Domain Adaptation Based on Correlation Subspace Dynamic Distribution Alignment for Remote Sensing Image Scene Classification

Jun Zhang, Jiao Liu, Bin Pan, and Zhenwei Shi

Abstract

Remote sensing image scene classification refers to assigning semantic labels according to the content of the remote sensing scenes. Most machine learning-based scene classification methods assume that training and testing data share the same distributions. However, in real application scenarios, this assumption is difficult to guarantee. Domain adaptation(DA) is a promising approach to address this problem by aligning the feature distribution of training and testing data. Inspired by the idea DA, in this article, we propose a correlation subspace dynamic distribution alignment (CS-DDA) method for remote sensing image scene classification. Aiming at the characteristics of remote sensing scenes, we introduce two strategies to balance the effects of source and target domains: subspace correlation maximization (SCM) and dynamic statistical distribution alignment (DSDA). On the one hand, SCM tries to avoid mapping source domain data into irrelevant subspace to preserve the representation information of the source domain. On the other hand, DSDA is proposed to reduce the data distribution discrepancy between aligned source and target domains. Specifically, DSDA is a dynamic adjustment process where an adaptive factor is learned to balance the interclass and intraclass distribution between domains. Moreover, we integrate SCM and DSDA into a uniform optimization framework, and the optimal solution can be converted to the generalized eigendecomposition problem by derivation. The experimental results indicate that the proposed method can generate better results when compared with other feature distribution alignment methods.

Index Terms

Data shift, distribution alignment, domain adaptation(DA), remote sensing image scene classification.

Manuscript received December 2, 2019; revised February 7, 2020; accepted March 26, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC1405605, in part by the National Natural Science Foundation of China under Grant 61671037, in part by the National Natural Science Foundation of China under Grant 41804118, and in part by the Natural Science Foundation of Tianjin under Grant 19JCZDJC40000. *(Corresponding author: Bin Pan.)*

Jun Zhang and Jiao Liu are with the School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China, and also with the Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, Tianjin 300401, China (e-mail:zhangjun@scse.hebut.edu.cn; liujiao.hebut@hotmail.com).

Bin Pan is with the School of Statistics and Data Science, Nankai University, Tianjin 300071, China (e-mail: panbin@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

I. INTRODUCTION

With the development of satellite sensing technology, the task of remote sensing image interpretation has drawn significant attention, such as scene classification, hyperspectral classification, semantic segmentation [1]–[4]. As a basic image understanding work, remote sensing image scene classification aims to infer semantic labels according to the content of the remote sensing scenes [5], [6], which is beneficial to resource management, urban planning, environment monitoring, and so on [7]–[9].

During the past decades, extensive efforts have been made to develop remote sensing image scene classification methods. Early works on remote sensing scene classification are mainly based on handcrafted feature representations, such as scale-invariant feature transform [10] and bag-of-visual-words model [11]. Nevertheless, due to the highly complex geometrical structures and spatial patterns in remote sensing scene images, handcrafted features may fail to capture high-level semantic information of remote sensing scene images [12], [13]. To address this problem, convolutional neural networks (CNNs) were used for scene classification of remote sensing images. References [14]–[17] trained new CNNs models from scratch using remote sensing images. Cheng et al. [1], Castelluccio et al. [18], and Scott et al. [19] fine-tuned the pre-trained CNNs on the remote sensing images. Fang et al. [20], Weng et al. [21], Han et al. [22], and Dai et al. [23] exploit the pre-trained CNNs as a feature extractor to extract the high-level features for remote sensing images. Especially, features can be extracted from any layer of a pre-trained network; references [24]–[29] fuse the feature of different layers.

The aforementioned methods usually assumed that the training and testing data shared the same distribution. However, in a real application, due to the influence of sensors, geographic locations, imaging conditions, and other factors, the distribution of training and testing data may be different. This phenomenon is referred to as the data shift. When there is a data shift between the training set and test set, classification models have to be reconstructed from scratch using the newly collected training data [30]. Obviously, it is expensive to collect annotation data and rebuild the model [31]. To weaken the influence of data shift, domain adaptation (DA) algorithms were proposed. DA is one of transfer learning approaches which tries to remove the data shift between source and target domains [32], [33]. In DA, we define one data set with plenty of labeled samples as the source domain, and another data set, which is collected under different imaging conditions, as the target domain.

In the remote sensing scene classification, researchers have proposed several methods to alleviate data shifts based on DA. Othman et al. [34] presented a domain adaptation network to deal with the data shift problem in remote sensing image scene classification. Reference [35] introduced an asymmetric adaptation neural network method for cross-domain classification. Teng et al. [36] proposed a classifier-constrained deep adversarial domain adaptation method for cross-domain semisupervised classification. In [37], subspace alignment (SA) is designed and a CNN-based framework was used to solve the data shifts problem. The abovementioned methods tried to reduce the distribution difference by aligning the global distribution of the two domains; however, they did not consider the distribution alignment between categories. Mathematically, the global distribution is called marginal distribution, and the class distribution is called conditional distribution.

Zhang et al. [38] proposed the joint geometrical and statistical alignment (JGSA)-based method to reduce the

data shift. Different from the methods that only considering marginal distribution, JGSA learned two projections that transform the source domain and target domain data into respective subspaces where the marginal distribution, conditional distribution, and geometrical divergence are aligned simultaneously. Inspired by the idea of multi-distribution ensemble, in this article, we develop a new domain adaptation method under the framework of JGSA.

However, JGSA is originally designed for the task of natural scene object recognition, which is quite different from scene classification. In remote sensing image scene classification task, JGSA has at least two shortcomings that may hinder its application:

1) Remote sensing scene images present strong interclass similarity. Besides removing the diversity between source and target domains, it is quite necessary to enhance the discrepancy of corresponding categories.

2) The weights between marginal and conditional distributions in JGSA are fixed. Although JGSA considers both of the distributions, it cannot evaluate the importance of them.

In this article, to overcome these limitations, we propose a correlation subspace dynamic distribution alignment (CS-DDA) method for remote sensing image scene classification. Two strategies are developed to improve JGSA: subspace correlation maximization (SCM) and dynamic statistical distribution alignment (DSDA). SCM attempts to prevent mapping the source domain data into irrelevant subspace, which preserves the information structure in the source domain. Meanwhile, we propose DSDA to balance the influence of marginal and conditional distributions. DSDA exploits an adaptive factor to adjust class distribution alignment. In particular, this balanced factor can be estimated according to the data distribution between the source domain and the target domain.

The main contributions of this paper can be summarized as follows:

1) A new DA method is proposed for remote sensing image scene classification, where CS-DDA is developed.

2) We propose SCM to avoid mapping the source domain data to unrelated subspaces.

3) We design DSDA that aims to eliminate the influence of distributions weights via learning an adaptive factor.

The rest of this paper is organized as follows. Section II introduces the overall framework of the proposed method as well as our two contributions. The description of data sets, the experimental setup, experimental results, and feature distribution analysis are presented in Section III. Section IV concludes this paper.

II. Methodology

This section describes CS-DDA for remote sensing scene classification, which contains algorithm background, SCM, DSDA, and final objective function. The process of remote sensing scene classification is shown in Fig. 1.

A. Algorithm Background

JGSA [38] is a representative algorithm for DA, so we first introduce the work of JGSA.

The labeled source domain data are denoted by $\mathcal{D}_s = \{x_i, y_i\}_{i=1}^{n_s}$, where $x_i \in \mathbb{R}^d$ is the deep feature of source domain image I_i, n_s is the number of source domain samples, $y_i \in \{1, 2, \ldots, C\}$ is corresponding category label, and C is the number of categories. The source domain matrix $X_s = \{x_i\}_{i=1}^{n_s}$ is drawn from the distribution $P_s(X_s)$. The unlabeled target domain data can be defined as follows: $\mathcal{D}_t = \{x_j\}_{j=1}^{n_t}$, and its matrix $X_t = \{x_j\}_{j=1}^{n_t}$ is drawn from the distribution $P_t(X_t)$. The feature spaces and label spaces between domains are the same: $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$. Due to

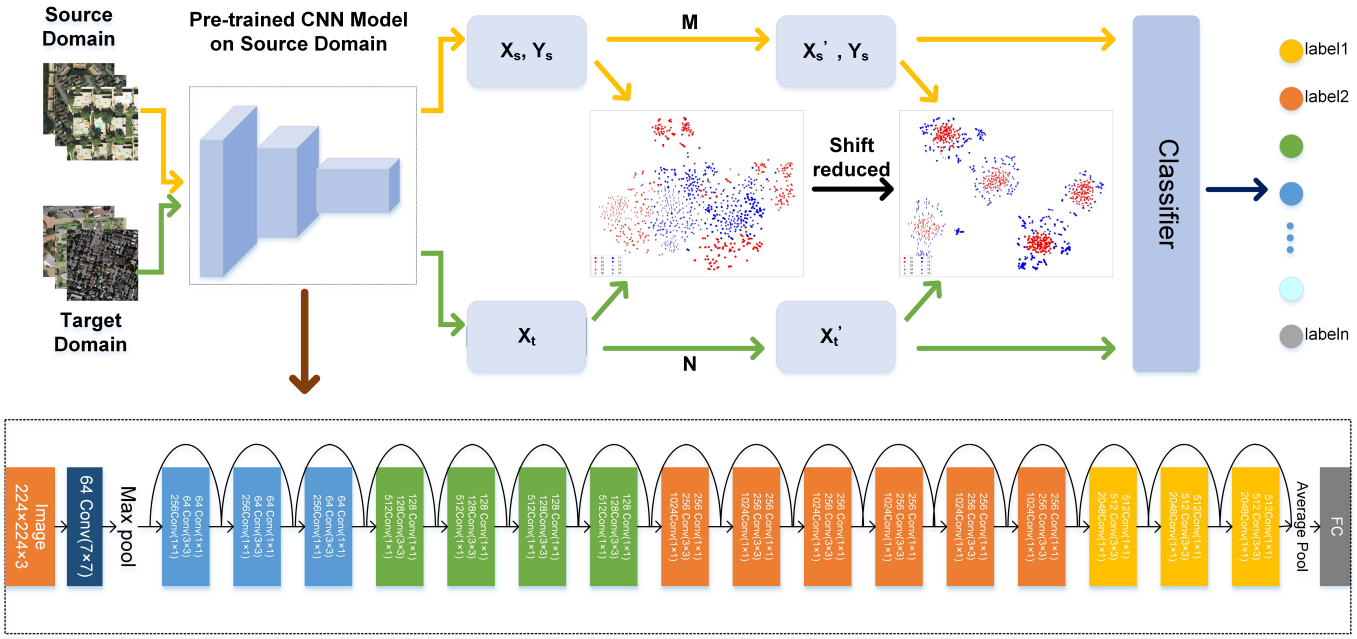

Fig. 1. Overall architecture of the proposed method. First, a pre-trained CNN on source domain is applied to extract features of the source domain and target domain. Then, to alleviate the data shift, we leverage CS-DDA to handle the original features X_s and X_t. Further, a classifier is trained by using the new feature representation $X_s^{'}$ in the source domain. Finally, the trained classifier f is utilized to classify $X_t^{'}$.

the data shifts, $P_s\left(X_s\right) \neq P_t\left(X_t\right)$. The previous method assumes a unified transformation $P_s\left(\phi\left(X_s\right)\right)=P_t\left(\phi\left(X_t\right)\right)$, $P_s\left(Y_s|\phi\left(X_s\right)\right)=P_t\left(Y_t|\phi\left(X_t\right)\right)$. Nevertheless, when the divergence of the two domains is large, such a unified transformation may not exist. Therefore, JGSA finds two projections M, N to generate a new feature representation of the source and target domains.

To prevent mapping features of the target domain into irrelevant dimensions, the variance of the target domain is maximized.

$$\max_{N} \operatorname{Tr}\left(N^T S_t N\right) \tag{1}$$

where $S_t=X_t H_t {X_t}^T$ is the scattering matrix of the target domain , $H_t=I_t-\frac{1}{n_t}1_t 1_t^T$ is the centering matrix, and $1_t \in \mathbb{R}^{n_t}$ is the column vector with all elements one.

Especially, JGSA exploits the label information of the source domain to make the features discriminative

$$\max_{M} \operatorname{Tr}\left(M^T S_b M\right) \tag{2}$$

$$\min_{M} \operatorname{Tr}\left(M^T S_w M\right) \tag{3}$$

where S_b is the interclass scattering matrix of the source domain and S_w is the intraclass scattering matrix of the source domain, which is defined as follows:

$$S_w = \sum_{c=1}^{C} {X_s}^{(c)} H_s^{(c)} \left({X_s}^{(c)}\right)^T \tag{4}$$

$$S_b = \sum_{c=1}^{C} n_s^{(c)} \left(a_s^{(c)} - \overline{a}_s\right)\left(a_s^{(c)} - \overline{a}_s\right)^T \tag{5}$$

where ${X_s}^{(c)} \in \mathbb{R}^{d \times n_s^{(c)}}$ is the set of source samples belonging to class c, $a_s^{(c)} = \frac{1}{n_s^{(c)}} \sum_{i=1}^{n_s^{(c)}} x_i^{(c)}$, $\overline{a}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} x_i$, $H_s^{(c)} = I_s^{(c)} - \frac{1}{n_s^{(c)}} 1_s^{(c)} \left(1_s^{(c)}\right)^T$ is the centering matrix of data within class c, $I_s^{(c)} \in \mathbb{R}^{n_s^{(c)} \times n_s^{(c)}}$ is the identity matrix, $1_s \in \mathbb{R}^{n_s}$ is the column vector with all elements one, and $n_s^{(c)}$ is the number of source samples in class c.

To reduce the difference between marginal distributions of source and target domains, JGSA adopts maximum mean discrepancy (MMD) [39] as the distance measure to compare different distributions, which measures the distance between the sample means of the source and target data in the subspace.

$$\min_{M,N} \left\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} M^T \mathbf{x}_i - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} N^T \mathbf{x}_j \right\|_F^2. \tag{6}$$

However, reducing the discrepancy in the marginal distributions cannot guarantee that the conditional distributions are also be removed [40]. Hence, JGSA leverages the pseudo labels of the target domain to minimize the conditional distribution differences between the source domain and the target domain

$$\min_{M,N} \sum_{c=1}^{C} \left\| \frac{1}{n_s^{(c)}} \sum_{\mathbf{x}_i \in {X_s}^{(c)}} M^T \mathbf{x}_i - \frac{1}{n_t^{(c)}} \sum_{\mathbf{x}_j \in {X_t}^{(c)}} N^T \mathbf{x}_j \right\|_F^2. \tag{7}$$

Moreover, JGSA removes the geometric discrepancy between domains by making the subspace of the source domain and the target domain subspace close

$$\min_{M,N} \|M - N\|_F^2. \tag{8}$$

B. SCM

JGSA utilizes the label information to constrain the new representation of source domain data. Nevertheless, for the remote sensing image scene, source domain data may be mapped into unrelated subspaces. The reason is that the category of remote sensing image scene is related to not only the category of objects but also related to the background and spatial layout.

To avoid projecting source domain data into unrelated subspaces, we maximize the source domain variance. The variance maximization can be calculated as follows:

$$\max_{M} \operatorname{Tr}\left(M^T S_s M\right) \tag{9}$$

where

$$S_s = X_s H_s {X_s}^T \tag{10}$$

is the scattering matrix of the source domain. In (10), H_s is the central matrix, which is defined as follows:

$$H_s = I_s - \frac{1}{n_s} 1_s 1_s^T \tag{11}$$

where $I_s \in \mathbb{R}^{n_s \times n_s}$ is the identity matrix, and $1_s \in \mathbb{R}^{n_s}$ is the column vector with all elements one.

C. DSDA

Different from natural object images, remote sensing scene images often exhibit complex spatial structures with intraclass diversity and interclass similarity. Therefore, for remote sensing scene images, it is not appropriate to alleviate the marginal distribution and the conditional distribution discrepancy with equal weight as JGSA.

In this article, we propose a DSDA algorithm to adaptively adjust the importance of the conditional distribution according to the feature distribution of the remote sensing scene. We add into the factor α to control the conditional distribution. It can be represented as

$$\begin{aligned}&\min_{M,N} \Big\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} M^T \mathbf{x}_i - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} N^T \mathbf{x}_j \Big\|_F^2 + \\ &\alpha \sum_{c=1}^{C} \Big\| \frac{1}{n_s^{(c)}} \sum_{\mathbf{x}_i \in {X_s}^{(c)}} M^T \mathbf{x}_i - \frac{1}{n_t^{(c)}} \sum_{\mathbf{x}_j \in {X_t}^{(c)}} N^T \mathbf{x}_j \Big\|_F^2\end{aligned} \tag{12}$$

where the first term denotes the marginal distribution distance between domains and the second term is the conditional distribution distance. According to the relationship between the matrix norm and the matrix trace, (12) can be rewritten as follows:

$$\min_{M,N} \mathrm{Tr}\left(\left[M^T N^T\right] \begin{bmatrix} R_s & R_{st} \\ R_{ts} & R_t \end{bmatrix} \begin{bmatrix} M \\ N \end{bmatrix} \right) \tag{13}$$

where

$$R_s = X_s\big(Q_s + \alpha \sum_{c=1}^{C} Q_s^{(c)}\big) {X_s}^T, Q_s = \frac{1}{n_s^2} 1_s 1_s^T$$

$$\big(Q_s^{(c)}\big)_{ij} = \begin{cases} \frac{1}{\left(n_s^{(c)}\right)^2} & \mathbf{x}_i, \mathbf{x}_j \in {X_s}^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

$$R_t = X_t\big(Q_t + \alpha \sum_{c=1}^{C} Q_t^{(c)}\big) {X_t}^T, Q_t = \frac{1}{n_t^2} 1_t 1_t^T$$

$$\big(Q_t^{(c)}\big)_{ij} = \begin{cases} \frac{1}{\left(n_t^{(c)}\right)^2} & \mathbf{x}_i, \mathbf{x}_j \in {X_t}^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

$$R_{st} = X_s\big(Q_{st} + \alpha \sum_{c=1}^{C} Q_{st}^{(c)}\big) {X_t}^T, Q_{st} = -\frac{1}{n_s n_t} 1_s 1_t^T$$

$$\big(Q_{st}^{(c)}\big)_{ij} = \begin{cases} -\frac{1}{n_s^{(c)} n_t^{(c)}} & \mathbf{x}_i \in {X_s}^{(c)}, \mathbf{x}_j \in {X_t}^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

$$R_{ts} = X_t\big(Q_{ts} + \alpha \sum_{c=1}^{C} Q_{ts}^{(c)}\big) {X_s}^T, Q_{ts} = -\frac{1}{n_t n_s} 1_t 1_s^T$$

$$\big(Q_{ts}^{(c)}\big)_{ij} = \begin{cases} -\frac{1}{n_s^{(c)} n_t^{(c)}} & \mathbf{x}_j \in {X_s}^{(c)}, \mathbf{x}_i \in {X_t}^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

However, there is a distinct disadvantage in (12): the factor α has to be manually fixed, which cannot balance the influence of marginal and conditional distribution in remote sensing scenes. Aiming at this problem, in CS-DDA, we propose an adaptive factor adjustment approach based on $\mathcal{A}$-distance [41], [42]. The $\mathcal{A}$-distance is defined as follows:

$$d_A\left(\mathcal{D}_s, \mathcal{D}_t\right) = 2\left(1 - 2\epsilon\left(h\right)\right) \tag{18}$$

where h is the classifier and $\epsilon\left(h\right)$ denotes the error of classifier discriminating the source domain and the target domain.

Thus, we estimate α based on the $\mathcal{A}$-distance. In this process, we utilize the global and local information of the domain, which can be calculated as follows:

$$\hat{\alpha} \approx \frac{\sum_{c=1}^{C} d_c\left(\mathcal{D}_s^{(c)}, \mathcal{D}_t^{(c)}\right)}{d_m\left(\mathcal{D}_s, \mathcal{D}_t\right)} \tag{19}$$

where d_c denotes the conditional $\mathcal{A}$-distance, d_m denotes the marginal $\mathcal{A}$-distance, $\mathcal{D}_s^{(c)}$denotes samples from class c in $\mathcal{D}_s$, and $\mathcal{D}_t^{(c)}$ denotes samples from class c in $\mathcal{D}_t$.

D. Final objective function

In remote sensing image scene classification, we require to generate an effective and robust feature representation. Hence, we incorporate (1)-(3), (8), (9), (12) and (19) to determine the final objective function. The final objective function can be formalized as follows

$$\max_{M,N} \frac{\beta(9) + \gamma(1) + \delta(2)}{(12) + \lambda(8) + \delta(3)} \tag{20}$$

In (20), minimizing the denominator has two meanings: one is to reduce the statistical and geometric differences between the source domain and the target domain and the other is to minimize the intraclass variance of the source domain. In addition, maximizing the numerator of (20) encourages large target domain variance, source domain variance, and the inter-class variance of the source domain. Among them, $\beta, \gamma, \delta, \lambda$ are factors that balance the importance of each part.

Furthermore, we follow [43] to constrain N. To optimize (20), we write $\begin{bmatrix} M^T & N^T \end{bmatrix}$ as Z^T, so the objective function can be written as follows:

$$\max_{Z} \frac{\operatorname{Tr}\left(Z^T \begin{bmatrix} \beta S_s + \delta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} Z\right)}{\operatorname{Tr}\left(Z^T \begin{bmatrix} R_s + \lambda I + \delta S_w & R_{st} - \lambda I \\ R_{ts} - \lambda I & R_t + (\lambda + \delta) I \end{bmatrix} Z\right)} \tag{21}$$

We discover that scaling Z does not affect the objective function, so (21) can be rewritten as follows:

$$\max_{Z} \operatorname{Tr}\left(Z^T \begin{bmatrix} \beta S_s + \delta S_b & 0 \\ 0 & \gamma S_t \end{bmatrix} Z\right) \tag{22}$$

$$s.t \quad \mathrm{Tr}\left(Z^T\left[\begin{array}{cc} R_s+\lambda I+\delta S_w & R_{st}-\lambda I \\ R_{ts}-\lambda I & R_t+(\lambda+\delta)I \end{array}\right]Z\right)=1$$

Actually, (22) is a constrained optimization problem, and it is convenient to apply Lagrange techniques. Hence, the corresponding Lagrangian function for (22) is given as follows:

$$\begin{aligned} \mathcal{L} = &\mathrm{Tr}\left(Z^T\left[\begin{array}{cc} \beta S_s+\delta S_b & 0 \\ 0 & \gamma S_t \end{array}\right]Z\right)+ \\ &\mathrm{Tr}\left(\left(Z^T\left[\begin{array}{cc} R_s+\lambda I+\delta S_w & R_{st}-\lambda I \\ R_{ts}-\lambda I & R_t+(\lambda+\delta)I \end{array}\right]Z-I\right)\Theta\right) \end{aligned} \tag{23}$$

By setting the derivative of $\frac{\partial \mathcal{L}}{\partial Z}=0$, we can convert it into an eigendecomposition problem

$$\left[\begin{array}{cc} \beta S_s+\delta S_b & \mathbf{0} \\ \mathbf{0} & \gamma S_t \end{array}\right]Z=\left[\begin{array}{cc} R_s+\lambda I+\delta S_w & R_{st}-\lambda I \\ R_{ts}-\lambda I & R_t+(\lambda+\delta)I \end{array}\right]Z\Theta \tag{24}$$

where Θ is the eigenvalues matrix and Z is the corresponding eigenvectors matrix. Therefore, we can acquire the map matrices M and N by Z. Finally, the source domain data and the target domain data are mapped into the new subspaces to get the new feature representation $X_s{}' = M^T X_s$ and $X_t{}' = N^T X_t$. The pseudocode of CS-DDA is shown in Algorithm 1.

III. Experiment

In this section, we introduce the data set description, experimental setup, experimental results, and feature distribution analysis.

A. Data Sets

To verify our algorithm, we select the UC Merced, AID, NWPU-RESISC45, and RSSCN7 to build the cross-domain remote sensing image scene data sets.

1) UC Merced Data set [11] : The UC Merced data set is a widely used data set for remote sensing image scene classification and can be downloaded from the United States Geological Survey National Map of 20 U.S. regions. It consists of 2100 remote sensing images from 21 scene classes: agricultural, airplane, baseball diamond, beach, buildings, chaparral, dense residential, forest, freeway, golf course, harbor, intersection, medium density residential, mobile home park, overpass, parking lot, river, runway, sparse residential, storage tanks, and tennis courts. Each scene class contains 100 red-green-blue (RGB) images with a pixel resolution of 1 ft and an image size of 256 $\times$ 256 pixels.

2) AID Data set [44]: The AID data set is a large scale aerial image data set and acquired from Google Earth. It contains 10,000 images with a size of 600$\times$600 pixels, which are divided into 30 classes, including airport, bare land, baseball field, beach, bridge, center, church, commercial, dense residential, desert, farmland, forest, industrial, meadow, medium residential, mountain, park, parking, playground, pond, port, railway station, resort, river, school, sparse residential, square, stadium, storage tanks, and viaducts. The number of each class varies from 220 to 420 and the pixel resolution changes from 8 m to half a meter.

Algorithm 1 CS-DDA Algorithm.

Input:

Data: Source domain matrix X_s, target domain matrix X_t, source domain label Y_s

Parameter: $\gamma = 1, \delta = 1, \lambda = 1, \beta = 1, \alpha$

Output:

Projection matrix: M, N

New feature representation: $X_s{}'$, $X_t{}'$

Target domain label Y_t

Train a classifier using original X_s, and apply prediction on X_t to get the initial pseudo-label $Y_{t'}$

Calculate

$S_t{=}X_t H_t {X_t}^T$, $S_s{=}X_s H_s {X_s}^T$

$S_w{=}\sum_{c=1}^{C} {X_s}^{(c)} H_s^{(c)} \big({X_s}^{(c)}\big)^T$

$S_b{=}\sum_{c=1}^{C} n_s^{(c)} \big(a_s^{(c)} - \overline{a}_s\big)\big(a_s^{(c)} - \overline{a}_s\big)^T$

repeat

Estimate α by $\hat{\alpha} \approx \frac{\sum_{c=1}^{C} d_c\left(\mathcal{D}_s^{(c)}, \mathcal{D}_t^{(c)}\right)}{d_m(\mathcal{D}_s, \mathcal{D}_t)}$

Compute

$R_s = X_s\big(Q_s + \alpha \sum_{c=1}^{C} Q_s^{(c)}\big){X_s}^T$

$R_t = X_t\big(Q_t + \alpha \sum_{c=1}^{C} Q_t^{(c)}\big){X_t}^T$

$R_{st} = X_s\big(Q_{st} + \alpha \sum_{c=1}^{C} Q_{st}^{(c)}\big){X_t}^T$

$R_{ts} = X_t\big(Q_{ts} + \alpha \sum_{c=1}^{C} Q_{ts}^{(c)}\big){X_s}^T$

Solve the generalized eigendecomposition problem in Eq. (24), and obtain projection matrices M, N by Z

Construct the new feature representation $X_s{}' = M^T X_s$ and $X_t{}' = N^T X_t$

Train a new classifier on $X_s{}'$ to update the pseudo-label of $Y_{t'}$

Update α,R_s,R_t,R_{st},R_{ts}

until Convergence

return Target domain label Y_t

3) NWPU-RESISC45 Data set [1]: The NWPU-RESISC45 data set consists of 31,500 remote sensing images divided into 45 scene classes. These 45 scene classes are as follows: airplane, airport, baseball diamond, basketball court, beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium residential, mobile home park, mountain, overpass, palace, parking lot, railway, railway station, rectangular farmland, river, roundabout, runway, sea ice, ship, snowberg, sparse residential, stadium, storage tank, tennis court, terrace, thermal power station, and wetland. Each class includes 700 images with a size of 256×256 pixels in the red green blue (RGB) color space. The spatial resolution varies from about 30 m to 0.2 m per pixel for most of the scene classes.

Fig. 2. Sample images from four public data sets used in our experiments (five classes in total).(Top row to the bottom row) UC Merced data set, the AID data set, the NWPU-RESISC45 data set, and the RSSCN7 data set. Each column presents the corresponding class of these four data sets. (From left to right) Dense residential, farmland, forest, parking lot, river.

4) RSSCN7 Data set [45]: The RSSCN7 data set contains 2800 remote sensing scene images, which are from seven typical scene categories, namely, the grassland, forest, farmland, parking lot, residential region, industrial region, and river and lake. There are 400 images in each scene type, and each image has a size of 400 $\times$400 pixels. It is worth noticing that the sample images in each class are sampled on four different scales with 100 images per scale with different imaging angles.

From the above description of the data set, we discover that these four data sets have five public categories, namely farmland, forests, dense residential areas, rivers, and parking lot. Hence, we choose these five categories to do the cross-domain remote sensing image scene data sets. Some example images are shown in Fig. 2.

B. Experiment Setup

In this article, we establish three cross-domain scenarios termed as UCM→RSSCN7, AID→RSSCN7 and NWPU→RSSCN7 referring to source domain→target domain. To analyze the generalization ability of CS-DDA on the three cross-domain remote sensing scene data sets, two classic pre-trained CNN models (i.e., AlexNet [46] and ResNet50 [47]) are utilized to extract the features of the source and target domains. Specifically, the features of both source domain data and target domain data before the final fully connected (FC) layer are extracted [37]. The dimensions

of features extracted from AlexNet and ResNet50 are 4096 and 2048 respectively. For the parameter setting of CS-DDA, we follow JGSA.

To evaluate the proposed algorithm, we select overall accuracy (OA), class accuracy (CA), kappa coefficient and confusion matrix as evaluation metrics.

In addition, to verify the CS-DDA, we compare it to the following.

1) 1-Nearest Neighbor (NN): A basic baseline that learns NN classifier on the source domain and applies to the target, which could be viewed as a lower bound.

2) Transfer component analysis (TCA) [32]: TCA tries to learn some transfer components across domains to reduce the marginal distribution difference.

3) SA [48]: SA aligns the base of the source domain with the base of the target domain by finding a transformation.

4) CORrelation ALignment (CORAL) [49]: CORAL minimizes domain shift by aligning the second-order statistics of source and target distributions.

5) Joint distribution analysis (JDA) [50]: JDA attempts to jointly adopt both the marginal distribution and conditional distribution.

6) Transfer Joint Matching (TJM) [51]: TJM aims to reduce the domain discrepancy by jointly matching the features and reweighting the instances.

7) JGSA: A baseline of CS-DDA.

C. Experimental Results

In this section, we comprehensively evaluate CS-DDA and seven baseline methods on three cross-domain scenarios. All experimental results are illustrated in Table I, where the best results are marked in bold.

1) Results on NWPU→RSSCN7: Our first experiment was conducted on NWPU→RSSCN7, including two cases. In the first case of AlexNet as a feature extractor, CS-DDA achieves much better performance than all the other baselines, with the OA of 84.40. Especially, the OA and Kappa of CS-DDA improve by 2.25 and 2.81, respectively, compared with the best baseline JGSA. There is a reasonable explanation of why CS-DDA performs better than JGSA. CS-DDA searches for relevant subspace as much as possible and adjusts the conditional distribution according to the data distribution of the source domain and target domain. To analyze the generalization ability of CS-DDA on the NWPU→RSSCN7, we also adopt ResNet50 as a feature extractor. Similar to the first case, CS-DDA achieves favorable results in the case of ResNet50 as a feature extractor.

It is worth noting that almost all feature distribution alignment methods are better than the methods of NN. This phenomenon indicates the importance of mitigating the data shift when the source domain and target domain are drawn from different distributions. The OA and Kappa of the JDA are better than TCA. The reason is that JDA simultaneously adapts both marginal and conditional distributions. CS-DDA and JGSA are better than JDA because they preserve the discrimination information of the source domain. In particular, TJM is better than TCA because TJM combines marginal distribution matching and reweighting the instances. In addition, the OA and Kappa of SA have not shown good performance. The reason is that SA fails to minimize the distributions between domains after aligning the subspaces.

TABLE I: CLASSIFICATION ACCURACY OF DIFFERENT METHODS ON THREE CROSS-DOMAIN DATA SETS (%).

Data set	Feature	Class	Methods							
			NN	TCA	SA	CORAL	JDA	TJM	JGSA	CS-DDA
NWPU→RSSCN7	AlexNet	Dense residential	74.25	82.25	80.25	77.00	81.75	82.25	**89.50**	83.50
		Farmland	49.50	69.50	72.00	73.50	72.25	81.75	81.75	**87.50**
		Forest	**96.25**	84.25	86.50	88.50	90.75	87.00	83.25	83.25
		Parking lot	68.25	78.25	81.00	**84.25**	79.75	76.25	71.25	77.50
		River	32.25	72.50	72.00	66.00	80.75	79.00	85.00	**90.25**
		Kappa	55.13	71.69	72.94	72.31	76.44	76.56	77.69	**80.50**
		OA	64.10	77.35	78.35	77.85	81.15	81.25	82.15	**84.40**
	ResNet50	Dense residential	79.75	**86.25**	79.50	78.00	84.50	83.50	82.25	81.00
		Farmland	73.00	84.25	83.75	86.50	85.25	85.25	**88.75**	88.25
		Forest	**93.25**	79.75	89.50	87.75	83.25	85.50	91.00	93.00
		Parking lot	92.50	89.75	95.00	94.75	89.00	86.50	**95.75**	95.00
		River	47.75	60.75	61.75	73.75	72.25	67.50	81.25	**85.75**
		Kappa	71.56	75.19	77.38	80.19	78.56	77.06	84.75	**85.75**
		OA	77.25	80.15	81.9	84.15	82.85	81.65	87.80	**88.60**
AID→RSSCN7	AlexNet	Dense residential	89.00	88.25	85.50	87.75	79.75	84.75	**90.75**	82.25
		Farmland	81.75	74.25	70.00	78.75	83.25	76.75	**91.00**	90.25
		Forest	77.50	77.75	80.75	77.25	83.75	84.25	87.00	**88.25**
		Parking lot	30.25	47.50	53.50	46.00	**76.75**	67.25	58.75	74.00
		River	51.00	77.75	80.75	84.75	86.75	86.5	90.50	**91.50**
		Kappa	57.38	66.38	67.63	68.63	77.56	74.88	79.50	**81.56**
		OA	65.90	73.10	74.10	74.90	82.05	79.90	83.60	**85.25**
	ResNet50	Dense residential	78.75	82.00	83.00	**89.00**	80.25	83.75	78.75	79.00
		Farmland	53.50	83.25	81.50	**92.00**	89.00	87.50	83.50	87.00
		Forest	**94.75**	83.00	73.50	86.00	77.25	77.50	89.50	89.50
		Parking lot	68.00	67.75	71.50	64.00	77.75	74.25	**93.25**	93.00
		River	52.00	74.50	72.75	78.25	82.25	82.75	82.75	**84.00**
		Kappa	61.75	72.63	70.56	77.31	76.63	76.44	81.94	**83.13**
		OA	69.40	78.10	76.45	81.85	81.30	81.15	85.55	**86.50**
UCM→RSSCN7	AlexNet	Dense residential	28.50	48.50	47.00	59.50	68.75	61.50	74.75	**80.50**
		Farmland	58.50	54.50	73.75	39.75	79.00	79.25	85.50	**87.75**
		Forest	**95.50**	84.75	83.75	63.75	89.50	88.00	84.50	83.00
		Parking lot	27.50	49.00	47.00	34.25	65.25	**66.25**	52.75	53.25
		River	48.75	89.25	80.50	**95.50**	84.50	79.00	92.75	91.50
		Kappa	39.69	56.50	58.00	48.19	71.75	68.50	72.56	**74.00**
		OA	51.75	65.20	66.40	58.55	77.40	74.80	78.05	**79.20**
	ResNet50	Dense residential	49.75	56.25	49.75	68.50	83.25	69.25	**84.75**	**84.75**
		Farmland	61.75	69.00	74.00	58.00	**85.00**	81.75	77.75	78.25
		Forest	90.50	84.75	74.00	71.00	87.00	90.25	**91.50**	91.25
		Parking lot	50.00	69.75	53.50	48.25	81.00	69.00	79.50	**79.75**
		River	72.75	81.75	86.75	88.75	74.75	77.75	85.50	**85.75**
		Kappa	56.19	65.38	59.50	58.63	77.75	72.00	79.75	**79.94**
		OA	64.85	73.20	67.60	66.90	82.20	77.60	83.80	**83.95**

In the first case of AlexNet as a feature extractor, for ”Farmland” and ”River”, CS-DDA acquires a better CA compared with JGSA. In Fig. 3, we provide the confusion matrix for both cases in NWPU→RSSCN7. From Fig. 3, we find that the performance of the ResNet50 is better than that of AlexNet. However, some categories are

confused, such as dense residential/parking lot, forest/ river. As shown in Fig. 4, the forest and river share similar backgrounds. The dense residential and parking lots of scenes in Fig. 4 are composed of the same objects, such as roads and buildings, which can affect the feature representation of the scenes.

Fig. 3. Confusion matrix on the NWPU→RSSCN7. (a) AlexNet as a feature extractor. (b) ResNet50 as a feature extractor. The rows and columns of the matrix denote the actual and predicted classes, respectively. The corresponding category labels of 1-5 are dense residential, farmland, forest, parking lot, river. The color bar indicates the proportion of samples over the actual total class samples.

Fig. 4. Examples of major confusion in the RSSCN7 data set.

2) Results on AID→RSSCN7: Our second experiment was conduced on the AID→RSSCN7. Table I shows the OA and Kappa gained by two different features. A similar tendency is observed: CS-DDA still outperforms over the state-of-the-art. Furthermore, from Table I, we can discover that the OA and Kappa of almost all methods improve compared to NN. This phenomenon illustrates that domain shift also appears in AID and RSSCN7 data sets.

In the first case of AlexNet as feature extractor, for ”Forest” and ”River”, CS-DDA achieves the best CA compared with state-of-the-art methods. In Fig. 5, we provide the confusion matrix for both scenarios in AID→RSSCN7. The confusion on AID→RSSCN7 is similar to the confusion on NWPU→RSSCN7.

3) Results on UC Merced→RSSCN7: Our third experiment was conducted on the UC Merced→RSSCN7, and the results are provided in Table I. From Table I, it can be seen that CS-DDA achieves significantly better performance than the state-of-the-art methods.

In the first scenario of AlexNet as a feature extractor, for ”Dense Residential” and ”Farmland”, CS-DDA realizes the best CA compared with state-of-the-art methods. In Fig. 6, we provide the confusion matrix for both scenarios

Fig. 5. Confusion matrix on the AID→RSSCN7. (a) AlexNet as a feature extractor. (b) ResNet50 as a feature extractor. The rows and columns of the matrix denote the actual and predicted classes, respectively. The corresponding category labels of 1-5 are dense residential, farmland, forest, parking lot and river. The color bar indicates the proportion of samples over the actual total class samples.

in UC Merced →RSSCN7. Compared with the previous two groups of experiments, the result of UC Merced as the source domain is a little worse. The reason is that UC Merced is much more simple than AID and NWPU.

Fig. 6. Confusion matrix on the UC Merced→RSSCN7. (a) AlexNet as a feature extractor. (b) ResNet50 as a feature extractor. The rows and columns of the matrix denote the actual and predicted classes, respectively. The corresponding category labels of 1-5 are dense residential, farmland, forest, parking lot and river. The color bar indicates the proportion of samples over the actual total class samples.

Generally speaking, the experiments on three cross-domain remote sensing image scene data sets show similar results. CS-DDA is designed to not only align the source and target domains as much as possible but also dynamically adjust the marginal distribution and conditional distribution.

D. Feature distribution analysis

In this section, to understand more deeply about CS-DDA, we provide the feature distributions of UCM→RSSCN7 in the 2-D space. We leverage the t-distributed stochastic neighbor embedding (t-SNE) [52] as a visualization tool to observe the data shift between the source domain and the target domain.

The feature distribution of Fig. 7 is original without domain adaptation. From Fig 7, we can observe that the source domain and the target domain have large data shifts. The feature distribution of Fig. 8 is processed by CS-DDA. It is clear that CS-DDA alleviates the distribution discrepancy between the two domains. In Fig. 8, we can observe that CS-DDA is more discriminative, which implies the source classifier will predict the target data more correctly. In addition, CS-DDA aligns the corresponding categories of the source domain and the target, which is also beneficial to classify the target domain.

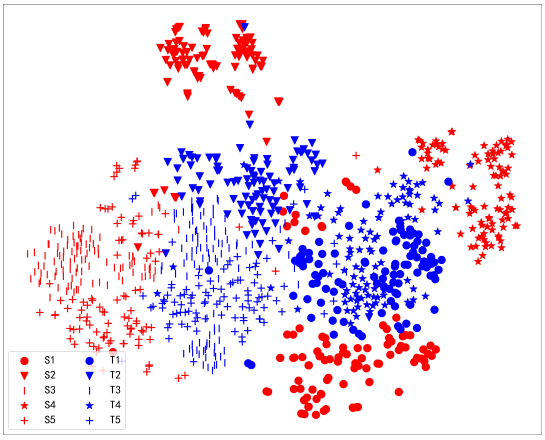

Fig. 7. Feature visualization: data embedding by t-SNE of original features. Red: source domain. Blue: target domain. Samples in different shapes represent they are in different categories. The corresponding label categories in the legend are dense residential, farmland, forest, parking lot, river.

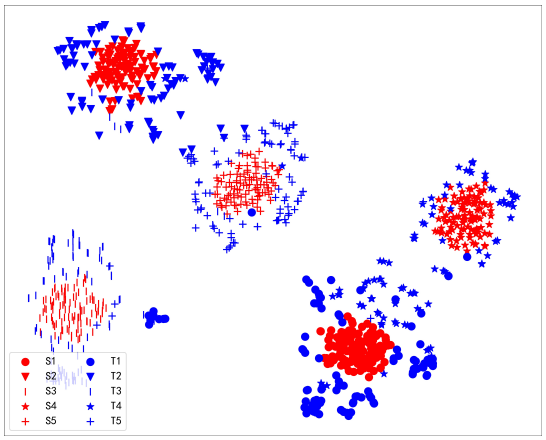

Fig. 8. Feature visualization: data embedding by t-SNE of CS-DDA features. Red: source domain. Blue: target domain. Samples in different shapes represent they are in different categories. The corresponding label categories in the legend are dense residential, farmland, forest, parking lot, river.

IV. Conclusion

In this article, we introduce a CS-DDA method to alleviate the data shift in remote sensing scene classification. The algorithm of CS-DDA includes SCM and DSDA. Among them, SCM generates a beneficial feature representation, and DSDA eliminates distribution discrepancies between domains. The major contributions of this article consist of three folds: 1) we present a new feature distribution alignment method based on DA; 2) to prevent mapping the source domain data to unrelated subspaces, we propose SCM; 3) since the remote sensing image has the characteristics of intraclass diversity and interclass similarity, we leverage a balance factor to adjust the class condition distribution. We exploit the features of two classic pre-trained CNN models to experiment on three cross-domain data sets. The

experimental results demonstrate that the proposed method is superior to several state-of-the-art methods.

In the future, we will extend the CS-DDA method to confront different DA scenarios in remote sensing scene classification.

ACKNOWLEDGMENT

The authors would like to thank Shawn Newsam from the University of California at Merced, Guisong Xia from Wuhan University, Qin zou from Wuhan University, and Gong Cheng from Northwestern Polytechnical University, for providing the UC Merced, AID, RSSCN7 and NWPU-RESISC45 datasets in their study, respectively.

REFERENCES

[1] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[2] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 911–923, 2019.

[3] X. Xu, Z. Shi, B. Pan, and X. Li, "A classification-based model for multi-objective hyperspectral sparse unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9612–9625, 2019.

[4] B. Pan, Z. Shi, X. Xu, T. Shi, N. Zhang, and X. Zhu, "Coinnet: Copy initialization network for multispectral imagery semantic segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 5, pp. 816–820, 2019.

[5] X. Lu, H. Sun, and X. Zheng, "A feature aggregation convolutional neural network for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 7894–7906, 2019.

[6] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 2, pp. 1155–1167, 2019.

[7] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14 680–14 707, 2015.

[8] R. M. Anwer, F. S. Khan, J. van de Weijer, M. Molinier, and J. Laaksonen, "Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification," *ISPRS journal of photogrammetry and remote sensing*, vol. 138, pp. 74–85, 2018.

[9] B. Pan, X. Xu, Z. Shi, N. Zhang, H. Luo, and X. Lan, "Dssnet: A simple dilated semantic segmentation network for hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, 2020.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010, pp. 270–279.

[12] Y. Yuan, J. Fang, X. Lu, and Y. Feng, "Remote sensing image scene classification using rearranged local features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1779–1792, 2019.

[13] J. Xie, N. He, L. Fang, and A. Plaza, "Scale-free convolutional neural network for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[14] Y. Liu and C. Huang, "Scene classification via triplet networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 1, pp. 220–237, 2017.

[15] K. Nogueira, O. A. Penatti, and J. A. dos Santos, "Towards better exploiting convolutional neural networks for remote sensing scene classification," *Pattern Recognition*, vol. 61, pp. 539–556, 2017.

[16] Y. Liu, Y. Zhong, F. Fei, Q. Zhu, and Q. Qin, "Scene classification based on a deep random-scale stretched convolutional neural network," *Remote Sensing*, vol. 10, no. 3, p. 444, 2018.

[17] F. Zhang, B. Du, and L. Zhang, "Scene classification via a gradient boosting random convolutional network framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1793–1802, 2016.

[18] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *arXiv preprint arXiv:1508.00092*, 2015.

[19] G. J. Scott, M. R. England, W. A. Starms, R. A. Marcum, and C. H. Davis, "Training deep convolutional neural networks for land–cover classification of high-resolution imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 4, pp. 549–553, 2017.

[20] Z. Fang, W. Li, J. Zou, and Q. Du, "Using cnn-based high-level features for remote sensing scene classification," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2016, pp. 2610–2613.

[21] Q. Weng, Z. Mao, J. Lin, and W. Guo, "Land-use classification via extreme learning classifier based on deep convolutional features," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 5, pp. 704–708, 2017.

[22] 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 Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 23–43, 2018.

[23] X. Dai, X. Wu, B. Wang, and L. Zhang, "Semisupervised scene classification for remote sensing images: A method based on convolutional neural networks and ensemble learning," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 869–873, 2019.

[24] S. Chaib, H. Liu, Y. Gu, and H. Yao, "Deep feature fusion for vhr remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4775–4784, 2017.

[25] G. Wang, B. Fan, S. Xiang, and C. Pan, "Aggregating rich hierarchical features for scene classification in remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 9, pp. 4104–4115, 2017.

[26] J. Fang, Y. Yuan, X. Lu, and Y. Feng, "Robust space-frequency joint representation for remote sensing image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2019.

[27] J. Zhang, M. Zhang, L. Shi, W. Yan, and B. Pan, "A multi-scale approach for remote sensing scene classification based on feature maps selection and region representation," *Remote Sensing*, vol. 11, no. 21, 2019.

[28] E. Li, J. Xia, P. Du, C. Lin, and A. Samat, "Integrating multilayer features of convolutional neural networks for remote sensing scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5653–5665, 2017.

[29] Q. Liu, R. Hang, H. Song, and Z. Li, "Learning multiscale deep features for high-resolution satellite image scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 117–126, 2018.

[30] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[31] Q. Wang, J. Gao, and X. Li, "Weakly supervised adversarial domain adaptation for semantic segmentation in urban scenes," *IEEE Transactions on Image Processing*, vol. 28, no. 9, pp. 4376–4386, 2019.

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

[33] L. Zhang, S. Wang, G.-B. Huang, W. Zuo, J. Yang, and D. Zhang, "Manifold criterion guided transfer learning via intermediate domain generation," *IEEE transactions on neural networks and learning systems*, 2019.

[34] E. Othman, Y. Bazi, F. Melgani, H. Alhichri, N. Alajlan, and M. Zuair, "Domain adaptation network for cross-scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4441–4456, 2017.

[35] N. Ammour, L. Bashmal, Y. Bazi, M. M. Al Rahhal, and M. Zuair, "Asymmetric adaptation of deep features for cross-domain classification in remote sensing imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 4, pp. 597–601, 2018.

[36] 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 Geoscience and Remote Sensing Letters*, 2019.

[37] 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 Geoscience and Remote Sensing Letters*, vol. 16, no. 8, pp. 1324–1328, 2019.

[38] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1859–1867.

[39] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. Mar, pp. 723–773, 2012.

[40] S. Li, S. Song, G. Huang, Z. Ding, and C. Wu, "Domain invariant and class discriminative feature learning for visual domain adaptation," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4260–4273, 2018.

[41] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Advances in neural information processing systems*, 2007, pp. 137–144.

[42] J. Wang, W. Feng, Y. Chen, H. Yu, M. Huang, and P. S. Yu, "Visual domain adaptation with manifold embedded distribution alignment," in *2018 ACM Multimedia Conference on Multimedia Conference*. ACM, 2018, pp. 402–410.

[43] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 7, pp. 1414–1430, 2017.

[44] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[45] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2321–2325, 2015.

[46] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

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

[48] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2960–2967.

[49] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[50] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 2200–2207.

[51] ——, "Transfer joint matching for unsupervised domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1410–1417.

[52] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.