attempts adopting the partial domain adaptation in the remote sensing community.

## III. METHODOLOGY

### A. Preliminary and Overview

In the standard domain adaptation scenario, the source domain dataset ( $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$ ) has annotations and we can access the label-free target domain dataset ( $\mathcal{D}_t = \{(\mathbf{x}_i^t)\}_{i=1}^{n_t}$ ), in which  $n_s$  and  $n_t$  are the number of images in the source domain dataset and the target domain dataset, respectively. Here,  $\mathbf{x}_i^s$  is an sample in  $\mathcal{D}_s$  and  $y_i^s$  denotes

the corresponding annotation, while  $\mathbf{x}_i^t$  is an sample in  $\mathcal{D}_t$  without a label. In real-world applications, the distribution of the source domain ( $\mathcal{P}_s(\mathbf{x}^s, y^s)$ ) and the target domain ( $\mathcal{P}_t(\mathbf{x}^t)$ ) is usually different so that directly adopting the model trained from  $\mathcal{D}_s$  to test  $\mathcal{D}_t$  may cause severe deterioration. Furthermore, in partial domain adaptation, the label space of the target domain is a subspace of the label space of the source domain (i.e.,  $\mathcal{C}_t \subseteq \mathcal{C}_s$ ), which is common to be seen in practice, as we usually need to transfer a model from a larger scale dataset to a smaller scale dataset. However, there are two major challenges in partial domain adaptation scenarios: 1) the source domain may have some outlier classes that are not appearing in the target domain causing the nonnegligible negative transfer effect and 2) there is no way for us to identify the outlier source classes since the target classes are unknown during the training progress.

Therefore, we propose a new partial domain adaptation algorithm for remote sensing scene classification to resolve the domain gaps and the category misalignments brought by the outlier classes in the source domain. Fig. 3 shows the framework of our proposed method, including three major parts, i.e., a PADM, an improved DANN with multiweights, and an ACER. We summarized the three main parts of our proposed method as follows.

- 1) *A PADM:* We employ a progressive strategy for the auxiliary domain directly using original source samples to balance the difference between label distributions across domains. As a result, the negative transfer effect caused by outlier classes from the source domain can be greatly reduced.
- 2) *Improved DANN With Multiweights:* We improve the DANN by adopting the normalized estimation of class-level weights of the target domain to the classification loss of source domain and set different weights for hard and easy samples during the domain adversarial alignment according to the results of domain discriminator.
- 3) *ACER:* We design an ACER to improve the prediction confidence for samples that are easy being confused with other types and avoid those untransferable samples (such as the samples belonging to outlier classes in the source domain) to be mistakenly classified because of forceful entropy minimization.

### B. Progressive Auxiliary Domain Module

Adversarial learning methods have been widely and successfully utilized in previous DA studies [24], [25], [26], [57]. They align feature distributions by incorporating a classifier in the source domain with gradient reversal to confuse a domain discriminator. For example, the objective of the well-established DANN [24] can be formulated as

$$C_{\text{DANN}}(\theta_f, \theta_y, \theta_d) = \frac{1}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} L_y(G_y(G_f(\mathbf{x}_i)), y_i) - \frac{\lambda}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_s \cup \mathcal{D}_t} L_d(G_d(G_f(\mathbf{x}_i)), d_i) \quad (1)$$

where  $n = n_s + n_t$  and  $\lambda$  denotes a hyperparameter that balances the domain loss ( $L_d$ ) and the classification loss ( $L_y$ ),

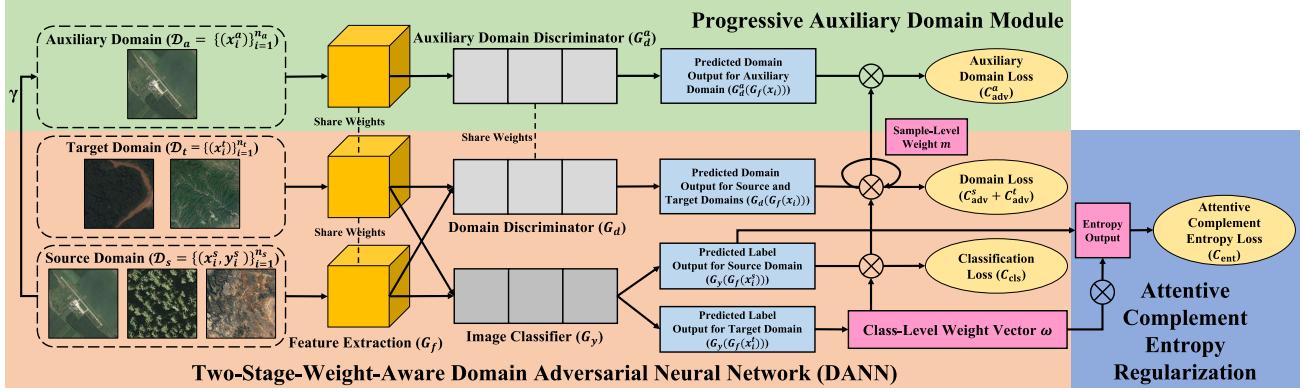


Fig. 3. Flowchart of our proposed method, including three major parts, i.e., a PADM, an improved DANN with multiweights, and an ACER. The PADM directly uses original source samples to balance the difference between label distributions across domains at the early stage of training phase. The improved DANN with multiweights adopts the normalized estimated class-level weights of the target domain to the classification loss of source domain and sets different weights for “hard” and “easy” samples during the domain adversarial alignment according to the results of domain discriminator. The ACER improves the prediction confidence for samples that are easy to be confused with other types and avoids those untransferrable samples (such as the samples belonging to outlier classes in the source domain) to be mistakenly classified because of forceful entropy minimization.

which is calculated by the domain discriminator ( $G_d$ ) and the classifier of source domain ( $G_y$ ), respectively. For more explanation, the output of  $G_d(G_f(\mathbf{x}_i))$  is the probability of the feature map in sample  $i$  belonging to the source domain. When  $G_d(G_f(\mathbf{x}_i))$  is larger than 0.5, it denotes that the sample  $i$  belongs to the target domain, and when  $G_d(G_f(\mathbf{x}_i))$  is smaller than 0.5, it represents that the sample  $i$  belongs to the source domain. To maximally confuse the domain discriminator ( $G_d$ ),  $d_i$  is equal to 1 for the source domain samples and 0 for the target domain samples.

In the partial domain adaptation scenario, the label distributions are not symmetric between the source and the target domains, thus matching the target classes may be hampered by the outlier classes from the source domain, leading to an unavoidable negative transfer problem. In general, a class-level reweighting method for the source domain seems a significant scheme for the partial domain adaptation because we hope to decrease the weights of the outlier classes in the source domain and increase that for the shared classes. This method can effectively reduce the negative transfer effect and promote the positive transfer effect across the source and the target domains. Previous methods attempt to adopt the predictions of the target domain from the image classifier to generate the class-level weight mechanism to suppress the negative effect of outlier classes [30], [32], which are considered as a quite effective way. However, this method heavily depends on high-accuracy “pseudo” target prediction results to acquire suitable weights. Rare attention has been paid to augmenting the samples of outlier classes during the adversarial learning module, pursuing the balance between the label distributions across the source and the target domains. To this end, we propose the PADM devoted to create a new auxiliary domain, borrowing fewer and fewer samples from the source domain to the auxiliary domain within an iterative adversarial learning framework. We expect that the augmented auxiliary domain and the target domain look much more similar to the source domain with respect to the label distribution,

and the challenging partial problem can be transformed to a well-studied standard domain adaptation task. Therefore, the PADM deals with partial domain adaptation by augmenting the auxiliary domain and transforming it into a standard domain adaptation-like problem.

In our PADM, we resort to augmenting an auxiliary domain by using the original samples from the source domain rather than adopting the previous reweighting mechanisms. Obviously, our purpose is to balance the different label distributions between the source and the target domains by increasing the target domain from original source samples rather than using a weighting mechanism in the source domain. This is quite reasonable because we will approximate the partial domain adaptation issue to a large standard domain adaptation-like issue so that the negative transfer effects induced by outlier classes will be effectively alleviated, which means that the performance of “pseudo” target prediction is not very well. Especially in remote sensing scene classification, some classes may have similar texture and spectral characteristics, such as bare land and dessert, grassland and farmland, and overpass and intersection. To this end, if we use previous popular partial domain adaptation methods that heavily depend on high-accuracy “pseudo” target prediction results to acquire suitable weights, the adaptation model may cause deterioration to some extent. Equation (2) is our new objective of adversarial learning, adding an item that represents an auxiliary domain using the original source domain samples. It should be emphasized that we set a progressive hyperparameter  $\gamma$  in (2). In our PADM, as the training iteration increases,  $\gamma$  gradually decreases to 0 as the number of training steps increases. The reason is that the learned representative features are not quite transferable in early iterations so that we borrow more source samples into the target domain for avoiding class mismatching. Along with the model training, the learned representative features become more and more discriminative and transferable. At this time, the estimation of class-level weights becomes more and more reliable and accurate so

that the original augmentation of source samples in the target domain is no more important

$$C_{\text{adv}}(\theta_f, \theta_d) = \frac{1}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_s \cup \mathcal{D}_t} L_d(G_d(G_f(\mathbf{x}_i)), d_i) + \frac{\gamma}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} L_d(G_d(G_f(\mathbf{x}_i)), 0). \quad (2)$$

To this end, the benefits of our auxiliary domain can be concluded from two aspects: 1) we can effectively avoid class mismatching problem since we extend the label space in the auxiliary domain through using original source domain samples and 2) by augmenting a large number of source domain samples, we can effectively reduce the new negative transfer effect thanks to the less transferable representative features in early training iterations because of the progressive procedure, thus preventing further ruining the original distribution of the target domain.

### C. Improved DANN With Multiweights

In this section, we introduce two kinds of reweight mechanisms to improve the well-established DANN: an entropy-aware weight and a target-class-aware weight, which is a sample-level weight scheme and a class-level weight scheme, respectively. In original DANN [see (1)], it is unreasonable that each sample from both source and target domains performs identically in the adversarial domain loss  $L_d$ . Those hard-to-transfer samples with uncertain predictions may badly deteriorate the adversarial learning procedure. Because of this, we hope that the model can decrease weights for those hard-to-transfer samples and increase weights for those easy-to-transfer samples during the domain adversarial alignment. Inspired by conditional domain adversarial network (CDAN) [25], we quantify the prediction uncertainty by the entropy criterion  $H(h) = -\sum_{c=1}^C h_c \log(h_c)$ , where  $C$  denotes the number of classes and  $h_c$  represents the probability of predicting a sample to class  $c$ . Therefore, we reweight the domain discriminator by each training sample with an entropy-aware weight  $m(\mathbf{x}_i) = 1+e^{-H(G_y(G_f(\mathbf{x}_i)))}$ . Notably, the entropy-aware weight is a sample-level weight and the adversarial objective can be formulated as follows:

$$C_{\text{adv}}(\theta_f, \theta_d) = \frac{1}{n} \sum_{\mathbf{x}_i \in \mathcal{D}_s \cup \mathcal{D}_t} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), d_i) + \frac{\gamma}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 0). \quad (3)$$

On the other hand, besides adopting the PADM mentioned in Section III-B, we employ a simple reweighting mechanism that decreases the contribution of the source samples belonging to the outlier classes. We average the label predictions  $\hat{y}$  for all samples in the target domain to effectively eliminate the influence of source classifier mistakes derived from some target domain samples. To this end, our weight vector can be calculated as  $\omega = \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{y}_i$ , where  $\omega$  is a class-level weight vector with  $|\mathcal{C}_s|$ -dimension. Equations (4) and (5) are the improved adversarial learning objective and classification objective, respectively. It is noted that the weight vector  $\omega$

only acts on the source domain and the auxiliary domain in the domain adversarial alignment

$$\begin{aligned} C_{\text{adv}}(\theta_f, \theta_d) &= \frac{1}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 1) \\ &\quad + \frac{1}{n_t} \sum_{\mathbf{x}_i \in \mathcal{D}_t} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 0) \\ &\quad + \frac{\gamma}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 0) \end{aligned} \quad (4)$$

$$C_{\text{cls}}(\theta_f, \theta_y) = \frac{1}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} L_y(G_y(G_f(\mathbf{x}_i)), y_i). \quad (5)$$

### D. Attentive Complement Entropy Regularization

Entropy minimization regularization is very popular to increase the prediction confidence of classifier, which is first proposed in semisupervised learning [71] and first applied in domain adaptation domain to refine the adaptive classifier [72]. Inspired by the idea of entropy function in information theory, entropy loss is proposed to reduce the uncertainty of probabilities for output classes, making the classifier more accessible to the unlabeled target data. Although transferable attention domain adaptation (TADA) [26] utilizes the entropy regularization both on the source and target domain samples, it may incur some problem for the target samples because they are unlabeled with the high uncertainty, and it does not consider the partial domain adaptation scenarios. Clearly, entropy minimization regularization increases the certainty of the classifier predictions. On the contrary, not all samples in the target domain are transferable partial domain adaptation scenarios. For instance, the input samples may probably belong to the outlier classes. A negative effect could easily appear because of the forceful minimization for the entropy of these samples. Due to these dissimilar samples that are easier to be mistakenly classified, it may cause a negative effect to directly strengthen their certainty of the classifier predictions since increasing their confidence will further confuse the classifier.

Especially in remote sensing scene classification, due to different sensors and surface environment, the spectral and texture characteristics for source and target remote sensing images are quite different. Therefore, the prediction from the classifier for the target remote sensing images may have low prediction scores for correct classes. To this end, we adopt the complement entropy [73] that expects uniform and low prediction scores for incorrect classes for labeled source samples. To accurately suppress uncertainty and enhance confidence, we further place more emphasis on the uncertain samples that own smaller confidence. Different from the original complementary training strategy in [73], we adopt the adaptive ACER as a regularized objective ( $C_{\text{ACER}}$ ) and exploit the class-level weight ( $\omega_{y_i}$ ) for each sample like that in improved DANN (I-DANN) (see Section III-C), which effectively suppresses the outlier classes in the source domain and makes our entropy regularization more efficient in partial domain adaptation

**Algorithm 1** Our Proposed Partial Domain Adaptation Algorithm for Remote Sensing Scene Classification

**Input:** Source domain dataset  $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{n_s}$  and target domain dataset  $\mathcal{D}_t = \{(\mathbf{x}_i^t)\}_{i=1}^{n_t}$ , where  $n_s$  and  $n_t$  are the numbers of images in the source and target dataset respectively.  $C$  denotes the number of classes in the source domain.  $\mathbf{x}_i^s$  is an example in  $\mathcal{D}_s$  and  $y_i^s$  is the corresponding label.  $\mathbf{x}_i^t$  is an example in  $\mathcal{D}_t$  while it does not access to the labels. Feature extractor  $G_f$ , image classifier  $G_y$  and domain discriminator  $G_d$ .  $N$  and  $N_u$  are the total training iterations and the updating interval iterations, respectively.  $\gamma$  is the progressive ratio for the auxiliary domain and  $B$  is the batch size.  $\omega$  is the class-level weight vector.

**Output:** Well-trained  $G_f^*$ ,  $G_y^*$ ,  $G_d^*$ .

- 1: Initialize the model parameters  $\theta_f$ ,  $\theta_y$  and  $\theta_d$ .
- 2: Initialize the class-level weight vector  $\omega$ ,  $\omega_i = 1/C$ .
- 3: **for**  $i = 1 : N$  **do**
- 4: Acquire  $B$  samples from  $\mathcal{D}_s$  and  $\mathcal{D}_t$ , respectively.
- 5: Random acquire  $\gamma B$  samples from  $\mathcal{D}_s$ .
- 6: Update  $G_f$ ,  $G_y$  and  $G_d$  according to Eq. (8).
- 7: **if**  $i \% N_u == 0$  **then**
- 8: Update the class-level weight vector  $\omega = \frac{1}{n_t} \sum_{i=1}^{n_t} \hat{\mathbf{y}}_i$ .
- 9: Update the progressive ratio for the auxiliary domain  $\gamma = \gamma(1 - \frac{N_u}{N})$ .
- 10: **end if**
- 11: **end for**
- 12: **return**  $G_f^* = G_f$ ,  $G_y^* = G_y$ ,  $G_d^* = G_d$ .

scenarios

$$C_{\text{ACER}}(\theta_f, \theta_y, \theta_d) = -\frac{1}{n_s \log(C-1)} \times \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} L_{\text{ACER}}(G_y(G_f(\mathbf{x}_i)), y_i) \quad (6)$$

where  $L_{\text{ACER}}$  is formulated as (7).  $\delta$  is a hyperparameter and  $g$  is the index of ground-truth class in  $y$ .  $C$  denotes the total number of classes. We conduct performance comparisons among different entropy regularization in Section V-C3

$$L_{\text{ACER}} = (1 - \hat{y}_g)^\delta \sum_{j=1, j \neq g}^C \left( \frac{\hat{y}_j}{1 - \hat{y}_j} \right) \log \left( \frac{\hat{y}_j}{1 - \hat{y}_j} \right). \quad (7)$$

#### E. Minimax Optimization Problem

Therefore, our proposed method contains a PADM, an I-DANN with multiweights, and an attentive entropy regularization. Overall, our final learning objective includes the classification objective, adversarial objective, and entropy objective, which can be formulated as follows:

$$\begin{aligned} C_{\text{Overall}}(\theta_f, \theta_y, \theta_d) &= C_{\text{cls}}(\theta_f, \theta_y) - \lambda C_{\text{adv}}(\theta_f, \theta_d) \\ &\quad + \alpha C_{\text{ACER}}(\theta_f, \theta_y, \theta_d) \\ &= \frac{1}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} L_y(G_y(G_f(\mathbf{x}_i)), y_i) \end{aligned}$$

TABLE I

DETAILED INFORMATION OF THREE PUBLIC REMOTE SENSING DATASETS

Index	AID	NWPU-RESISC45	UC Merced
Year	2017	2017	2010
Classes	31	45	21
Images per class	220 ~ 420	700	100
Images	10,000	31,500	2,100
Resolution (m)	0.5 ~ 8	0.2 ~ 30	0.3
Size (pixel)	600 × 600	256 × 256	256 × 256
Source	Google Earth	Google Earth	USGS

$$\begin{aligned} &- \frac{\lambda}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 1) \\ &- \frac{\lambda}{n_t} \sum_{\mathbf{x}_i \in \mathcal{D}_t} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 0) \\ &- \frac{\lambda \gamma}{n_s} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} m(\mathbf{x}_i) L_d(G_d(G_f(\mathbf{x}_i)), 0) \\ &- \frac{\alpha}{n_s \log(C-1)} \sum_{\mathbf{x}_i \in \mathcal{D}_s} \omega_{y_i} L_{\text{ACER}}(G_y(G_f(\mathbf{x}_i)), y_i) \quad (8) \end{aligned}$$

where  $C_{\text{cls}}$ ,  $C_{\text{adv}}$ , and  $C_{\text{ent}}$  represent the learning objective of the classification for the source domain, the domain adversarial alignment, and the entropy loss, respectively, and  $\lambda$  and  $\alpha$  represent the hyperparameters that tradeoff the adversarial alignment objective and attentive entropy objective with the classification objective in the unified optimization problem, respectively.  $\gamma$  is a gradually decreasing ratio. The minimax optimization problem aims at finding the network parameters  $\theta_f$ ,  $\theta_y$ , and  $\theta_d$  that jointly satisfy

$$\begin{aligned} (\hat{\theta}_f, \hat{\theta}_y) &= \arg \min_{\theta_f, \theta_y} C_{\text{Overall}}(\theta_f, \theta_y) \\ (\hat{\theta}_d) &= \arg \max_{\theta_d} C_{\text{Overall}}(\theta_d). \end{aligned} \quad (9)$$

#### IV. DATASETS

We collect a remote sensing dataset to validate the performance of our proposed method. The dataset is based on three different open-source remote sensing datasets, i.e. AID [7], NWPU-RESISC45 [8], and UC Merced [70]. Table I lists the detailed information of these three public remote sensing datasets. They are derived from different platforms and regions with different resolutions and acquisition dates, which are suitable for validating domain adaptation approaches [62], [66], [74], [75], [76]. We first review the three publicly available remote sensing datasets and then introduce the collected dataset and the partial domain adaptation scenarios.

- 1) AID [7] is made up of 30 aerial scene types collected from Google Earth imagery. All the sample images per class in AID are chosen from different countries and regions at different times and seasons under different imaging conditions. There are 10 000 images in the AID dataset, with 220–420 images of size 600 × 600.
- 2) NWPU-RESISC45 [8] consists of 31 500 images divided into 45 scene classes with spatial resolution ranging from 30 to 0.2 m per pixel. Each class has 700 images of size 256 × 256 pixels. The scale of NWPU-RESISC45 is larger than UC Merced and AID, along with rich image

TABLE II

LABEL INFORMATION FOR SIX TRANSFER TASKS. THE CLASS NAMES IN THE PARENTHESES DENOTE THE CLASS NAMES FOR THE TARGET DOMAIN. NOTABLY, ALL CLASSES IN UC MERCED ARE INCLUDED IN NWPU-RESISC45 AND IT IS THE STANDARD DOMAIN ADAPTATION SCENARIO. OTHER FIVE TRANSFER TASKS ARE THE PARTIAL DOMAIN ADAPTATION SCENARIOS

Source Domain	Target Domain	Shared Classes	Outlier Classes
AID	NWPU-RESISC45	Airport, Baseball field, Beach, Bridge, Church, Forest, Commercial, Dense residential, Medium residential, Desert, Farmland (Circular farmland & Rectangular farmland), Playground (Ground track field), Port (Harbor), Industrial, Sparse residential, Storage tank, viaduct (Overpass), Meadow, Mountain, River, Railway station, Stadium, Parking	Bare land, Center, School, Park, Pond, Resort, Square
AID	UC Merced	Baseball field, Beach, Commercial (Buildings), Forest, Medium residential, Parking, Port (Harbor), Sparse residential, Storage tanks, Viaduct (Overpass), Dense residential, Farmland (Agricultural), River	Airport, Bare land, Bridge, Square, Stadium, Center, Church, Desert, School, Mountain, Desert, Industrial, Meadow, Railway station, Park, Playground, Pond, Resort
NWPU-RESISC45	AID	Airport, Baseball diamond, Beach, Church, Commercial area, Circular farmland & Rectangular farmland (Farmland), Harbor (Port), Railway station, River, Sparse residential, Dense residential, Parking lot, Ground track field (Playground), Forest, Industrial area, Meadow, Medium residential, Mountain, Stadium, Storage tank, Overpass (Viaduct), Desert, Bridge	Basketball court, Chaparral, Cloud, Thermal power station, Freeway, Golf course, Intersection, Island, Terrace, Airplane, Lake, Mobile home park, Palace, Tennis court, Roundabout, Runway, Sea ice, Ship, Snowberg, Wetland
NWPU-RESISC45	UC Merced	Airplane, Baseball diamond, Beach, Chaparral, Commercial area, Dense residential, Forest, Freeway, Golf course, Harbor, Intersection, Circular farmland & Rectangular farmland (Farmland), Mobile home park, Parking lot, River, Runway, Sparse residential, Storage tank, Tennis court, Medium residential, Overpass	Airport, Basketball court, Bridge, Church, Island, Lake, Cloud, Desert, Ground track field, Industrial area, Meadow, Mountain, Palace, Railway, Railway station, Sea ice, Ship, Snowberg, Stadium, Terrace, Wetland, Roundabout, Thermal power station
UC Merced	AID	Agricultural (Farmland), Baseball diamond, Buildings (Commercial), Dense residential, Forest, Harbor (Port), Medium residential, Beach, Overpass (Viaduct), Parking lot, River, Sparse residential, Storage tanks	Airplane, Chaparral, Golf course, Mobile home park, Tennis court, Intersection, Runway, Freeway
UC Merced	NWPU-RESISC45	Airplane, Baseball diamond, Beach, Buildings (Commercial area), Agricultural (Circular farmland & Rectangle farmland), Forest, Freeway, Golf course, Harbor, Intersection, Medium residential, Mobile home park, Overpass, Parking lot, River, Runway, Tennis court, Storage tanks, Sparse residential, Chaparral	None

variations, high within-class diversity and between-class similarity.

- 3) UC Merced [70] comprises 21 land use classes selected from aerial orthoimagery with a pixel resolution of 1 ft. UC Merced is the first publicly available remote sensing evaluation dataset and has been widely used to develop and evaluate remote sensing image classification. The images, downloaded from the United States Geological Survey (USGS), are cropped from large aerial images.

According to the classes in AID, NWPU-RESISC45, and UC Merced, we collect six transfer tasks and the label information can be shown in Table II. Note that the class names in the parentheses denote the class names for the target domain. For example, in the transfer task of AID → NWPU-RESISC45, the farmland in the AID corresponds to the circular farmland and the rectangular farmland in the NWPU-RESISC45; the playground in the AID corresponds to the ground track field in the NWPU-RESISC45. It is easy to understand that the more the outlier classes are, the harder the transfer task is. Fig. 4 shows some example classes of these six transfer tasks. All of these transfer tasks are partial domain adaptation scenarios except UC Merced → NWPU-RESISC45. Since all the classes in the UC Merced are included in the NWPU-RESISC45, the transfer task of UC Merced → NWPU-RESISC45 is actually a standard domain adaptation scenario. We still retain this standard domain adaptation scenario because we can proof that our proposed method not only improves in the partial domain adaptation scenarios but also performs better in the standard domain adaptation scenarios.

Different from standard domain adaptation dataset such as Office-31 [77] and Office-Home [78] that set common classes and outlier classes according to the alphabetic order,

we evaluate our partial domain adaptation methods cross different remote sensing datasets rather than only one dataset. Each dataset has its own collecting purpose and setting of corresponding classes. Any two datasets may include different types of classes for land cover and land use (see Tables I and II). To this end, it is difficult to unify the categories among different remote sensing datasets. In this article, we select three public remote sensing datasets, which is listed in Table I. For example, in AID → NWPU-RESISC45, the number of shared classes is 23 and the number of private classes in the source domain is 7 (see Table II). Noticeably, we remove private classes in the target domain in partial domain adaptation scenarios. Compared to the standard domain adaptation dataset such as Office-31 [77] and Office-Home [78], this kind of setting is more practical in real-world applications that we cannot know the exact number of classes in advance. Furthermore, it can evaluate the effectiveness of our method under varying numbers of outlier classes in the source domain, while the standard domain adaptation dataset usually evaluates the performance under fixed outlier classes [30], [34].

## V. EXPERIMENTAL RESULTS

### A. Experimental Setup

In our experiments, we use all labeled source domain samples and all unlabeled target domain samples and compare the average classification accuracy. We set tradeoff hyperparameters  $\lambda$ ,  $\alpha$ , and the original  $\gamma$  as 1.0, 0.1, and 0.25, respectively. We implement our method based on PyTorch [84] and conduct them on GeForce 2080 Ti. The backbone architectures in our experiments are ResNet-50 [79] pretrained on ImageNet dataset [85]. The learning rate is 0.001 with the same annealing strategy as [24], and we adopt mini-batch SGD [86] with



Fig. 4. Example classes of six partial domain adaptation scenarios, including AID  $\rightarrow$  NWPU-RESISC45, AID  $\rightarrow$  UC Merced, NWPU-RESISC45  $\rightarrow$  AID, NWPU-RESISC45  $\rightarrow$  UC Merced, UC Merced  $\rightarrow$  AID, and UC Merced  $\rightarrow$  NWPU-RESISC45. Notably, all classes in UC Merced are included in NWPU-RESISC45 and it is the standard domain adaptation scenario. Other five transfer tasks are the partial domain adaptation scenarios.

a momentum of 0.9 as our optimizer, following the batch size of 36. Our method is validated by our collected datasets described in Section IV. We evaluate the accuracy for six types of transfer tasks for partial domain adaptation scenarios and their average accuracy. The six transfer tasks are  $A \rightarrow N$ ,  $A \rightarrow U$ ,  $N \rightarrow A$ ,  $N \rightarrow U$ ,  $U \rightarrow A$ , and  $U \rightarrow N$ , where  $A$ ,  $N$ , and  $U$  denote AID, NWPU-RESISC45, and UC Merced, respectively.

#### B. Comparisons Between Our Proposed Method and Other State-of-the-Art Domain Adaptation Methods

We compare our proposed method with other five state-of-the-art partial domain adaptation approaches, including PADA [32], SAN [30], ETN [33], CPADA [69], BA<sup>3</sup>US [34], DRCN [35], AR [82], and CAL [83]. PADA [32] adds the average predictions of the target domain simultaneously to the source classifier and the domain discriminator. SAN [30] employs a multidiscriminator weighted by the percentages of predicted results to the target domain. ETN [33] adopts an entropy minimization principle to reweight the source classifier

and the domain discriminator. CPADA [69] transfers relevant examples in the shared classes and ignores irrelevant ones in the specific classes with the aid of the coordinate loss. BA<sup>3</sup>US [34] emphasizes uncertain samples and exploits and adaptive weighted complement entropy objective to encourage incorrect classes to have uniform and low prediction scores. DRCN [35] plugs one residual block into the source network along with the task-specific feature layer, enhancing the adaptation from source to target, and explicitly weakens the impacts from the outlier classes. AR [82] proposes a training algorithm that alternately updates the parameters of the network and optimizes the weights of source domain data. CAL [83] designs a contrastive learning-aided alignment method for partial domain adaptation scenarios to reweight source samples to reduce the contribution of outlier samples. At the same time, we also compare our proposed method with other seven standard domain adaptation approaches to indicate the severe negative transfer effect. These seven standard domain adaptation approaches include deep domain confusion (DDC) [19], DANN [24], Deep Coral [21], CDAN [25], TADA [26], transferable ResNet (TransResNet) [80], and category contrast technique (Caco) [81]. DDC introduces an adaptation layer and an additional domain confusion loss through maximum mean discrepancy (MMD). DANN adopts a gradient reversal layer to facilitate adaptation so that the model does not perform well in discriminating between the source domain and the target domain, which has been demonstrated in Section III. Deep Coral learns a transfer network through a linear transformation to align the second-order statistics of the source and target distributions. CDAN is designed with two conditioning strategies, i.e., multilinear conditioning and entropy conditioning. TADA focuses the adaptation model on more transferable regions and images. TransResNet [80] remedies the residual block in the ResNet [79], separating source and target input features and highlighting more transferable channels in each block. Caco [81] constructs a semantics-aware dictionary with samples from both source and target domains where each target sample is assigned a (pseudo) category label based on the category priors of source samples. We also list the straightforward CNN model that leverages only classification loss without any domain adaptation approaches in Table III. In this article, we conduct our experiments on popular architectures, ResNet-50 [79], which contains the residual block so that empirically eases the problem of gradient vanishing and explosion. The ResNet family has been considered as the most versatile backbone across a wide array of deep learning applications.

Table III lists the accuracy of our collected dataset for six partial domain adaptation scenarios under the backbone of ResNet-50. Among standard domain adaptation methods, we can observe that only DANN [24] and Deep Coral [21] perform better than the straightforward CNN model (i.e., ResNet-50), with only 0.13% and 1.05% improvement, respectively. Other standard domain adaptation methods achieve worse performance than the naive ResNet-50 model, indicating that there exists severe negative transfer effect for partial domain adaptation scenarios (see details in VI-A). As for partial domain adaptation approaches, our

TABLE III  
ACCURACY (%) ON OUR COLLECTED DATASET FOR PARTIAL DOMAIN ADAPTATION SCENARIOS (RESNET-50)

Method	A → N	A → U	N → A	N → U	U → A	U → N	Avg
ResNet-50 [79]	72.32	51.38	80.43	73.67	57.74	60.21	65.96
Standard domain adaptation	DDC [19]	74.37	46.92	77.12	63.76	64.42	65.14
	DANN [24]	75.68	50.46	77.04	63.00	65.58	64.76
	Deep Coral [21]	73.43	54.46	81.33	73.00	57.91	61.93
	CDAN [25]	74.62	46.62	76.46	62.71	64.66	64.81
	TADA [26]	74.34	49.08	75.88	63.71	63.52	65.47
	TransResNet [80]	75.38	51.15	81.11	73.29	64.64	63.93
Partial domain adaptation	Caco [81]	73.45	56.46	82.75	72.81	55.67	62.26
	PADA [32]	72.02	51.85	83.14	74.14	61.08	59.23
	SAN [30]	79.55	58.08	85.93	76.43	76.46	68.81
	ETN [33]	78.67	52.08	84.90	74.76	71.45	68.50
	CPADA [69]	72.92	59.58	85.49	73.88	73.47	67.01
	BA <sup>3</sup> US [34]	83.98	58.77	89.21	75.86	73.74	73.70
	DRCN [35]	79.18	64.15	85.81	80.67	63.89	66.71
	AR [82]	<b>84.67</b>	63.23	86.10	79.95	77.11	71.80
Ours	CAL [83]	84.17	63.62	85.66	<b>80.82</b>	76.31	74.19
	Ours	84.61	<b>64.62</b>	<b>89.54</b>	80.76	<b>79.71</b>	<b>76.92</b>

TABLE IV  
ACCURACY (%) OF ABLATION STUDIES ON OUR COLLECTED DATASET FOR OUR PROPOSED METHOD (RESNET-50)

PADM	I-DANN	WCER	A → N	A → U	N → A	N → U	U → A	U → N	Avg
✗	✗	✗	75.68	50.46	77.04	63.00	65.58	64.76	66.09
✓	✗	✗	79.55	<b>62.77</b>	80.94	72.90	73.12	72.42	73.62
✗	✓	✗	80.51	55.15	87.05	77.62	67.54	69.90	72.96
✗	✗	✓	74.80	60.18	83.50	72.20	70.22	66.96	71.31
✓	✓	✗	81.28	62.54	87.85	<b>81.14</b>	76.75	73.67	77.21
✓	✗	✓	82.17	62.42	86.90	78.55	<b>77.77</b>	74.77	77.10
✗	✓	✓	82.98	57.73	86.71	74.79	73.73	75.40	74.92
✓	✓	✓	<b>84.61</b>	<b>64.62</b>	<b>89.54</b>	80.76	<b>79.71</b>	<b>76.92</b>	<b>79.36</b>

TABLE V  
COMPARISON BETWEEN  $e^{-H(h)}$  AND  $1 + e^{-H(h)}$  IN “m” OF THE SAMPLE-LEVEL WEIGHT IN (3)

Method	A → N	A → U	N → A	N → U	U → A	U → N	Avg
$e^{-H(h)}$	81.45	62.00	88.51	79.95	76.31	74.85	77.18
$1 + e^{-H(h)}$ (Ours)	<b>84.61</b>	<b>64.62</b>	<b>89.54</b>	<b>80.76</b>	<b>79.71</b>	<b>76.92</b>	<b>79.36</b>

proposed method achieves the highest average accuracy of 79.36%, with 1.90%–12.45% improvement compared to other state-of-the-art partial domain adaptation methods, which shows superior classification accuracy on each transfer tasks. Also, our proposed method attains 13.67% gains compared to the straightforward CNN model.

### C. Ablation Studies

Table IV shows the accuracy of ablation studies on our collected dataset for our proposed method under the backbone of ResNet-50. Notably, PADM, I-DANN, and ACER represents the strategies of PADM, improved DANN with multiweights, and ACER, respectively. The Baseline method (the first line in Table IV) here is the naive DANN [24], which is described in (1).

1) *Effectiveness of the PADM:* Unlike previous partial domain adaptation methods [30], [32], we resort to augment an auxiliary domain by using the original samples from the source domain instead of employing the previous reweighting scheme. Our PADM attains 7.53% improvement compared to DANN. It is desirable that PADM promotes the classification accuracy substantially on harder transfer tasks, such as  $\mathbf{A} \rightarrow \mathbf{U}$  (+12.31%) and  $\mathbf{N} \rightarrow \mathbf{U}$  (+9.90%), which have more outlier classes. Furthermore, the PADM achieves +4.25%, +3.48%, and +2.15%, respectively, based

on I-DANN, DANN + ACER, and I-DANN + ACER with respect to the average accuracy, indicating the superiority and effectiveness of the PADM.

2) *Effectiveness of the I-DANN:* According to the original DANN, our I-DANN contains two kinds of reweight mechanisms to improve the performance in the partial domain adaptation scenarios: an entropy-aware weight and a target-class-aware weight, which is a sample-level weight scheme and a class-level weight scheme, respectively. As shown in Table IV, our I-DANN performs +6.88% better than the original DANN, especially attaining significant improvement on more challenging transfer tasks, such as  $\mathbf{N} \rightarrow \mathbf{U}$  (+14.62%) and  $\mathbf{N} \rightarrow \mathbf{A}$  (+10.01%). In addition, the I-DANN achieves +3.59%, +3.61%, and +2.26% improvements, respectively, on the basis of DANN + PADM, DANN + ACER, and DANN + ACER + PADM with respect to the average accuracy.

In addition, we further compare the performance between  $e^{-H(h)}$  and  $1 + e^{-H(h)}$  in “m” of the sample-level weight in (3), which is listed in Table V. We can find that  $1 + e^{-H(h)}$  can be more stable than  $e^{-H(h)}$  for our model, as well as avoiding the value of exponential form is too small. Furthermore, we also conduct the performance comparisons among different weight mechanisms in I-DANN. As can be seen in Table VI, simultaneously adopting these two kinds of weights

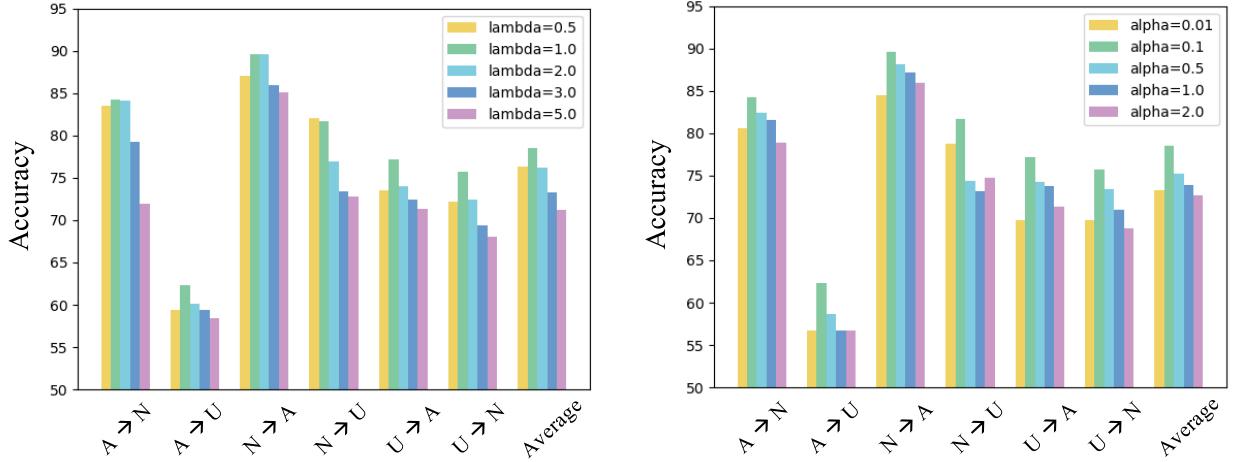


Fig. 5. Accuracy (%) of our method for our collected dataset in the partial domain adaptation scenarios under different hyperparameters of (Left)  $\lambda$  and (Right)  $\alpha$  (ResNet-50).

TABLE VI  
PERFORMANCE COMPARISONS AMONG DIFFERENT WEIGHT MECHANISMS IN I-DANN

Method	$A \rightarrow N$	$A \rightarrow U$	$N \rightarrow A$	$N \rightarrow U$	$U \rightarrow A$	$U \rightarrow N$	Avg
Only sample-level weight	81.57	62.92	88.71	78.86	76.51	73.43	77.00
Only class-level weight	83.73	60.23	87.73	79.19	78.05	73.19	77.03
Both sample-level and class-level weight (I-DANN) (Ours)	<b>84.61</b>	<b>64.62</b>	<b>89.54</b>	<b>80.76</b>	<b>79.71</b>	<b>76.92</b>	<b>79.36</b>

(i.e., sample- and class-level weight mechanisms) achieves the best accuracy, with over 2.30% better than only one weight mechanism.

3) *Effectiveness of the ACER*: Our ACER is proposed not only to reduce the uncertainty of probabilities for output classes, making the classifier more accessible to the unlabeled target data, but also to alleviate the forceful minimization of the entropy of nontransferable samples (such as the samples belonging to outlier classes in the source domain). We can observe from Table IV that although the performance drops on the task of  $A \rightarrow N$  (-0.89%), our ACER improves the average accuracy of +5.22%. Furthermore, the ACER attains +3.48%, +1.96%, and +2.15% gains, respectively, on the basis of DANN + PADM, I-DANN, and I-DANN + PADM with the respect of the average accuracy.

Furthermore, we compare different entropy regularizations on our collected dataset for partial domain adaptation scenarios in Table VII. We can observe that ACER achieves higher accuracy than other entropy regularization, including attentive entropy regularization in TADA [26], original complement entropy regularization [73], confidence-weighted complement entropy in BA<sup>3</sup>US [34], and guided complement entropy [87], with 0.66%-2.15% improvement. To this end, we adopt ACER as our entropy regularization in our partial domain adaptation method.

4) *Sensitive Analysis*: Here, we conduct ablation studies for two hyperparameters, i.e.,  $\lambda$  and  $\alpha$ . Fig. 5 shows the results of our proposed method for our collected dataset in the partial domain adaptation scenarios under two hyperparameters of  $\lambda$  (see the left in Fig. 5) and  $\alpha$  (see the right in Fig. 5) for ResNet-50. When  $\alpha = 0.1$ , we evaluate different values of  $\lambda$  ranging from 0.5 to 5.0 (on the left in Fig. 5). It is evident that when  $\lambda > 1.0$ , the accuracy drops dramatically. Also, the performance of  $\lambda = 1.0$  is slightly better than  $\lambda = 0.5$ . When

$\lambda = 1.0$ , we evaluate different values of  $\alpha$  ranging from 0.01 to 2.0 (on the right in Fig. 5). We observe that  $\alpha = 0.1$  performs better than others with an obvious improvement. In conclusion, we set the hyperparameters  $\lambda$  and  $\alpha$  as 1.0 and 0.1 in all our experiments, respectively.

## VI. DISCUSSION

### A. Comparisons Among ResNet-50, Standard Domain Adaptation Methods, and Partial Domain Adaptation Methods

In this section, we compare the straightforward CNN model (ResNet-50), standard domain adaptation methods, and partial domain adaptation methods in respect of their accuracy. In particular, we point out the accuracy degradation for standard domain adaptation methods. We calculate the accuracy gap between the aforementioned domain adaptation methods and the straightforward CNN model (ResNet-50), which only uses the source dataset without any domain adaptation algorithm. In practice, the accuracy gap can be calculated as

$$\text{Accuracy Gap} = \text{ACC}_{\text{DA}} - \text{ACC}_{\text{ResNet-50}} \quad (10)$$

where  $\text{ACC}_{\text{DA}}$  and  $\text{ACC}_{\text{ResNet-50}}$  are the accuracy of approaches that using domain adaptation and the straightforward CNN method (ResNet-50), respectively. If  $\text{Accuracy Gap} > 0$ , it means that the domain adaptation method performs a positive effect in the partial domain adaptation scenarios. Conversely,  $\text{Accuracy Gap} < 0$  denotes that the domain adaptation method brings about an accuracy degradation in the partial domain adaptation scenario without any improvement. Table VIII lists the performance of Accuracy Gap for all domain adaptation methods above, including standard domain adaptation approaches and partial domain adaptation

TABLE VII  
PERFORMANCE COMPARISONS AMONG DIFFERENT ENTROPY REGULARIZATIONS ON OUR COLLECTED DATASET  
FOR PARTIAL DOMAIN ADAPTATION SCENARIOS (RESNET-50)

Method	A → N	A → U	N → A	N → U	U → A	U → N	Avg
Attentive entropy regularization in TADA [26]	84.27	62.38	89.64	<b>81.71</b>	77.21	75.69	78.48
Complement entropy regularization [73]	84.17	63.62	88.51	81.52	78.09	76.28	78.70
Confidence-weighted complement entropy in BA <sup>3</sup> US [34]	84.05	62.00	86.81	76.81	78.76	74.82	77.21
Guided complement entropy [87]	85.38	63.92	<b>89.72</b>	77.19	78.80	75.13	78.36
Attentive complement entropy regularization (ACER) (Ours)	<b>84.61</b>	<b>64.62</b>	89.54	80.76	<b>79.71</b>	<b>76.92</b>	<b>79.36</b>

TABLE VIII  
ACCURACY GAP (%) ON OUR COLLECTED DATASET FOR PARTIAL DOMAIN ADAPTATION SCENARIOS (RESNET-50).  
THE RED NUMBERS DENOTE THE ACCURACY GAP < 0

Method	A → N	A → U	N → A	N → U	U → A	U → N	Avg
Standard domain adaptation	DDC [19]	2.05	<b>-4.46</b>	<b>-3.31</b>	<b>-9.91</b>	6.68	4.93
	DANN [24]	3.36	<b>-0.92</b>	<b>-3.39</b>	<b>-10.67</b>	7.84	4.55
	Deep Coral [21]	1.11	3.08	0.90	<b>-0.67</b>	0.17	1.72
	CDAN [25]	2.30	<b>-4.76</b>	<b>-3.97</b>	<b>-10.96</b>	6.92	4.60
	TADA [26]	2.02	<b>-2.30</b>	<b>-4.55</b>	<b>-9.96</b>	5.78	5.26
	TransResNet [80]	3.06	<b>-0.23</b>	0.68	<b>-0.38</b>	6.90	3.72
Partial domain adaptation	Caco [81]	1.13	5.08	2.32	<b>-0.86</b>	<b>-2.07</b>	2.05
	PADA [32]	<b>-0.30</b>	0.47	2.71	0.47	3.34	<b>-0.98</b>
	SAN [30]	7.23	6.70	5.50	6.76	18.72	8.6
	ETN [33]	6.35	0.70	4.47	1.09	13.71	8.29
	CPADA [69]	0.60	8.20	5.06	0.21	15.73	6.80
	BA <sup>3</sup> US [34]	11.66	7.39	8.78	2.19	16.00	13.49
	DRCN [35]	6.86	12.77	5.38	7.00	6.15	6.50
	AR [82]	<b>12.35</b>	11.85	5.67	6.28	19.37	11.59
	CAL [83]	11.85	12.24	5.23	<b>7.15</b>	18.57	13.98
	Ours	12.29	<b>13.24</b>	<b>9.11</b>	7.09	<b>21.97</b>	<b>16.71</b>
13.67%							

approaches. Unfortunately, although employing domain adaptation algorithms, most standard domain adaptation methods have serious accuracy deficiency in more difficult partial domain adaptation scenarios, such as  $A \rightarrow U$ ,  $N \rightarrow A$ , and  $N \rightarrow U$ . For example, in the task of  $N \rightarrow U$ , all standard domain adaptation methods encounter serious accuracy degradation ranging from  $-0.38\%$  to  $-10.96\%$ . In the task of  $A \rightarrow U$ , all standard domain adaptation approaches incur the performance degradation except Deep Coral [21] and Caco [81], with  $-0.23$  to  $-4.76$ . For other transfer tasks, Accuracy Gap  $> 0$  since the advantages of domain adaptation strategies win the effect of outlier classes from the source domain in the partial domain adaptation scenarios. To this end, directly adopting standard domain adaptation approaches (consider that the label space of the source domain is identical to that of the target domain) will lead to relatively severe deficiency in the partial domain adaptation scenarios. Although PADA [32] adopts partial domain adaptation strategies, it still causes slight accuracy degradation, with  $-0.30\%$  in  $A \rightarrow N$  and  $-0.98\%$  in  $U \rightarrow N$ , which indicates that directly employing partial domain adaptation algorithms in computer vision domain may not satisfy in the remote sensing community. On the other hand, it is convinced that our proposed method greatly improves the transferability in partial domain adaptation scenarios, with  $+13.67\%$  improvement compared to the straightforward ResNet-50 model. The results also prove the necessity for our strategies of our proposed method to alleviate the accuracy degradation in partial domain adaptation scenarios.

#### B. Accuracy for Varying Numbers of Outlier Classes

In this section, we investigate a wider spectrum of partial domain adaptation by varying the number of outlier

classes. Fig. 6 shows that when the number of outlier classes increases, the performance of the straightforward CNN model (ResNet-50) and the standard domain adaptation method (CDAN) degrades quickly, meaning that negative transfer becomes severer when the domain gap is enlarged. The performance of state-of-the-art partial domain adaptation methods (such as PADA and SAN) still has a certain measure of degradation. From Fig. 6, our proposed method always achieves the highest accuracy and lowest performance degradation as the number of outlier classes increases, indicating that our proposed method can effectively alleviate the negative transfer effect in partial domain scenarios. To this end, the margins that our proposed method outperforms other domain adaptation methods become larger and larger when the number of outlier classes increases.

#### C. Features and Weights Visualization

To display the feature transferability, we visualize the network representations of the last convolutional layer from three transfer tasks in Fig. 7 (from top to bottom:  $A \rightarrow U$ ,  $N \rightarrow U$ , and  $U \rightarrow A$ ) learned by Baseline (ResNet-50), DANN, Deep Coral, SAN, and Ours (from left to right) using t-SNE visualization [89], [90]. From left (ResNet-50) to right (Ours), the target domains are made more and more indistinguishable. For example, the representations generated by our proposed method formed exactly 13 clusters with clearer boundaries in the transfer task of  $U \rightarrow A$ . The better visualization results of our method indicate that our strategies are able to learn more transferable features and eliminate the negative transfer effect for the partial domain adaptation.

In addition, we illustrate the estimation of class-level weights ( $\omega$  in (8)) for two transfer tasks (top line:  $A \rightarrow N$

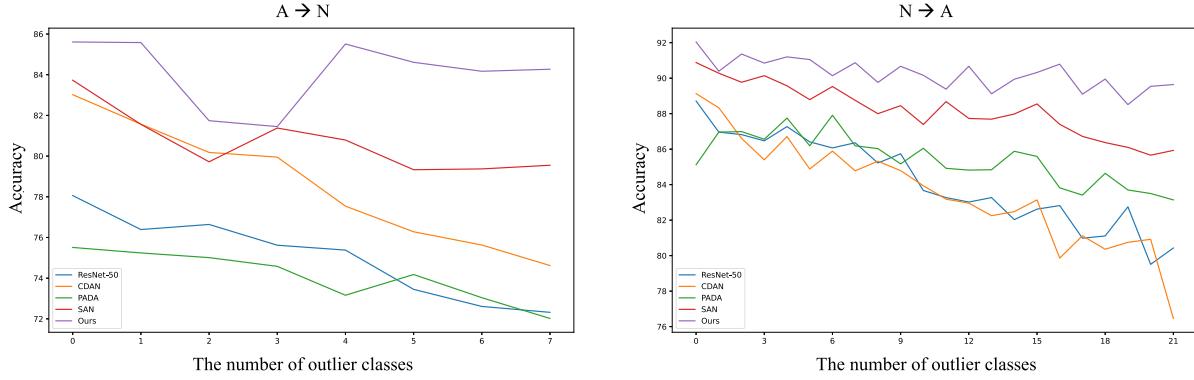


Fig. 6. Accuracy with varying number of outlier classes [(Left)  $A \rightarrow N$  and (Right)  $N \rightarrow A$ ]. Our proposed method always achieves the highest accuracy and the lowest performance degradation as the number of outlier classes increasing, indicating that our proposed method can effectively alleviate the negative transfer effect in partial domain scenarios. To this end, the margins that our proposed method outperforms other domain adaptation methods become larger and larger when the number of outlier classes increases.

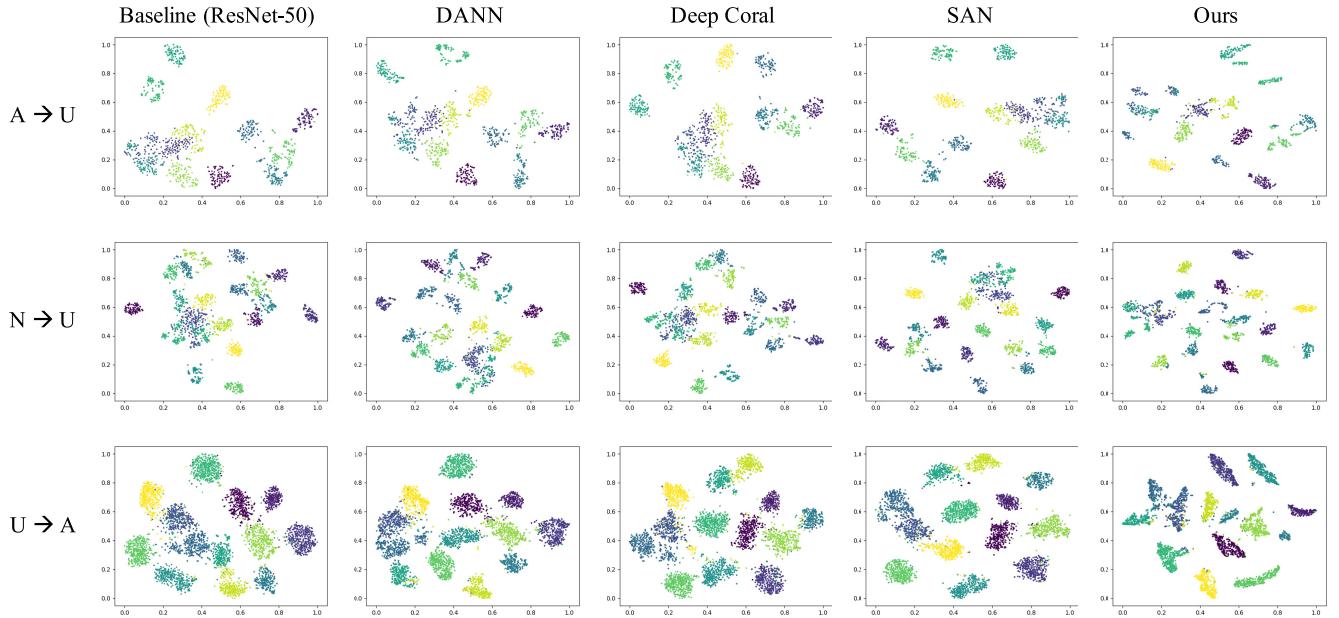


Fig. 7. t-SNE visualization of features for three transfer tasks (from top to bottom:  $A \rightarrow U$ ,  $N \rightarrow U$ , and  $U \rightarrow A$ ) learned by Baseline (ResNet-50), DANN, Deep Coral, SAN, and Ours (from left to right). From left to right, the target domains are made more and more indistinguishable.

and bottom line:  $N \rightarrow A$  in Fig. 8. Fig. 8 shows the class-level weights at the iteration of 0, 400, 800, 1200, 1600, and 2000 (from left to right). The  $x$ -axis denotes the value of weights and the  $y$ -axis represents the class index. In the task of  $A \rightarrow N$ , class indexes from 0 to 23 are shared classes and others are outlier classes in the source domain. We can observe that as the iteration number increases, the weights of shared classes become higher, while the weights of outlier classes become lower. In the task of  $N \rightarrow A$ , the outlier classes' indexes are 24–44, which is a more challenging partial domain adaptation task. It is also evident that the weight estimation sounds accurate, resulting in high-accuracy classification performance.

#### D. Effectiveness of Our Proposed Method on Standard Partial Domain Adaptation Datasets

We also evaluate our proposed method on standard partial domain adaptation datasets, such as Office-31 [77] and

DomainNet [88] in Tables IX and X. Office-31 contains three domains with 31 categories: Amazon (A), DSLR (D), and Webcam (W). Following the same data protocol as previous partial domain adaptation publications [32]. DomainNet [88] is a large-scale challenging domain adaptation dataset. We pick four domains [Clipart (C), Painting (P), Real (R), and Sketch (S)] with 126 classes. We use the first 40 categories to build the target domain according to the alphabetical order [82], [91]. As listed in Tables IX and X, our proposed method achieves the highest average accuracy in both traditional domain adaptation datasets, meaning that our proposed method is more effective and robust than other state-of-the-art partial domain adaptation methods.

#### E. Potential Practical Application Scenarios in Remote Sensing Community Based on Our Proposed Method

Partial domain adaptation problem is a more practical and challenging transfer task. Despite that the issue of domain

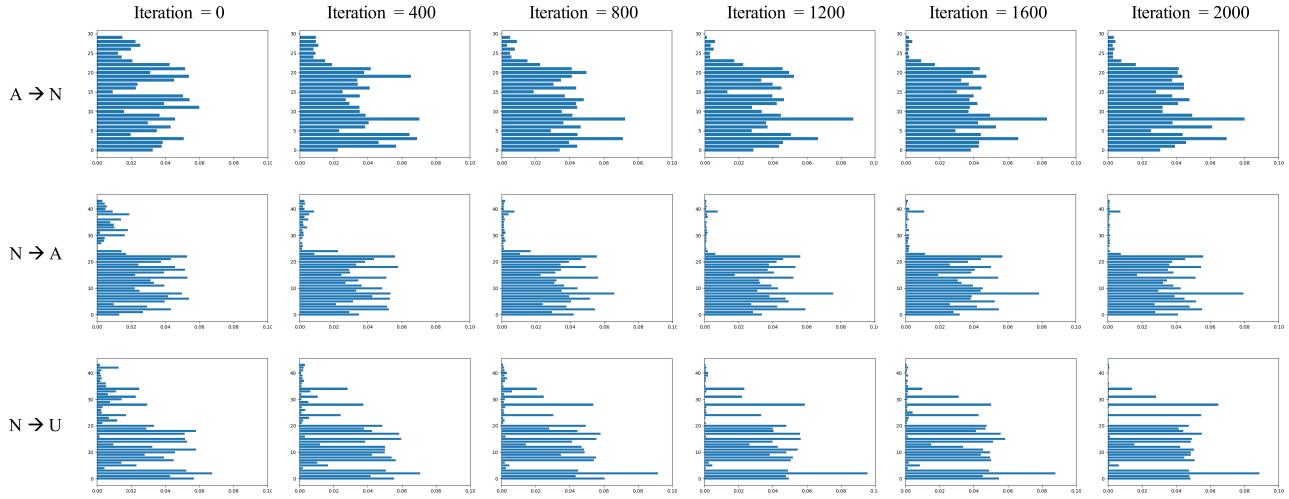


Fig. 8. Illustration of class-level weights ( $\omega$  in (8)) for two transfer tasks (top line:  $A \rightarrow N$  and bottom line:  $N \rightarrow A$ ). We display the class-level weights at the iteration of 0, 400, 800, 1200, 1600, and 2000 (from left to right). The  $x$ -axis denotes the value of weights and the  $y$ -axis represents the class index. We can observe that as the iteration increasing, the weights of shared classes become higher, while the weights of outlier classes become lower.

TABLE IX  
ACCURACY (%) ON OFFICE-31 [77] FOR PARTIAL DOMAIN ADAPTATION SCENARIOS (RESNET-50)

Method	$A \rightarrow D$	$A \rightarrow W$	$D \rightarrow A$	$D \rightarrow W$	$W \rightarrow A$	$W \rightarrow D$	Avg
ResNet-50 [79]	83.44	75.59	83.92	96.27	84.97	98.09	87.05
Standard domain adaptation	DANN [24]	81.53	73.56	82.78	96.27	86.12	98.73
	CDAN [25]	77.07	80.51	93.58	98.98	91.65	98.09
Partial domain adaptation	PADA [32]	82.17	86.54	92.69	99.32	<b>95.41</b>	<b>100.00</b>
	SAN [30]	94.27	93.90	94.15	99.32	88.73	99.36
	ETN [33]	95.03	94.52	<b>96.21</b>	<b>100.00</b>	94.64	<b>100.00</b>
	Ours	<b>99.25</b>	<b>98.87</b>	94.71	<b>100.00</b>	94.88	98.51
<b>97.70</b>							

TABLE X  
ACCURACY (%) ON DOMAINNET [88] FOR PARTIAL DOMAIN ADAPTATION SCENARIOS (RESNET-50)

Method	$C \rightarrow P$	$C \rightarrow R$	$C \rightarrow S$	$P \rightarrow C$	$P \rightarrow R$	$P \rightarrow S$	$R \rightarrow C$	$R \rightarrow P$	$R \rightarrow S$	$S \rightarrow C$	$S \rightarrow P$	$S \rightarrow R$	Avg
ResNet-50 [79]	41.21	60.01	42.13	54.52	70.80	48.32	63.10	58.63	50.26	45.43	39.30	49.75	51.96
Standard domain adaptation	DANN [24]	27.83	36.64	29.91	31.79	41.98	36.58	47.64	46.81	40.85	25.82	29.54	35.68
	CDAN [25]	37.46	48.26	46.61	45.50	60.96	52.63	62.01	60.63	54.74	35.37	38.50	43.63
Partial domain adaptation	PADA [32]	22.49	32.85	29.95	25.71	56.47	30.45	65.28	63.35	54.17	17.45	23.89	26.91
	SAN [30]	34.35	51.62	46.23	57.13	70.21	58.25	69.61	67.49	67.88	41.69	45.15	48.44
	ETN [33]	<b>44.99</b>	<b>57.90</b>	53.66	59.11	70.97	63.35	69.39	63.59	62.40	44.48	<b>48.80</b>	48.29
	Ours	42.87	54.72	<b>53.79</b>	<b>65.03</b>	<b>76.39</b>	<b>64.69</b>	<b>79.99</b>	<b>74.31</b>	<b>75.02</b>	<b>50.36</b>	42.69	<b>49.65</b>
<b>60.63</b>													

adaptation has been discussed and developed in remote sensing image classification during recent years, seldom studies conducted experiments to tackle the problem in domain shift scenario that the label space of the source domain subsumes that of the target domain. However, in the remote sensing community, we often meet the scenario where we want to transfer from a larger scale dataset to a smaller scale dataset. If we directly adopt standard domain adaptation methods, the outlier classes (which do not exist in the target domain) may seriously deteriorate the transferability and generate a negative transfer effect since the target label distribution disjoint with the source label distribution. For example, in land cover and land use mapping, the samples of some classes (such as snow/ice and tundra) from the source domain may not be included in the target domain (such as the tropical areas), which may inevitably cause a negative transfer effect. On the other hand, if we want to conduct a forest inventory (tree species classification) for a new study area with unknown tree species, a reasonable way is that we utilize the rich tree species data bank collected from other regions and adopt

partial domain adaptation algorithms to map the tree species for the new study area.

Furthermore, it is hard to be informed of the exact label space of the target domain and relatively labor-exhausting and time-consuming to collect samples for the target domain. Therefore, our proposed method is a step into the empty slots in the partial domain adaptation for remote sensing image classification. Compared to other standard domain adaptation methods, our method can effectively address the negative transfer effect generated by the outlier classes from the source domain. Compared to other state-of-the-art partial domain adaptation methods, our method also keeps high improvement in standard domain adaptation scenarios, indicating that our proposed method is a more robust and practical method in real-world scenarios and applications.

## VII. CONCLUSION

In this context, we propose a new partial domain adaptation algorithm for remote sensing scene classification, which tackles the problem that the label space of the source domain

subsumes that of the target domain. Our proposed method constitutes three main parts. First, we employ a PADM directly using original source samples to balance the difference between the label distributions across domains. To this end, the negative transfer effect caused by outlier classes from the source domain can be greatly alleviated. Second, we not only adopt the normalized estimated class-level weights of the target domain to the classification loss but also set different weights for hard and easy samples during the domain adversarial alignment according to the results of domain discriminator. Finally, we design an ACER to improve the prediction confidence for samples that are easy to be confused with other types and avoid those untransferable samples (such as the samples belonging to the outlier classes in the source domain) to be mistakenly classified because of forceful entropy minimization. To perform an evaluation of our proposed method, we integrate a test dataset that includes three common remote sensing datasets (i.e., AID, NWPU-RESISC45, and UC Merced). Our method achieves an average accuracy of 79.36%, considerably outperforming other state-of-the-art partial domain adaptation methods with an average accuracy improvement of 1.90–12.45% and attaining a 13.67% gain compared to the straightforward CNN model (i.e., ResNet-50). The experiment results indicate that our approach shows promising potential for solving a more general and practical problem with fewer annotations and human resources. In the future, we will strive to explore the effectiveness of our proposed method to practical partial domain adaptation scenarios and applications using multiregional, multisensor, and multitemporal remote sensing images with limited annotations, e.g., land cover and land use mapping, and crop monitoring.

## REFERENCES

- [1] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
- [2] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, “Deep learning classification of land cover and crop types using remote sensing data,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.
- [3] X. X. Zhu et al., “Deep learning in remote sensing: A comprehensive review and list of resources,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.
- [4] K. Li, G. Wan, G. Cheng, L. Meng, and J. Han, “Object detection in optical remote sensing images: A survey and a new benchmark,” *ISPRS J. Photogramm. Remote Sens.*, vol. 159, pp. 296–307, Jan. 2020.
- [5] Y. Li, Y. Zhang, and Z. Zhu, “Error-tolerant deep learning for remote sensing image scene classification,” *IEEE Trans. Cybern.*, vol. 51, no. 4, pp. 1756–1768, Apr. 2020.
- [6] 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.
- [7] G.-S. Xia et al., “AID: A benchmark data set for performance evaluation of aerial scene classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, Apr. 2017.
- [8] G. Cheng, J. Han, and X. Lu, “Remote sensing image scene classification: Benchmark and state of the art,” *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
- [9] J. Quinonero-Candela, M. Sugiyama, A. Schwaighofer, and N. D. Lawrence, *Dataset Shift in Machine Learning*. Cambridge, MA, USA: MIT Press, 2009.
- [10] 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.
- [11] S. Yang, K. Yu, F. Cao, H. Wang, and X. Wu, “Dual-representation-based autoencoder for domain adaptation,” *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 7464–7477, Aug. 2021.
- [12] Y. Li, D. Kong, Y. Zhang, Y. Tan, and L. Chen, “Robust deep alignment network with remote sensing knowledge graph for zero-shot and generalized zero-shot remote sensing image scene classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 179, pp. 145–158, Sep. 2021.
- [13] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [14] P. Gong et al., “Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data,” *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
- [15] W. Li, R. Dong, H. Fu, J. Wang, L. Yu, and P. Gong, “Integrating Google Earth imagery with Landsat data to improve 30-m resolution land cover mapping,” *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111563.
- [16] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.
- [17] S. Zhao et al., “Emotional semantics-preserved and feature-aligned cyclegan for visual emotion adaptation,” *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 10000–10013, Oct. 2021.
- [18] A. Gretton et al., “Optimal kernel choice for large-scale two-sample tests,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1205–1213.
- [19] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep domain confusion: Maximizing for domain invariance,” 2014, *arXiv:1412.3474*.
- [20] M. Long, Y. Cao, J. Wang, and M. Jordan, “Learning transferable features with deep adaptation networks,” in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.
- [21] B. Sun and K. Saenko, “Deep coral: Correlation alignment for deep domain adaptation,” in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 443–450.
- [22] M. Long, H. Zhu, J. Wang, and M. I. Jordan, “Deep transfer learning with joint adaptation networks,” in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.
- [23] L. Luo, L. Chen, S. Hu, Y. Lu, and X. Wang, “Discriminative and geometry-aware unsupervised domain adaptation,” *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3914–3927, Sep. 2020.
- [24] Y. Ganin et al., “Domain-adversarial training of neural networks,” *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 2030–2096, 2016.
- [25] M. Long, Z. Cao, J. Wang, and M. I. Jordan, “Conditional adversarial domain adaptation,” in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 1647–1657.
- [26] X. Wang, L. Li, W. Ye, M. Long, and J. Wang, “Transferable attention for domain adaptation,” in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 5345–5352.
- [27] X. Wu, J. Chen, F. Yu, M. Yao, and J. Luo, “Joint learning of multiple latent domains and deep representations for domain adaptation,” *IEEE Trans. Cybern.*, vol. 51, no. 5, pp. 2676–2687, May 2019.
- [28] J. Jiang, X. Wang, M. Long, and J. Wang, “Resource efficient domain adaptation,” in *Proc. 28th ACM Int. Conf. Multimedia*, 2020, pp. 2220–2228.
- [29] 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.
- [30] Z. Cao, M. Long, J. Wang, and M. I. Jordan, “Partial transfer learning with selective adversarial networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2724–2732.
- [31] J. Zhang, Z. Ding, W. Li, and P. Ogunbona, “Importance weighted adversarial nets for partial domain adaptation,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8156–8164.
- [32] Z. Cao, L. Ma, M. Long, and J. Wang, “Partial adversarial domain adaptation,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 135–150.
- [33] Z. Cao, K. You, M. Long, J. Wang, and Q. Yang, “Learning to transfer examples for partial domain adaptation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2985–2994.
- [34] J. Liang, Y. Wang, D. Hu, R. He, and J. Feng, “A balanced and uncertainty-aware approach for partial domain adaptation,” 2020, *arXiv:2003.02541*.
- [35] S. Li et al., “Deep residual correction network for partial domain adaptation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2329–2344, Jul. 2020.
- [36] Z. Chen, C. Chen, Z. Cheng, B. Jiang, K. Fang, and X. Jin, “Selective transfer with reinforced transfer network for partial domain adaptation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 12706–12714.
- [37] J. Chen, X. Wu, L. Duan, and S. Gao, “Domain adversarial reinforcement learning for partial domain adaptation,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 2, pp. 1–15, Feb. 2020.

- [38] J. Hu et al., "Discriminative partial domain adversarial network," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 632–648.
- [39] S. Yang, Y. Kim, D. Jung, and C. Kim, "Partial domain adaptation using graph convolutional networks," 2020, *arXiv:2005.07858*.
- [40] D. Tuia, D. Marcos, and G. Camps-Valls, "Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization," *ISPRS J. Photogramm. Remote Sens.*, vol. 120, pp. 1–12, Oct. 2016.
- [41] 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.
- [42] L. Yan, R. Zhu, Y. Liu, and N. Mo, "TrAdaBoost based on improved particle swarm optimization for cross-domain scene classification with limited samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3235–3251, Sep. 2018.
- [43] X. Zhou and S. Prasad, "Deep feature alignment neural networks for domain adaptation of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, Oct. 2018.
- [44] L. Yan, R. Zhu, N. Mo, and Y. Liu, "Cross-domain distance metric learning framework with limited target samples for scene classification of aerial images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, Jun. 2019.
- [45] J. Lin, T. Yu, L. Mou, X. Zhu, R. K. Ward, and Z. J. Wang, "Unifying top-down views by task-specific domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, Jun. 2020.
- [46] X. Ma, X. Mou, J. Wang, X. Liu, J. Geng, and H. Wang, "Cross-dataset hyperspectral image classification based on adversarial domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, May 2020.
- [47] E. Alvarez-Vanhard, T. Houet, C. Mony, L. Lecoq, and T. Corpetti, "Can UAVs fill the gap between in situ surveys and satellites for habitat mapping?" *Remote Sens. Environ.*, vol. 243, Jun. 2020, Art. no. 111780.
- [48] Y. Hamrouni, E. Pailllassa, V. Chéret, C. Monteil, and D. Sheeren, "From local to global: A transfer learning-based approach for mapping poplar plantations at national scale using Sentinel-2," *ISPRS J. Photogramm. Remote Sens.*, vol. 171, pp. 76–100, Jan. 2021.
- [49] Y. Li, Z. Zhu, J.-G. Yu, and Y. Zhang, "Learning deep cross-modal embedding networks for zero-shot remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, Dec. 2021.
- [50] W. Liu and F. Su, "Unsupervised adversarial domain adaptation network for semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 11, pp. 1978–1982, Nov. 2020.
- [51] G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova, "Transferring deep learning models for cloud detection between Landsat-8 and Proba-V," *ISPRS J. Photogramm. Remote Sens.*, vol. 160, pp. 1–17, Feb. 2020.
- [52] P. Shamsolmoali, M. Zarepoor, H. Zhou, R. Wang, and J. Yang, "Road segmentation for remote sensing images using adversarial spatial pyramid networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, Jun. 2020.
- [53] Q. Xu, X. Yuan, and C. Ouyang, "Class-aware domain adaptation for semantic segmentation of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Nov. 2020, Art. no. 4500317.
- [54] Y. Li, T. Shi, Y. Zhang, W. Chen, Z. Wang, and H. Li, "Learning deep semantic segmentation network under multiple weakly-supervised constraints for cross-domain remote sensing image semantic segmentation," *ISPRS J. Photogramm. Remote Sens.*, vol. 175, pp. 20–33, May 2021.
- [55] M. Luo and S. Ji, "Cross-spatiotemporal land-cover classification from VHR remote sensing images with deep learning based domain adaptation," *ISPRS J. Photogramm. Remote Sens.*, vol. 191, pp. 105–128, Sep. 2022.
- [56] 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 Sens.*, vol. 12, no. 3, p. 575, 2020.
- [57] 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*, Jun. 2020, pp. 56–57.
- [58] 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.
- [59] X. Li, M. Luo, S. Ji, L. Zhang, and M. Lu, "Evaluating generative adversarial networks based image-level domain transfer for multi-source remote sensing image segmentation and object detection," *Int. J. Remote Sens.*, vol. 41, no. 19, pp. 7343–7367, 2020.
- [60] H. Zuo, J. Lu, and G. Zhang, "Multiple-source domain adaptation in rule-based neural network," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–6.
- [61] C. Geiß, H. Schrade, P. A. Pelizari, and H. Taubenböck, "Multistrategy ensemble regression for mapping of built-up density and height with Sentinel-2 data," *ISPRS J. Photogramm. Remote Sens.*, vol. 170, pp. 57–71, Dec. 2020.
- [62] X. Lu, T. Gong, and X. Zheng, "Multisource compensation network for remote sensing cross-domain scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, Apr. 2019.
- [63] A. Elshamli, G. W. Taylor, and S. Areibi, "Multisource domain adaptation for remote sensing using deep neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, May 2019.
- [64] J. Zheng et al., "Unsupervised mixed multi-target domain adaptation for remote sensing images classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 1381–1384.
- [65] J. Zheng et al., "A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation ( $S^2M^2T$  DA) scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2021, Art. no. 5609213.
- [66] R. Adayel, Y. Bazi, H. Alhichri, and N. Alajlan, "Deep open-set domain adaptation for cross-scene classification based on adversarial learning and Pareto ranking," *Remote Sens.*, vol. 12, no. 11, p. 1716, May 2020.
- [67] J. Zhang, J. Liu, L. Shi, B. Pan, and X. Xu, "An open set domain adaptation network based on adversarial learning for remote sensing image scene classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 1365–1368.
- [68] J. Zheng, W. Wu, S. Yuan, H. Fu, W. Li, and L. Yu, "Multisource-domain generalization-based oil palm tree detection using very-high-resolution (VHR) satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.
- [69] J. Hu, H. Tuo, C. Wang, H. Zhong, H. Pan, and Z. Jing, "Unsupervised satellite image classification based on partial transfer learning," *Aerospace Syst.*, vol. 3, no. 1, pp. 21–28, Mar. 2020.
- [70] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst. (GIS)*, 2010, pp. 270–279.
- [71] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," in *Proc. 17th Int. Conf. Neural Inf. Process. Syst.*, 2004, pp. 529–536.
- [72] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 136–144.
- [73] H.-Y. Chen et al., "Complement objective training," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–11.
- [74] 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 J. Photogramm. Remote Sens.*, vol. 155, pp. 72–89, Sep. 2019.
- [75] W. Liu and F. Su, "A novel unsupervised adversarial domain adaptation network for remotely sensed scene classification," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6099–6116, Aug. 2020.
- [76] S. Zhu, B. Du, L. Zhang, and X. Li, "Attention-based multiscale residual adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2021, Art. no. 5400715.
- [77] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2010, pp. 213–226.
- [78] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 5018–5027.
- [79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
- [80] J. Zheng, W. Wu, Y. Zhao, and H. Fu, "TransresNet: Transferable ResNet for domain adaptation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 764–768.
- [81] J. Huang, D. Guan, A. Xiao, S. Lu, and L. Shao, "Category contrast for unsupervised domain adaptation in visual tasks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 1203–1214.
- [82] X. Gu et al., "Adversarial reweighting for partial domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 14860–14872.
- [83] C. Yang, Y.-M. Cheung, J. Ding, K. C. Tan, B. Xue, and M. Zhang, "Contrastive learning assisted-alignment for partial domain adaptation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 7, 2022, doi: [10.1109/TNNLS.2022.3145034](https://doi.org/10.1109/TNNLS.2022.3145034).

- [84] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 8024–8035.
- [85] O. Russakovsky et al., "Imagenet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
- [86] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proc. COMPSTAT*. Springer, 2010, pp. 177–186.
- [87] H.-Y. Chen et al., "Improving adversarial robustness via guided complement entropy," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jun. 2019, pp. 4881–4889.
- [88] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1406–1415.
- [89] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 1–27, 2008.
- [90] J. Donahue et al., "DeCAF: A deep convolutional activation feature for generic visual recognition," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 647–655.
- [91] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8050–8058.



**Juepeng Zheng** (Graduate Student Member, IEEE) received the bachelor's degree in surveying engineering from the College of Surveying and Geo-informatics, Tongji University, Shanghai, China, in 2019. He is currently pursuing the Ph.D. degree in ecological science with the Department of Earth System Science, Tsinghua University, Beijing, China.

His research interests lie in remote sensing image understanding, high-performance computing, and agricultural mapping.



**Yi Zhao** received the bachelor's degree in computer science from the School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu, China, in 2019. He is currently pursuing the Ph.D. degree in ecological science from the Department of Earth System Science, Tsinghua University, Beijing, China.

His research interests include deep learning framework optimization and transfer learning.



**Wenzhao Wu** received the master's degree from the Department of Earth System Science, Tsinghua University, Beijing, China, in 2020.

He is currently working at the National Supercomputing Center, Wuxi, China. His research interests include transfer learning for remote sensing and high-performance deep learning systems.



**Mengxuan Chen** is currently pursuing the bachelor's degree in mechanical engineering with the University of Michigan–Shanghai Jiaotong University Joint Institute, Shanghai, China.

Her research interests include deep learning algorithms and applications.



**Weijia Li** received the bachelor's degree from the Department of Computer Science, Sun Yat-Sen University, Guangzhou, China, in 2014, and the Ph.D. degree from the Department of Earth System Science, Tsinghua University, Beijing, China, in 2019.

From 2019 to 2021, she was a Post-Doctoral Researcher at CUHK-Sensetime Joint Laboratory, Department of Information Engineering, CUHK. She is an Associate Professor with the School of Geospatial Engineering and Science, Sun Yat-Sen University. Her research interests include remote sensing image understanding, computer vision, and deep learning.



**Haohuan Fu** (Member, IEEE) received the Ph.D. degree in computing from Imperial College London, London, U.K., in 2009.

He is currently a Professor with the Ministry of Education Key Laboratory for Earth System Modeling and the Department of Earth System Science, Tsinghua University, Beijing, China. He is also the Deputy Director of the National Supercomputing Center, Wuxi, China. His research interests include design methodologies for highly efficient and highly scalable simulation applications that can take advantage of emerging multicore, manycore, and reconfigurable architectures, and make full utilization of current Peta-Flops and future Exa-Flops supercomputers; and intelligent data management, analysis, and data mining platforms that combine the statistics methods and machine learning technologies.