Domain Adaptation in Remote Sensing Image Classification: A Survey

Jiangtao Peng ⓘ , *Senior Member, IEEE*, Yi Huang, Weiwei Sun ⓘ , *Senior Member, IEEE*, Na Chen, Yujie Ning, and Qian Du ⓘ , *Fellow, IEEE*

Abstract—Traditional remote sensing (RS) image classification methods heavily rely on labeled samples for model training. When labeled samples are unavailable or labeled samples have different distributions from that of the samples to be classified, the classification model may fail. The cross-domain or cross-scene remote sensing image classification is developed for this case where an existing image for training and an unknown image from different scenes or domains for classification. The distribution inconsistency problem may be caused by the differences in acquisition environment conditions, acquisition scene, acquisition time, and/or changing sensors. To cope with the cross-domain remote sensing image classification problem, many domain adaptation (DA) techniques have been developed. In this article, we review DA methods in the fields of RS, especially hyperspectral image classification, and provide a survey of DA methods into traditional shallow DA methods (e.g., instance-based, feature-based, and classifier-based adaptations) and recently developed deep DA methods (e.g., discrepancy-based and adversarial-based adaptations).

Index Terms—Cross-domain classification, distribution difference, domain adaptation (DA), remote sensing (RS) image.

I. BACKGROUND

AS a technique to acquire information on the Earth's surface, remote sensing (RS) has been successfully applied in many fields [1], such as ecology, environmental geology, atmospheric science, etc. During the past few decades, many satellites were launched, such as MODIS, LandSat, WorldView, GaoFen, etc. Except for the spaceborne sensors, there are many airborne and unmanned aerial vehicle (UAV) borne sensors. Different types of sensors (e.g., hyperspectral, multispectral, LiDAR) provide a plenitude of RS images for multitemporal and multisource

observations, which can be used for large-scale and long-term applications.

Although different types of RS images are available, they are difficult to be collaboratively processed due to the existence of large distribution differences (e.g., different sensors and acquisition conditions) among these images. In particular, for RS image classification, we usually want to build a model on a known image to classify an unknown one. If these two images have different data distributions, traditional classification methods may not provide satisfying results. Fortunately, if distribution of two images are related, we can use a domain adaptation (DA) technique to build connection between images and transfer knowledge from one image to the other. For RS image processing, there are various cases where two images have different but related distributions [2], [3]: 1) Difference in sensors: two images are acquired by two different sensors on the same scene; 2) difference in spatial locations (bias sampling): two images correspond to two disjoint regions in a large scene; 3) difference in scenes: two images correspond to two different scenes with similar materials; 4) difference in acquisition conditions (atmospheric, illumination, or acquisition angle): two images are acquired in different imaging conditions; 5) difference in acquisition times: two images are acquired in different times and the ground materials have been changed. In all these cases, two images have some related characteristics, such as the same or similar scenes, the same sensor or similar acquisition condition. Thanks to such correlation, the data inconsistency problem can be solved by the DA technique, which forms the differences in sensors or imaging environmental conditions into a data or feature transfer problem.

DA aims to solve the distribution discrepancy between domains [4], [5]. Depending on the availability of target labels, the DA methods can be categorized into unsupervised, semisupervised, and supervised methods, in which an unsupervised DA method with no label in the target domain is a hotspot because it matches many actual situations. In early research of unsupervised DA, scholars mainly focus on traditional DA methods that aim to align distributions from the aspects of instance, feature, or classifier. The instance-based methods mainly consider to adjust the marginal distribution of source or target samples. Feature-based methods align the subspace features of different domains to minimize their distribution differences. Classifier-based methods mainly aim to adapt a classifier trained on the source domain to the target domain. In recent years, many deep learning-based DA methods are proposed [6].

Jiangtao Peng, Yi Huang, Na Chen, and Yujie Ning are with the Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China (e-mail: pengjt1982@126.com; 202021104010021@stu.hubu.edu.cn; chenna0407@aliyun.com; 201722110411029@stu.hubu.edu.cn).

Weiwei Sun is with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China (e-mail: nbsww@outlook.com).

Qian Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: du@ece.msstate.edu).

TABLE I
TAXONOMY OF DA METHODS

CATEGORY	SUBCATEGORY	TYPICAL EXAMPLES
Shallow DA	Instance-based	nonparametric KMM [7], RHM [8], IWATL [9], CSSPL [10], ELM-based transfer [11]
	Feature-based	(1) Subspace-based Adaptation: SGF [12], GFK [13], coclustering [14], SA [15]–[17], TSSA [18], TA [19], TPCA [20], SCA [21], SDA [22], DCA [23], CS-DDA [24], KSA [25], IRDMKSA [26], HFAA [27], IRHTL [28], invariant-feature-based [29]–[32], dictionary learning-based [33], PCDA-NSD [34], DSTL [35], SS-DDNMF [36], GFP [37], (2) Transformation-based Adaptation: TCA and SSTCA [38], [39], TJM [40], DTJM [41], JDA [42], JGSA [43], DATL [44], [45], LPJT [46], GEDA [47], DADFL [48], CORAL [49], [50], CCCA [51], [52], CCA-based [53]–[56], MA-based [57]–[62], MRDA [63], DABL [64], SSWK-MEDA [65], NFNalign [66], OT [67], 3-D Gabor transformation [68]
	Classifier-based	ML-based [69], multiple cascade-classifier [70]–[72], invariant SVM [73], SCT-SVM [74], multiple-kernel learning [75]–[78], ELM-based [79]–[83], MDAF and MBCF [84], open set DA [85], EasyTL [86], BHC [87], DASVM [88], MRC [89], AL-based [90]–[97]
Deep DA	Discrepancy-based	DAN [98], JAN [99], MRAN [100], DSAN [101], DeepCORAL [102], DNN with class centroid alignment [103], TCANet [104], class-wise distribution alignment based deep DA [105], DDA-Net [106], TDDA [107], TSTnet [108], GNN [109], AMF-FSL [110], MSCN [111], AMRAN [112], DJDANs [113]
	Adversarial-based	GAN [114], [115], adversarial CNN [116], MADA [117], DAAN [118], MCD [119], DWL [120], GAN with VAE-based generator [121], [122], content-wise alignment [123], class reconstruction driven adversarial [124], class-wise adversarial [125], ADADL [126], DABAN [127], UDAD [123], deep metric learning [128], DCFSL [129], CDADA [130], DFENet [131], TriADA [132], SCDAL [133], TSAN [134], DANN [135], [136], StandardGAN [137]
	Others	two-branch CNN [138], FANN [139], DATL [140], MSSN [141], active transfer learning network [140], AALDA [142], attention-based residual network [143], 3CN [144], bayesian-inspired CNN [145]

By means of deep network architecture, deep DA methods can automatically extract deep features from domains and further learn transferrable features by adding feature adaptation layers in an original deep network architecture or constructing feature learning modules (e.g., adversarial learning). Considering different network architectures, deep DA methods are mainly categorized into discrepancy-based methods and adversarial-based methods [5]. Although these traditional and deep DA methods are widely applied for computer vision tasks, their applicability on RS images is not clear. In this article, we provide a review on these unsupervised DA methods and test their performance on cross-domain RS image classification.

The rest of this article is organized as follows. Section II introduces notations. Sections III, IV, and V describe the shallow DA methods, such as instance-based, feature-based, and classifier-based methods, respectively. Section VI presents the deep DA methods, such as discrepancy-based and adversarial-based methods. Section VII provides experimental results. Finally, Section VIII concludes this article.

II. NOTATIONS

Denote source and target domains as D_s and D_t, respectively. The sample and corresponding label sets in D_s are $X_s = \{\mathbf{x}_1^s, \ldots, \mathbf{x}_{n_s}^s\}$ and $Y_s = \{y_1, y_2, \ldots, y_{n_s}\}$, with $\mathbf{x}_i^s \in \mathcal{R}^D$ and $y_i^s \in \{1, 2, \ldots, C\}$, where n_s is the number of labeled samples,

D is the dimensionality and C is the number of classes. The target domain only has the unlabeled dataset $X_t = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_{n_t}^t\}$, where n_t is the number of target samples.

DA considers the classification problems where the class space of source and target domains are the same but the distribution between domains are different but related. The objective of DA is to classify target samples using the model built on source samples.

The DA methods can be categorized as traditional shallow DA methods and recent deep DA methods. The traditional DA methods can be instance-based, feature-based, and classifier-based DA methods. In the following, they are introduced in detail. Table I shows the taxonomy of DA methods discussed in this article.

III. INSTANCE-BASED METHODS

Instance-based DA methods mainly adjust the marginal distribution of source or target samples such that the distribution of domains are aligned. Let $p_s(x)$ and $p_t(x)$ be the marginal density distribution of source and target samples, respectively, an importance weight can be defined as [5]

$$w(x) = \frac{p_t(x)}{p_s(x)}. \tag{1}$$

By adjusting the reweighting factor $w(x)$, the sample selection bias and covariate shift problems can be alleviated to
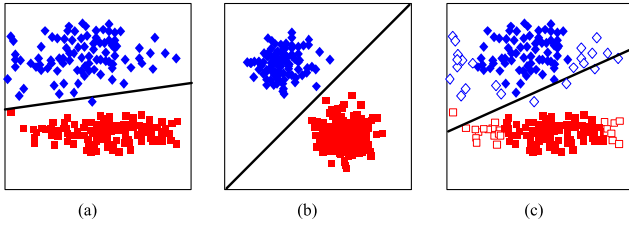
Fig. 1. Illustration of instance reweighting. Two classes are denoted as blue diamonds and red squares. The black line in each subfigure is the classification decision line. (a) Source domain. (b) Target domain. (c) Source instance reweighting.

certain extent [146], [147]. As shown in Fig. 1, the instance reweighting strategy reweights the source data [i.e., the solid points in Fig. 1(c) have large weights] to minimize the marginal distribution difference between domains and then a classifier built on the reweighted source data [i.e., the black line in Fig. 1(c)] can be used to classify target samples.

To solve the sample selection bias, Huang et al. [7] presented a nonparametric kernel mean matching (KMM) method to directly produce resampling weights without distribution estimation. Yaras et al. proposed a randomized histogram matching (RHM) method to augment training data to describe domain shifts of satellite images. In detail, they analyzed different reasons for the domain shift, such as changing sensors, illumination variations, and imaging conditions, and modeled these factors as nonlinear pixelwise transformations, and then training data augmentation with deep neural networks was employed to increase the model robustness to these transformations [8]. Cui et al. [9] proposed an iterative weighted active transfer learning framework (IWATL) for hyperspectral image (HSI) classification. It weighted the source samples by considering the distance between the samples and the classification hyperplane as well as the similarity between the source and target distributions. Li et al. [10] proposed a cost-sensitive self-paced learning (CSSPL) framework for the classification of multitemporal images, which automatically assigned sample weight via a mixture weight regularizer. To reuse a large number of existed labeled images, a historical and target training data weighting strategy was proposed in an extreme learning machine (ELM)-based RS image transfer classification framework [11].

In the instance-based DA methods, source or/and target sample reweighting and landmark selection are widely used strategies [146], [147]. These strategies can also be embedded into the feature-based and classifier-based adaptations for domain-invariant feature learning or classifier refinement [25], [46], [148], respectively.

IV. FEATURE-BASED METHODS

Feature-based DA methods transform source and target data into a feature space such that the data distribution of both domains in the feature space are similar. Then, source features and corresponding labels can be used to generate a classifier to predict label of target samples. Feature-based adaptation is usually realized by joint feature extraction, and typical methods are subspace-based and transformation-based adaptation methods.
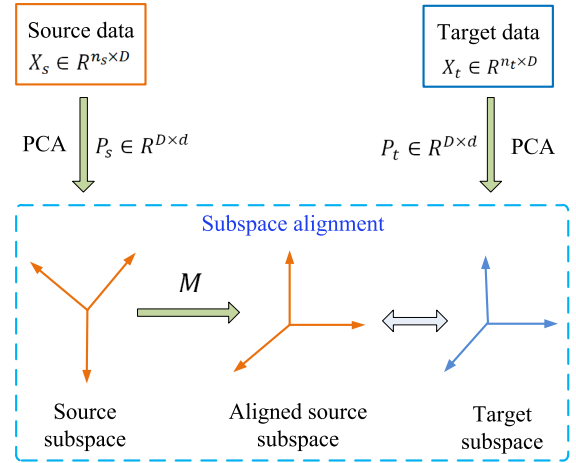

Fig. 2. Illustration of subspace alignment. The PCA is used to generate source and target subspaces, which are then aligned based on a linear transformation matrix M.

A. Subspace-Based Adaptation

Subspace-based DA methods usually project the source and target samples into individual subspaces according to subspace learning or dimensionality reduction methods, and then align the subspaces [5].

Goplan et al. [12] proposed a sampling geodesic flow (SGF) method for DA, which learns intermediate representations of source and target samples via Grassmann manifolds to describe domain shift. However, the SGF approach has several limitations, such as the difficulty in its sampling strategy and the high dimensionality of the new representations. To solve these problems, geodesic flow kernel (GFK) method was proposed [13]. It constructed a GFK to model domain shift and provided a simple solution to compute the kernel. So, the GFK method is easy to implement than SGF [13], [41]. Banerjee et al. [14] proposed a coclustering-based method for DA in the absence of source samples. The samples from both domains are projected in a shared space by the GFK-based projection, followed by a probabilistic support vector machine (SVM)-based iterative coclustering method.

Fernando et al. [15] proposed a subspace alignment (SA) method, which first employed the principal component analysis (PCA) to generate individual subspaces for source and target domains and then learned a linear transformation M to align these subspaces

$$M = \arg\min_{M} \left\| P_s M - P_t \right\|_F^2 = P_s^{\mathrm{T}} P_t \qquad (2)$$

where $\| \cdot \|_F^2$ is the Frobenius norm, and $P_s, P_t \in R^{D \times d}$ are the low-dimensional representations (i.e., basis vectors) of source and target data, respectively. The procedure of SA is illustrated in Fig. 2.

Sun et al. [16] directly applied the SA for cross-view RS scene classification, where the partial least squares (PLS) method was used to generate discriminative subspace of source domain. They further proposed a transfer sparse subspace analysis (TSSA)

algorithm for unsupervised cross-view RS scene classification [18]. It minimized the maximum mean discrepancy (MMD) distance between domains and preserved the self-expressiveness property of the data in a reproducing kernel Hilbert space (RKHS) according to sparse subspace clustering. Wei et al. [17] proposed a robust DA method for HSIs which employed SA to perform subspace feature-level alignment. The SA was extended to a tensor alignment (TA) for HSI classification [19], where tensors of source and target domains were constructed and SA was performed on the tensors. Gao et al. [20] proposed an unsupervised tensorized principal component alignment framework for multimodal RS image classification. Gui et al. [21] proposed a statistical scattering component-based SA for cross-domain polarimetric synthetic aperture radar (PolSAR) image classification.

It can be seen that SA only aligns the subspace bases without considering the distributions of subspaces. To incorporate the distribution alignment into SA, a subspace distribution alignment (SDA) method was proposed to align both subspace bases and subspace distributions [22]. For the cross-scene classification of HSIs, a discriminative cooperative alignment (DCA) method was proposed to alleviate spectral shift [23]. In the DCA, SA and distribution alignment work cooperatively through the subspace correlation constraint and MMD [23]. Zhang et al. [24] proposed a correlation subspace dynamic distribution alignment (CS-DDA) method for RS scene classification, which maximizes the correlation between source and target subspaces and meanwhile dynamically minimizes the statistical distribution difference between domains.

To handle nonlinearity, Aljundi et al. [25] extended the linear SA to kernel-based SA (KSA), The kernel of source and target domains are first constructed on the selected landmarks, and then the SA is performed on the source and target kernels to align the kernel-based subspaces [25]. To further exploit the source labels and multiple kernel representations, an ideal regularized discriminative multiple kernel subspace alignment (IRDMKSA) was proposed for HSI classification [26]. It performs SA in the composite-kernel-based spaces to reduce the distribution differences of domains.

Traditional subspace learning-based strategies usually assume the existence of a single subspace for both domains. However, such an assumption may not be true in many scenarios due to the diversity in the statistical properties of the underlying classes [149]. Banerjee et al. [149] proposed a hierarchical subspace learning-based unsupervised DA technique for multitemporal RS image classification, where node-specific subspaces are learned from a binary-tree. Shen et al. [27] presented a hyperspectral feature adaptation and augmentation (HFAA) method for cross-scene HSI classification, which iteratively learns a common subspace by introducing two separate projection matrices and augments it with a feature selection strategy. Li et al. [28] proposed an iterative reweighting heterogeneous transfer learning (IRHTL) framework, which iteratively learns a shared space of source and target data based on a weighted SVM and conducts an iterative reweighting strategy to reweight the source samples.

Invariant feature-based methods can be regarded as a special case of subspace-based adaptation. It aims to select a set of features that are not affected by shifting factors. The selected features can form a new subspace. Bruzzone et al. [29] proposed a multiobjective optimization framework to select spatially invariant features for the classification of spatially disjoint scenes. The multiobjective framework ensures the selected features with both high discrimination ability and high spatially invariance. The invariant feature selection can also be performed in an RKHS [30]. Paris et al. [31] presented an invariant-feature-based sensor-driven hierarchical DA method. Yan et al. [32] proposed a TrAdaBoost based on an improved particle swarm optimization (PSO) method for cross-domain scene classification, which can select an optimal feature subspace for classifying "harder" and "easier" instances.

There are other subspace-based adaptation methods. Ye et al. [33] proposed a dictionary learning-based feature-level DA technique, which learns a common dictionary to represent source and target data and then aligns their representation coefficient features to reduce the spectral shifts between domains. Wang et al. [34] proposed a pairwise constraint discriminant analysis and nonnegative sparse divergence (PCDA-NSD) method for HSI classification. The PCDA learned potential discriminant information of sample sets in the source and target domains by using pairwise constraints and NSD measured the divergence between different distributions. Lin et al. [35] proposed a dual space unsupervised structure preserving transfer learning (DSTL) framework for HSI classification. It first transfers the data on both domains to a specific subspace, on which the initial classification results for the target HSI are obtained. Then, the initial results on the original target data space are optimized by applying the Markov random field (MRF) approach. Chen et al. [36] proposed a semisupervised dual-dictionary nonnegative matrix factorization (SS-DDNMF) method for heterogeneous transfer learning on cross-scene HSIs, where two different dictionaries are designed for source and target scenes to project two different feature spaces into a shared subspace. Gui et al. [37] proposed a general feature paradigm (GFP) for PolSAR image classification, where information scattering and statistical information are used to reduce the domain shifts.

B. Transformation-Based Adaptation

The transformation-based DA methods transform the original data into new representations to minimize the statistical distribution (i.e., marginal and conditional distributions) discrepancy and geometrical divergence between domains while preserving the underlying structure of original data [5], as shown in Fig. 3.

The MMD, Kullback–Leibler divergence (KL-divergence), or Bregman divergence are usually used to measure the domain's distribution discrepancy [38]. The MMD is defined as

$$\text{MMD}(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} \phi(\mathbf{x}_i) - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} \phi(\mathbf{x}_j) \right\|_F^2 \quad (3)$$

where ϕ is a nonlinear map induced by a universal kernel. Pan et al. [38] introduced the MMD to measure the distribution
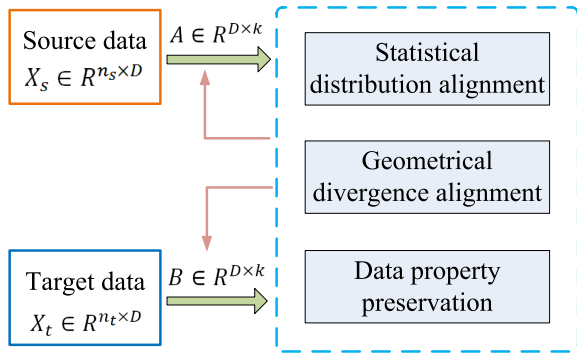
Fig. 3. Illustration of transformation-based adaptation. The transformation matrices A and B can be the same or different.

discrepancy and proposed a transfer component analysis (TCA) method. TCA intends to learn a set of transfer components in an RKHS using the MMD such that marginal distribution differences between domains are reduced and data variance is maximized [38]. The transformation matrix $W \in R^{(n_s+n_t) \times d}$ in the TCA can be solved by the following optimization problem:

$$\min_{W} \text{tr}(W^{\text{T}} \mathbf{KLKW}) + \mu\, \text{tr}(W^{\text{T}}W) \tag{4}$$

$$\text{s.t. } W^{\text{T}} \mathbf{KHKW} = I_m \tag{5}$$

where μ is a regularization parameter, $I_m \in R^{m \times m}$ is an identity matrix, H is the centering matrix, K is the kernel matrix defined on all source and target data. Here, $L_{i,j} = 1/n_s^2$ if $\mathbf{x}_i, \mathbf{x}_j \in X_s$; $L_{i,j} = 1/n_t^2$ if $\mathbf{x}_i, \mathbf{x}_j \in X_t$; otherwise, $L_{i,j} = -1/(n_s n_t)$.

The unsupervised TCA can be extended to semisupervised TCA (SSTCA) by using the source labels [38]. Matasci et al. [39] directly applied the TCA and SSTCA for the DA of RS image classification. Long et al. [40] proposed a transfer joint matching (TJM) method, which performs feature matching and instance reweighting simultaneously in a unified optimization framework to reduce marginal distribution differences between domains. Peng et al. [41] proposed a discriminative transfer joint matching (DTJM) for HSI classification by considering the label information of source domain.

To align both the marginal and conditional distributions between domains, Long et al. [42] further proposed a joint distribution adaptation (JDA) method. JDA finds a linear transformation $A \in R^{D \times d}$ to align the marginal distribution based on the MMD

$$\text{MMD}(X_s, X_t) = \left\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} A^{\text{T}} \mathbf{x}_i - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} A^{\text{T}} \mathbf{x}_j \right\|_F^2$$

$$= \text{tr}(A^{\text{T}} X M_0 X^{\text{T}} A) \tag{6}$$

and to align the conditional distribution based on a conditional MMD

$$\text{CMMD}(X_s^c, X_t^c) = \left\| \frac{1}{n_s^c} \sum_{\mathbf{x}_i \in X_s^c} A^{\text{T}} \mathbf{x}_i - \frac{1}{n_t^c} \sum_{\mathbf{x}_j \in X_t^c} A^{\text{T}} \mathbf{x}_j \right\|_F^2$$

$$= \text{tr}(A^{\text{T}} X M_c X^{\text{T}} A), \quad c = 1, \dots, C \tag{7}$$

where M_0 and M_c are the MMD matrices [42].

The above TCA, TJM, and JDA methods assume that there exists a unified transformation A to map source and target samples into a common space. However, if domain shift is large, it is very different to find a common transformation [43]. Zhang et al. [43] proposed a joint geometrical and statistical alignment (JGSA) method to learn two coupling mappings A and B for source and target domains, respectively. The distribution divergence minimization in the JGSA can be represented as

$$\min_{A,B} \left\| \frac{1}{n_s} \sum_{\mathbf{x}_i \in X_s} A^{\text{T}} \mathbf{x}_i - \frac{1}{n_t} \sum_{\mathbf{x}_j \in X_t} B^{\text{T}} \mathbf{x}_j \right\|_F^2 \tag{8}$$

$$\min_{A,B} \sum_{c=1}^{C} \left\| \frac{1}{n_s^c} \sum_{\mathbf{x}_i \in X_s^c} A^{\text{T}} \mathbf{x}_i - \frac{1}{n_t^c} \sum_{\mathbf{x}_j \in X_t^c} B^{\text{T}} \mathbf{x}_j \right\|_F^2. \tag{9}$$

Inspired by the JGSA, Zhou et al. [44], [45] proposed a DA technique based on transformation learning (DATL) for HSI classification. It learned two different transformations using the idea of linear discriminant analysis (LDA) to minimize the the ratio of within-class distance to between-class distance. A distance-based objective function is designed to optimize the transformations and meanwhile to preserve stochastic neighborhood and discriminative information of domains in a latent space [44], [45]. Li et al. [46] proposed a locality preserving joint transfer (LPJT) method to improve the JGSA by considering the local discriminative information preservation and landmark selection in a unified optimization framework. Huang et al. [47] proposed a graph embedding and distribution alignment (GEDA) method for HSI classification, which used the graph embedding method to preserve discriminative information of source and target domains and a pseudo-label learning method to refine the target pseudo labels [47]. Similarly, they further proposed a distribution alignment and discriminative feature learning (DADFL) method [48], which performs classwise discriminative information preservation and uses a structural prediction method to learn pseudo-label of target samples.

Sun et al. [49] proposed a correlation alignment (CORAL) method, which finds a linear transformation A to transform the source data such that the variances of transformed source data and target data can be minimized

$$\min_{A} \|C_{\hat{S}} - C_T\|_F^2 = \min_{A} \|A^{\text{T}} C_S A - C_T\|_F^2 \tag{10}$$

where $C_{\hat{S}}$ is the covariance of the transformed source features $A^{\text{T}} X_S$. The solution to (10) has the closed-form: $A = C_S^{-1/2} C_T^{1/2}$. Peng et al. [50] proposed a sparse matrix transform-based CORAL method for HSI classification. Zhu et al. [51] proposed a class centroid alignment method, which aligns the class centroids by moving the target domain samples toward the source domain. To consider the first- and second-order statistical alignment, a class centroid and covariance alignment (CCCA) method was developed for classification of RS images [52]. The proposed method included three main steps: 1) spatial filtering preprocessing, 2) overall centroid alignment-based coarse adaptation, and 3) CCCA-based refined adaptation.

Many canonical correlation analysis (CCA)-based transformation methods were proposed for DA. Qin et al. proposed a

cross-domain collaborative learning (CDCL) method for heterogeneous DA of HSIs. It consisted of three parts, i.e., random walker (RW)-based pseudo-labeling, cross-domain learning via cluster canonical correlation analysis (C-CCA), and final classification based on extended RW (ERW) algorithm [53]. Samat et al. [54] proposed a supervised and semisupervised multiview CCA ensemble method for heterogeneous DA in RS image classification. Li et al. [55] proposed a sparse subspace correlation analysis-based supervised classification (SSCA-SC) method for HSI classification, which integrated the idea of CCA into a sparse representation subspace learning framework and directly classified the target samples based on the sparse representation reconstruction residuals. Volpi et al. [56] proposed a kernel CCA transformation (kCCA) method to align spectral characteristics of multitemporal cross-sensor images for change detection.

To correct nonlinear variation between domains, Tuia et al. constructed a nonlinear transform based on vector quantization and graph matching to describe the data changes under different acquisition conditions [57]. They further proposed a semisupervised manifold alignment (SS-MA) method to align the manifolds of RS images [58] by solving a standard Rayleigh quotient

$$F^{\text{opt}} = \arg\min_F \left\{ \text{tr}\left((F^{\text{T}}VF)^{-1}F^{\text{T}}UF \right) \right\} \qquad (11)$$

where the affinity matrix V can be used to maximize the distances between samples of different classes, and U enhances class similarity between the labeled instances between domains. The matrices U and V can be constructed based on the graph Laplacian. The SS-MA method can be used for multitemporal, multisource, and multiangular classification. Yang et al. [59] proposed a global aligned local manifold (GALM) method to align two globally similar manifolds and to minimize the impact of spectral changes at the local scale. They further extended the MA and proposed spectral and spatial proximity-based MA for multitemporal HSI classification [60]. Hong et al. [61] proposed a learnable MA (LeMA) for semisupervised cross-modality hyperspectral–multispectral classification. Ma et al. [62] proposed a unsupervised MA method for cross-domain classification of RS images, which used an SVM-prediction-based cross-domain similarity matrix and a per-class MMD constraint. To exploit the manifold structure of data, Luo et al. [63] proposed a manifold regularized distribution adaptation (MRDA) algorithm to minimize the per-class MMD and meanwhile preserve the manifold structure of source and target data in the subspace. Wang et al. [64] proposed a DA broad learning (DABL) method for HSI classification, which combined the DA and broad learning system (BLS) to perform MMD-based distribution alignment and manifold structure preservation. Dong et al. proposed a spectral–spatial weighted kernel manifold embedded distribution alignment (SSWK-MEDA) for RS image classification [65], which applied spatial filter to preprocess the hyperspectral data and constructed a spatial–spectral composite kernel for kernel-based adaptation. Gross et al. [66] proposed a nonlinear feature normalization alignment (NFNalign) transformation to mitigate nonlinear effects in hyperspectral data.

There are also some other transform-based methods. Chakraborty et al. [150] proposed an artificial neural network-based DA strategy, which unified the common data transformation and transfer learning methods. Tardy et al. [67] applied the optimal transport (OT) for land-cover mapping of high-resolution satellite images time series. Jia et al. [68] applied the 3-D Gabor transformation to extract spatial–spectral features of HSI for DA.

V. CLASSIFIER-BASED ADAPTATION

The classifier-based DA methods adapt a classifier trained on a source domain to a target domain by considering unlabeled samples of the target domain.

The classifier-based DA can be performed by the adaptation of classifier parameters [69], [70], [71], [72], [151]. Bruzzone et al. [69] proposed a classifier-based DA method to solve the data distribution difference between multitemporal RS images by updating the parameters of a trained maximum-likelihood (ML) classifier on the basis of the distribution of a new image to be classified. The ML-based DA technique was further extended to the Bayesian cascade classifier, multiple-classifier, and multiple cascade-classifier [70], [71], [72]. Zhong et al. [152] proposed a classifier updating method by considering spectral features and guided-filter-based *posteriori* spatial features. Izquierdo-Verdiguier et al. [73] updated the SVM classifier by adding virtual support vectors (VSVs) for training, where the VSVs contained the invariances to rotations, reflections, and object scale. An SVM-based sequential classifier training (SCT-SVM) approach was proposed for multitemporal RS image classification [74]. By casting the DA as a multitask or multiple-kernel learning problem, many multiple-kernel learning-based DA methods were proposed [75], [76], [77], [78]. Xu et al. [79] proposed a DA method through transferring the parameters of ELM. Considering the simplicity of ELM, many ELM-based classifier adaptation methods were proposed, such as cross domain ELM (CDELM) [80], ELM-based heterogeneous DA [81], interpretable rule-based fuzzy ELM (IRF-ELM) [82], ensemble transfer learning based on ELM (TL-ELM) [83]. Wei et al. investigated the combination of multiple classifiers for DA of RS image classification. The multiple domain adaptation fusion (MDAF) method and the multiple base classifier fusion (MBCF) method were proposed to obtain a more stable classification performance [84]. Zhang et al. [85] considered the open set DA problem for RS scene classification via updating the classifier by exploring transferability and discriminability. Wang et al. [86] proposed an easy transfer learning (EasyTL) approach by exploiting intradomain structures to learn both nonparametric transfer features and classifiers.

Semisupervised learning (SSL) and active learning (AL) techniques can also be used to solve the domain shift problem by updating a source classifier with the use of unlabeled target samples [153], [154]. Rajan et al. [87] proposed a binary hierarchical classifier (BHC) framework for knowledge transfer from an existing labeled source domain to a spatially separate and multitemporal target domain, where an SSL technique was used to update the BHC to reflect the characteristics of new
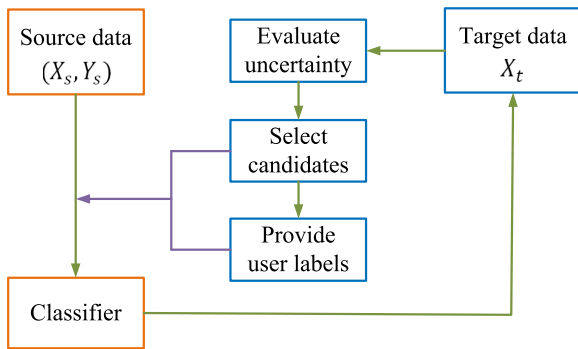
Fig. 4.   Flow chart of the AL method for DA.

Fig. 5.   Flow chart of discrepancy-based deep DA method.

data. The classical supervised SVM was also extended to a semisupervised case to solve the DA problem [88], [155]. A typical semisupervised SVM is the domain adaptation SVM (DASVM) [88], which built a standard SVM on a source domain and then iteratively adjusted the SVM model using unlabeled target samples. Kim et al. proposed an adaptive manifold classifier (MRC) in a semisupervised setting, where a kernel machine was first trained with labeled data and then iteratively adapted to new data using manifold regularization [89]. The AL technique was also adopted to update existing classifiers [90], [91], [92], [93], [94], [95], [96], [97]. As shown in Fig. 4, the AL method first built a classifier on the source data and then classified target samples. By selecting some candidates with the highest uncertainty and providing user labels for them, these samples can be used to expand the training set to update the classifier. Deng et al. [140] proposed an active multikernel DA method for HSI classification, which combines the AL with multikernel learning for DA. Kalita et al. [156] proposed a standard deviation (SD)-based AL technique to exploit the labeled source images to generate the "most-informative" target samples. Saboori et al. [157] proposed an active multiple kernel Fredholm learning (AMKFL) method, where a Fredholm kernel regularized model was presented to label samples.

VI. DEEP DOMAIN ADAPTATION

Unlike hand-crafted features in traditional DA methods, deep learning methods can automatically learn features using deep neural networks (DNNs) [158]. Most of the current deep DA methods add adaptation layers to an original deep network architecture to realize the source-to-target adaptation or adopt an adversarial learning strategy to minimize the cross-domain discrepancy. Deep DA methods are mainly divided into discrepancy-based methods, adversarial-based methods, and others [5], [159].

A. Discrepancy-Based Adaptation

The discrepancy-based deep DA methods mainly aim to match marginal or/and conditional distributions between domains by adding adaptation layers (e.g., MMD-based metric) in deep neural networks (DNN) for task-specific representations, as shown in Fig. 5.
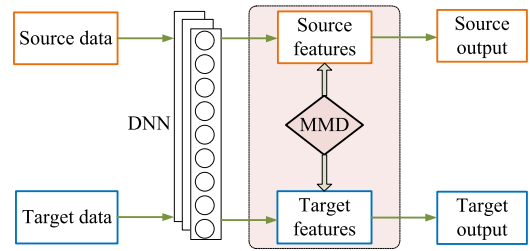
Long et al. [98] proposed a DNN which was the first time to utilize the DNNs to learn transferable features across domains for DA. In DAN, three adaptation layers based on multiple kernel variant of MMD (MK-MMD) are designed to align marginal distributions between domains. In order to reduce both marginal and conditional distribution differences, they further proposed a joint adaptation network (JAN) which used source labels and target pseudo labels to construct MMD and conditional MMD to align the joint distribution according to an adversarial training strategy. Based on DAN, Zhu et al. [99] proposed a multirepresentation adaptation network (MRAN) [100], which performs cross-domain classification tasks through multirepresentation alignment. They further proposed a deep subdomain adaptation network (DSAN) using the idea of subdomain adaptation [101]. Zhu et al. [160] developed a weakly pseudo-supervised decorrelated subdomain adaptation (WPS-DSA) network for cross-domain land-use classification. Sun et al. [102] proposed a DeepCORAL method to extend the CORAL to deep learning.

For HSI classification, Ma et al. [103] designed a class centroid alignment module in the DNN for cross-domain HSI classification. Garea et al. [104] proposed a TCA-based network (TCANet) for DA of HSIs, which used the TCA to construct an adaptation layer. Wang et al. [105] proposed a deep DA with MMD-based classwise distribution alignment and manifold structure preservation in the target domain. Ma et al. [106] proposed a deep DA network (DDA-Net) for cross-dataset HSI classification, which minimized the domain discrepancy and transferred the task-relevant knowledge from source to target in an unsupervised way. Li et al. [107] proposed a two-stage deep DA (TDDA) method, where in the first stage, the distribution distance between domains is minimized based on the MMD to learn a deep embedding space, and in the second stage, a spatial–spectral Siamese network is constructed to learn discriminative spatial–spectral features to further decrease the distribution discrepancy. Zhang et al. [108] proposed a topological structure and semantic information transfer network (TSTnet). It employs the graph structure to characterize topological relationships and combines the graph convolutional network (GCN) and CNN for cross-scene HSI classification. The optimal transmission (OT)-based graph alignment and MMD-based distribution alignment work cooperatively. Wang et al. [109] proposed a graph neural network (GNN) DA method for multitemporal HSIs, which incorporated the domainwise and classwise CORAL into the GNN network to align the joint distributions of domains. Liang et al. [110] proposed an attention multisource fusion-based

deep few-shot learning (AMF-FSL) method for small-sized HSI classification, which contains three modules, namely, the target-based class alignment, domain attention assignment, and multisource data fusion. It can transfer the learned ability of classification from multiple source data to target data.

Othman et al. [161] proposed a DA network for cross-scene classification. It uses a pretraining and fine-tuning strategy to ensure that the network can correctly classify the source samples, and align source and target distributions and preserve the geometrical structure of target data [161]. Similarly, Lu et al. [111] proposed a multisource compensation network (MSCN) for cross-scene classification task. In the network, a cross-domain alignment module and a classifier complement module are designed to reduce the domain shift and to align categories in multiple sources, respectively. Zhu et al. [112] proposed an attention-based multiscale residual adaptation network (AMRAN) for cross-scene classification, which contains a residual adaptation module for marginal distribution alignment, an attention module for robust feature extraction, and a multiscale adaptation module for multiscale feature extraction and conditional distribution alignment [112]. Geng et al. [113] proposed a deep joint distribution adaptation networks (DJDANs) for transfer learning in SAR image classification, where marginal and conditional distribution adaptation networks are developed.

B. Adversarial-Based Adaptation

Inspired by generative adversarial nets (GAN) [114], [115], adversarial DA approaches learn transferable and domain invariant features through adversarial learning. GAN contains a generator model G and a discriminator model D. The generator aims to produce samples similar to the source domain, and to confuse the discriminator to make a wrong decision. Among them, the training purpose and process can be summarized

$$\min_{G,C} \mathcal{L}_{\text{cls}}(\mathbf{X}_s, Y_s)$$

$$= -\mathbb{E}_{(\mathbf{x}_s, y_s) \sim (\mathbf{X}_s, Y_s)} \sum_{k=1}^{K} \mathbf{1}_{[k=y_s]} \log C(G(\mathbf{x}_s))$$

$$\min_{D} \mathcal{L}_{\text{adv}_D}(\mathbf{X}_s, \mathbf{X}_t, G) = -\mathbb{E}_{\mathbf{x}_s \sim \mathbf{X}_s}[\log D(G(\mathbf{x}_s))]$$
$$- \mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log(1 - D(G(\mathbf{x}_t)))]$$

$$\min_{G} \mathcal{L}_{\text{adv}_G}(\mathbf{X}_s, \mathbf{X}_t, D) = -\mathbb{E}_{\mathbf{x}_t \sim \mathbf{X}_t}[\log D(G(\mathbf{x}_t))]$$

where C is the classifier, $\mathcal{L}_{\text{cls}}$ is the classification loss on labeled source data, $\mathcal{L}_{\text{adv}_G}$ and $\mathcal{L}_{\text{adv}_D}$ are the loss functions of the adversarial training of G and D. The discriminator then tends to discriminate between the true source data and the counterfeits generated by model G [5]. After GAN-based adaptation, a task-specific classifier built on the source domain can be used to classify target samples. Fig. 6 illustrates the GAN-based DA.

Tzeng et al. [116] introduced an adversarial CNN-based architecture that aligned distributions between domains by minimizing the classification loss, soft label loss, domain classifier
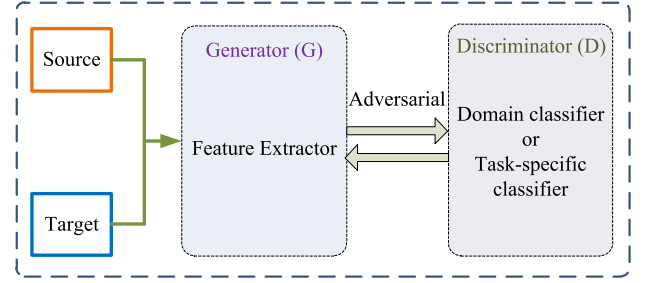

Fig. 6. GAN for domain adaptation.

loss, and domain confusion loss. Pei et al. [117] proposed a multiadversarial DA (MADA) method that constructed multiple classwise domain discriminators to reduce the joint distribution difference between domains. Yu et al. [118] proposed a dynamic adversarial adaptation network (DAAN), which can dynamically assess the relative importance of global and local domain distributions. Saito et al. [119] proposed a deep DA method based on the maximum classifier discrepancy (MCD) that leverages task-specific decision boundaries and adversarial learning ideas to adjust the distribution of source and target domains. The MCD method aims to learn domain-invariant features, which may have lower discriminative ability. To solve this problem, a dynamic weighted learning (DWL) method was proposed to adjust the weights of domain alignment learning and class discrimination learning in the MCD framework [120].

Recently, many adversarial-based DA approaches were developed for HSI classification. Ma et al. [121] proposed an adversarial learning-based DA method for the classification of HSIs, which included a variational autoencoder (VAE)-based generator and a multiclassifier-based discriminator. The generator learns features such that the source classification error is minimized and the classification disagreement on the target dataset is maximized. The discriminator deceives the generator by adjusting classifiers such that the classification disagreement on the target data set is minimized. Miao et al. [122] also used the VAE module to construct a generative model and further designed a joint distributions alignment module to perform coarse-to-fine joint distributions alignment for HSI classification. Yu et al. [123] proposed a contentwise alignment method within an adversarial learning framework. Pande et al. [124] proposed a class reconstruction driven adversarial DA method, which incorporates an additional class-level cross-sample reconstruction loss to make the learned space classwise compactness and an additional orthogonality constraint over the source domain to avoid any redundancy within the encoded features. Liu et al. [125] proposed a classwise adversarial adaptation network for HSI classification, which performed classwise adversarial learning. Saboori et al. [126] proposed an adversarial discriminative active deep learning (ADADL) method for HSI classification. Similar to MCD, it incorporates two different land-cover classifiers as a discriminator to consider class boundaries when aligning feature distributions, and combines the entropy measure along with the cross-entropy loss during training to use the information in unlabeled target data [126]. Wang et al. [127] proposed a

domain adversarial broad adaptation network (DABAN) for HSI classification. It included a domain adversarial adaptation network (DAAN) and a conditional adaptation broad network (CBAN), which can align the statistical distribution between domains and also enhance the representation ability of domain-invariant features. Yu et al. [123] proposed an unsupervised DA architecture with dense-based compaction (UDAD) for cross-scene HSI classification. It incorporated spectral–spatial feature compaction, unsupervised DA, and classifier training into an integrated framework and utilized adversarial domain learning to reduce the domain discrepancy. Deng et al. [128] proposed a deep metric learning-based feature embedding method for HSI classification, which uses an adversarial learning strategy to align source and target features and to preserve the similar clustering structure of source and target features. Fang et al. [162] developed a confident learning (CL)-based DA (CLDA) for HSI classification, where the CL module is designed to select high-confidence pseudo-labeled target samples. Li et al. [129] proposed a deep cross-domain few-shot learning (DCFSL) method for HSI classification, which combines FSL and DA, where a conditional adversarial DA is employed to reduce domain shift and FSL is used to learn transferable knowledge from source to target for classification.

Adversarial-based DA approaches also used for the scene classification of RS images. Teng et al. [130] presented a classifier-constrained deep adversarial domain adaptation (CDADA) method exploiting the idea of MCD for cross domain semisupervised classification of RS scene images, where a deep convolutional neural network (DCNN) is used to build feature representations and adversarial DA is used to align the feature distribution of domains. Zhang et al. [131] proposed a domain feature enhancement network (DFENet) to enhance the discriminative ability of the learned features for dealing with the domain variances of scene classification. Specifically, a context-aware feature refinement module is first designed to recalibrate global and local features by explicitly modeling interdependencies between the channel and spatial for each domain. Then, a multilevel adversarial dropout module is further designed to strengthen the generalization capability of the network. Yan et al. [132] proposed a triplet adversarial domain adaptation (TriADA) method for pixel-level classification of very high resolution (VHR) RS images, which learned a domain invariant classifier by a domain similarity discriminator. Zhu et al. [133] proposed a semisupervised center-based discriminative adversarial learning (SCDAL) method for cross-domain scene classification of aerial images using adversarial learning with center loss. Liu et al. [163] proposed an unsupervised adversarial DA network for remotely sensed scene classification, where a GAN model-based feature extractor makes the source and target distributions closer, and a transferred classifier trained by transferred source domain features is able to acquire a better classification accuracy on the target domain. Zheng et al. [134] proposed a two-stage adaptation network (TSAN) for RS scene classification considering single source domain and multiple target domains, which utilizes the adversarial learning to align single source features with mixed-multiple-target features and self-supervised learning to distinguish the mixed-multiple-target

domain. Adayel et al. [164] developed a deep open-set DA method for cross-scene classification using adversarial learning and pareto ranking. To exploit the classification information in target domain, Zheng et al. [165] proposed a DA via a task-specific classifier (DATSNET) method for RS scene classification, where an adversarial learning strategy is used to adjust task-specific classification decision boundaries.

Adversarial-based approaches were also applied for other DA tasks of RS images. Bejiga et al. [135] proposed a domain adversarial neural network (DANN) for large-scale land cover classification of multispectral images, where the network consisted of a feature extractor, a class predictor, and domain classifier blocks. Rahhal et al. [166] proposed an adversarial learning method for DA from multiple remote sensing sources. It aligns the source and target distributions using a min-max entropy optimization method. Elshamli et al. [136] employed the denoising autoencoders (DAE) and domain-adversarial neural networks (DANN) to tackle the DA problem for multispatial and multitemporal RS images. Martini et al. [3] developed self-attention-based domain-adversarial networks for land cover classification using multitemporal satellite images, where the deep adversarial network can reduce the domain discrepancy between distinct geographical zones. Ji et al. [167] proposed an end-to-end GAN-based DA method for land cover classification from multiple-source RS images, where the source images are translated to the style of the target images through adversarial learning for training a fully convolutional network (FCN) for semantic segmentation of target images. Tasar et al. [137] proposed a multisource DA method (i.e., StandardGAN) for semantic segmentation of VHR satellite images. They further designed a unsupervised, multisource, multitarget, and life-long DA method for semantic segmentation of satellite images [168]. Wittich et al. [169] deployed a deep adversarial DA network using semantically consistent appearance adaptation for the classification of aerial images. A color mapping generative adversarial network (ColorMapGAN) was built for DA of RS image semantic segmentation [170]. Makkar et al. [171] adopted adversarial learning to extract discriminative target domain features that are aligned with source domain for geospatial image analysis. Mateo et al. [172] investigated a cross-sensor adversarial DA method of Landsat-8 and Proba-V images for cloud detection.

C. Others

There are some other deep DA methods. Yang et al. [138] proposed a transfer learning-based two-branch CNN model for HSI classification, where the spatial and spectral CNNs are used to extract joint spectral–spatial features from HSIs followed by target network training using transfer learning with limited labeled samples of target domain. Zhou and Prasad [139] proposed a deep feature alignment neural network (FANN) for HSI classification, where discriminative features for both domains were extracted using deep convolutional recurrent neural networks (CRNN) and then aligned layer-by-layer according to the transformation learning-based domain adaptation (DATL) method. Deng et al. [140] proposed an active transfer learning network for HSI classification, which exploited a hierarchical

stacked sparse autoencoder (SSAE) network to extract deep joint spectral–spatial features and an active TL strategy to transfer the pretrained SSAE network and the limited training samples from source to target domains. Song et al. [173] added an SA layer into CNN models for DA, which can fulfill domain alignment in feature subspace by fine-tuning the modified CNN models. Liu et al. [174] combined transfer learning and virtual samples in a 3D-CNN model to solve the problem of insufficient samples. Zhong et al. [141] proposed a cross-scene deep transfer learning network with spectral feature adaptation (SFA) for HSI classification, which designed a multiscale spectral–spatial unified network (MSSN) with two-branch architecture and a multiscale bank to extract discriminating features of HSI. Chen et al. [142] proposed an augmented associative learning-based DA (AALDA) method for HSI classification, which employs the criterion of cycle consistency to generate features that are domain-invariant and discriminative. Mdrafi et al. [143] proposed an attention-based DA using residual network for HSI classification, which considers different levels of attentions. Saha et al. [175] developed a graph neural network for multitarget DA in RS classification. Lasloum et al. [176] presented a multisource semisupervised DA method using a pretrained CNN for RS scene classification.

Othman et al. [144] designed a three-layer convex network termed as 3CN for DA in multitemporal VHR RS images. It is composed of three main layers: 1) mapping source training samples to the target domain via ELM; 2) target image classification via ELM; and 3) spatial regularization via the random-walker algorithm. Kellenberger et al. [177] combined the CNNs with AL for animal detection in UAV images, which used the OT to find corresponding regions between source and target data sets in the space of CNN activations. Kalita et al. [178] investigated the DA problem for land cover classification by utilizing the ensemble decision approach of deep neural networks to address the extra and missing class problem. Chakraborty et al. [179] proposed a multilevel weighted transformation based neuro-fuzzy DA method using stacked auto-encoder for land-cover classification. Lucas et al. [145] proposed a Bayesian-inspired CNN-based semisupervised DA method to produce land cover maps from satellite image time series data. Tong et al. [180] proposed a transferable deep model for land-cover classification of multisource high-resolution RS images, which used a pseudo-label learning strategy to automatically select training samples from the target domain and extracted multiscale contextual information of RS for classification.

VII. EXPERIMENTAL RESULTS AND ANALYSIS

Two images in the 2013 and 2018 IEEE GRSS data fusion contest, i.e., Houston2013 and Houston2018, are used in the experiment. The two images were acquired by the ITRES Compact Airborne Spectrographic Imager (CASI)-1500 sensor over the University of Houston campus and the neighboring urban area on June 23, 2012 and February 16, 2017, respectively [181], [182]. Houston2013 has the size of 349×1905 pixels, 144 spectral bands, and 15 categories. Houston2018 has the size of 4172×1202 pixels, 48 spectral bands, and 20 categories. For

(a)

(b)

(c)

(d)

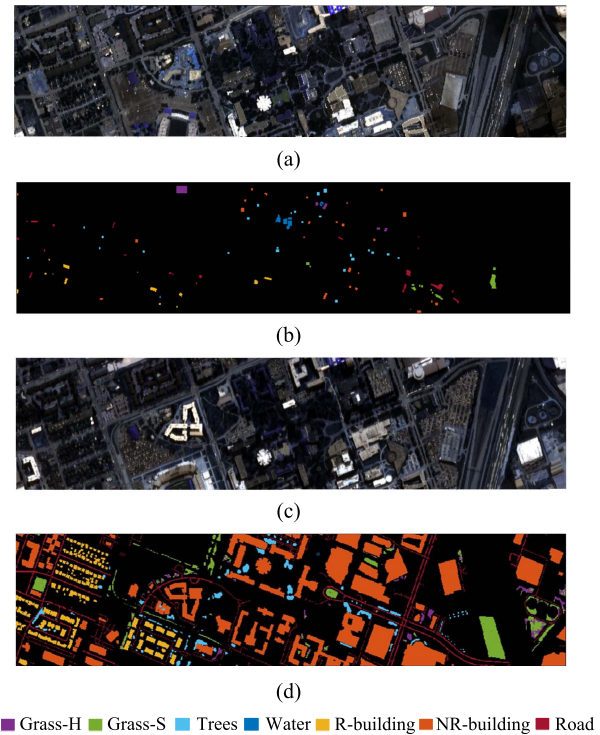■ Grass-H ■ Grass-S ■ Trees ■ Water ■ R-building ■ NR-building ■ Road

Fig. 7. RGB composite image and ground-truth map for Houston2013 in (a) and (b) and Houston2018 in (c) and (d).

TABLE II
NUMBER OF SAMPLES IN THE HOUSTON TASK

Houston2013 and Houston2018 data			
No.	Class	Source (Houston2013)	Target (Houston2018)
1	Grass healthy	345	1353
2	Grass stressed	365	4888
3	Trees	365	2766
4	Water	285	22
5	Residential buildings	319	5347
6	Non-residential buildings	408	32459
7	Road	443	6365
	Total	2530	53200

consistency, 48 spectral bands of Houston2013 are selected and seven common classes in these two images are considered for the DA task [108]. The Houston2013 and Houston2018 images are set as source and target domains, respectively. The RGB composite image and ground-truth map of two images are shown in Fig. 7. The number of samples are shown in Table II.

In the experiments, we compare some traditional shallow methods and recent deep DA methods, as shown in the upper and lower part of the line in Table III. The 1-nearest neighbor (1-NN) classifier is chosen as the base classifier. The NA (no adaptation) uses the 1-NN classifier built on the source domain to directly classify target samples. The GFK [13], SA [15], and KSA [25] are subspace-based adaptation methods. The TCA [38], [39], JDA [42], JGSA [43], LPJT [46], DADFL [48],

TABLE III
CLASSIFICATION ACCURACIES FOR THE HOUSTON DATA

| | Method | Class | | | | | | | OA(%) | κ |
		1	2	3	4	5	6	7		
Shallow DA	NA	78.94	94.82	66.52	100	64.00	68.10	60.91	69.49	0.5386
	GFK	77.38	95.50	66.27	100	63.89	67.66	60.52	69.17	0.5348
	SA	78.79	94.82	66.56	100	64.02	68.23	60.71	69.55	0.5392
	KSA	99.26	30.34	14.64	81.82	56.27	78.61	19.71	62.08	0.3588
	TCA	52.92	53.60	68.87	100	52.93	59.26	48.56	57.18	0.3755
	JDA	81.82	80.36	65.80	100	60.50	59.86	72.07	64.15	0.4865
	JGSA	92.68	81.87	72.52	100	71.97	73.71	70.02	74.28	0.6132
	LPJT	82.26	91.98	59.44	100	63.76	69.15	68.83	70.51	0.5602
	DADFL	88.40	84.49	72.52	100	70.75	67.98	74.93	71.38	0.5819
	CORAL	50.41	64.75	72.85	100	63.34	60.25	61.21	61.51	0.4443
	EasyTL	99.63	16.43	20.35	54.55	81.20	65.60	13.17	54.89	0.3114
Deep DA	DAN	63.78	75.57	62.80	100	46.04	69.89	73.44	67.92	0.5206
	MRAN	58.31	74.00	67.53	81.82	55.71	69.07	58.55	66.57	0.5091
	DSAN	26.46	78.07	73.17	100	41.05	72.16	61.30	67.18	0.5079
	DeepCoral	13.30	60.17	45.41	40.91	54.67	72.77	19.25	60.44	0.3776
	TSTnet	86.99	83.16	61.64	95.45	86.09	78.33	58.21	76.50	0.6378
	DAAN	68.00	89.85	68.33	100	49.24	75.20	65.95	72.30	0.5751
	MCD	65.85	87.77	55.60	100	85.17	72.57	75.04	74.49	0.6139
	DWL	86.47	83.53	63.77	100	55.86	76.49	67.37	73.58	0.5937

and CORAL [49] are transformation-based adaptation methods. EasyTL [86] is a classifier-based method. In the LPJT method, the landmark selection for source and target domains are considered. For the deep DA methods, the DAN [98], MRAN [100], DSAN [101], DeepCORAL [102], and TSTnet [108] are discrepancy-based methods. The DAAN [118], MCD [119], and DWL [120] are adversarial-based methods. For subspace-based DA algorithms, the dimensionality of the subspace is set to 20. The optimal learning rate lr of all deep learning algorithms is chosen from $\{0.0001, 0.001, 0.01, 0.1\}$, the batch size of the network is all set to 128, and the number of training iterations is 100. For DAN, DAAN, MRAN, and DSAN, there is a regularization parameter λ whose value is chosen from $\{0.001, 0.01, 0.1\}$. For all compared deep learning algorithms, to ensure the fairness of the comparison experiments, the backbone network is the ResNet18. The classification evaluation indicators are overall accuracy (OA) and kappa coefficient (κ).

From Table III, we can see that some traditional DA methods provide poor adaptation performance and their OAs are even worse than NA. It demonstrates that not all of the DA methods can reduce distribution discrepancy, and it is likely to produce negative transfer when there is significant spectral difference between domains. The JGSA, LPJT, and DADFL provide relatively better results than NA because these methods align the statistical and geometrical discrepancy between domains by learning two projections for source and target domains, respectively, and taking into account the local or global discriminative

information of domains. The Houston2013 and Houston2018 datasets have great spectral differences, so there may not exist a shared subspace generated by a unified transformation. In addition, the discriminative information of source or/and target domain can be used to improve the DA performance.

For deep DA methods, the adversarial methods, such as DAAN, MCD, and DWL, show better results than NA. Through an adversarial learning, the ability of generator and discriminator are simultaneously improved. The feature generator is likely to produce target features that are highly similar to the source features, and then a task-specific classifier built on the source domain can be used to classify target samples. Among all methods, the recently proposed TSTnet produces the best results. In the feature extraction part of TSTnet, the GCN and CNN are used to extract convolutional and topological structure features. In the adaptation part, the optimal transmission-based graph alignment and MMD-based distribution alignment work cooperatively. By exploiting the topological structure and semantic information of HSIs and considering the distribution alignment and topological relationship alignment, TSTnet generates excellent results.

The classification map on the target domain of different methods are shown in Fig. 8. It can be seen that some methods, such as TCA, CORAL, EasyTL, misclassify the class "Grass stressed" in green color to the class "Grass healthy" in purple color due to the high spectral similarity between these two classes. In addition, many methods misclassify the class "Nonresidential buildings" to the class "Road."
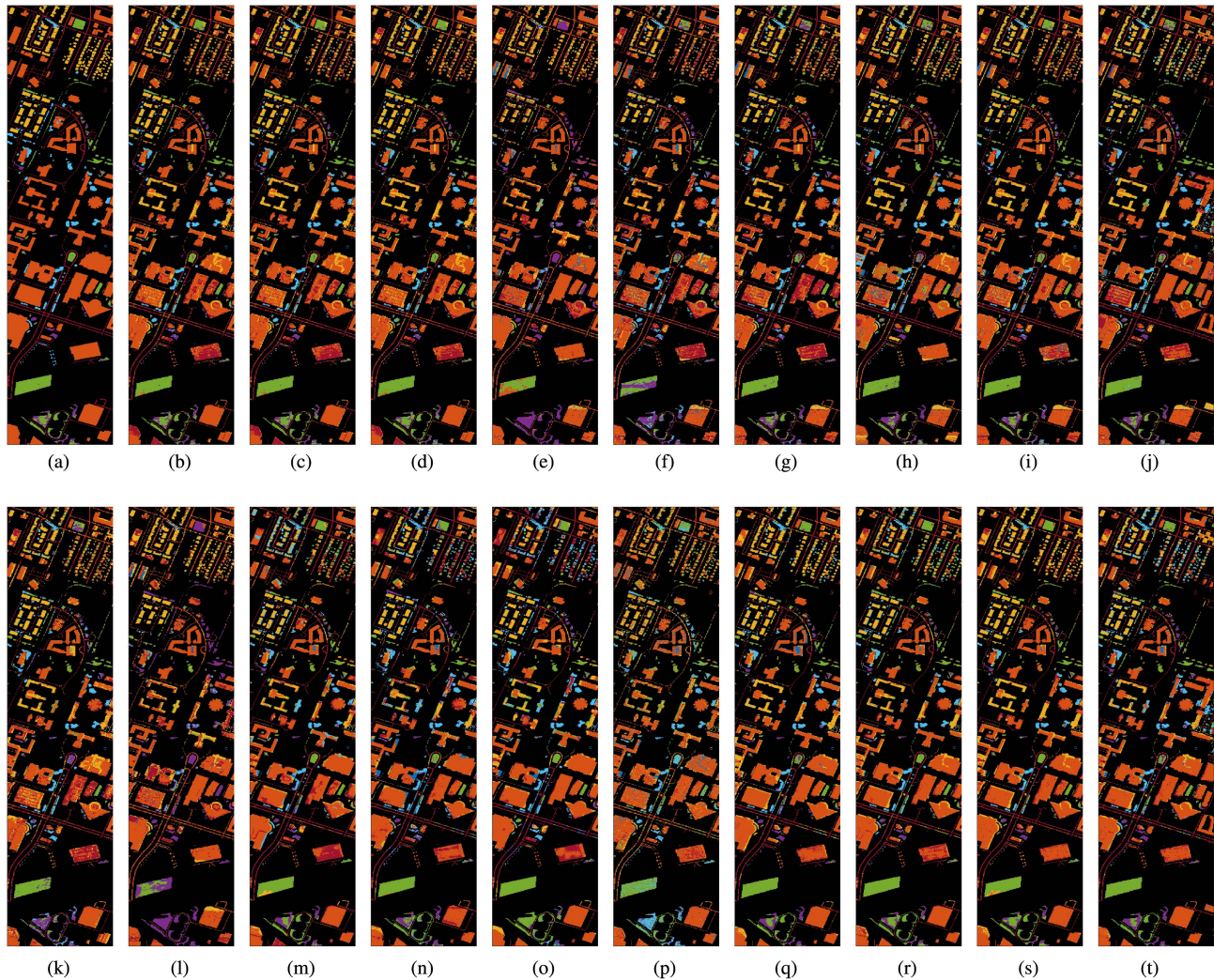
Fig. 8. Houston task. (a) Ground-truth. (b) NA (69.49%). (c) GFK (69.17%). (d) SA (69.55%). (e) KSA (62.08%). (f) TCA (57.18%). (g) JDA (64.15%). (h) JGSA (74.28%). (i) LPJT (70.51%). (j) DADFL (71.38%). (k) CORAL (61.51%). (l) EasyTL (54.89%). (m) DAN (67.92%). (n) MRAN (66.57%). (o) DSAN (67.18%). (p) DeepCoral (60.44%). (q) TSTnet (76.50%),. (r) DAAN (72.30%). (s) MCD (74.49%). (t) DWL (73.58%).

TABLE IV
EXECUTION TIME (IN SECONDS) OF TRADITIONAL DA METHODS

Method	NA	GFK	SA	KSA	TCA	JDA
Time(s)	1.18	0.49	0.46	806.77	118.92	210.76

Method	JGSA	LPJT	DADFL	CORAL	EasyTL
Time(s)	298.26	6095.4	1369.70	0.31	67.63

TABLE V
EXECUTION TIME (IN SECONDS) FOR ONE EPOCH TRAINING OF DEEP
DA METHODS

Method	DAN	MRAN	DSAN	DeepCORAL
Time(s)	1.26	1.23	1.45	0.74

Method	TSTnet	DAAN	MCD	DWL
Time(s)	6.58	1.32	3.24	0.54

Here, we list the computational time of each algorithm on the Houston task. All experiments are conducted on a computer with a 2.7 GHz Intel Xeon Gold 6258R CPU and two NVIDIA GeForce RTX 2080 Ti GPUs. The traditional methods are implemented on Matlab and their running times are shown in Table IV. The deep learning methods are implemented on the Pytorch deep learning framework. The execution time for one epoch training of deep DA methods are shown in Table V. Some traditional methods, such as JGSA, LPJT, and DADFL, need to iteratively update the transformation matrix and pseudo labels, so their running times are relatively long.

VIII. CONCLUSION

The early DA methods focus on either instance reweighting or subspace adaptation or transformation-based adaptation.

For RS image classification, due to the existence of large spectral drift between domains, it usually needs to simultaneously consider instance reweighting, subspace learning, and feature transformation. For traditional DA methods, we can incorporate landmark selection or feature weighting, target pseudo-label learning, local discriminative preservation into a subspace-based transformation framework to improve the discriminative ability of DA models.

For deep DA methods, the feature extraction module can be further improved by considering the data characteristics of RS images. The adversarial learning strategy can be combined with the discrepancy-based adaptation. In addition, the target pseudo-label learning can be used in the deep DA methods to iteratively update the network and improve the discriminative ability.

Currently, many existing RS DA methods focus on the general homogeneous unsupervised DA problem where the source and target domains have similar or same dimensionality feature spaces and there are no labeled instances in the target domain. In real situations, the RS classification problem may be more complex. The feature space and class space of source and target domains may be different. The classical DA problem can be extended to the following cases.

1) *Heterogeneous DA [183]*: The dimensionality of source and target domains are different and features of two domains are disjoint. For example, due to the difference in hyperspectral sensors, different HSIs usually have different spectral bands.

2) *Multisource DA [184]*: There are multiple source domains. The challenges lie in the unavailability of target labels and complex composition of multiple source domains [185]. For long-term RS image series analysis, there may exist multiple historical labeled images as sources. Compared with a single source domain, the joint use of multiple sources is likely to improve the DA performance.

3) *Open set DA [186]*: Only a few categories of interest are shared between source and target data. That is, the class space of source and target domains are different and intersect. Due to the changes of ground materials and acquisition regions, the source and target domains usually have some different classes especially for large-scale RS classification.

4) *Partial DA [187]*: It is assumed that the target label space is a subspace of the source label space. For example, if we only focus on some special classes in the target domain, the rich information of source domain can be used to perform partial DA.

5) *Few-shot DA [188], [189]*: The combination of DA with few-shot learning for using very few labeled target samples in training. When the source and target domains have great distribution differences and the number of classes is large, the unsupervised DA will fail. In this case, the limited labeled target samples can play a great role in building a connection between source and target classes.

6) *Domain Generalization [190]*: It aims to achieve out-of-distribution generalization by using only source data for model learning. There are many RS images obtained by different sensors or acquired in different conditions. It is likely to learn a model with high generalization ability from available RS images, and then applies the model to classify other images in real-time RS analysis.

The DA techniques can be used for large scene and long-term RS image processing. The labeling process for a large scene is costly and time-consuming. The DA technique can help to transfer the labels from a small region to the whole scene. For long-term image processing, historical images can be used to predict unseen images, and change analysis among RS images of different times can be performed.

REFERENCES

[1] P. Liu, "A survey of remote-sensing Big Data," *Front. Environ. Sci.*, vol. 3, 2015, Art. no. 45.

[2] F. Pacifici, J. Chanussot, and Q. Du, "2011 GRSS data fusion contest: Exploiting worldview-2 multi-angular acquisitions," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2011, pp. 1163–1166.

[3] M. Mauro, V. Mazzia, A. Khaliq, and M. Chiaberge, "Domain-adversarial training of self-attention based networks for land cover classification using multi-temporal Sentinel-2 satellite imagery," 2021, *arXiv:2104.00564*.

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

[5] A. Farahani, S. Voghoei, K. Rasheed, and H. R. Arabnia, "A brief review of domain adaptation," *Adv. Data Sci. Inf. Eng.*, pp. 877–894, 2021.

[6] N. Wambugu et al., "Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: A review," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 105, p. 102603, 2021.

[7] J. Huang, A. Gretton, K. Borgwardt, B. Schölkopf, and A. Smola, "Correcting sample selection bias by unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 601–608.

[8] C. Yaris, B. Huang, K. Bradbury, and J. M. Malof, "Randomized histogram matching: A simple augmentation for unsupervised domain adaptation in overhead imagery," 2021, *arXiv:2104.14032*.

[9] Y. Cui, L. Wang, J. Su, S. Gao, and L. Wang, "Iterative weighted active transfer learning hyperspectral image classification," *J. Appl. Remote Sens.*, vol. 15, no. 3, 2021, Art. no. 032207.

[10] H. Li, J. Li, Y. Zhao, M. Gong, Y. Zhang, and T. Liu, "Cost-sensitive self-paced learning with adaptive regularization for classification of image time series," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 11713–11727 , 2021.

[11] Y. Zhou, J. Lian, and M. Han, "Remote sensing image transfer classification based on weighted extreme learning machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1405–1409, Oct. 2016.

[12] R. Gopalan, R. Li, and R. Chellappa, "Domain adaptation for object recognition: An unsupervised approach," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2011, pp. 999–1006.

[13] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2066–2073.

[14] B. Banerjee and K. M. Buddhiraju, "Domain adaptation in the absence of source domain labeled samples—A coclustering-based approach," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1905–1909, Dec. 2016.

[15] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2013, pp. 2960–2967.

[16] H. Sun, S. Liu, S. Zhou, and H. Zou, "Unsupervised cross-view semantic transfer for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 13–17, Jan. 2016.

[17] W. Wei, W. Li, L. Zhang, C. Wang, P. Zhang, and Y. Zhang, "Robust hyperspectral image domain adaptation with noisy labels," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1135–1139, Jul. 2019.

[18] H. Sun, S. Liu, S. Zhou, and H. Zou, "Transfer sparse subspace analysis for unsupervised cross-view scene model adaptation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2901–2909, Jul. 2016.

[19] Y. Qin, L. Bruzzone, and B. Li, "Tensor alignment based domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9290–9307, Nov. 2019.

[20] G. Gao and Y. Gu, "Tensorized principal component alignment: A unified framework for multimodal high-resolution images classification," *IEEE Trans. Geoscie. Remote Sens.*, vol. 57, no. 1, pp. 46–61, Jan. 2019.

[21] R. Gui, X. Xu, R. Yang, L. Wang, and F. Pu, "Statistical scattering component-based subspace alignment for unsupervised cross-domain polsar image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5449–5463, Jul. 2021.

[22] B. Sun and K. Saenko, "Subspace distribution alignment for unsupervised domain adaptation," in *Proc. Brit. Mach. Vis. Conf., Nottingham*, vol. 4, 2015, pp. 24.1–24.10.

[23] Y. Zhang, W. Li, R. Tao, J. Peng, Q. Du, and Z. Cai, "Cross-scene hyperspectral image classification with discriminative cooperative alignment," *IEEE Trans. Geoscie. Remote Sens.*, vol. 59, no. 11, pp. 9646–9660, Nov. 2021.

[24] 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. Geoscie. Remote Sens.*, vol. 58, no. 11, pp. 7920–7930, Nov. 2020.

[25] R. Aljundi, R. Emonet, D. Muselet, and M. Sebban, "Landmarks-based kernelized subspace alignment for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 56–63.

[26] W. Yang, J. Peng, and W. Sun, "Ideal regularized discriminative multiple kernel subspace alignment for domain adaptation in hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5833–5846, 2020.

[27] J. Shen, X. Cao, Y. Li, and D. Xu, "Feature adaptation and augmentation for cross-scene hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 622–626, Apr. 2018.

[28] X. Li, L. Zhang, B. Du, L. Zhang, and Q. Shi, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2022–2035, May 2017.

[29] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.

[30] C. Persello and L. Bruzzone, "Kernel-based domain-invariant feature selection in hyperspectral images for transfer learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2615–2626, May 2016.

[31] C. Paris and L. Bruzzone, "A sensor-driven hierarchical method for domain adaptation in classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1308–1324, Mar. 2018.

[32] L. Yan, R. Zhu, Y. Liu, and N. Mo, "Tradaboost based on improved particle swarm optimization for cross-domain scene classification with limited samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3235–3251, Sep. 2018.

[33] M. Ye, Y. Qian, J. Zhou, and Y. Y. Tang, "Dictionary learning-based feature-level domain adaptation for cross-scene hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1544–1562, Mar. 2017.

[34] X. Wang, Y. Kong, Y. Gao, and Y. Cheng, "Dimensionality reduction for hyperspectral data based on pairwise constraint discriminative analysis and nonnegative sparse divergence," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1552–1562, Apr. 2017.

[35] J. Lin, C. He, Z. J. Wang, and S. Li, "Structure preserving transfer learning for unsupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1656–1660, Oct. 2017.

[36] 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, 2020.

[37] R. Gui, X. Xu, R. Yang, Z. Xu, L. Wang, and F. Pu, "A general feature paradigm for unsupervised cross-domain PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 4013305.

[38] 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.

[39] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, "Semisupervised transfer component analysis for domain adaptation in remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3550–3564, Jul. 2015.

[40] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 1410–1417.

[41] J. Peng, W. Sun, L. Ma, and Q. Du, "Discriminative transfer joint matching for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 6, pp. 972–976, Jun. 2019.

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

[43] J. Zhang, W. Li, and P. Ogunbona, "Joint geometrical and statistical alignment for visual domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1859–1867.

[44] X. Zhou and S. Prasad, "Transformation learning based domain adaptation for robust classification of disparate hyperspectral data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 3640–3643.

[45] X. Zhou and S. Prasad, "Domain adaptation for robust classification of disparate hyperspectral images," *IEEE Trans. Comput. Imag.*, vol. 3, no. 4, pp. 822–836, Dec. 2017.

[46] J. Li, M. Jing, K. Lu, L. Zhu, and H. T. Shen, "Locality preserving joint transfer for domain adaptation," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 6103–6115, Dec. 2019.

[47] Y. Huang, J. Peng, Y. Ning, W. Sun, and Q. Du, "Graph embedding and distribution alignment for domain adaptation in hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7654–7666, 2021.

[48] Y. Huang, J. Peng, Y. Ning, Q. Cao, and W. Sun, "Distribution alignment and discriminative feature learning for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 6004505.

[49] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2058–2065.

[50] J. Peng, W. Sun, T. Wei, and W. Fan, "A modified correlation alignment algorithm for the domain adaptation of GF-5 hyperspectral image," *J. Remote Sens. (Chin.)*, vol. 24, no. 4, pp. 417–426, 2020.

[51] L. Zhu and L. Ma, "Class centroid alignment based domain adaptation for classification of remote sensing images," *Pattern Recognit. Lett.*, vol. 83, pp. 124–132, 2016.

[52] L. Ma, M. M. Crawford, L. Zhu, and Y. Liu, "Centroid and covariance alignment-based domain adaptation for unsupervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2305–2323, Apr. 2019.

[53] Y. Qin, L. Bruzzone, B. Li, and Y. Ye, "Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3952–3966, Jun. 2019.

[54] A. Samat, C. Persello, P. Gamba, S. Liu, J. Abuduwaili, and E. Li, "Supervised and semi-supervised multi-view canonical correlation analysis ensemble for heterogeneous domain adaptation in remote sensing image classification," *Remote Sens.*, vol. 9, 2017, Art. no. 337.

[55] X. Li, L. Zhang, B. Du, and L. Zhang, "On gleaning knowledge from cross domains by sparse subspace correlation analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3204–3220, Jun. 2019.

[56] M. Volpi, G. Camps-Valls, and D. Tuia, "Spectral alignment of multi-temporal cross-sensor images with automated kernel canonical correlation analysis," *ISPRS J. Photo. Remote Sens.*, vol. 107, pp. 50–63, 2015.

[57] D. Tuia, J. Munoz-Mari, L. Gomez-Chova, and J. Malo, "Graph matching for adaptation in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 329–341, Jan. 2013.

[58] D. Tuia, M. Volpi, M. Trolliet, and G. Camps-Valls, "Semisupervised manifold alignment of multimodal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7708–7720, Dec. 2014.

[59] H. L. Yang and M. M. Crawford, "Domain adaptation with preservation of manifold geometry for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 543–555, Feb. 2016.

[60] H. L. Yang and M. M. Crawford, "Spectral and spatial proximity-based manifold alignment for multitemporal hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 51–64, Jan. 2016.

[61] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LEMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photo. Remote Sens.*, vol. 147, pp. 193–205, 2019.

[62] L. Ma, C. Luo, J. Peng, and Q. Du, "Unsupervised manifold alignment for cross-domain classification of remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 10, pp. 1650–1654, Oct. 2019.

[63] C. Luo and L. Ma, "Manifold regularized distribution adaptation for classification of remote sensing images," *IEEE Access*, vol. 6, pp. 4697–4708, 2018.

[64] H. Wang, X. Wang, C. P. Chen, and Y. Cheng, "Hyperspectral image classification based on domain adaptation broad learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3006–3018, 2020.

[65] Y. Dong, T. Liang, Y. Zhang, and B. Du, "Spectral–spatial weighted kernel manifold embedded distribution alignment for remote sensing image classification," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 3185–3197, Jun. 2021.

[66] W. Gross, D. Tuia, U. Soergel, and W. Middelmann, "Nonlinear feature normalization for hyperspectral domain adaptation and mitigation of nonlinear effects," *IEEE Trans. Geoscie. Remote Sens.*, vol. 57, no. 8, pp. 5975–5990, Aug. 2019.

[67] B. Tardy, J. Inglada, and J. Michel, "Assessment of optimal transport for operational land-cover mapping using high-resolution satellite images time series without reference data of the mapping period," *Remote Sens.*, vol. 11, no. 9, 2019, Art. no. 1047.

[68] S. Jia et al., "Gradient feature-oriented 3-D domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5505517.

[69] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 456–460, Feb. 2001.

[70] L. Bruzzone and D. F. Prieto, "A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images," *Pattern Recognit. Lett.*, vol. 23, no. 9, pp. 1063–1071, 2002.

[71] L. Bruzzone and R. Cossu, "A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 1984–1996, Sep. 2002.

[72] L. Bruzzone, R. Cossu, and G. Vernazza, "Combining parametric and non-parametric algorithms for a partially unsupervised classification of multitemporal remote-sensing images," *Inf. Fusion*, vol. 3, no. 4, pp. 289–297, 2002.

[73] E. Izquierdo-Verdiguier, V. Laparra, L. Gómez-Chova, and G. Camps-Valls, "Encoding invariances in remote sensing image classification with SVM," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 981–985, Sep. 2013.

[74] Y. Guo, X. Jia, and D. Paull, "Effective sequential classifier training for SVM-based multitemporal remote sensing image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 3036–3048, Jun. 2018.

[75] J. M. Leiva-Murillo, L. Gómez-Chova, and G. Camps-Valls, "Multitask remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 151–161, Jan. 2013.

[76] Z. Sun, C. Wang, H. Wang, and J. Li, "Learn multiple-kernel SVMs for domain adaptation in hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1224–1228, Sep. 2013.

[77] Z. Deng, H. Sun, and S. Zhou, "Semi-supervised ground-to-aerial adaptation with heterogeneous features learning for scene classification," *ISPRS Int. J. Geo-Information*, vol. 7, 2018, Art. on. 182.

[78] A. Saboori and H. Ghassemian, "Robust transfer joint matching distributions in semi-supervised domain adaptation for hyperspectral images classification," *Int. J. Remote Sens.*, vol. 41, no. 23, pp. 9283–9307, 2020.

[79] 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.

[80] D. Shen and L. Ma, "Cross-domain extreme learning machine for classification of hyperspectral images," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3305–3308.

[81] L. Zhou and L. Ma, "Extreme learning machine-based heterogeneous domain adaptation for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1781–1785, Nov. 2019.

[82] S. K. Meher and N. S. Kothari, "Interpretable rule-based fuzzy ELM and domain adaptation for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5907–5919, Jul. 2021.

[83] X. Liu, Q. Hu, Y. Cai, and Z. Cai, "Extreme learning machine-based ensemble transfer learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3892–3902, 2020.

[84] H. Wei, L. Ma, Y. Liu, and Q. Du, "Combining multiple classifiers for domain adaptation of remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1832–1847, 2021.

[85] J. Zhang, J. Liu, B. Pan, Z. Chen, X. Xu, and Z. Shi, "An open set domain adaptation algorithm via exploring transferability and discriminability for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609512.

[86] J. Wang, Y. Chen, H. Yu, M. Huang, and Q. Yang, "Easy transfer learning by exploiting intra-domain structures," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2019, pp. 1210–1215.

[87] S. Rajan, J. Ghosh, and M. M. Crawford, "Exploiting class hierarchies for knowledge transfer in hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3408–3417, Nov. 2006.

[88] L. Bruzzone and M. Marconcini, "Domain adaptation problems: A DASVM classification technique and a circular validation strategy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 5, pp. 770–787, May 2009.

[89] W. Kim and M. M. Crawford, "Adaptive classification for hyperspectral image data using manifold regularization kernel machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4110–4121, Nov. 2010.

[90] S. Rajan, J. Ghosh, and M. M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[91] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, 2011.

[92] C. Persello and L. Bruzzone, "Active learning for domain adaptation in the supervised classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4468–4483, Nov. 2012.

[93] G. Matasci, D. Tuia, and M. Kanevski, "SVM-based boosting of active learning strategies for efficient domain adaptation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1335–1343, Oct. 2012.

[94] M. M. Crawford, D. Tuia, and H. L. Yang, "Active learning: Any value for classification of remotely sensed data?," *Proc. IEEE*, vol. 101, no. 3, pp. 593–608, Mar. 2013.

[95] N. Alajlan, E. Pasolli, F. Melgani, and A. Franzoso, "Large-scale image classification using active learning," *IEEE Geosc. Remote Sens. Lett.*, vol. 11, no. 1, pp. 259–263, Jan. 2014.

[96] C. Persello, "Interactive domain adaptation for the classification of remote sensing images using active learning," *IEEE Geosc. Remote Sens. Lett.*, vol. 10, no. 4, pp. 736–740, Jul. 2013.

[97] B. Demir, F. Bovolo, and L. Bruzzone, "Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 300–312, Jan. 2013.

[98] M. Long, Y. Cao, J. Wang, and M. Jordan, "Learning transferable features with deep adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 97–105.

[99] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2208–2217.

[100] Y. Zhu et al., "Multi-representation adaptation network for cross-domain image classification," *Neural Netw.*, vol. 119, pp. 214–221, 2019.

[101] Y. Zhu et al., "Deep subdomain adaptation network for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1713–1722, Apr. 2021.

[102] B. Sun and K. Saenko, "Deep coral: Correlation alignment for deep domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 443–450.

[103] L. Ma and J. Song, "Deep neural network-based domain adaptation for classification of remote sensing images," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042612.

[104] A. Garea, D. B. Heras, and F. Argüello, "Tcanet for domain adaptation of hyperspectral images," *Remote Sens.*, vol. 11, 2019, Art. no. 2289.

[105] Z. Wang, B. Du, Q. Shi, and W. Tu, "Domain adaptation with discriminative distribution and manifold embedding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1155–1159, Jul. 2019.

[106] X. Ma, X. Mou, J. Wang, X. Liu, H. Wang, and B. Yin, "Cross-data set hyperspectral image classification based on deep domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10164–10174, Dec. 2019.

[107] Z. Li, X. Tang, W. Li, C. Wang, C. Liu, and J. He, "A two-stage deep domain adaptation method for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 7, 2020, Art. no. 1054.

[108] Y. Zhang, W. Li, M. Zhang, Y. Qu, R. Tao, and H. Qi, "Topological structure and semantic information transfer network for cross-scene hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3109872.

[109] W. Wang, L. Ma, M. Chen, and Q. Du, "Joint correlation alignment-based graph neural network for domain adaptation of multitemporal hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3170–3184, 2021.

[110] X. Liang, Y. Zhang, and J. Zhang, "Attention multisource fusion-based deep few-shot learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8773–8788, 2021.

[111] 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.

[112] S. Zhu, B. Du, L. Zhang, and X. Li, "Attention-based multiscale residual adaptation network for cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5400715.

[113] J. Geng, X. Deng, X. Ma, and W. Jiang, "Transfer learning for SAR image classification via deep joint distribution adaptation networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5377–5392, Aug. 2020.

[114] I. Goodfellow et al., "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 27, pp. 1–9, 2014.

[115] D. Wittich and F. Rottensteiner, "Adversarial domain adaptation for the classification of aerial images and height data using convolutional neural networks," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 4, pp. 197–204, 2019.

[116] E. Tzeng, J. Hoffman, T. Darrell, and K. Saenko, "Simultaneous deep transfer across domains and tasks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2015, pp. 4068–4076.

[117] Z. Pei, Z. Cao, M. Long, and J. Wang, "Multi-adversarial domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 3934–3941.

[118] C. Yu, J. Wang, Y. Chen, and M. Huang, "Transfer learning with dynamic adversarial adaptation network," in *Proc. IEEE Int. Conf. Data Mining*, 2019, pp. 778–786.

[119] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3723–3732.

[120] N. Xiao and L. Zhang, "Dynamic weighted learning for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 15242–15 251.

[121] X. Ma, X. Mou, J. Wang, X. Liu, J. Geng, and H. Wang, "Cross-dataset hyperspectral image classification based on adversarial domain adaptation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4179–4190, May 2021.

[122] J. Miao, B. Zhang, and B. Wang, "Coarse-to-fine joint distribution alignment for cross-domain hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12415–12428, 2021.

[123] C. Yu, C. Liu, M. Song, and C.-I . Chang, "Unsupervised domain adaptation with content-wise alignment for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5511705.

[124] S. Pande, B. Banerjee, and A. Pižurica, "Class reconstruction driven adversarial domain adaptation for hyperspectral image classification," in *Proc. Iberian Conf. Pattern Recognit. Image Anal.*, 2019, pp. 472–484.

[125] Z. Liu, L. Ma, and Q. Du, "Class-wise distribution adaptation for unsupervised classification of hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 508–521, Jan. 2021.

[126] A. Saboori and H. Ghassemian, "Adversarial discriminative active deep learning for domain adaptation in hyperspectral images classification," *Int. J. Remote Sens.*, vol. 42, no. 10, pp. 3981–4003, 2021.

[127] H. Wang, Y. Cheng, C. P. Chen, and X. Wang, "Hyperspectral image classification based on domain adversarial broad adaptation network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517813.

[128] B. Deng, S. Jia, and D. Shi, "Deep metric learning-based feature embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1422–1435, Feb. 2020.

[129] Z. Li, M. Liu, Y. Chen, Y. Xu, W. Li, and Q. Du, "Deep cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501618.

[130] 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.

[131] X. Zhang, X. Yao, X. Feng, G. Cheng, and J. Han, "DFENet for domain adaptation based remote sensing scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5611611.

[132] L. Yan, B. Fan, H. Liu, C. Huo, S. Xiang, and C. Pan, "Triplet adversarial domain adaptation for pixel-level classification of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3558–3573, May 2020.

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

[134] J. Zheng et al., "A two-stage adaptation network (TSAN) for remote sensing scene classification in single-source-mixed-multiple-target domain adaptation (S^2 M^2T DA) scenarios," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609213.

[135] M. B. Bejiga, F. Melgani, and P. Beraldini, "Domain adversarial neural networks for large-scale land cover classification," *Remote Sens.*, vol. 11, 2019, Art. no. 1153.

[136] A. Elshamli, G. W. Taylor, A. Berg, and S. Areibi, "Domain adaptation using representation learning for the classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 9, pp. 4198–4209, Sep. 2017.

[137] O. Tasar, Y. Tarabalka, A. Giros, P. Alliez, and S. Clerc, "Standardgan: Multi-source domain adaptation for semantic segmentation of very high resolution satellite images by data standardization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 192–193.

[138] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.

[139] X. Zhou and S. Prasad, "Deep feature alignment neural networks for domain adaptation of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5863–5872, Oct. 2018.

[140] C. Deng, X. Liu, C. Li, and D. Tao, "Active multi-kernel domain adaptation for hyperspectral image classification," *Pattern Recognit.*, vol. 77, pp. 306–315, 2018.

[141] C. Zhong, J. Zhang, S. Wu, and Y. Zhang, "Cross-scene deep transfer learning with spectral feature adaptation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2861–2873, 2020.

[142] M. Chen, L. Ma, W. Wang, and Q. Du, "Augmented associative learning-based domain adaptation for classification of hyperspectral remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6236–6248, 2020.

[143] R. Mdrafi, Q. Du, A. C. Gurbuz, B. Tang, L. Ma, and N. H. Younan, "Attention-based domain adaptation using residual network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 6424–6433, 2020.

[144] E. Othman, Y. Bazi, N. Alajlan, H. AlHichri, and F. Melgani, "Three-layer convex network for domain adaptation in multitemporal VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 354–358, Mar. 2016.

[145] B. Lucas, C. Pelletier, D. Schmidt, G. I. Webb, and F. Petitjean, "A bayesian-inspired, deep learning-based, semi-supervised domain adaptation technique for land cover mapping," *Mach. Learn.*, pp. 1–33, 2021.

[146] H. Shimodaira, "Improving predictive inference under covariate shift by weighting the log-likelihood function," *J. Statist. Plan. Inference*, vol. 90, no. 2, pp. 227–244, 2000.

[147] B. Zadrozny, "Learning and evaluating classifiers under sample selection bias," in *Proc. Int. Conf. Mach. Learn.*, 2004, Art. no. 114.

[148] J. Zhao, L. Li, F. Deng, H. He, and J. Chen, "Discriminant geometrical and statistical alignment with density peaks for domain adaptation," *IEEE Trans. Cybern.*, vol. 52, no. 2, pp. 1193–1206, Feb. 2022.

[149] B. Banerjee and S. Chaudhuri, "Hierarchical subspace learning based unsupervised domain adaptation for cross-domain classification of remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5099–5109, Nov. 2017.

[150] S. Chakraborty and M. Roy, "A neural approach under transfer learning for domain adaptation in land-cover classification using two-level cluster mapping," *Appl. Soft Comput.*, vol. 64, pp. 508–525, 2018.

[151] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, "A novel domain adaptation bayesian classifier for updating land-cover maps with class differences in source and target domains," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2810–2826, Jul. 2012.

[152] S. Zhong and Y. Zhang, "An iterative training sample updating approach for domain adaptation in hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 10, pp. 1821–1825, Oct. 2021.

[153] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.

[154] G. Jun and J. Ghosh, "Semisupervised learning of hyperspectral data with unknown land-cover classes," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 273–282, Jan. 2013.

[155] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[156] I. Kalita, R. N. S. Kumar, and M. Roy, "Deep learning-based cross-sensor domain adaptation under active learning for land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6005005.

[157] A. Saboori, H. Ghassemian, and F. Razzazi, "Active multiple kernel fredholm learning for hyperspectral images classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 356–360, Feb. 2021.

[158] J. E. Ball, D. T. Anderson, and C. S. Chan Sr, "Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 0 42609.

[159] B. Kellenberger, O. Tasar, B. Bhushan Damodaran, N. Courty, and D. Tuia, "Deep domain adaptation in earth observation," in *Deep Learn. for the Earth Sci.: A Comprehensive Approach Remote Sens., Climate Sci., and Geosciences*, Hoboken, NJ, USA: Wiley, 2021, pp. 90–104.

[160] Q. Zhu, Y. Sun, Q. Guan, L. Wang, and W. Lin, "A weakly pseudo-supervised decorrelated subdomain adaptation framework for cross-domain land-use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5623913.

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

[162] Z. Fang et al., "Confident learning-based domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527116.

[163] W. Liu and F. Su, "A novel unsupervised adversarial domain adaptation network for remotely sensed scene classification," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6099–6116, 2020.

[164] 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, 2020, Art. no. 1716.

[165] Z. Zheng, Y. Zhong, Y. Su, and A. Ma, "Domain adaptation via a task-specific classifier framework for remote sensing cross-scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5620513.

[166] M. M. Al Rahhal, Y. Bazi, H. Al-Hwiti, H. Alhichri, and N. Alajlan, "Adversarial learning for knowledge adaptation from multiple remote sensing sources," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1451–1455, Aug. 2021.

[167] S. Ji, D. Wang, and M. Luo, "Generative adversarial network-based full-space domain adaptation for land cover classification from multiple-source remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 3816–3828, May 2021.

[168] O. Tasar, A. Giros, Y. Tarabalka, P. Alliez, and S. Clerc, "Daugnet: Unsupervised, multisource, multitarget, and life-long domain adaptation for semantic segmentation of satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1067–1081, Feb. 2021.

[169] D. Wittich and F. Rottensteiner, "Appearance based deep domain adaptation for the classification of aerial images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 180, pp. 82–102, 2021.

[170] O. Tasar, S. Happy, Y. Tarabalka, and P. Alliez, "ColormapGAN: Unsupervised domain adaptation for semantic segmentation using color mapping generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7178–7193, Oct. 2020.

[171] N. Makkar, L. Yang, and S. Prasad, "Adversarial learning based discriminative domain adaptation for geospatial image analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 150–162, 2022.

[172] G. Mateo-García, V. Laparra, D. López-Puigdollers, and L. Gómez-Chova, "Cross-sensor adversarial domain adaptation of landsat-8 and proba-v images for cloud detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 747–761, 2021.

[173] 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.

[174] X. Liu, Q. Sun, Y. Meng, M. Fu, and S. Bourennane, "Hyperspectral image classification based on parameter-optimized 3D-CNNs combined with transfer learning and virtual samples," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1425.

[175] S. Saha, S. Zhao, and X. X. Zhu, "Multitarget domain adaptation for remote sensing classification using graph neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6506505.

[176] T. Lasloum, H. Alhichri, Y. Bazi, and N. Alajlan, "Ssdan: Multi-source semi-supervised domain adaptation network for remote sensing scene classification," *Remote Sens.*, vol. 13, no. 19, 2021, Art. no. 3861.

[177] B. Kellenberger, D. Marcos, S. Lobry, and D. Tuia, "Half a percent of labels is enough: Efficient animal detection in UAV imagery using deep CNNs and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9524–9533, Dec. 2019.

[178] I. Kalita and M. Roy, "Deep neural network-based heterogeneous domain adaptation using ensemble decision making in land cover classification," *IEEE Trans. Artif. Intell.*, vol. 1, no. 2, pp. 167–180, Oct. 2020.

[179] S. Chakraborty and M. Roy, "A multi-level weighted transformation based neuro-fuzzy domain adaptation technique using stacked auto-encoder for land-cover classification," *Int. J. Remote Sens.*, vol. 41, no. 17, pp. 6831–6857, 2020.

[180] X.-Y . Tong et al., "Land-cover classification with high-resolution remote sensing images using transferable deep models," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111322.

[181] C. Debes et al., "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.

[182] B. Le Saux, N. Yokoya, R. Hansch, and S. Prasad, "2018 IEEE GRSS data fusion contest: Multimodal land use classification [technical committees]," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 1, pp. 52–54, Mar. 2018.

[183] F. Liu, G. Zhang, and J. Lu, "Heterogeneous domain adaptation: An unsupervised approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5588–5602, Dec. 2020.

[184] S. Sun, H. Shi, and Y. Wu, "A survey of multi-source domain adaptation," *Inf. Fusion*, vol. 24, pp. 84–92, 2015.

[185] H. Liu, M. Shao, and Y. Fu, "Structure-preserved multi-source domain adaptation," in *Proc. IEEE Int. Conf. Data Mining*, 2016, pp. 1059–1064.

[186] P.P. Busto and J. Gall, "Open set domain adaptation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2017, pp. 754–763.

[187] Z. Cao, L. Ma, M. Long, and J. Wang, "Partial adversarial domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 135–150.

[188] M. Rußwurm, S. Wang, M. Korner, and D. Lobell, "Meta-learning for few-shot land cover classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 200–201.

[189] S. Motiian, Q. Jones, S. Iranmanesh, and G. Doretto, "Few-shot adversarial domain adaptation," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 1–11, 2017.

[190] D. Li, Y. Yang, Y.-Z . Song, and T. M. Hospedales, "Deeper, broader and artier domain generalization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2017, pp. 5542–5550.

Jiangtao Peng (Senior Member, IEEE) received the B.S. degree in information and computing science and the M.S. degree in applied mathematics from Hubei University, Wuhan, China, in 2005 and 2008, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently a Professor with the Faculty of Mathematics and Statistics, Hubei University, China. His research interests include machine learning and hyperspectral image processing.

Yi Huang received the B.S. degree in mathematics and applied mathematics, in 2020, from the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China, where she is currently working toward the Ph.D. degree in applied mathematics.

Her research interests include machine learning and hyperspectral image processing.
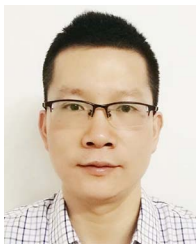
Weiwei Sun (Senior Member, IEEE) received the B.S. degree in surveying and mapping, and Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2007 and 2013, respectively.

From 2011 to 2012, he was with the Department of Applied Mathematics, University of Maryland, College Park, MD, USA, working as a Visiting Scholar with the famous professor John Benedetto to study on the dimensionality reduction of Hyperspectral Image. From 2014 to 2016, he was with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, working as a Postdoc to study intelligent processing in Hyperspectral imagery. From 2017 to 2018, he worked as a Visiting Scholar with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. He is currently a Full Professor with Ningbo University, China. He has authored or coauthored more than 70 journal papers. His research interests include hyperspectral image processing with manifold learning, anomaly detection and target recognition of remote sensing imagery using compressive sensing.

Na Chen received the B.S. degree in information and computing science and the M.S. degree in applied mathematics from Hubei University, Wuhan, China, in 2006 and 2009, respectively, and the Ph.D. degree in computational mathematics from Huazhong University of Science and Technology, Wuhan, China, in 2012.

She is currently an Associate Professor with the Faculty of Mathematics and Statistics, Hubei University. Her research interests include statistical learning theory and image processing.

Yujie Ning received the B.S. degree in mathematics and applied mathematics, in 2021, from the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China, where she is currently working toward the Ph.D. degree in applied mathematics.

Her research interests include machine learning and hyperspectral image processing.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland-Baltimore County, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of SPIE-International Society for Optics and Photonics. She was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. From 2016 to 2020, she was the Editor-in-Chief for the IEEE JSTARS. She is currently a member of IEEE Periodicals Review and Advisory Committee. She served as a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair with the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014. She was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012.