An Open Set Domain Adaptation Algorithm via Exploring Transferability and Discriminability for Remote Sensing Image Scene Classification

Jun Zhang, Jiao Liu, Bin Pan, *Member, IEEE*, Zongqing Chen, Xia Xu, *Graduate Student Member, IEEE*, and Zhenwei Shi, *Member, IEEE*

Abstract—Remote sensing image scene classification aims to automatically assign semantic labels for remote sensing images. Recently, to overcome the distribution discrepancy of training data and test data, domain adaptation has been applied to remote sensing image scene classification. Most domain adaptation approaches usually explore transferability under the assumption that the source domain and target domain have common classes. However, in real applications, new categories may appear in the target domain. Besides, only considering the transferability will degrade the classification performance due to the strong interclass similarity of remote sensing images. In this article, we present an open set domain adaptation algorithm via exploring transferability and discriminability (OSDA-ETD) for remote sensing image scene classification. To be specific, we propose the transferability technology, which aims at the high interdomain variations and high intraclass diversity of remote sensing images. The purpose of transferability is to reduce the global distribution difference of domains and the local distribution discrepancy of the same classes in different domains. For high interclass similarity in remote sensing images, we adopt the discriminability strategy. The discriminability intends to enlarge the distribution discrepancy of different classes in different domains. To further promote the effectiveness of scene classification, we integrate the transferability and the discriminability into a framework. Moreover, we prove that the algorithm has a unique optimizer.

Index Terms—Discriminability, open set domain adaptation, remote sensing, scene classification, transferability.

NOMENCLATURE

$\mathcal{D}_s$ and $\mathcal{D}_t$	Source and target domains.
$\mathcal{Y}_s$ and $\mathcal{Y}_t$	Source and target label spaces.
$\mathbf{x}_{s_i}$ and $\mathbf{x}_{t_j}$	Source and target samples.
$\mathbf{x}_{t_j}^k$	Predicted known target samples.
n_s and n_t	Number of source or target samples.
n_t^k	Number of predicted known target samples.
n_s^c	Number of source samples with label c.
$n_t^{\hat{c}}$	Number of target samples with pseudolabel $\hat{c}$.
C	Number of known classes.
$C + 1$	Number of target domain classes.
Q_s	Source domain marginal distribution.
$Q_t^{\mathcal{Y}_s}$	Known class marginal distribution in target domain.
$Q_s^{\mathbf{x}_s \mid \mathbf{y}_c}$ and $Q_t^{\mathbf{x}_t \mid \mathbf{y}_{\hat{c}}}$	Source and target conditional distributions.
$\lambda, \alpha, \eta, \gamma, \sigma, \mu, \rho,$ and p	Parameters.
$\boldsymbol{\theta}$	Parameter.
K	Kernel.
$\mathbf{Y}, \tilde{\mathbf{Y}}$	Label matrix.

I. INTRODUCTION

THE purpose of remote sensing image scene classification is to assign each image with a specific semantic label, which has been widely applied to urban planning, environmental monitoring, and geographic image retrieval [1]–[3]. In recent years, many deep learning approaches have been proposed for scene classification, such as autoencoder- [4], [5], convolutional neural networks- [6]–[8], and generative adversarial networks-based methods [9]–[11].

The previous methods assume that the training and test data are drawn from the same distribution [12]. However, this assumption is hard to guarantee in remote sensing. Remote sensing images are taken under different conditions. Sensor types, resolutions, geographic locations, and other factors can

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
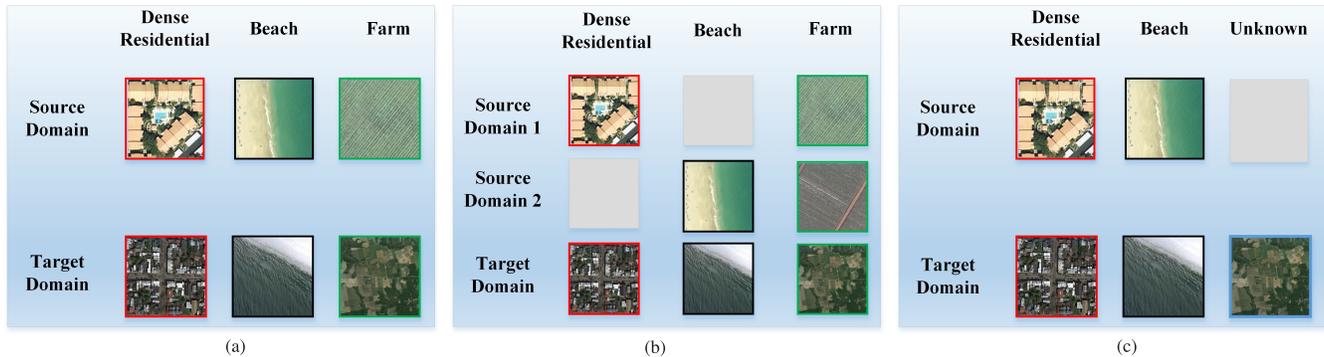

Fig. 1.   Different domain adaptation scenarios, where gray modules indicate that this class is absent in the domain. (a) It assumes that the source domain and the target domain have common categories, where $\mathcal{Y}_t$ denotes the target domain label space and $\mathcal{Y}_s$ denotes the source domain label space ($\mathcal{Y}_t = \mathcal{Y}_s$). (b) Target domain and the combination of multiple-source domains share the same categories, where $\mathcal{Y}_{sk}$ denotes the label space in the kth source domain ($\mathcal{Y}_t = \cup_{k=1}^{K} \mathcal{Y}_{sk}$). (c) Scenario of open set domain adaptation, and the target domain includes unknown classes ($\mathcal{Y}_t \supset \mathcal{Y}_s$).

cause distribution discrepancy between the training and test data [13], [14]. To deal with this problem, domain adaptation has been introduced, which aims to reduce the distribution discrepancy between the labeled data (source domain) and the unlabeled data (target domain) [15].

In remote sensing scene classification, based on the number of available source domains, domain adaptation can be roughly divided into the single-source domain and multi-source domain adaptation methods. The single-source domain adaptation means that there is one available source domain. In single-source domain adaptation, the source domain and target domain have common classes, but the data distribution is different, as shown in Fig. 1(a). The literature [16]–[20] designs domain adaptive algorithm to reduce the global or local distribution difference between domains. Recently, to improve the classification performance of the target domain, multisource domain adaptation has been proposed. In multi-source domain adaptation, there have multiple source domains. The literature [21] proposes a situation that multiple source domains and target domain have common classes and designs a multibranch neural network to learn invariant features between domains. To further deal with the case that the classes of target domain more than any source domain. Recently, the literature [22] utilizes multiple complementary source domains to combine the classes of the target domain, as shown in Fig. 1(b), and proposes the multisource compensation network to mitigates the domain shifts.

However, it still has the following two bottlenecks in the domain adaptation of scene classification.

1) It is difficult to ensure that the source domain and the target domain have common classes. The new classes will inevitably appear in the target domain.
2) The traditional domain adaptation methods reduce the global difference of interdomain or the local discrepancy of intraclass to improve transferability. When small interclass variations occur, only improving transferability may deteriorate the decision boundary.

To cope with the first challenges, a novel scenario of an open set domain adaptation has been proposed. As shown in Fig. 1(c), the target domain contains unknown classes

that do not present in the source domain. In the literature [23], the proposed semisupervised dual-dictionary nonnegative matrix factorization algorithm is extended to open set domain adaptation for hyperspectral image classification. In computer vision, the literature [24]–[26] achieves open set domain adaptation via recognizing unknown classes and improving the transferability of interdomain or intraclass. However, the low interclass discrepancy of remote sensing images determines that only considering the transferability will degrade the classification performance.

In this article, to address the abovementioned challenges, we present an open set domain adaptation algorithm via exploring transferability and discriminability (OSDA-ETD) for remote sensing image scene classification. As shown in Fig. 2, based on the remote sensing images characteristics of large interdomain variations, large intraclass diversity, and small interclass discrepancy, OSDA-ETD adopts two strategies. The first technology is transferability, which not only minimizes the global distribution discrepancy between domains but also reduces the local distribution difference of the same class in different domains. The second strategy is discriminability, which tries to maximize the distribution difference of different categories in different domains. Finally, we embed the transferability and discriminability into a framework and provide theoretical proof about the optimizer.

The main contributions of this article are summarized as follows.

1) We present a new open set domain adaptation algorithm called OSDA-ETD, which considers the remote sensing images characteristics of large interdomain variations, large intraclass diversity, and large interclass similarity.
2) Aiming at the remote sensing characteristics, we propose the transferability and discriminability strategy and prove that the OSDA-ETD has a unique optimizer.

The remainder of this article is structured as follows. In Section II, we introduce the proposed method in detail. In Section III, we conduct experiments on cross-domain datasets. Section IV concludes this article.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

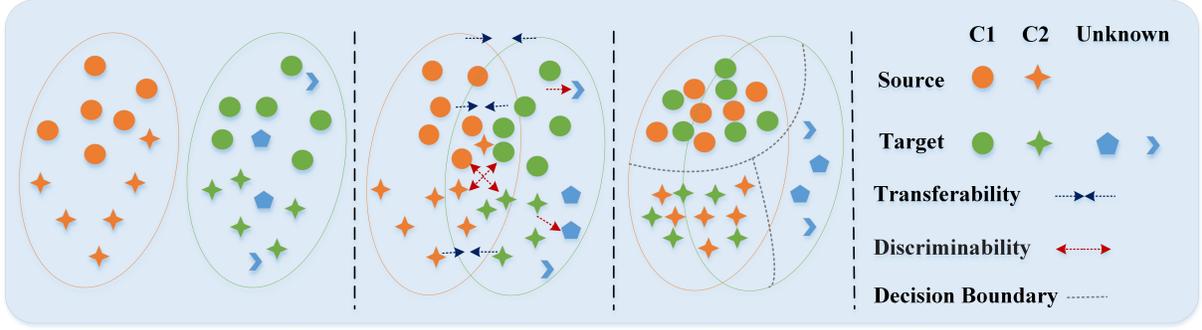ZHANG *et al.*: OSDA-ETD FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

3

Fig. 2. Motivation and strategy diagram of OSDA-ETD. In the left column, the source domain and the target domain show the characteristics of large interdomain variations, large intraclass diversity, and small interclass discrepancy. In the second column, the proposed method not only improves the transferability of interdomain and interclass but also enhances the discriminability of interclass. In the third column, with the proposed method, the same class can be well-aligned, and different classes are more discriminative. The right column represents the legend.

II. METHODOLOGY

A. Open Set Domain Adaptation

In open set domain adaptation, let us define $\mathcal{D}_s = \{\mathbf{x}_{s_i}, \mathbf{y}_{s_i}\}_{i=1}^{n_s}$ as the labeled source domain and $\mathcal{D}_t = \{\mathbf{x}_{t_j}\}_{j=1}^{n_t}$ as the unlabeled target domain. $\mathcal{Y}_s = \{\mathbf{y}_c\}_{c=1}^{C}$ denotes the label space of the source domain, and $\mathcal{Y}_t = \{\mathbf{y}_{\hat{c}}\}_{\hat{c}=1}^{C+1}$ denotes the label space of target domain. The relationship of label set between the source domain and target domain can be expressed as $\mathcal{Y}_s \subset \mathcal{Y}_t$. Generally speaking, the known classes refer to the shared classes between domains, while the unknown classes are the private classes of the target domain. In addition, the marginal distribution of source domain and target domain is Q_s and Q_t, respectively. In the standard domain adaptation, we have $Q_s \neq Q_t$. In open set domain adaptation, we further have $Q_s \neq Q_t^{\mathcal{Y}_s}$ and $Q_s^{\mathbf{x}_s|\mathbf{y}_c} \neq Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}$. In this article, $Q_t^{\mathcal{Y}_s}$ denotes the marginal distribution of the known classes in target domain, and $Q_s^{\mathbf{x}_s|\mathbf{y}_c}$ and $Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}$ represent the class conditional distribution of source domain and target domain, respectively.

Notably, some frequently used notations and their corresponding descriptions are listed in Nomenclature. Then, we will present the proposed approach.

B. General Framework of OSDA-ETD

In computer vision, Fang *et al.* [25] design distribution alignment with open difference (DAOD) to solve open set domain adaptation. However, DAOD cannot get advanced results for remote sensing images. The reason is that remote sensing images have the characteristic of high interdomain variations, high intraclass diversity, and low interclass discrepancy. In this article, we propose OSDA-ETD for remote sensing image scene classification. Peculiarly, OSDA-ETD tries to optimize four complementary objectives as follows.

1) In open set domain adaptation, the labeled source domain and the unlabeled target domain have different classes. To identify known classes, OSDA-ETD minimizes the source domain structure risk. For the unknown classes, based on DAOD, OSDA-ETD minimizes open set difference.
2) Aiming at the remote sensing images characteristic of high interdomain variations and high intraclass diversity, OSDA-ETD adopts the transferability strategy, which can reduce the global interdomain difference and the local intraclass discrepancy.

3) For the remote sensing images characteristic of a low interclass discrepancy, OSDA-ETD utilizes the discriminability strategy, which can enhance the interclass difference in different domains.
4) Maximizing the manifold consistency to extract the geometric relationship between domains.

The overall objective function of OSDA-ETD can be defined as follows:

$$
\begin{aligned}
\boldsymbol{g}^* &= \underset{\boldsymbol{g}\in\mathcal{H}}{\arg\min}\, R_s(\boldsymbol{g}) + \rho\|\boldsymbol{g}\|_k^2 + \gamma\, R_t^{u,C+1}(\boldsymbol{g}) - \sigma\, R_s^{u,C+1}(\boldsymbol{g}) \\
&\quad + \lambda(1-\alpha)D_{\boldsymbol{g},k}^t\left(Q_s, Q_t^{\mathcal{Y}_s}\right) + \lambda\alpha\sum_{\hat{c}=1}^{C}\sum_{c=\hat{c}}^{C} D_{\boldsymbol{g},k}^t\left(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}\right) \\
&\quad - \eta\sum_{\hat{c}=1}^{C}\sum_{c\neq\hat{c}} D_{\boldsymbol{g},k}^d\left(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}\right) + \mu R_m(\mathcal{D}_s, \mathcal{D}_t)
\end{aligned}
\tag{1}
$$

where $\mathcal{H}$ is the hypothesis space, $R_s(\boldsymbol{g}) + \rho\|\boldsymbol{g}\|_k^2$ is the term of source domain structural risk, $R_t^{u,C+1}(\boldsymbol{g})$ and $R_s^{u,C+1}(\boldsymbol{g})$ are the risks that the samples are regarded as the unknown classes, $\gamma\, R_t^{u,C+1}(\boldsymbol{g}) - \sigma\, R_s^{u,C+1}(\boldsymbol{g})$ is the open set difference, $D_{\boldsymbol{g},k}^t(Q_s, Q_t^{\mathcal{Y}_s})$ represents the distribution discrepancy of interdomain, $D_{\boldsymbol{g},k}^t(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}})$ denotes the distribution difference of same class between different domains, $D_{\boldsymbol{g},k}^d(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}})$ implies the distribution difference of different class between different domains, and $R_m(\mathcal{D}_s, \mathcal{D}_t)$ represents the term of manifold regularization. In addition, λ, μ, η, α, γ, ρ, and σ are the parameters.

According to the representer theorem [27], the minimizer $\boldsymbol{g}^*$ in (1) is defined as

$$
\boldsymbol{g}^*(\mathbf{x}) = \sum_{i=1}^{n_s+n_t} \boldsymbol{\theta}_i K(\mathbf{x}_i, \mathbf{x}) \quad \forall \mathbf{x}\in\mathcal{X}
\tag{2}
$$

where $\mathbf{x}_i \in \mathcal{D}_s \cup \mathcal{D}_t$, $\boldsymbol{\theta}_i \in \mathbb{R}^{(C+1)\times 1}$ is the parameter, and K is the kernel. With the form of $\boldsymbol{g}^*$, in Sections II-C–II-F, we will describe the computation of terms in (1).

C. Open Set Loss of OSDA-ETD

In open set domain adaptation, our ultimate goal is to learn an adaptative model for the target domain. However, the target domain is composed of known and unknown classes; the identification of the unknown class is much more difficult than

the known class. Fortunately, some incomplete multisource domain adaptation works [28], [29] consider the recognition of unknown categories in the target domain. Recently, based on the open set learning bound, DAOD enables the unknown target samples to be separated from samples using an open set difference. In our work, to identify the target domain samples as the corresponding known class and unknown class, following DAOD, we connect the structural risk of the source domain and the open set difference to construct the open set loss of OSDA-ETD. In this process, we utilize two label matrices. The first label matrix $\mathbf{Y} \in \mathbb{R}^{(C+1) \times (n_s+n_t)}$ can be defined as follows:

$$\mathbf{Y}_{ij} = \begin{cases} 1, & \mathbf{x}_j \in \mathcal{D}_{s,i} \\ 0, & \text{otherwise} \end{cases} \quad \text{when } i \leq C \quad (3)$$

$$\mathbf{Y}_{ij} = \begin{cases} 1, & \mathbf{x}_j \in \mathcal{D}_{t,C+1} \\ 0, & \text{otherwise} \end{cases} \quad \text{when } i = C+1 \quad (4)$$

where $D_{s,i}$ denotes the set of source samples in the ith class and $D_{t,C+1}$ denotes target samples.

The other label matrix $\tilde{\mathbf{Y}} \in \mathbb{R}^{(C+1) \times (n_s+n_t)}$ is

$$\tilde{\mathbf{Y}}_{ij} = 1 \quad \text{iff} \quad i = C+1 \quad \text{and} \quad \mathbf{x}_j \in \mathcal{D}_s, \text{ otherwise } \tilde{\mathbf{Y}}_{ij} = 0. \quad (5)$$

Based on the label matrices $\mathbf{Y}$ and $\tilde{\mathbf{Y}}$, the open set loss in OSDA-ETD can be defined as

$$R_s(\boldsymbol{g}) + \gamma R_t^{u,C+1}(\boldsymbol{g}) - \sigma R_s^{u,C+1}(\boldsymbol{g}) + \rho \|\boldsymbol{g}\|_k^2$$
$$= \|(\mathbf{Y} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\mathbf{B}\|_F^2 - \sigma \|(\tilde{\mathbf{Y}} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\tilde{\mathbf{B}}\|_F^2 + \rho \operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\boldsymbol{\theta}) \quad (6)$$

where $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n_s+n_t}]^{\mathrm{T}} \in \mathbb{R}^{(n_s+n_t) \times (C+1)}$, $\mathbf{B}$ and $\tilde{\mathbf{B}}$ are the $(n_s + n_t) \times (n_s + n_t)$ diagonal matrix, and $\mathbf{K}$ is a $(n_s + n_t) \times (n_s + n_t)$ kernel matrix. If $\mathbf{x}_i \in \mathcal{D}_s$, then $\mathbf{B}_{ii} = (1/n_s)^{1/2}$ and $\tilde{\mathbf{B}}_{ii} = (1/n_s)^{1/2}$. If $\mathbf{x}_i \in \mathcal{D}_t$, then $\mathbf{B}_{ii} = (\gamma/n_t)^{1/2}$ and $\tilde{\mathbf{B}}_{ii} = 0$.

Unfortunately, due to the distribution difference between the source domain and target domain, the model inferred by (6) cannot generalize well to the target domain. In the next section, we will address this issue.

D. Transferability in OSDA-ETD

In this section, following the literature [25], [30], we reduce the global interdomain difference and the local intraclass discrepancy. Generally, reducing the distribution differences denotes improving transferability. However, for open set domain adaptation, the target domain includes unknown classes. If we directly increase transferability, it will generate a negative transfer. Therefore, we enhance transferability after excluding recognized unknown target samples. In this process, we utilize the target pseudolabels predicted by classifiers. In particular, we rethink transferability that involves inter-domain and intraclass transfer. In terms of enhancing the interdomain transferability, we utilize the maximum mean discrepancy (MMD) criteria to measure the distribution distance between Q_s and $Q_t^{\mathcal{Y}_s}$, which can be achieved as follows:

$$D_{\boldsymbol{g},k}^t \left(Q_s, Q_t^{\mathcal{Y}_s} \right) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} \boldsymbol{g}(\mathbf{x}_{s_i}) - \frac{1}{n_t^k} \sum_{j=1}^{n_t^k} \boldsymbol{g}(\mathbf{x}_{t_j}^k) \right\|_{\mathcal{H}_K}^2$$
$$= \operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\mathbf{M_0}\mathbf{K}\boldsymbol{\theta}) \quad (7)$$

where n_t^k is the number of predicted known target samples, $\mathbf{x}_{t_j}^k$ represents the predicted known target samples, and $\mathbf{M_0}$ can be expressed as

$$\mathbf{M_0} = \begin{bmatrix} \mathbf{e}_s \mathbf{e}_s^{\mathrm{T}} & \mathbf{e}_s \mathbf{e}_t^{\mathrm{T}} \\ \mathbf{e}_t \mathbf{e}_s^{\mathrm{T}} & \mathbf{e}_t \mathbf{e}_t^{\mathrm{T}} \end{bmatrix} \quad (8)$$

where

$$\mathbf{e}_s = \frac{1}{n_s} \mathbf{I}_{n_s \times 1}^s, \quad \mathbf{e}_t = \frac{1}{n_t^k} \mathbf{I}_{n_t \times 1}^t. \quad (9)$$

In (9), all elements in $\mathbf{I}_{n_s \times 1}^s$ are one. For $\mathbf{I}_{n_t \times 1}^t$, only elements with known class are one, and the rest are zero.

For the intraclass transfer, we employ the classwise MMD to approximately measure $Q_s^{\mathbf{x}_s|\mathbf{y}_c}$ and $Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}$, and it can be defined as follows:

$$\sum_{\hat{c}=1}^{C} \sum_{c=\hat{c}}^{C} D_{\boldsymbol{g},k}^t \left(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}} \right)$$
$$= \sum_{\hat{c}=1}^{C} \sum_{c=\hat{c}}^{C} \left\| \frac{1}{n_s^c} \sum_{i=1}^{n_s^c} \boldsymbol{g}(\mathbf{x}_{s_i}^c) - \frac{1}{n_t^{\hat{c}}} \sum_{j=1}^{n_t^{\hat{c}}} \boldsymbol{g}(\mathbf{x}_{t_j}^{\hat{c}}) \right\|_{\mathcal{H}_K}^2 \quad (10)$$

where n_s^c refers to the number of source samples with label c, $n_t^{\hat{c}}$ denotes the count of target samples with pseudolabel $\hat{c}$, and $\mathbf{x}_{s_i}^c$ and $\mathbf{x}_{t_j}^{\hat{c}}$ are the samples of class c.

For the convenience of calculation, following the literature [31], we utilize the one-hot coding label matrix to calculate the transferability of intraclass. In particular, the source domain and predicted target domain one-hot coding label matrices are written as $\mathbf{Y}_s = [\mathbf{y}_{s,1}, \ldots, \mathbf{y}_{s,n_s}]$ and $\hat{\mathbf{Y}}_t = [\hat{\mathbf{y}}_{t,1}, \ldots, \hat{\mathbf{y}}_{t,n_t^k}]$, where $\mathbf{y}_{s,i} \in \mathbb{R}^{1 \times C}$ and $\hat{\mathbf{y}}_{t,i} \in \mathbb{R}^{1 \times C}$. Therefore, (10) can be expressed as

$$\sum_{\hat{c}=1}^{C} \sum_{c=\hat{c}}^{C} D_{\boldsymbol{g},k}^t \left(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}} \right) = \operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\mathbf{M_l}\mathbf{K}\boldsymbol{\theta}) \quad (11)$$

where

$$\mathbf{M_l} = \begin{bmatrix} \mathbf{A}_s \mathbf{A}_s^{\mathrm{T}} & -\mathbf{A}_s \mathbf{A}_t^{\mathrm{T}} \\ -\mathbf{A}_t \mathbf{A}_s^{\mathrm{T}} & \mathbf{A}_t \mathbf{A}_t^{\mathrm{T}} \end{bmatrix}. \quad (12)$$

In (12), $\mathbf{A}_s$ and $\mathbf{A}_t$ can be expressed as follows:

$$\mathbf{A}_s = \left[\frac{\mathbf{Y}_s(:, 1)}{n_s^1}, \ldots, \frac{\mathbf{Y}_s(:, C)}{n_s^C} \right] \quad (13)$$

$$\mathbf{A}_t = \left[\frac{\hat{\mathbf{Y}}_t(:, 1)}{n_t^1}, \ldots, \frac{\hat{\mathbf{Y}}_t(:, C)}{n_t^C} \right] \quad (14)$$

where $\mathbf{Y}_s(:, c)$ and $\hat{\mathbf{Y}}_t(:, c)$ denote the cth column of $\mathbf{Y}_s$ and $\hat{\mathbf{Y}}_t(:, c)$, respectively.

In this section, we consider the transferability from both global and local perspectives. However, the remote sensing image characteristics of high interclass similarity determine that only considering the transferability will degrade the classification performance. In the next section, we address this problem.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: OSDA-ETD FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

5

E. Discriminability in OSDA-ETD

In this section, to improve the classification performance, we explore the discriminability based on the remote sensing scene images peculiarity of large interclass similarity. To achieve discriminability, we leverage classwise MMD to enlarge the distance of different known class, which can be expressed as

$$
\sum_{\hat{c}=1}^{C}\sum_{c\neq\hat{c}} D_{g,k}^{d}\left(Q_s^{\mathbf{x}_s|\mathbf{y}_c}, Q_t^{\mathbf{x}_t|\mathbf{y}_{\hat{c}}}\right)
$$

$$
= \sum_{\hat{c}=1}^{C}\sum_{c\neq\hat{c}}\left\|\frac{1}{n_s^c}\sum_{i=1}^{n_s^c}\mathbf{g}\left(\mathbf{x}_{s_i}^c\right)-\frac{1}{n_t^{\hat{c}}}\sum_{j=1}^{n_t^{\hat{c}}}\mathbf{g}\left(\mathbf{x}_{t_j}^{\hat{c}}\right)\right\|_{\mathcal{H}_K}^2
$$

$$
= \operatorname{tr}\left(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\mathbf{F}\mathbf{K}\boldsymbol{\theta}\right) \tag{15}
$$

where

$$
\mathbf{F} = \begin{bmatrix} \mathbf{F}_s\mathbf{F}_s^{\mathrm{T}} & -\mathbf{F}_s\mathbf{F}_t^{\mathrm{T}} \\ -\mathbf{F}_t\mathbf{F}_s^{\mathrm{T}} & \mathbf{F}_t\mathbf{F}_t^{\mathrm{T}} \end{bmatrix}. \tag{16}
$$

In (16), $\mathbf{F}_s$ and $\mathbf{F}_t$ can be defined as

$$
\mathbf{F}_s = \left[\mathbf{A}_s(:,1)\bullet(C-1),\ldots,\mathbf{A}_s(:,C)\bullet(C-1)\right] \tag{17}
$$

$$
\mathbf{F}_t = \left[\mathbf{A}_t(:,1:C)_{\hat{c}\neq1},\ldots,\mathbf{A}_t(:,1:C)_{\hat{c}\neq C}\right] \tag{18}
$$

where $\mathbf{A}_s(:,1)$ represents the 1st column of $\mathbf{A}_s$, $\mathbf{A}_s(:,1)\bullet(C-1)$ repeats $\mathbf{A}_s(:,1)$ for $C-1$ times, and $\mathbf{A}_t(:,1:C)_{\hat{c}\neq1}$ is composed of the 1st to the Cth (except the first) columns of $\mathbf{A}_t$.

F. Manifold Regularization

To learn the geometrical relation between the source domain and the target domain, following the literature [25], [30], [32], OSDA-ETD employs the manifold regularization, which can be computed as follows:

$$
R_m(\mathcal{D}_s, \mathcal{D}_t) = \sum_{i,j=1}^{n_s+n_t}\|\mathbf{g}(\mathbf{x}_i)-\mathbf{g}(\mathbf{x}_j)\|_2^2\mathbf{W}_{ij}
$$

$$
= \sum_{i,j=1}^{n_s+n_t}\mathbf{g}(\mathbf{x}_i)\mathbf{L}_{ij}\mathbf{g}(\mathbf{x}_j) \tag{19}
$$

where $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{D}_s \cup \mathcal{D}_t$, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix, $\mathbf{D}_{ii} = \sum_{j=1}^{n_s+n_t}\mathbf{W}_{ij}$ is the diagonal matrix, and $\mathbf{W}$ is the pairwise affinity matrix. The pairwise affinity matrix estimates the similarity between samples, and it can be defined as follows:

$$
\mathbf{W}_{ij} = \begin{cases} \operatorname{sim}(\mathbf{x}_i, \mathbf{x}_j), & \mathbf{x}_i \in \mathcal{N}_p(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_p(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{20}
$$

where $\operatorname{sim}(\mathbf{x}_i, \mathbf{x}_j)$ represents the similarity function, $\mathcal{N}_p(\mathbf{x}_i)$ denotes the set of p-nearest neighbors to point $\mathbf{x}_i$, and p is the parameter.

G. Overall Reformulation of OSDA-ETD

Finally, we combine (6), (7), (11), (15), and (19) to reformulate the optimization problem in (1), which can be achieved as follows:

$$
\boldsymbol{\theta}^* = \operatorname*{argmin}_{\boldsymbol{\theta}\in\mathbb{R}^{(n_s+n_t)\times(C+1)}} \mathcal{L}(\boldsymbol{\theta}) \tag{21}
$$

where

$$
\mathcal{L}(\boldsymbol{\theta}) = \|(\mathbf{Y}-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\mathbf{B}\|_F^2 - \sigma\|(\tilde{\mathbf{Y}}-\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\tilde{\mathbf{B}}\|_F^2
$$
$$
+ \operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}(\lambda(1-\alpha)\mathbf{M_0}+\lambda\alpha\mathbf{M_l}-\eta\mathbf{F}+\mu\mathbf{L})\mathbf{K}\boldsymbol{\theta})
$$
$$
+ \rho\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\boldsymbol{\theta}). \tag{22}
$$

There are some negative term in $\mathcal{L}(\boldsymbol{\theta})$. Therefore, the optimizer by solving $(\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) = \mathbf{0}$ may be a maximum point or a saddle point. Excitingly, based on the literature [25], we prove that there exists a unique optimizer in $\mathcal{L}(\boldsymbol{\theta})$. It is worth noting that the optimize can be solved by $(\partial\mathcal{L}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}) = \mathbf{0}$. In the next section, we discuss in detail.

H. Theoretical Proof of the Unique Optimizer

In this section, we provide the theoretical proof for the optimizer of $\mathcal{L}(\boldsymbol{\theta})$, as shown in Lemma 1.

Lemma 1: If the parameters $\sigma < 1$, $\eta \leq (\lambda\alpha/C - 1)$, and the kernel function K is universal, then the $\mathcal{L}(\boldsymbol{\theta})$ has a unique optimizer. It can be written as follows:

$$
\boldsymbol{\theta} = \left((\mathbf{B}^2 - \sigma\tilde{\mathbf{B}}^2 + \lambda(1-\alpha)\mathbf{M_0}+\lambda\alpha\mathbf{M_l}-\eta\mathbf{F}+\mu\mathbf{L})\mathbf{K}+\rho\mathbf{I}\right)^{-1}
$$
$$
\left(\mathbf{B}^2\mathbf{Y}^{\mathrm{T}}-\sigma\tilde{\mathbf{B}}^2\tilde{\mathbf{Y}}^{\mathrm{T}}\right). \tag{23}
$$

According to the literature [25], we prove Lemma 1 from three parts.

Claim 1:

$$
\lim_{\|\boldsymbol{\theta}\|_{\ell^2}\to+\infty} \mathcal{L}(\boldsymbol{\theta}) = +\infty. \tag{24}
$$

In (22), it is obvious that

$$
\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}(\lambda(1-\alpha)\mathbf{M_0}+\lambda\alpha\mathbf{M_l}+\mu\mathbf{L})\mathbf{K}\boldsymbol{\theta})+\rho\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\boldsymbol{\theta}) \geq 0. \tag{25}
$$

Then, we assume that the following equation holds:

$$
\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}(\lambda\alpha\mathbf{M_l} - \eta\mathbf{F})\mathbf{K}\boldsymbol{\theta}) \geq 0. \tag{26}
$$

To verify the hypothesis of (26), we directly discuss the relationship between matrix $\lambda\mu\mathbf{M_l}$ and $\eta\mathbf{F}$; the details are given as follows:

$$
\lambda\alpha\mathbf{A}_s\mathbf{A}_s^{\mathrm{T}} - \eta\mathbf{F}_s\mathbf{F}_s^{\mathrm{T}} = \lambda\alpha\mathbf{A}_s\mathbf{A}_s^{\mathrm{T}} - \eta(C-1)\,\mathbf{A}_s\mathbf{A}_s^{\mathrm{T}} \tag{27}
$$

$$
\lambda\alpha\mathbf{A}_t\mathbf{A}_t^{\mathrm{T}} - \eta\mathbf{F}_t\mathbf{F}_t^{\mathrm{T}} = \lambda\alpha\mathbf{A}_t\mathbf{A}_t^{\mathrm{T}} - \eta(C-1)\mathbf{A}_t\mathbf{A}_t^{\mathrm{T}} \tag{28}
$$

$$
-\lambda\alpha\mathbf{A}_s\mathbf{A}_t^{\mathrm{T}} - \left(-\eta\mathbf{F}_s\mathbf{F}_t^{\mathrm{T}}\right) > -\lambda\alpha\mathbf{A}_s\mathbf{A}_t^{\mathrm{T}} \tag{29}
$$

$$
-\lambda\alpha\mathbf{A}_t\mathbf{A}_s^{\mathrm{T}} - \left(-\eta\mathbf{F}_t\mathbf{F}_s^{\mathrm{T}}\right) > -\lambda\alpha\mathbf{A}_t\mathbf{A}_s^{\mathrm{T}}. \tag{30}
$$

From (27) to (30), we observe that, if $\eta \leq (\lambda\alpha/C - 1)$, then (26) holds.

By combining (25) and (26), we can get the following formula:

$$
\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}(\lambda(1-\alpha)\mathbf{M_0}+\lambda\alpha\mathbf{M_l}-\eta\mathbf{F}+\mu\mathbf{L})\mathbf{K}\boldsymbol{\theta})+\rho\operatorname{tr}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{K}\boldsymbol{\theta}) \geq 0. \tag{31}
$$

Next, we consider the partial term of open set loss function

$$
\|(\mathbf{Y} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\mathbf{B}\|_F^2 - \sigma\|(\tilde{\mathbf{Y}} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\tilde{\mathbf{B}}\|_F^2. \tag{32}
$$

Since the kernel $\mathbf{K}$ is universal and $\sigma < 1$, the matrix $\mathbf{K}(\mathbf{B}^2 - \sigma\tilde{\mathbf{B}})\mathbf{K}$ is symmetric and positive definite. It can be achieved as follows:

$$
\mathbf{K}(\mathbf{B}^2 - \gamma\tilde{\mathbf{B}}^2)\mathbf{K} = \mathbf{O}\boldsymbol{\Lambda}\mathbf{O}^{\mathrm{T}} \tag{33}
$$

where $\mathbf{O}$ is the orthogonal matrix and $\mathbf{\Lambda}$ is the diagonal matrix. Then,

$$
\begin{aligned}
\|(\mathbf{Y} &- \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\mathbf{B}\|_F^2 - \sigma\|(\tilde{\mathbf{Y}} - \boldsymbol{\theta}^{\mathrm{T}}\mathbf{K})\widetilde{\mathbf{B}}\|_F^2 \\
&= \mathrm{tr}[(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{O})\mathbf{\Lambda}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{O})^{\mathrm{T}}] - 2\,\mathrm{tr}[(\mathbf{Y}\mathbf{B}^2\mathbf{K} - \sigma\tilde{\mathbf{Y}}\tilde{\mathbf{B}}^2\mathbf{K})\boldsymbol{\theta}] + m \\
&\geq v\,\mathrm{tr}[(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{O})\mathbf{I}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{O})^{\mathrm{T}}] - O(\|\boldsymbol{\theta}\|_{\ell^2}) + m \\
&= v\,\mathrm{tr}[\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{\theta}] - O(\|\boldsymbol{\theta}\|_{\ell^2}) + m \qquad (34)
\end{aligned}
$$

where v is the smallest diagonal element of the diagonal matrix $\mathbf{\Lambda}$, and m is the constant.

Therefore,

$$
\begin{aligned}
\lim_{\|\boldsymbol{\theta}\|_{\ell^2} \to +\infty} \mathcal{L}(\boldsymbol{\theta}) &\geq \lim_{\|\boldsymbol{\theta}\|_{\ell^2} \to +\infty} v\|\boldsymbol{\theta}\|_{\ell^2}^2 - O(\|\boldsymbol{\theta}\|_{\ell^2}) + m \\
&= +\infty. \qquad (35)
\end{aligned}
$$

Claim 2: There exists optimizers for $\mathcal{L}(\boldsymbol{\theta})$.

In Claim 1, (24) indicates that there have a constant $r > 0$ such that $\mathcal{L}(\boldsymbol{\theta}) > \mathcal{L}(0)$ for any $\boldsymbol{\theta} \in \mathbb{R}^{(n_s+n_t) \times (C+1)} \setminus B_r(0)$, where $B_r(0)$ is an open ball, r denotes the radius of open ball, and the center of open ball is $\mathbf{0}$.

As $\mathcal{L}$ is a continuous function and the closed ball $\bar{B}_r(0)$ is a compact set, hence, there exist optimizers for $\mathcal{L}$, and these points belong to the $B_r(0)$.

Claim 3: The optimizer of $\mathcal{L}(\boldsymbol{\theta})$ is unique.

If a point $\boldsymbol{\theta_0}$ is a minimizer, then

$$
\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta_0}} = 0. \qquad (36)
$$

If the solution of (36) is unique, then the solution is the unique minimizer. The calculation of (36) can be achieved as follows:

$$
\begin{aligned}
-2\big(\mathbf{K}\mathbf{B}^2\mathbf{Y}^{\mathrm{T}} &- \sigma\mathbf{K}\tilde{\mathbf{B}}^2\tilde{\mathbf{Y}}^{\mathrm{T}}\big) + 2\rho\mathbf{K}\boldsymbol{\theta} + 2\big(\mathbf{K}\mathbf{B}^2\mathbf{K} - \sigma\mathbf{K}\tilde{\mathbf{B}}^2\mathbf{K}\big)\boldsymbol{\theta} \\
&+ 2\mathbf{K}(\lambda(1-\alpha)\mathbf{M_0} + \lambda\alpha\mathbf{M_l} - \eta\mathbf{F} + \mu\mathbf{L})\mathbf{K}\boldsymbol{\theta} = 0. \quad (37)
\end{aligned}
$$

The solution of (37) can be written as in (23).

Based on Claims 1–3, we illustrate that $\mathcal{L}(\boldsymbol{\theta})$ has a unique optimizer. The pseudocode of OSDA-ETD is provided in Algorithm 1.

III. EXPERIMENTS

In this section, we evaluate the effectiveness of OSDA-ETD. First, we simply introduce the datasets. Next, we describe the experimental setup. After that, we show the experimental results. Finally, we discuss parameter sensitivity. Our code has been released.

A. Datasets

To verify the performance of OSDA-ETD, we utilize the UC Merced dataset, the AID dataset, and the NWPU-RESISC45 dataset to construct the cross-domain datasets. The details of these datasets are given as follows.

1) *UC Merced [33]:* The UC Merced dataset can be downloaded from the United States Geological Survey (USGS) National Map. This dataset is composed of 21 land-use scene categories. Each class contains 100 images with a spatial resolution of 0.3 m and a size of 256×256 pixels.

2) *AID [34]:* The AID dataset is collected from Google Earth imagery and has a number of 10 000 images with

---

Algorithm 1 OSDA-ETD Algorithm

Input:
  Data: source domain data matrix $\mathbf{X}_s$, target domain data matrix $\mathbf{X}_t$, source domain label $\mathbf{Y}_s$;
  Parameter: λ, σ, ρ, α, γ, μ, η, p, and iterations T;
1: Train the classifier OSNN$^{\mathrm{cv}}$ utilizing $\mathbf{X}_s$ and $\mathbf{Y}_s$, then apply prediction on $\mathbf{X}_t$ to get the initial pseudo label $\tilde{\mathbf{Y}}_t$;
2: Calculate Laplacian matrix $\mathbf{L}$;
3: Choose kernel function and compute kernel matrix $\mathbf{K}$;
4: $i = 1$;
5: **while** $i < T + 1$ **do**
6:   Construct discriminability matrix $\mathbf{F}$ by Equations (13), (14), (16), (17) and (18);
7:   Compute $\mathbf{M_1}$ by Equations (12), (13), (14);
8:   Calculate $\mathbf{M_0}$ by Equations (8), (9);
9:   Update the pseudo labels by $\tilde{\mathbf{Y}}_t = \boldsymbol{\theta}^T K$;
10:   $i = i + 1$
11: **end while**
Output:
  Predicted target labels $\tilde{\mathbf{Y}}_t$ and classifier $\boldsymbol{\theta}^T K$.

TABLE I
COMMON CLASSES EXTRACTED FROM THREE DATASETS

Class	Data set		
	UC Merced	AID	NWPU-RESISC45
Airfield	100	360	1400
Anchorage	100	380	700
Beach	100	400	700
Dense Residential	100	410	700
Farm	100	370	1400
Flyover	100	420	700
Forest	100	250	700
Game Space	100	660	1400
Parking Space	100	390	700
River	100	410	700
Sparse Residential	100	300	700
Storage Cisterns	100	360	700

a size of 600×600 pixels. These images are divided into 30 classes. In addition, the spatial resolution of the image ranges from 8 m to about half a meter.

3) *NWPU-RESISC45 [35]:* The NWPU-RESISC45 dataset includes 45 scene categories. Each class contains 700 images with a size of 256×256 pixels. The spatial resolution ranges from about 30 to 0.2 m.

In the above three heterogeneous datasets, the same class may have different names as they are labeled by different experts. Hence, following the literature [22], [36], we select 12 public classes from the three datasets to construct the cross-domain datasets. This means that there have six domain adaptation tasks: UC Merced → AID, UC Merced → NWPU-RESISC45, AID → UC Merced, AID → NWPU-RESISC45, NWPU-RESISC45 → UC Merced, and NWPU-RESISC45 → AID. The category information of cross-domain datasets is shown in Table I, and the sample images of cross-domain datasets are shown in Fig. 3. In addition, we make an open set protocol for the cross-domain scene datasets, which chooses

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: OSDA-ETD FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

7

Fig. 3. Sample images of 12 common classes extracted from three datasets. Each column represents the corresponding categories of these datasets, from left to right are airfield, anchorage, beach, dense residential, farm, flyover, forest, game space, parking space, river, sparse residential, and storage cisterns. The top row to the bottom row are from the UC Merced dataset, the AID dataset, and the NWPU-RESISC45 dataset.

TABLE II

CLASSIFICATION ACCURACY OF THE AID DATASET AS SOURCE DOMAIN (%)

Data set	Class	Methods						
		OSNN$^{\text{CV}}$	TCA	SA	JDA	JPDA	DAOD	OSDA-ETD
AID→UC Merced	Airfield	57.00	48.00	53.00	44.00	**80.00**	48.00	69.00
	Anchorage	55.00	49.00	55.00	61.00	**65.00**	55.00	55.00
	Beach	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
	Dense Residential	30.00	29.00	36.00	32.00	41.00	45.00	**63.00**
	Farm	23.00	25.00	31.00	43.00	**80.00**	32.00	45.00
	Flyover	2.00	4.00	8.00	**58.00**	0.00	7.00	7.00
	Forest	98.00	94.00	98.00	**99.00**	98.00	94.00	98.00
	Game Space	78.00	66.00	75.00	72.00	79.00	91.00	**92.00**
	Parking Space	98.00	98.00	94.00	98.00	98.00	**100.00**	**100.00**
	Unknown	46.67	48.00	49.00	43.00	42.67	**75.67**	70.00
	OS*	60.11	57.00	61.11	67.44	**71.22**	63.56	69.89
	OS	58.77	56.10	59.90	65.00	68.37	64.77	**69.90**
	ALL*	60.11	57.00	61.11	67.44	**71.22**	63.56	69.89
	ALL	56.75	54.75	58.08	61.33	64.08	66.58	**69.92**
AID→NWPU-RESISC45	Airfield	28.29	38.21	32.07	56.79	58.29	68.93	**76.29**
	Anchorage	81.29	82.57	81.71	81.29	85.29	**98.29**	98.00
	Beach	68.29	73.14	69.71	85.57	86.00	95.86	**97.00**
	Dense Residential	75.71	77.71	69.57	78.29	85.86	94.43	**96.00**
	Farm	71.00	66.29	67.36	79.29	74.79	80.36	**92.00**
	Flyover	50.71	56.71	53.29	90.43	86.00	97.43	**97.71**
	Forest	91.00	88.14	97.86	98.43	97.86	96.86	**99.29**
	Game Space	77.07	67.50	73.29	78.07	77.21	**95.79**	**95.79**
	Parking Space	64.29	69.57	65.86	76.43	74.57	87.57	**89.43**
	Unknown	53.19	56.19	51.19	46.86	39.86	**94.05**	82.19
	OS*	67.52	68.87	67.86	80.51	80.65	90.61	**93.50**
	OS	66.08	68.87	66.19	77.14	76.57	90.95	**92.37**
	ALL*	65.33	65.99	65.29	78.23	78.01	88.38	**92.13**
	ALL	62.90	64.03	62.47	71.95	70.38	89.51	**90.14**

the top nine categories as the shared classes in alphabetical order and the last three classes as unknown classes in the target domain.

B. Experimental Setup

1) Evaluation Metrics: To evaluate the proposed method, we adopt known classes average accuracy (OS*), all classes average accuracy (OS), known classes overall accuracy (ALL*), all classes overall accuracy (ALL), and classes accuracy (CA) as the evaluation criteria based on former works [24], [37] and reality. The details are given as follows:

$$\text{ALL*} = \frac{|\mathbf{x}: \ \mathbf{x} \in \mathcal{D}_{tk} \wedge \hat{\mathbf{g}}(\mathbf{x}) = y_k|}{|(\mathbf{x}, y_k): (\mathbf{x}, y_k) \in \mathcal{D}_{tk}|} \tag{38}$$

$$\text{ALL} = \frac{|\mathbf{x}: \mathbf{x} \in \mathcal{D}_t \wedge \hat{\mathbf{g}}(\mathbf{x}) = y|}{|(\mathbf{x}, y): (\mathbf{x}, y) \in \mathcal{D}_t|} \tag{39}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

8

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

TABLE III

CLASSIFICATION ACCURACY OF THE UC MERCED DATASET AS SOURCE DOMAIN (%)

Data set	Class	Methods						
		OSNN$^{\mathrm{CV}}$	TCA	SA	JDA	JPDA	DAOD	OSDA-ETD
UC Merced→AID	Airfield	11.67	13.61	13.89	13.61	**34.72**	2.78	4.44
	Anchorage	12.37	29.21	14.21	61.58	7.63	51.58	**63.16**
	Beach	89.75	84.75	93.25	91.75	93.25	98.50	**99.50**
	Dense Residential	33.90	28.78	40.24	34.39	68.54	87.56	**91.71**
	Farm	4.86	6.49	16.22	12.16	**86.76**	8.92	14.86
	Flyover	32.62	36.43	15.48	**75.71**	4.05	0.00	0.00
	Forest	88.00	82.80	94.00	92.80	97.60	**99.20**	**99.20**
	Game Space	69.85	52.27	**76.36**	60.30	68.18	55.76	64.55
	Parking Space	87.95	83.59	78.46	89.74	89.74	97.44	**98.21**
	Unknown	48.50	**51.12**	44.95	34.95	20.75	48.60	39.16
	OS*	47.89	46.44	49.12	59.12	**61.16**	55.75	59.51
	OS	47.95	46.91	48.71	56.70	57.12	55.03	**57.48**
	ALL*	48.52	45.93	49.78	58.63	**60.16**	54.62	58.85
	ALL	48.51	47.11	48.68	53.25	51.21	53.25	**54.37**
UC Merced→NWPU-RESISC45	Airfield	33.07	33.36	29.79	**45.64**	34.36	38.50	40.36
	Anchorage	50.67	56.57	38.86	33.14	44.57	95.86	**97.14**
	Beach	68.71	71.14	72.43	92.43	96.71	95.86	**97.43**
	Dense Residential	67.71	58.57	56.86	68.86	75.00	86.29	**89.71**
	Farm	3.14	3.86	8.50	43.93	**78.43**	37.79	39.43
	Flyover	72.73	74.71	60.00	91.71	**92.57**	81.57	91.57
	Forest	97.14	89.43	99.00	99.00	**99.57**	98.86	99.43
	Game Space	60.14	52.14	**78.57**	60.14	58.71	70.57	75.29
	Parking Space	46.14	43.43	38.57	78.71	46.71	89.00	**90.71**
	Unknown	44.38	50.05	39.10	31.57	29.57	**92.00**	85.95
	OS*	55.67	53.69	53.62	68.17	69.63	77.14	**80.12**
	OS	54.55	53.33	52.17	64.51	65.62	78.63	**80.70**
	ALL*	49.79	48.18	49.95	63.61	66.51	70.10	**73.01**
	ALL	48.70	47.71	47.78	57.20	59.12	74.48	**75.60**

$$\text{CA} = \frac{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i \wedge \hat{\boldsymbol{g}}(\mathbf{x}) = i\right|}{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i\right|} \qquad (40)$$

$$\text{OS*} = \frac{1}{C} \sum_{i=1}^{C} \frac{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i \wedge \hat{\boldsymbol{g}}(\mathbf{x}) = i\right|}{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i\right|} \qquad (41)$$

$$\text{OS} = \frac{1}{C+1} \sum_{i=1}^{C+1} \frac{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i \wedge \hat{\boldsymbol{g}}(\mathbf{x}) = i\right|}{\left|\mathbf{x} : \mathbf{x} \in \mathcal{D}_t^i\right|} \qquad (42)$$

where $\mathcal{D}_{tk}$ denotes the known samples in the target domain and $\mathcal{D}_t^i$ is the target samples of the ith class.

2) Implementation Detail: To analyze the classification performance on the cross-domain dataset, we leverage ResNet50 [38] to extract deep features from each image scene. In this article, there have several parameters to set. As suggested in [30], we choose the Gaussian kernel as the kernel function. To balance the positive term $R_{u,C+1}^t(\boldsymbol{g})$ and the negative term $R_{u,C+1}^s(\boldsymbol{g})$ in the open set difference, we follow the rule of DAOD. In general, we set $\gamma = 0.4$ and $\sigma \in [0, 0.4]$. As for the other parameters $\lambda, \rho, p, \mu,$ and T, we also follow DAOD. Different from DAOD, OSDA-ETD considers the discriminability of interclass. Therefore, we set

that the discriminability parameter is searched in $\eta \in [1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0]$.

3) Compared Methods: We compare the performance of OSDA-ETD with the following methods.

1) *Open Set Nearest Neighbor Class Verification (OSNN$^{\mathrm{CV}}$) [39]:* The OSNN$^{\mathrm{CV}}$ is an open set classifier, which identifies unknown classes and known classes based on the labels consistency of test samples and two nearest neighbor samples.

2) *Transfer Component Analysis (TCA) [40] +OSNN$^{\mathrm{CV}}$:* TCA aligns the global distribution discrepancy between domains by using MMD, which is designed for closed set domain adaptation. In this article, we only replace classifiers.

3) *Subspace Alignment [41] + OSNN$^{\mathrm{CV}}$:* SA aligns the subspace between domains by learning a mapping function. To implement SA in open set domain adaptation, we utilize OSNN$^{\mathrm{CV}}$ as the classifier.

4) *Joint Distribution Adaptation (JDA) [42] + OSNN$^{\mathrm{CV}}$:* We extend JDA to the open set setting and try to reduce distribution discrepancy of known samples predicted by OSNN$^{\mathrm{CV}}$.

5) *Joint Probability Domain Adaptation (JPDA) [43]+OSNN$^{\mathrm{CV}}$:* JPDA leverage discriminative joint

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: OSDA-ETD FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

9

TABLE IV
CLASSIFICATION ACCURACY OF THE NWPU-RESISC45 DATASET AS SOURCE DOMAIN (%)

Data set	Class	Methods						
		OSNN^{CV}	TCA	SA	JDA	JPDA	DAOD	OSDA-ETD
NWPU-RESISC45→AID	Airfield	55.00	59.44	60.00	63.33	68.61	89.44	**94.44**
	Anchorage	60.00	61.05	56.84	74.47	65.26	59.21	**82.63**
	Beach	90.75	88.50	93.75	90.25	90.00	92.50	**99.50**
	Dense Residential	75.85	75.61	80.49	81.22	83.17	84.88	**93.17**
	Farm	71.62	75.68	74.59	75.95	77.84	96.76	**98.11**
	Flyover	87.38	86.43	90.24	91.19	83.57	94.05	**98.33**
	Forest	94.40	90.00	97.20	97.20	98.00	97.60	**100.00**
	Game Space	57.73	62.42	61.36	76.36	**84.55**	70.30	76.67
	Parking Space	97.69	94.87	96.92	95.13	94.36	97.44	**98.97**
	Unknown	48.41	47.85	43.46	43.83	38.13	**84.49**	73.64
	OS*	76.71	77.33	79.04	82.79	82.82	86.91	**93.54**
	OS	73.88	74.39	75.49	78.89	78.35	86.67	**91.55**
	ALL*	75.00	75.96	77.42	82.06	82.58	85.33	**92.09**
	ALL	68.96	69.58	69.70	73.38	72.48	85.14	**87.90**
NWPU-RESISC45→UC Merced	Airfield	**98.00**	96.00	**98.00**	90.00	97.00	97.00	97.00
	Anchorage	91.00	91.00	92.00	**94.00**	**94.00**	82.00	82.00
	Beach	90.00	91.00	96.00	93.00	**99.00**	96.00	96.00
	Dense Residential	41.00	41.00	38.00	**58.00**	32.00	29.00	31.00
	Farm	55.00	56.00	52.00	46.00	**59.00**	33.00	34.00
	Flyover	95.00	93.00	90.00	**97.00**	88.00	89.00	89.00
	Forest	96.00	95.00	95.00	**98.00**	92.00	90.00	91.00
	Game Space	26.00	38.00	39.00	49.00	**58.00**	37.00	37.00
	Parking Space	**98.00**	96.00	94.00	96.00	89.00	94.00	94.00
	Unknown	43.00	44.33	42.67	40.67	42.00	**72.67**	71.33
	OS*	76.67	77.44	77.11	**80.11**	78.67	71.89	72.33
	OS	73.30	74.13	73.67	**76.17**	75.00	71.97	72.23
	ALL*	76.67	77.44	77.11	**80.11**	78.67	71.89	72.33
	ALL	68.25	69.17	68.50	70.25	69.50	**72.08**	**72.08**

probability MMD to execute closed set domain adaptation. For the convenience of comparison, we also extend JPDA into the open set setting.

6) *DAOD+OSNN^{cv}:* A baseline of OSDA-ETD.

C. Experimental Results

1) Results of AID as Source Domain: Table II summarizes the results of AID → UC Merced and AID → NWPU-RESISC45, where the highest accuracy is boldfaced. In Table II, we can discover that SA and TCA perform poorly, even worse than the standard classifier OSNN^{cv}. The reason is that TCA and SA are designed for closed set domain adaptation. If we directly apply the two methods to open set domain adaptation, it will cause a negative transfer. In contrast, the experimental results of almost all open set methods are superior to OSNN^{cv}. This phenomenon implies that there have distribution differences between domains. More importantly, OSDA-ETD achieves the best performance of all classes' overall accuracy. Compared with JDA and DAOD, OSDA-ETD exploits the discriminability of different classes and, thus, can lead to better performance. Moreover, different from JPDA, OSDA-ETD emphasizes interdomain transferability, which helps to implement intraclass transferability and interclass discriminability. In Table II, we also find that some open set domain adaptation methods are inferior to OSDA-ETD and DAOD. The reason is that they utilize the common space to reduce the distribution discrepancy, and this may cause the known class and the unknown class to mix together. In DAOD and OSDA-ETD, the negative term $R_{u,C+1}^s(\boldsymbol{g})$ is utilized to solve the abovementioned problem.

2) Results of UC Merced as Source Domain: To validate the robustness of OSDA-ETD, we perform experiments on UC Merced→AID and UC Merced→NWPU-RESISC45; the results are list in Table III. As shown in Table III, the average classification accuracy (OS*, OS) and the overall accuracy (ALL*, ALL) of OSDA-ETD are superior to DAOD. This phenomenon demonstrates the effectiveness of increasing discriminability. Furthermore, for recognizing all target samples, OSDA-ETD can achieve the highest classification accuracy among all methods. However, in UC Merced → AID, JPDA is better than OSDA-ETD in identifying known classes. The reason is that OSDA-ETD balances the recognition of known and unknown classes via the positive term $R_{u,C+1}^t(\boldsymbol{g})$ and the negative term $R_{u,C+1}^s(\boldsymbol{g})$.

3) Results of NWPU-RESISC45 as Source Domain: The experimental results on NWPU-RESISC45 → AID and NWPU-RESISC45 → UC Merced are reported in Table IV.
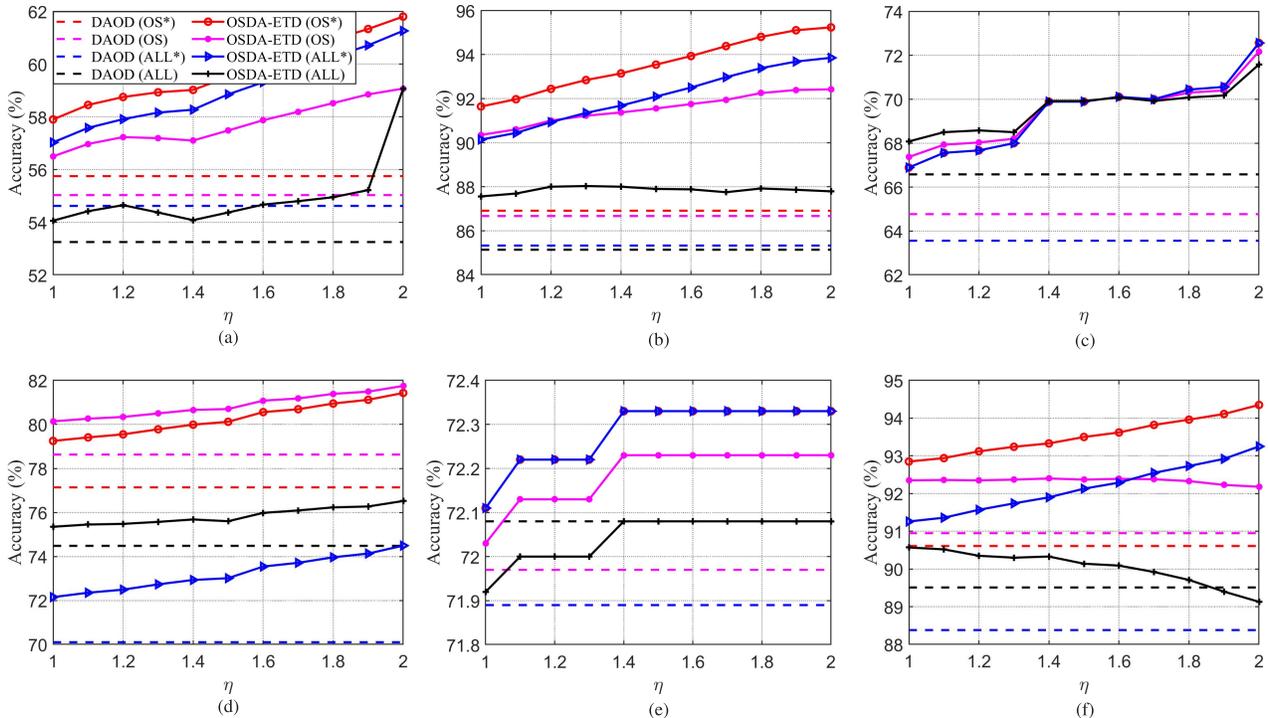
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Fig. 4. Effectiveness of discriminability on six cross dataset. (a) UC Merced → AID. (b) NWPU-RESISC45 → AID. (c) AID → UC Merced. (d) UC Merced → NWPU-RESISC45. (e) NWPU-RESISC45 → UC Merced. (f) AID → NWPU-RESISC45.

On the NWPU-RESISC45 → AID dataset, OSDA-ETD achieves good results in most classes, which verifies the effectiveness of our proposal. On NWPU-RESISC45 → UC Merced, JDA performs better than other methods in recognizing known classes. For identifying unknown classes, DAOD and OSDA-ETD gain a huge performance. Therefore, DAOD and OSDA-ETD achieve the highest overall accuracy (ALL) among all methods. In addition, Table IV shows that the average classification accuracy (OS*, OS) of OSDA-ETD is better than DAOD.

D. Parameter Sensitivity

We implement experiments on six cross datasets; experimental results demonstrate that the OSDA-ETD performs better on most tasks. To further verify the effectiveness of discriminability, we conduct sensitivity analysis about η, as shown in Fig. 4. To be specific, we investigate the sensitivity of η in a range [1, 2]. From Fig. 4, we can find that, with the increase in η, the OS*, OS, ALL*, and ALL are also improved in most datasets. This phenomenon indicates that discriminability is extremely important in remote sensing scene classification. In Fig. 4(f), with the increase in η, the evaluation metrics of ALL* increase, while ALL decreases. This phenomenon is caused by the low recognition rate of unknown samples. In Fig. 4(f), we utilize the AID dataset as the source domain to classify the samples in the NWPU-RESISC45 dataset. Due to the complexity of AID and NWPU-RESISC45 datasets, the discriminability of unknown classes is difficult to measure. In the future, we will further optimize discriminability to improve the recognition rate of unknown classes. In addition, even though the ALL

of OSDA-ETD is decreasing, most of the performance of OSDA-ETD is better than DAOD. Note that, we report the experimental results by fixing η to 1.5 in this article, which also demonstrates the potential of OSDA-ETD.

IV. Conclusion

In this article, we introduce a realistic open set domain adaptation setting in remote sensing image scene classification, where the target domain can contain the unknown classes. To realize open set domain adaptation, the OSDA-ETD algorithm is proposed. Specifically, based on the characteristics of remote sensing images, OSDA-ETD considers transferability and discriminability. To improve transferability, OSDA-ETD reduces the global and local discrepancy. For the discriminability, OSDA-ETD enlarges the difference of interclass. Finally, we integrate the transferability and the discriminability into a framework. It is worth noting that we prove that the OSDA-ETD has a unique optimizer. Extensive experiments validate the superiority of the OSDA-ETD.

In the future, we will improve the accuracy of unknown classes and extend the OSTA-ETD to a more general domain adaptation.

References

[1] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

[2] R. Cao, L. Fang, T. Lu, and N. He, "Self-attention-based deep feature fusion for remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 1, pp. 43–47, Jan. 2021.

[3] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," 2020, *arXiv:2005.01094*. [Online]. Available: http://arxiv.org/abs/2005.01094

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: OSDA-ETD FOR REMOTE SENSING IMAGE SCENE CLASSIFICATION

11

[4] E. Othman, Y. Bazi, N. Alajlan, H. Alhichri, and F. Melgani, "Using convolutional features and a sparse autoencoder for land-use scene classification," *Int. J. Remote Sens.*, vol. 37, no. 10, pp. 2149–2167, 2016.

[5] X. Han, Y. Zhong, B. Zhao, and L. Zhang, "Scene classification based on a hierarchical convolutional sparse auto-encoder for high spatial resolution imagery," *Int. J. Remote Sens.*, vol. 38, no. 2, pp. 514–536, Jan. 2017.

[6] Y. Liu and C. Huang, "Scene classification via triplet networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 220–237, Jan. 2018.

[7] J. Zhang, M. Zhang, B. Pan, and Z. Shi, "Semisupervised center loss for remote sensing image scene classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1362–1373, Mar. 2020.

[8] Y. Yuan, J. Fang, X. Lu, and Y. Feng, "Remote sensing image scene classification using rearranged local features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1779–1792, Mar. 2019.

[9] W. Han, R. Feng, L. Wang, and Y. Cheng, "A semi-supervised generative framework with deep learning features for high-resolution remote sensing image scene classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 145, pp. 23–43, Nov. 2018.

[10] S. Xu, X. Mu, D. Chai, and X. Zhang, "Remote sensing image scene classification based on generative adversarial networks," *Remote Sens. Lett.*, vol. 9, no. 7, pp. 617–626, 2018.

[11] G. Wang and P. Ren, "Delving into classifying hyperspectral images via graphical adversarial learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2019–2031, 2020.

[12] S. Xu, X. Mu, D. Chai, and S. Wang, "Adapting remote sensing to new domain with ELM parameter transfer," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1618–1622, Sep. 2017.

[13] 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, pp. 3840–3857, Jun. 2019.

[14] 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.

[15] W. Mei and D. Weihong, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Jul. 2018.

[16] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4441–4456, Aug. 2017.

[17] N. Ammour, L. Bashmal, Y. Bazi, M. M. A. Rahhal, and M. Zuair, "Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 597–601, Apr. 2018.

[18] W. Teng, N. Wang, H. Shi, Y. Liu, and J. Wang, "Classifier-constrained deep adversarial domain adaptation for cross-domain semisupervised classification in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 5, pp. 789–793, May 2020.

[19] S. Song, H. Yu, Z. Miao, Q. Zhang, Y. Lin, and S. Wang, "Domain adaptation for convolutional neural networks-based remote sensing scene classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1324–1328, Aug. 2019.

[20] J. Zhang, J. Liu, B. Pan, and Z. Shi, "Domain adaptation based on correlation subspace dynamic distribution alignment for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7920–7930, Nov. 2020.

[21] M. M. Al Rahhal, Y. Bazi, T. Abdullah, M. L. Mekhalfi, H. AlHichri, and M. Zuair, "Learning a multi-branch neural network from multiple sources for knowledge adaptation in remote sensing imagery," *Remote Sens.*, vol. 10, no. 12, p. 1890, 2018.

[22] 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, pp. 2504–2515, Apr. 2020.

[23] H. Chen, M. Ye, L. Lei, H. Lu, and Y. Qian, "Semisupervised dual-dictionary learning for heterogeneous transfer learning on cross-scene hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3164–3178, Jun. 2020.

[24] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 153–168.

[25] Z. Fang, J. Lu, F. Liu, J. Xuan, and G. Zhang, "Open set domain adaptation: Theoretical bound and algorithm," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 3, 2020, doi: 10.1109/TNNLS.2020.3017213.

[26] H. Liu, Z. Cao, M. Long, J. Wang, and Q. Yang, "Separate to adapt: Open set domain adaptation via progressive separation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2927–2936.

[27] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *J. Mach. Learn. Res.*, vol. 7, no. 11, pp. 2399–2434, 2006. [Online]. Available: http://jmlr.org/papers/v7/belkin06a.html

[28] Z. Ding, M. Shao, and Y. Fu, "Incomplete multisource transfer learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 310–323, Feb. 2018.

[29] T. Gong, X. Zheng, and X. Lu, "Cross-domain scene classification by integrating multiple incomplete sources," *IEEE Trans. Geosci. Remote Sens.*, early access, Nov. 16, 2021, doi: 10.1109/TGRS.2020.3034344.

[30] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *Proc. 26th ACM Int. Conf. Multimedia*, Oct. 2018, pp. 402–410.

[31] W. Zhang and D. Wu, "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," 2019, *arXiv:1912.00320*. [Online]. Available: http://arxiv.org/abs/1912.00320

[32] M. Long, J. Wang, G. Ding, S. J. Pan, and P. S. Yu, "Adaptation regularization: A general framework for transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 5, pp. 1076–1089, May 2014.

[33] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst. (GIS)*, 2010, pp. 270–279.

[34] 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, pp. 3965–3981, Jul. 2017.

[35] 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.

[36] 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.

[37] P. P. Busto and J. Gall, "Open set domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 754–763.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[39] P. R. M. Júnior *et al.*, "Nearest neighbors distance ratio open-set classifier," *Mach. Learn.*, vol. 106, no. 3, pp. 359–386, 2017.

[40] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.

[41] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Subspace alignment for domain adaptation," 2014, *arXiv:1409.5241*. [Online]. Available: http://arxiv.org/abs/1409.5241

[42] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.

[43] W. Zhang and D. Wu, "Discriminative joint probability maximum mean discrepancy (DJP-MMD) for domain adaptation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–8.
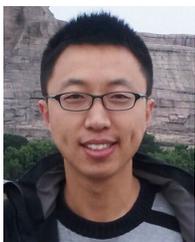
Jun Zhang received the B.S. and Ph.D. degrees from Hebei University of Technology, Tianjin, China, in 1999 and 2011, respectively.

He is currently an Associate Professor with the School of Artificial Intelligence, Hebei University of Technology. His research interests include machine learning and intelligent computing.
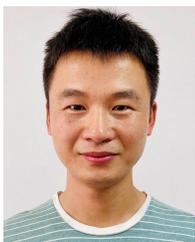
Jiao Liu received the B.S. degree in computer science and technology from Taishan Medical University, Tai'an, China, in 2018, and the M.S. degree in computer science and technology from Hebei University of Technology, Tianjin, China, in 2021.

Her research interests include machine learning and intelligent computing.

Xia Xu (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Astronautics, Beihang University, Beijing, China, in 2019.

She is currently an Assistant Professor with the College of Computer Science, Nankai University, Tianjin, China. Her research interests include hyperspectral unmixing, multiobjective optimization, and remote sensing image processing.

Bin Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote sensing image processing, and multiobjective optimization.

Zhenwei Shi (Member, IEEE) received the Ph.D. degree in mathematics from Dalian University of Technology, Dalian, China, in 2005.

From 2005 to 2007, he was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored more than 100 scientific articles in related journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE Conference on Computer Vision and Pattern Recognition. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi received the Best Reviewer Awards for his service to IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2017. He has been an Associate Editor of *Infrared Physics and Technology* since 2016.

Zongqing Chen received the Ph.D. degree in mathematics from Nankai University, Tianjin, China, in 2016.

Since 2018, he has been a Post-Doctoral Researcher with the School of Statistics and Data Science, Nankai University. His research interests include graph theory and its applications, statistical inference, and machine learning theory.