Graph Embedding and Distribution Alignment for Domain Adaptation in Hyperspectral Image Classification

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

***Abstract*—Recent studies in cross-domain classification have shown that discriminant information of both source and target domains is very important. In this article, we propose a new domain adaptation (DA) method for hyperspectral image (HSI) classification, called graph embedding and distribution alignment (GEDA). GEDA uses the graph embedding method and a pseudo-label learning method to learn interclass and intraclass divergence matrices of source and target domains, which preserves the local discriminant information of both domains. Meanwhile, spatial and spectral features of HSI are used, and distribution alignment and subspace alignment are performed to minimize the spectral differences between domains. We perform DA tasks on Yancheng, Botswana, University of Pavia, and Center of Pavia, Shanghai and Hangzhou data sets. Experimental results show that the classification performance of the proposed GEDA is better than that of existing DA methods.**

***Index Terms*—Distribution adaptation, domain adaptation, graph embedding, hyperspectral image classification.**

I. Introduction

HYPERSPECTRAL remote sensing technique can simultaneously obtain spatial and spectral information of ground objects, and has the ability to identify subtle difference between different materials [1]–[3]. Nowadays, hyperspectral remote sensing has become a hot spot in the development of Earth observation. A large amount of hyperspectral remote sensing images are available due to the launch of new satellites and the development of hyperspectral sensors [4]. Although there exist many hyperspectral images (HSIs), the collaborative processing of different HSIs is still very difficult due to the differences in sensors and acquisition conditions [5]–[8]. HSIs usually have differences in spectral coverage, spectral resolution, spatial resolution, number of bands, etc. In addition, the acquisition conditions of different HSIs are usually different, which bring obstacles to the long-time sequence analysis and collaborative processing of HSIs [5].

Recently, a new domain adaptation (DA) technique has emerged in the field of machine learning [5], [9]. The main idea of DA is to use the rich knowledge of a labeled source domain to improve the model performance of limited or no labeled target domain. DA can transfer the differences in the imaging environment and hardware conditions of multisource HSIs into a data or feature transformation problem [5], [10]–[15]. By mining data correlation, it can realize the transfer of common knowledge between domains. DA provides a theoretical feasibility for the cross-domain classification problems caused by the inconsistent characteristics of HSIs. The existing DA methods can be roughly classified as sample-based methods, feature-based methods, and classifier-based methods [5], [16]. In this article, we focus on the feature-based DA method, which either performs subspace learning by exploiting the subspace geometrical and statistical structures [16]–[19], or distribution alignment to reduce the marginal or/and conditional distribution divergence between domains [20]–[22].

In previous years, many feature-based DA algorithms have been proposed, such as subspace alignment (SA) [17], correlation alignment [23], transfer joint matching (TJM) [19], geodesic flow kernel (GFK) [24], transfer component analysis (TCA) [18], joint distribution alignment [20], scatter component analysis (SCA) [21], joint geometric and statistical alignment [22], and locality-preserving joint transfer (LPJT) [16]. SA intends to learn a transformation matrix that maps source and target domains into individual subspaces, so that the distance between the resulting subspaces is reduced [17]. CORrelation ALignment (CORAL) uses a linear transformation to align second-order statistical features of source and target domains [23]. GFK learns domain-invariant features by integrating an infinite number of subspaces [7], [24]. TCA maps the data from the two domains together into a high-dimensional reproducing kernel Hilbert space (RKHS) such that the distance

Manuscript received June 7, 2021; revised July 8, 2021; accepted July 18, 2021. Date of publication July 26, 2021; date of current version August 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61871177, Grant 41971296, and Grant 11771130, in part by the National Key Research and Development Program of China under Grant 2020YFA0714200, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LR19D010001, and in part by the Open Fund of State Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University under Grant 18R05. *(Corresponding authors: Jiangtao Peng; Weiwei Sun.)*

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

Digital Object Identifier 10.1109/JSTARS.2021.3099805

between the two domains in the RKHS is minimized [18]. TJM mainly reduces the differences between domains and constructs new features through feature matching and instance weighting [19]. Joint distribution adaptation (JDA) is designed to learn a transformation so that the transformed data align both the marginal and conditional distributions [20]. SCA takes the between- and within-class scatter matrices of source domain into consideration [21]. The above transformation-based DA methods, such as TCA, TJM, JDA, and SCA, only learn a unified transformation to map source and target domains into a shared subspace. When the distribution shift between the two domains is large, it is very different to adapt the distribution [22]. To reduce the shift both statistically and geometrically, joint geometrical and statistical alignment (JGSA) learns two coupling mappings A and B for source and target domains, respectively. It simultaneously performs the distribution alignment and SA between transformed domains and considers the discriminant information of source domain and global information of target domain [22]. However, JGSA does not consider the data manifold structure. To preserve the local manifold structure of data, LPJT jointly exploits feature adaptation with distribution matching and sample adaptation with landmark selection [16].

The aforementioned methods have shown good performance for DA in computer vision. However, directly applying these methods for HSI cross-domain classification usually produces poor results due to the existence of great spectral drafts between domains. The spectral drafts mainly come from the differences in imaging environment and hardware conditions and the variation of materials. In the case of large domain differences, it is very difficult to select effective landmarks with high domain matching degrees for the sample-based DA methods, and also unlikely to exist invariant features or shared latent feature subspace for the feature-based DA methods.

For cross-scene remote sensing image classification, the characteristics of remote sensing images can be exploited to improve the performance of DA. Sun *et al.* [25] constructed discriminative cross-view subspaces and applied the SA method for the unsupervised cross-view remote sensing image classification. Qin *et al.* proposed a tensor alignment method for HSI classification which performed SA between tensor domains [26]. Yang *et al.* [8] proposed an ideal regularized discriminative multiple kernel SA for HSI domain adaptation. Ma *et al.* [13] improved the CORAL by aligning both the class-wise centroid and covariance. Matasci *et al.* [11] investigated the performance of TCA for domain adaptation in remote sensing image classification. Garea *et al.* [27] combined TCA with convolutional neural networks and proposed a TCANet for DA of HSIs. Peng *et al.* [7] incorporated the label dependence constraint into the TJM model and proposed a discriminative TJM method for DA of HSIs. Recently, some deep learning-based DA methods have been proposed for cross-scene HSI classification, such as deep metric learning model [28], class-wise distribution adaptation (CDA) network [29], and deep cross-domain few-shot learning (DCFSL) [30]. In Ref. [28], a deep metric learning-based feature embedding model was proposed for HSI classification. It projected input features into a well-defined metric space, where the mapping features have small intraclass distance and large interclass distance [28]. In CDA, a class-wise adversarial adaptation network was constructed and a probability-prediction-based maximum mean discrepancy (MMD) method was introduced to measure the distribution distance [29]. DCFSL incorporated the few-shot learning (FSL) and DA in a unified framework for HSI classification, where a conditional adversarial DA strategy was utilized to overcome domain shift, and FSL was executed to discover transferable knowledge in the source classes and to learn a discriminative embedding model to the target classes [30].

From the above traditional and deep-learning-based DA methods, we can see that feature learning and distribution alignment are key factors for domain adaptation. Although it is very difficult to select effective landmarks, invariant features, or shared latent feature subspace in the case of large domain differences, it is feasible to project source and target domains into individual subspaces and then to minimize the subspace distribution distance statistically and geometrically to reduce the spectral drafts. Meanwhile, previous studies have shown that spatial-spectral features and local manifold structure are useful for HSI DA tasks [10], [12], [13], [15], [31]. Therefore, we propose a graph embedding and distribution alignment (GEDA) method for the cross-domain classification of HSIs in this article. By using the characteristics of HSI, domain relation and data label information, the proposed GEDA can simultaneously reduce the distributional shift and geometrical shift between domains. In the GEDA, spatial filtering is used to increase the spatial consistency of HSI, and then two coupling projections A and B are learned to project source and target domains into subspaces, respectively. The data after projection meets the following requirements: 1) The discriminant information of source and target domains is maintained; 2) the marginal and conditional distribution differences between source and target domains are minimized; and 3) the subspace offset between domains is minimized. While keeping the discriminant information in the two domains, an easy transfer learning (EasyTL) method is used to predict the pseudo-labels of the target domain [32], and a graph embedding method is used to learn the intraclass and interclass scatter matrices of both domains. By alternatively updating subspace features and target pseudo labels, the proposed GEDA can effectively align source and target data. Our contribution is of threefolds.

1) A unified DA framework called GEDA is proposed. It simultaneously considers subspace discriminative feature learning, pseudo-label learning, and distribution alignment.
2) The local spatial and spectral discriminative information of HSI are used by means of a simple spatial mean filtering and graph embedding.
3) An effective pseudo-label learning method (i.e., EasyTL) is employed to generate target labels and to promote the minimization of statistical distribution difference between source and target domains.

The rest of the article is organized as follows. Section II introduces the proposed GEDA method. Section III is the experimental part, which compares GEDA with existing DA methods for HSI cross-domain classification. Finally, Section IV concludes the article.

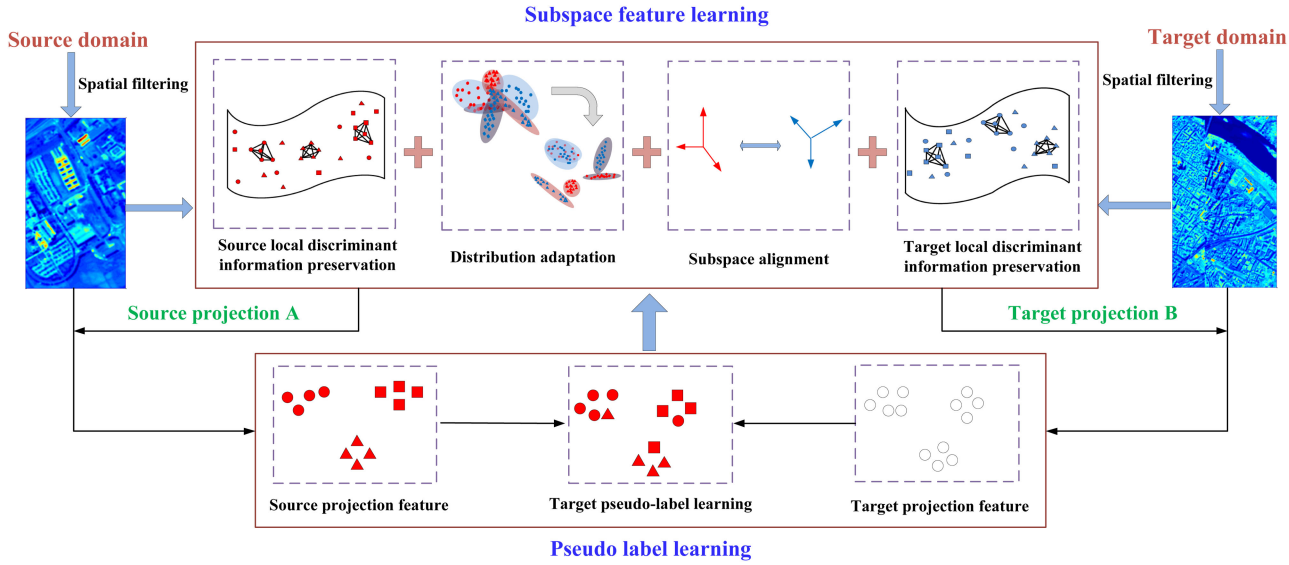

Fig. 1. Flowchart of GEDA.

II. The Proposed Method

Definition 1: A domain $D = \{\chi, P(X)\}$ is composed of a feature space χ and a marginal probability distribution of inputs $P(X)$, where $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \chi$ is a set of learning samples.

Definition 2: A task $T = \{Y, f(\mathbf{x})\}$ consists of classification results Y and a classifier $f(\mathbf{x})$, where $f(\mathbf{x}) = Q(y|\mathbf{x})$ can be interpreted as the conditional probability distribution [20].

For DA, there are two domains, i.e., source domain D_s and target domain D_t. In general, source domain has labels, and target domain has little or no label. HSI cross-domain classification usually focuses on the unsupervised DA problems where the target domain has no label. Given a labeled data set from the source domain, $\{(\mathbf{x}_1^s, y_1^s), (x_2^s, y_2^s), \ldots, (\mathbf{x}_{n_s}^s, y_{n_s}^s)\}$ with $\mathbf{x}_i^s \in \mathcal{R}^d$ and $y_i^s \in \{1, 2, \ldots, C\}$. The sample and label set of source domain are $X_s = \{\mathbf{x}_1^s, \ldots, \mathbf{x}_{n_s}^s\}$ and $Y_s = \{y_1, y_2, \ldots, y_{n_s}\}$, respectively. Let $X_t = \{\mathbf{x}_1^t, \ldots, \mathbf{x}_{n_t}^t\}$ denote the unlabeled data set from the target domain. DA considers the following situation: the class space of source and target domains are the same: $\mathcal{Y}_s = \mathcal{Y}_t$, but the marginal distribution and conditional probability distribution between the domains are inconsistent: $P_s(X_s) \neq P_t(X_t)$ and $Q_s(y_s|X_s) \neq Q_t(y_t|X_t)$. The aim of DA is to predict the target label using the model trained on the source samples.

To alleviate the effect of spectral shifts between source and target domains, the proposed GEDA method projects both domains into subspaces, and uses the potential shared features and intradomain structure information of two domains to reduce the domain differences both statistically and geometrically. By the aid of pseudo-label of target samples predicted by the EasyTL, GEDA can preserve the local spatial and spectral discriminant information by the spatial mean filtering and graph embedding, and meanwhile can effectively reduce the statistical distribution difference and subspace difference between source and target domains. The flowchart of GEDA is shown in Fig. 1. It mainly includes two modules: subspace learning module and pseudo-label learning module. These two modules have coupling interaction. On the one hand, if the features of source and target domains are aligned in the subspace learning module, the subsequent pseudo-label learning is more accurate. On the other hand, if the pseudo-label of target samples is accurate, the discriminative information of target domain will be well preserved and the subspace learning will be more effective.

A. Local Spatial Information Preservation

Due to spatial correlation, spatial mean filtering can be used to maintain the similarity between neighboring pixels and preserve the local neighborhood consistency. Mean filtering is carried out in both domains.

B. Local Spectral Information Preservation

Considering that the source domain D_s and target domain D_t may have great spectral differences, two projection matrices A and B are learned for source and target domains, respectively. The graph embedding method is used to learn the intraclass and interclass divergence matrices of each domain to preserve the intrinsic intraclass compactness and interclass variation of samples [33]. The preservation of local spectral information can be realized by solving the following Fisher-criterion-based optimization problem

$$\min_A \left\{ \frac{\mathrm{Tr}\left(A^{\mathrm{T}} X_s L_w^s X_s^{\mathrm{T}} A\right)}{\mathrm{Tr}\left(A^{\mathrm{T}} X_s L_b^s X_s^{\mathrm{T}} A\right)} = \frac{\mathrm{Tr}\left(A^{\mathrm{T}} S_w^s A\right)}{\mathrm{Tr}\left(A^{\mathrm{T}} S_b^s A\right)} \right\} \tag{1}$$

$$\min_B \left\{ \frac{\mathrm{Tr}\left(B^{\mathrm{T}} X_t L_w^t X_t^{\mathrm{T}} B\right)}{\mathrm{Tr}\left(B^{\mathrm{T}} X_t L_b^t X_t^{\mathrm{T}} B\right)} = \frac{\mathrm{Tr}\left(B^{\mathrm{T}} S_w^t B\right)}{\mathrm{Tr}\left(B^{\mathrm{T}} S_b^t B\right)} \right\} \tag{2}$$

where $L_w^s(L_w^t)$ and $L_b^s(L_b^t)$ are the Laplacian matrices of the intrinsic graph and penalty graph introduced in the source (target) domain [33], respectively; and $S_w^s(S_w^t)$ and $S_b^s(S_b^t)$ are the

intraclass and interclass divergence matrices of the source (target) domain, respectively. The above Fisher criteria can preserve the local spectral similarity and increase the class separability in both domains.

In graph embedding framework [33]–[35], each sample is regarded as a data node and the relation between nodes is described by a weight, such as

$$W_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}}, & \text{if nodes } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where t is a parameter and set to 2 as recommended in Ref. [16]. A graph Laplacian can be calculated by $L = D - W$, where D is a diagonal matrix with $\mathrm{D}_{ii} = \sum_{j \neq i} W_{ij}$. Graph Laplacian matrix L can be used to characterize the intraclass compactness and interclass variation of samples. To compute the graph Laplacian matrices in (1) and (2), it first needs to compute the weight matrix in each domain as follows.

a) Constructing the intrinsic weight matrix W_w: For each sample $\mathbf{x}_i$, connect the k_1-nearest neighbor pair $\mathbf{v}$ and $\mathbf{x}_i$ if $\mathbf{v}$ has the same label information with $\mathbf{x}_i$.
b) Constructing the penalty weight matrix W_b: For each sample $\mathbf{x}_i$, connect the k_2-nearest vertex pairs where samples in each pair belong to different classes.

Following Ref. [16], both the parameters k_1 and k_2 are set to 5 for simplicity.

For constructing the graph Laplacian matrices in the target domain, the pseudo label of target samples is needed. In this article, we use the EasyTL algorithm [32] to predict target domain pseudo labels. EasyTL aims to learn a probability annotation matrix $M \in R^{C \times n_t}$ with element $M_{cj} \in [0, 1]$ denoting the probability of target sample $\mathbf{x}_j^t$ belonging to class c based on the following optimization problem:

$$\min \mathcal{L} = \sum_{c=1}^{C} \sum_{j=1}^{n_t} D_{cj} M_{cj}$$
$$\text{s.t.} \begin{cases} 0 \leq M_{cj} \leq 1 \\ \sum_{c=1}^{C} M_{cj} = 1, & \forall j \in \{1, 2, \ldots, n_t\} \\ \sum_{j=1}^{n_t} M_{cj} \geq 1, & \forall c \in \{1, 2, \ldots, C\} \end{cases} \tag{4}$$

where D_{cj} measures the distance between the sample $\mathbf{x}_j^t$ and the c-th class center of source domain.

Based on the probability annotation matrix M obtained by solving (4), the pseudo label of target sample $\mathbf{x}_j^t$ can be given by

$$y_j^t = \arg\max_r \frac{M_{rj}}{\sum_{c=1}^{C} M_{cj}}, \quad \forall r \in \{1, 2, \ldots, C\}. \tag{5}$$

The advantages of EasyTL are that it takes no parameters and is easy to implement. There are also other algorithms for learning pseudo labels, such as the standard K-nearest neighbor or support vector machine (SVM) classifiers [20], [22], [36], label propagation [16], and label regression [37]. Taking into account both the accuracy and simplicity, EasyTL is used in this manuscript.

C. Distribution Difference Minimization

The joint distribution can be approximately described by the marginal distribution and conditional distribution [20]. The marginal distribution difference is described as the distance between the mean of samples after the projection [20], [38]

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

The conditional distribution difference is approximately expressed as the summation of the difference of projected mean sample of each class from different domains with EasyTL-generated pseudo label of target samples

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

where $X_s^{(c)} \in \mathcal{R}^{d \times n_s^{(c)}}$ and $X_t^{(c)} \in \mathcal{R}^{d \times n_t^{(c)}}$ are the c-th class sample sets of source and target domains, respectively.

D. Subspace Distance Minimization

The subspace difference is minimized based on the SA strategy [17]. Different from the SA that learns only one transformation matrix, here we learn two projection matrices A and B for source and target domains. Then, we directly minimize the distance between subspaces

$$\min_{A,B} \|A - B\|_F^2. \tag{8}$$

E. Objective Function

To simultaneously preserve local spatial and spectral discriminant information and reduce statistical distribution difference and subspace difference between two domains, the objective function of the proposed GEDA can be formulated as

$$\max \frac{\beta \{\text{Between Class}\}_{\mathrm{ST}}}{\left(\begin{array}{c} \{\text{Distribution shift}\} + \lambda \{\text{Subspace shift}\} \\ + \beta \{\text{Within Class}\}_{\mathrm{ST}} \end{array} \right)}$$

where β and λ are parameters, and $\{\text{Within Class}\}_{\mathrm{ST}}$ and $\{\text{Between Class}\}_{\mathrm{ST}}$ represent within-class and between-class divergence matrices of two domains, respectively.

To obtain the explicit form of the above objective function, we first combine (6) and (7) as follows:

$$\min_{A,B} \mathrm{Tr} \left(\begin{bmatrix} A^{\mathrm{T}} & B^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} K_s & K_{st} \\ K_{ts} & K_t \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} \right) \tag{9}$$

where

$$K_s = X_s \left(L_s + \sum_{c=1}^{C} L_s^{(c)} \right) X_s^{\mathrm{T}}, \quad L_s = \frac{1}{n_s^2} \mathbf{1}_s \mathbf{1}_s^{\mathrm{T}}$$
$$\left(L_s^{(c)} \right)_{ij} = \begin{cases} \frac{1}{\left(n_s^{(c)} \right)^2} & \mathbf{x}_i, \mathbf{x}_j \in X_s^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$K_t = X_t \left(L_t + \sum_{c=1}^{C} L_t^{(c)} \right) X_t^{\mathrm{T}}, \quad L_t = \frac{1}{n_t^2} \mathbf{1}_t \mathbf{1}_t^{\mathrm{T}}$$

$$\left(L_t^{(c)}\right)_{ij} = \begin{cases} \frac{1}{\left(n_t^{(c)}\right)^2} & \mathbf{x}_i, \mathbf{x}_j \in X_t^{(c)} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$K_{st} = X_s \left(L_{st} + \sum_{c=1}^{C} L_{st}^{(c)} \right) X_t^{\mathrm{T}}, \quad L_{st} = -\frac{1}{n_s n_t} \mathbf{1}_s \mathbf{1}_t^{\mathrm{T}}$$

$$\left(L_{st}^{(c)}\right)_{ij} = \begin{cases} -\frac{1}{n_s^{(c)} n_t^{(c)}} & \mathbf{x}_i \in X_s^{(c)}, \mathbf{x}_j \in X_t^{(c)} \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

In (9), $K_{ts} = K_{st}$, and $\mathbf{1}_n$ is the column vector with all ones.

Then, we combine (9), (1), (2), and (8) to generate the optimization function

$$\max_U \frac{\mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} \beta S_b^s & \mathbf{0} \\ \mathbf{0} & \beta S_b^t \end{bmatrix} U \right)}{\mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} K_s + \lambda I + \beta S_w^s & K_{st} - \lambda I \\ K_{ts} - \lambda I & K_t + \lambda I + \beta S_w^t \end{bmatrix} U \right)} \tag{13}$$

where $U = [A; B]$, $I \in \mathcal{R}^{d \times d}$ is the identity matrix. As shown in (13), GEDA maximizes the between-class divergence of source and target domains, and meanwhile minimizes their distribution differences, offsets, and within-class divergences.

F. Optimization

Let $N_s = K_s + \lambda I + \beta S_w^s$ and $N_t = K_t + \lambda I + \beta S_w^t$. Then, (13) can be rewritten as

$$\max_U \frac{\mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} \beta S_b^s & \mathbf{0} \\ \mathbf{0} & \beta S_b^t \end{bmatrix} U \right)}{\mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} N_s & K_{st} - \lambda I \\ K_{ts} - \lambda I & N_t \end{bmatrix} U \right)}. \tag{14}$$

The optimization problem (14) is equivalent to the following problem:

$$\begin{aligned} &\max_U \mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} \beta S_b^s & 0 \\ 0 & \beta S_b^t \end{bmatrix} U \right) \\ &\text{s.t. } \mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} N_s & K_{st} - \lambda I \\ K_{ts} - \lambda I & N_t \end{bmatrix} U \right) = 1. \end{aligned} \tag{15}$$

To solve model (15), a Lagrange function is constructed

$$\begin{aligned} L = &\ \mathrm{Tr}\left(U^{\mathrm{T}} \begin{bmatrix} \beta S_b^s & 0 \\ 0 & \beta S_b^t \end{bmatrix} U \right) \\ &+ \mathrm{Tr}\left(\left(U^{\mathrm{T}} \begin{bmatrix} N_s & K_{st} - \lambda I \\ K_{ts} - \lambda I & N_t \end{bmatrix} U - I \right) \Phi \right). \end{aligned} \tag{16}$$

Algorithm 1: GEDA

Input: Source data D_s and label Y_S, target data D_t; parameters: λ, β, subspace dimension k.

Output: Transformation matrices: A and B; Predicted labels Y_t for the target domain.

1. Spatial filtering for source and target data
2. Pseudo-label learning for target domain (EasyTL): $(X_s, X_t, Y_s) \to Y_{t0}$
3. **Repeat**:
   1) Construct S_b^s, S_w^s, S_b^t, S_w^t, K_s, K_t, K_{ts}, and K_{st}.
   2) Solve the generalized eigen-decomposition problem (17) and select the k eigenvectors corresponding to the k largest eigenvalues to form $U = [A; B]$.
   3) Map the original data to respective subspace to get the embeddings:
   $$Z_s = A^{\mathrm{T}} X_s$$
   $$Z_t = B^{\mathrm{T}} X_t$$
   4) Update the pseudo label Y_t of the target domain:
   $$(Z_s, Z_t, Y_s) \to Y_t$$
   Until *Convergence*
4. Obtain the final labels Y_t for target samples.

Taking the derivative of L with respect to U, we can get

$$\begin{bmatrix} \beta S_b^s & 0 \\ 0 & \beta S_b^t \end{bmatrix} U = \begin{bmatrix} N_s & K_{st} - \lambda I \\ K_{ts} - \lambda I & N_t \end{bmatrix} U \Phi \tag{17}$$

where $\Phi = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ are the k leading eigenvalues and $U = [U_1, \ldots, U_k]$ contains the corresponding eigenvectors, which can be solved analytically through generalized eigenvalue decomposition. Once the transformation matrix U is obtained, subspaces A and B can be obtained easily.

The pseudo code of GEDA is summarized in Algorithm 1. We adopt an iterative optimization strategy to alternatively update the target pseudo labels and subspace projection matrices A and B (i.e., U). It should be noted that the proposed GEDA method can be extended to solve the nonlinear problems in the RKHS by using kernel function.

III. Experimental Results

A. Data Sets

Six HSI data sets, i.e., University of Pavia, Center of Pavia, Yancheng, Botswana, Shanghai and Hangzhou, are used in the experiments.

University of Pavia and Center of Pavia: These two scenes were obtained by the ROSIS in 2003. The images have 115 spectral bands in the wavelength range of 0.43–0.86 μm, with a spatial resolution of 1.3 m. After removing noisy bands, there are 103 and 102 bands for the two images. These two scenes contain 610×340 and 1096×492 pixels, respectively.

Yancheng: This data was acquired in April 4, 2019 by the visible-shortwave infrared advanced hyperspectral imager

TABLE I
NUMBER OF SAMPLES IN THE FIRST THREE TASKS

Pavia University and Pavia Center				Yancheng				Botwana			
No.	Class	Source	Target	No.	Class	Source	Target	No.	Class	Source	Target
1	Asphalt	400	400	1	Off shore water	365	940	1	Floodplain grasses1	88	163
2	Meadows	400	400	2	Aquaculture	76	152	2	Floodplain grasses2	136	79
3	Trees	400	400	3	Paddy	427	405	3	Riparian	123	146
4	Bare Soil	400	400	4	River	137	80	4	Island interior	126	77
5	Bitumen	400	400	5	Fallow land	53	181	5	Acacia woodlands	210	104
6	Bricks	400	400	6	Dry land	82	122	6	Acacia shrublands	94	154

(AHSI) over the port of Yancheng City, China. AHSI is the main payload of the China GaoFen-5 (GF-5) satellite. The spectral resolution is 4.278–4.460 nm for VNIR and 8.420–8.450 nm for SWIR, and the spatial resolution is 30 m. The Yancheng image has 1175×585 pixels and 330 spectral bands ranging from 0.4 to 2.5 μm. After removing bad bands, the remaining 267 bands are used.

Botswana: This data was acquired at Botswana by the NASA's EO-1 satellite. The sensor on the EO-1 obtained 30-m pixel resolution data in 242 bands, covering bands of 400–2500 nm and spaced at 10-nm intervals. After removing noisy bands, the remaining 145 bands are used.

Shanghai–Hangzhou: The Shanghai and Hangzhou data sets were captured by the EO-1 Hyperion hyperspectral sensor [6], which retains 198 bands after removing the bad bands. The size of Shanghai image is 1600×230, which includes roads, buildings, plants, and the water of the Yangtze River and Huangpu River. The Hangzhou image size is 590×230, including roads, buildings, plants, West Lake, and Qiantang River basin.

For the DA problem, we need to construct source and target domains such that the data of source and target domains come from two different HSIs or different regions in one HSI. Based on the six HSI data sets, the following four DA tasks are constructed.

1) Task 1 (Pavia University and Center task): The source and target domains are chosen from the Pavia University and Pavia Center images, respectively. To keep the consistence of dimensionality, the first 102 bands of the University of Pavia data set are used for analysis. Six common classes in these two images, i.e., Asphalt, Meadows, Trees, Bare Soil, Bitumen, and Bricks, are used for the DA tasks. We randomly draw 400 samples from each class from each image to form source and target domains, respectively.
2) Task 2 (Yancheng task): The Yancheng image is divided into two disjoint regions for DA. The selected two disjoint regions have similar materials, in which six classes (Off shore water, Aquaculture, Paddy, River, Fallow land, Dry land) are chosen for the classification tasks. The selected six classes in the two regions constitute source and target domains.
3) Task 3 (Botswana task): Similar to Yancheng task, the Botswana image is also divided into two disjoint regions for DA. Six common classes in these two regions are chosen to form source and target domains.
4) Task 4 (Shanghai–Hangzhou task): The source and target domains are set as the Shanghai and Hangzhou images, respectively. Three common classes in these two images, i.e., Water, Land / Building, and Plant, are used for the DA task.

TABLE II
NUMBER OF SAMPLES IN THE SHANGHAI–HANGZHOU TASK

Shanghai-Hangzhou data			
No.	Class	Source	Target
1	Water	123123	18043
2	Land/building	161689	77450
3	Plant	83188	40207

The number of samples in each class for the above four DA tasks are shown in Tables I and II.

B. Comparison Methods

We compare our proposed GEDA with the following DA methods on the four DA tasks.

- NA (No adaptation): directly classifies the target samples based on the model built on the source samples.
- CORAL [23]: aligns source and target covariances.
- GFK [24]: transforms the domain by incorporating an infinite number of subspaces that represent changes in geometric and statistical features from the source domain to the target domain.
- TCA [18]: minimizes the MMD in the kernel principal component analysis feature space.
- TJM [19]: performs feature matching and instance reweighting simultaneously in the TCA framework.
- SA [17]: performs alignment between principal component analysis (PCA)-based source and target subspaces.
- JDA [20]: adapts both the marginal distribution and conditional distribution in a principled dimensionality reduction procedure.
- JGSA [22]: learns two coupled projections to project source and target domains into low-dimensional subspaces to reduce the geometric and statistical distribution differences.
- LPJT [16]: combines distribution matching and landmark selection-based sample adaptation in a feature adaptation framework.
- GEDA: uses spatial-spectral local discriminant information, intradomain structures, and data distribution to reduce geometrically and statistically the differences between domains.

TABLE III
CLASSIFICATION ACCURACIES FOR THE PAVIA UNIVERSITY AND CENTER TASK

	Class							
	1	2	3	4	5	6	OA	κ
NA	0.590	0.728	0.878	0.638	0.068	0.428	0.555	0.466
SA	0.813	0.582	0.903	0.450	0.128	0.478	0.559	0.471
CORAL	0.708	0.603	0.880	0.565	0.175	0.523	0.575	0.490
TJM	0.788	0.570	0.893	0.525	0.133	0.525	0.572	0.487
GFK	0.758	0.663	0.873	0.448	0.105	0.475	0.553	0.464
TCA	0.773	0.583	0.893	0.523	0.140	0.530	0.573	0.488
JDA	0.750	0.640	0.888	0.578	0.148	0.520	0.587	0.505
JGSA	0.703	0.738	0.795	0.590	0.690	0.568	0.680	0.617
LPJT	0.668	0.638	**0.915**	**0.828**	0.370	0.283	0.617	0.540
GEDA	**0.898**	**0.803**	0.900	0.740	**0.825**	**0.965**	**0.855**	**0.826**

Values in bold indicate the best accuracy among all methods.

Using the above DA methods, source and target data can be aligned at a certain extent, and then the classifier trained on the source features can be used to classify the target features. The final performance of each DA method is evaluated by the classification results. In the experiments, the 1-nearest neighbor (NN) classifier is chosen as the classifier because it is simple and parameter-free. For the subspace-based DA methods, the dimension of subspace is set to 20. In the case of randomly selecting samples, the experiment is run 20 times and the average results are reported. For JDA, JGSA, LPJT, and GEDA, the number of iterations T is set as 5. For GEDA, the size of spatial filter is determined by the characteristics of the data. The spatial window of Pavia and Botswana data sets are chosen as 5×5 and 7×7, respectively. The Yancheng and Shanghai–Hangzhou data sets do not need to be filtered. For each method, the overall accuracy (OA) and kappa coefficient (κ) on the target domain are used to evaluate the performance.

C. Experiments

1) Experiments on Pavia University and Center Task: In this task, the samples of source and target domains come from different images (i.e., Pavia University and Pavia Center, respectively) and the corresponding acquisition time is inconsistent, so there is significant spectral difference between source and target domains, and the DA task is challenging.

Table III shows the classification results of different DA methods on the Pavia University and Center task. It can be seen that the classical DA algorithms show poor results in this case and their classification accuracies are almost lower than 60%. Even if the category information of the source domain and the global information of the target domain are used, such as JGSA and LPJT, the classification performance is still not satisfactory. The main reason is that there have great spectral drifts in this cross-scene classification problem. By using the spatial filtering to increase spatial consistence and alleviate spectral variation of HSIs, and meanwhile employing the local discriminant information of source and target domains, the proposed GEDA method can achieve a good classification effect for this task.

Table IV provides the classification confusion matrix of NA, JGSA, and GEDA. NA shows very poor results on Classes 5 ("Bitumen") and 6 ("Bricks"). It misclassifies most of the samples in these two classes to Class 1 ("Asphalt"). As known, "Bitumen" and "Asphalt" are similar materials, so their spectra are very similar. "Bricks" is also similar to "Asphalt." Due to high spectral similarity, it is very likely to produce confusions on these classes (seeing the classification on Classes 1, 5, and 6 of NA). After the JGSA transformations, the classification performance on Classes 1, 5, 6 are improved at a certain extent. However, it still misclassifies most of samples in the "Bitumen" and "Asphalt" classes. Compared with JGSA, our proposed GEDA dramatically improves the classification performance on Classes 1, 5, and 6, and shows better overall results. It can discriminate subtle differences between similar classes using both local spatial-spectral information and distribution information between and within domains.

TABLE IV
CONFUSION MATRIX AND CA OF NA, JGSA, AND GEDA FOR PAVIA UNIVERSITY AND CENTER TASK

		Class predicted by NA						CA
		1	2	3	4	5	6	
True Class	1	**236**	0	0	3	91	70	59.00%
	2	0	**291**	103	6	0	0	72.75%
	3	0	49	**351**	0	0	0	**87.75%**
	4	4	74	1	**255**	1	65	**63.75%**
	5	343	0	0	6	**27**	24	6.75%
	6	78	0	0	132	19	**171**	42.75%
		Class predicted by JGSA						CA
		1	2	3	4	5	6	
True Class	1	**281**	1	1	0	86	31	70.25%
	2	0	**295**	69	36	0	0	73.75%
	3	0	75	**318**	7	0	0	79.50%
	4	5	41	1	**236**	7	110	59.00%
	5	91	1	0	1	**276**	31	69.00%
	6	66	12	0	40	55	**227**	56.75%
		Class predicted by GEDA						CA
		1	2	3	4	5	6	
True Class	1	**359**	0	0	0	1	40	**89.75%**
	2	0	**321**	74	5	0	0	**80.25%**
	3	0	40	**360**	0	0	0	90.00%
	4	0	2	1	**296**	0	101	74.00%
	5	51	0	0	0	**330**	19	**82.50%**
	6	14	0	0	0	0	**386**	**96.50%**

The bold values in the diagonal are the number of correctly classified samples in each class. The bold values in the last column are the highest accuracy in each class.

2) Experiments on Yancheng Task: The classification results on Yancheng data are shown in Table V. It can be seen that all DA methods obtain acceptable results with OA about 90%. As source and target domains come from two disjoint regions in the same HSI and the spectral resolution of Yancheng GF-5 image is very high, the spectral difference between domains is relative small. In particular, Class 3 ("Paddy") is obviously different from other five classes, and all DA methods provide correctly classification results on this class. However, most of DA methods completely misclassify Class 6. Comparing the three geometrical and statistical distribution alignment methods (i.e., JGSA, LPJT, and

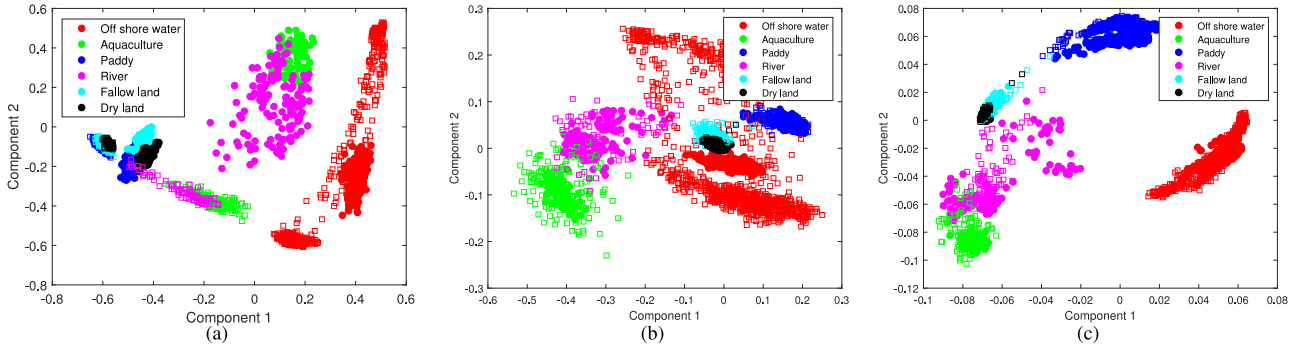

Fig. 2. Scatterplots after (a) original (NA), (b) JGSA feature extraction, (c) GEDA feature extraction, in the first versus second component spaces for the Yancheng task. (●) Source data. (□) Target data.

TABLE V
CLASSIFICATION ACCURACIES FOR YANCHENG TASK

	Class							
	1	2	3	4	5	6	OA	κ
NA	**1.000**	0.533	**1.000**	0.713	0.558	0.975	0.906	0.862
SA	0.996	0.868	**1.000**	0.688	0.978	0.000	0.907	0.863
TCA	0.946	0.816	**1.000**	0.738	**0.983**	0.000	0.880	0.827
CORAL	0.995	0.875	**1.000**	0.700	0.945	0.771	0.954	0.933
GFK	0.999	0.868	**1.000**	0.725	0.972	0.000	0.910	0.867
TJM	0.943	0.776	**1.000**	0.813	**0.983**	0.000	0.879	0.825
JDA	0.902	0.915	**1.000**	0.950	0.978	0.066	0.879	0.828
JGSA	0.999	0.868	**1.000**	**1.000**	**0.983**	0.000	0.922	0.886
LPJT	**1.000**	0.829	**1.000**	0.938	0.856	0.795	0.956	0.936
GEDA	**1.000**	**0.987**	**1.000**	0.838	0.967	**0.984**	**0.988**	**0.982**

Values in bold indicate the best accuracy among all methods.

TABLE VI
CONFUSION MATRIX AND CA OF NA, JGSA, AND GEDA FOR YANCHENG TASK

		Class predicted by NA						CA
		1	2	3	4	5	6	
True Class	1	**940**	0	0	0	0	0	**100.0%**
	2	0	**81**	0	71	0	0	53.29%
	3	0	0	**405**	0	0	0	**100.0%**
	4	0	12	4	**57**	7	0	71.25%
	5	0	0	1	0	**101**	79	55.80%
	6	0	0	0	0	3	**119**	97.54%

		Class predicted by JGSA						CA
		1	2	3	4	5	6	
True Class	1	**939**	0	0	1	0	0	99.89%
	2	0	**132**	0	20	1	0	86.84%
	3	0	0	**405**	0	0	0	**100.0%**
	4	0	0	0	**80**	0	0	**100.0%**
	5	0	0	3	0	**178**	0	**98.34%**
	6	0	0	1	0	121	**0**	0%

		Class predicted by GEDA						CA
		1	2	3	4	5	6	
True Class	1	**940**	0	0	0	0	0	**100.0%**
	2	0	**150**	0	2	0	0	**98.68%**
	3	0	0	**405**	0	0	0	**100.0%**
	4	0	1	2	**67**	9	1	83.75%
	5	0	0	1	0	**175**	5	96.69%
	6	0	0	0	0	2	**120**	**98.36%**

The bold values in the diagonal are the number of correctly classified samples in each class. The bold values in the last column are the highest accuracy in each class.

GEDA), JGSA preserves only the source global discriminant information by using the intraclass and interclass divergence matrices of the source domain, which cannot accurately classify Class 6. Different from the JGSA, LPJT considers both the local discriminant information in source and target domains, so it produces a relatively better results on Class 6. Notwithstanding, LPJT ignores the subspace difference and uses label propagation to learn pseudo label of target domain. The pseudo label may be inaccurate when the spectra of different classes are similar. In our proposed GEDA method, we use the EasyTL method to learn the pseudo label of target samples, and use the structural information of the two domains simultaneously. It can be seen that GEDA yields excellent results on Yancheng data.

Table VI provides the classification confusion matrices for NA, JGSA, and GEDA and the classification accuracy for each class. Although NA shows good classification results on Classes 1, 3, and 6, it mistakenly classifies nearly half of the samples in Class 2 ("Aquaculture") into Class 4 ("River") and at the same time nearly half of the samples in Class 5 ("Fallow land") into Class 6 ("Dry land"). It is clear that the classes "Aquaculture" and "River" are related to water, and "Fallow land" and "Dry land" are subclasses of land. Classes 2 and 4 and Classes 5 and 6 are similar in spectral, so they are difficult to be classified. By performing DA, JGSA improves the classification accuracy of Class 2 and Class 5. However, it misclassifies all the samples of Class 6 into Classes 5 and 3. Classes 6 and 5 are subgroups with similar spectral characteristics and are difficult to distinguish. By adding the learned target domain pseudo labels, GEDA improves the classification accuracy of JGSA on the sixth class, where its classification accuracy increase from 0% to 98.36%. This indicates that it is necessary to make proper use of the class information of target domain.

To visually show the feature transformation results, we display the first two components of PCA of the original data, the first two dimensions of JGSA transformed data, and GEDA

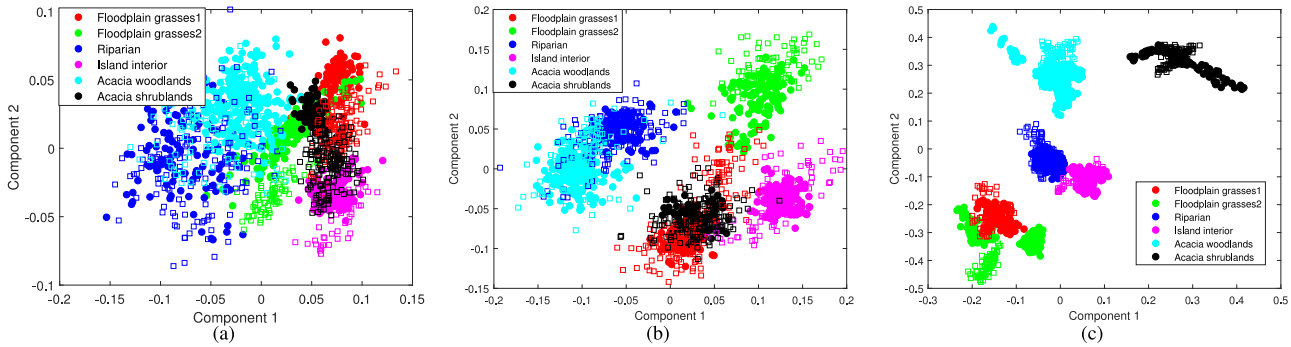

Fig. 3. Scatterplots after (a) original (NA), (b) JGSA feature extraction, (c) GEDA feature extraction, in the first versus second component spaces for the Botwana task. (●) Source data. (□) Target data. .

TABLE VII
CLASSIFICATION ACCURACIES FOR BOTSWANA TASK

	Class							
	1	2	3	4	5	6	OA	κ
NA	0.847	0.987	0.555	0.558	0.817	0.584	0.712	0.649
SA	0.675	0.924	0.870	0.961	0.587	0.812	0.788	0.744
CORAL	0.700	0.924	**0.918**	0.974	0.427	0.747	0.769	0.725
TJM	0.669	0.899	0.774	0.974	0.702	0.818	0.784	0.739
GFK	0.663	0.835	0.753	0.948	0.740	0.805	0.772	0.724
TCA	0.669	0.886	0.781	0.948	0.702	0.838	0.786	0.740
JDA	0.687	0.975	0.856	0.987	0.567	0.838	0.800	0.757
JGSA	0.847	**1.000**	0.774	0.935	0.673	0.974	0.860	0.830
LPJT	**0.969**	**1.000**	0.877	0.961	0.548	0.942	0.887	0.861
GEDA	0.951	**1.000**	**0.918**	**1.000**	**0.971**	**1.000**	**0.968**	**0.961**

Values in bold indicate the best accuracy among all methods.

TABLE VIII
CONFUSION MATRIX AND CA OF NA, JGSA, AND GEDA FOR BOTWANA TASK

		Class predicted by NA						CA
		1	2	3	4	5	6	
True Class	1	**138**	19	0	0	0	6	84.66%
	2	0	**78**	1	0	0	0	98.73%
	3	0	0	**81**	0	65	0	55.48%
	4	16	0	0	**43**	0	18	55.84%
	5	0	0	17	0	**85**	2	81.73%
	6	63	1	0	0	0	**90**	58.44%

		Class predicted by JGSA						CA
		1	2	3	4	5	6	
True Class	1	**138**	16	0	0	0	9	84.66%
	2	0	**79**	0	0	0	0	**100.0%**
	3	0	1	**113**	0	32	0	77.40%
	4	0	0	0	**72**	0	5	93.51%
	5	0	0	34	0	**70**	0	67.31%
	6	0	2	0	4	0	**150**	97.40%

		Class predicted by GEDA						CA
		1	2	3	4	5	6	
True Class	1	**155**	8	0	0	0	0	**95.09%**
	2	0	**79**	0	0	0	0	**100.0%**
	3	0	12	**134**	0	0	0	**91.78%**
	4	0	0	0	**77**	0	0	**100.0%**
	5	0	0	3	0	**101**	0	**97.12%**
	6	0	0	0	0	0	**154**	**100.0%**

The bold values in the diagonal are the number of correctly classified samples in each class. The bold values in the last column are the highest accuracy in each class.

transformed data in Fig. 2. It can be clearly seen that the same class of source and target domains distribute in different regions. That is, there are obvious distribution differences between disjointed source and target domains in the original data. JGSA can improve this situation, but within-class scatter is very large especially for the Class "Off shore water" in red. The proposed GEDA not only can reduce the distribution difference between source and target domains, but also can make the sample points in the same class closer to each other and samples belonging to different classes far away.

3) Experiments on Botswana Task: Table VII shows the classification results on the Botwana data set. The first six DA methods that do not utilize source labels and structural information provide similar results with OA less than 80%. The last four methods that consider the distribution alignment by using the source labels and target pseudo labels produce relatively better results. JGSA improves JDA because it also uses source discriminative information and target variance. By further considering discriminative information of the target domain and sample weight relations, LPJT improves JGSA slightly. The improvements of GEDA over LPJT are pseudo-label learning, SA, and spatial filtering processing.

Table VIII provides the classification confusion matrix of NA, JGSA, and GEDA. For DA problem, the distribution of source and target domains is different, so it is very likely to make a wrong classification of target samples without DA. Here, NA wrongly classifies many samples of Class 3 ("Riparian") to Class 5 ("Acacia woodlands"), and some samples of Class 6 ("Acacia shrublands") to Class 1 ("Floodplain grasses1"). JGSA dramatically improves the results by performing DA with the using of two coupling mappings and the source label information. By using the EasyTL to learn the pseudo label of target domain and the graph embedding method to learn the intraclass and interclass divergences, the proposed GEDA further improves the JGSA and almost correctly classifies all the samples.

TABLE IX
CLASSIFICATION ACCURACIES FOR SHANGHAI–HANGZHOU TASK

Class	NA	SA	CORAL	GFK	TCA	JDA	JGSA	LPJT	GEDA
1	0.996	0.996	0.997	0.993	0.999	**1.000**	0.999	**1.000**	0.988
2	0.737	0.642	0.707	0.641	0.547	0.586	0.775	0.608	**0.901**
3	0.456	0.456	0.606	0.396	0.274	0.590	0.827	0.635	**0.772**
OA	0.688	0.634	0.715	0.615	0.526	0.642	0.820	0.668	**0.874**
κ	0.471	0.401	0.531	0.365	0.239	0.440	0.704	0.479	**0.779**

Values in bold indicate the best accuracy among all methods.

Fig. 3 illustrates the scatterplots of NA, JGSA, and GEDA in first two principal component spaces. Fig. 3(a) shows the differences in the data distribution and mixed sample points are difficult to distinguish. After the JGSA transform in Fig. 3(b), the distribution difference is significantly reduced, and the class distance between sample points is large, and samples in the same class are more clustered. Compared with JGSA, our proposed GEDA in Fig. 3(c) can further reduce the distribution difference and increase the class separability. Even though only two dimensions are used, there is little overlap for different classes. It should be noted that the green class ("Floodplain grasses 2") in the 2-D projection space has three clusters because the original source and target samples in this class have been distributed in three subregions. Notwithstanding, this class can be well classified in a high-dimensional space, i.e., $k = 20$, as shown in Table VIII.

4) Experiments on Shanghai–Hangzhou Task: As the Shanghai–Hangzhou data contains a large number of samples [36], we divide the data into ten parts for the experiment. There are 10 data sets in source and target domains. For one data set in the target domain, we use each data set in the source domain to conduct an experiment and then report the average accuracy. The final classification accuracy is obtained when all the ten data sets in the target domain are classified.

Table IX shows the classification results of different DA methods. Without DA, NA shows very poor results with κ being only 0.471. The subspace learning methods, such as SA, GFK, and TCA, cannot show improvement in this case. Due to the scene difference, there is a large distribution difference between the two domains which cannot be reduced by learning the subspace alone. Rather than learning subspace, CORAL directly aligns the second-order statistical features of the two domains, which shows relatively better results than the subspace learning methods. However, it does not use the label information of source domain. Although JDA considers both the source labels and subspace learning strategy, it only learns one transformation matrix which cannot simultaneously reduce the great distribution difference between domains. To overcome the limitation of JDA, JGSA intends to learn two linear transformations for source and target domains, respectively, such that the transformed source and target domains have less distribution difference in the subspace. JGSA dramatically improves JDA. The proposed GEDA shows the best results for the Shanghai–Hangzhou task.

Fig. 4 displays the ground truth map and classification maps of the target scene by NA, JGSA, and GEDA. Although the classification OA of JGSA is 82%, it is not able to effectively classify "Water" (blue) and "Land/building" (cyan), as shown in the white rectangle region. Compared with JGSA, GEDA has higher classification accuracy in "Land/building" (cyan) and "Plant" (yellow).

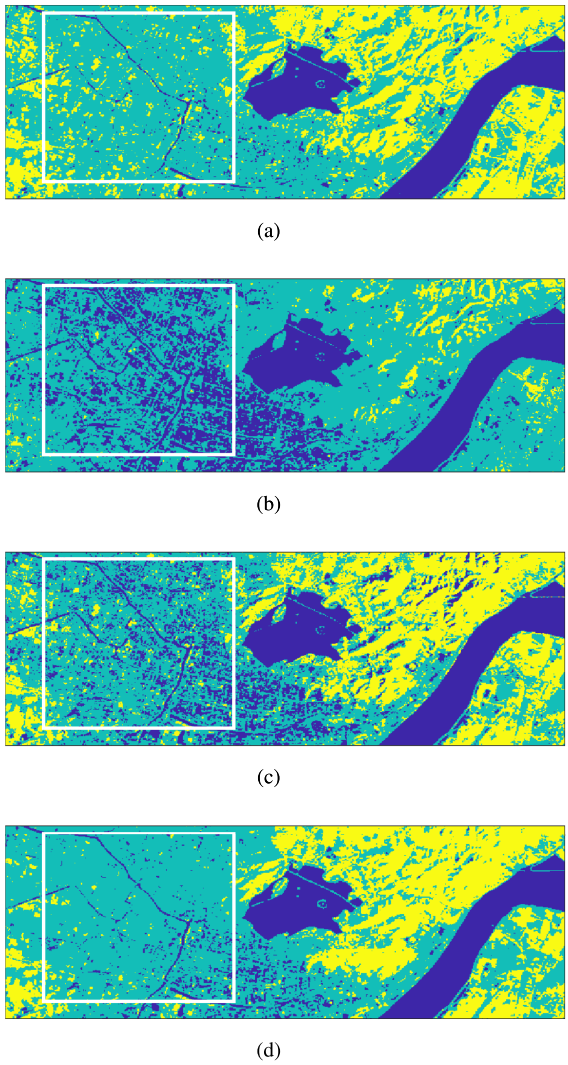

Fig. 4. Visualization and classification map of the target scene Hangzhou. (a) Ground truth map, (b) NA (68.80%), (c) JGSA (82%), and (d) GEDA (87.43%).

TABLE X
ABLATION STUDY OF GEDA FOR PAVIA UNIVERSITY AND CENTER TASK

	JGSA	JGSA + EasyTL	JGSA + GE	JGSA + GE + EasyTL	GEDA
OA	0.6737	0.7479	0.7025	0.7775	0.8550

D. Algorithm Analysis and Discussion

1) Ablation Analysis: In order to analyze the effect of each module in the proposed GEDA, we perform ablation experiments on the Pavia University and Center cross-scene DA task. Table X provides the classification OA. As our GEDA improves the original JGSA method, we first list the result of JGSA and the classification accuracy is 0.6737. Then, we incorporate the EasyTL-based pseudo-label learning and graph embedding into the JGSA, respectively. It can be seen that the resulting OAs are increased at a certain extent. If the EasyTL-based pseudo-label

TABLE XI
CLASSIFICATION ACCURACIES OF DIFFERENT DA METHODS ON SPATIAL FILTERED DATA FOR THE PAVIA UNIVERSITY AND CENTER TASK

	Class							
	1	2	3	4	5	6	OA	κ
NA	0.405	0.683	0.958	0.425	0.630	0.410	0.585	0.502
SA	0.800	0.708	0.908	0.850	0.838	0.710	0.802	0.763
CORAL	0.538	0.670	**0.963**	0.530	0.630	0.553	0.648	0.577
TJM	0.680	0.660	0.908	0.540	0.856	0.825	0.745	0.694
GFK	0.808	0.705	0.908	0.853	0.838	0.708	0.803	0.764
TCA	0.705	0.673	0.910	0.545	**0.860**	0.825	0.753	0.704
JDA	0.803	0.720	0.895	0.855	0.823	0.695	0.799	0.759
JGSA	0.850	0.655	0.958	0.843	0.753	0.850	0.818	0.782
LPJT	0.823	0.723	0.935	**0.918**	0.818	0.750	0.828	0.793
GEDA	**0.898**	**0.803**	0.900	0.740	0.825	**0.965**	**0.855**	**0.826**

Values in bold indicate the best accuracy among all methods.

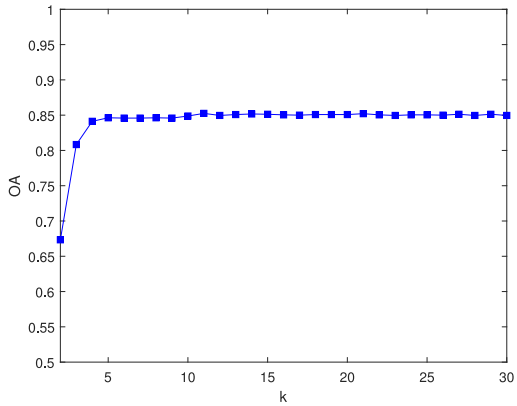

Fig. 5. OA versus subspace dimension k on Pavia University and Center task.

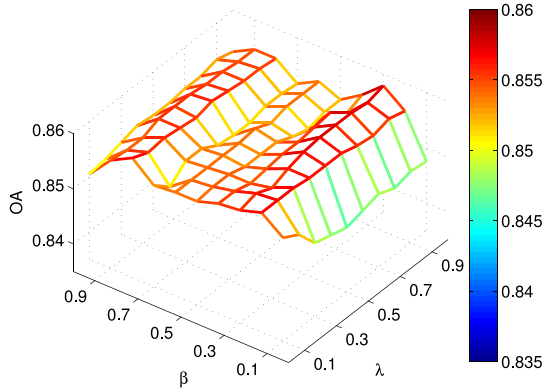

Fig. 6. The OA versus parameters β and λ on Pavia University and Center task.

TABLE XII
RUNNING TIME OF EACH METHOD FOR THE PAVIA UNIVERSITY AND CENTER TASK

Method	Times (s)
NA	1.08
SA	0.34
CORAL	0.27
GFK	0.51
TCA	9.47
TJM	90.03
JDA	9.04
JGSA	12.29
LPJT	143.13
GEDA	20.14

learning and graph embedding are simultaneously added into JGSA, the accuracy is further improved with OA of 0.7775. Finally, we consider all the above strategies and add spatial filtering into the model, and the classification accuracy is 0.8550. The results show that all modules in the proposed GEDA plays a role in classification improvement.

To further compare the performance of different DA methods when spatial mean filtering is performed, Table XI provides the classification accuracy of various algorithms on the spatial filtered data. When the results in Tables III and XI are compared, it can be seen that the performance of different DA methods has improved at a certain extent when spatial filtering is used, and the proposed GEDA provides the best results no matter the spatial filtering is used or not.

2) Parameter Analysis: In the proposed GEDA method, the subspace dimension k, the regularization parameter λ, and β are key parameters. We analyze the effect of these parameters on the Pavia University and Center data. The relationship between subspace dimension k and OA is shown in Fig. 5. It can be observed that the performance of GEDA is stable when the dimension of subspace is between 7 and 30. In the experiment, the dimension of the subspace is fixed as 20.

The OA versus parameters β and λ is provided in Fig. 6. It can be seen that the model is insensitive to parameters β and λ. In the experiment, parameters λ and β are set as 1 and 0.3, respectively.

3) Running Time: The running time of different algorithms on the Pavia University and Center task are shown in Table XII. Due to the iterative updating of target pseudo labels and subspace features, the proposed GEDA method is computationally less efficiency than SA, CORAL, and GFK. However, it is much more efficient than TJM and LPJT, and comparable with JGSA.

IV. CONCLUSION

In this article, we have proposed a new DA method based on the graph embedding and distribution alignment (GEDA). For cross-domain classification of HSIs, GEDA learns two coupling mappings by using the discriminant information of source and target domains, so that the mapped distributions of the two domains are close to each other, and the local discriminant information is maintained. The results of four DA tasks in the experiments demonstrate that GEDA can effectively perform cross-domain classification of HSI.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] J. Peng, W. Sun, and Q. Du, "Self-paced joint sparse representation for the classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1183–1194, Feb. 2019.
[3] J. Peng, L. Li, and Y. Y. Tang, "Maximum likelihood estimation based joint sparse representation for the classification of hyperspectral remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 1790–1802, Jun. 2019.
[4] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
[5] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 41–57, Jun. 2016.
[6] 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.
[7] 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.
[8] W. Yang, J. Peng, and W. Sun, "Ideal regularized discriminative multiple kernel subspace alignment for domain adaptation in hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5833–5846, 2020.
[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
[10] 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.
[11] 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.
[12] H. L. Yang and M. M. Crawford, "Domain adaptation with preservation of manifold geometry for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 543–555, Feb. 2016.
[13] 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.
[14] Q. Shi, B. Du, and L. Zhang, "Domain adaptation for remote sensing image classification: A low-rank reconstruction and instance weighting label propagation inspired algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5677–5689, Oct. 2015.
[15] 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.
[16] 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.
[17] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2960–2967.
[18] 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.
[19] 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.
[20] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 2200–2207.
[21] M. Ghifary, D. Balduzzi, W. B. Kleijn, and M. Zhang, "Scatter component analysis: A unified framework for domain adaptation and domain generalization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1414–1430, Jul. 2017.
[22] 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. 5150–5158.
[23] B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 2058–2065.
[24] 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.
[25] 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.
[26] 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.
[27] A. S. Garea, D. B. Heras, and F. Argüello, "TCANet for domain adaptation of hyperspectral images," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2289.
[28] 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.
[29] 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.
[30] 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.*, to be published, doi: 10.1109/TGRS.2021.3057066.
[31] 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.
[32] 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 and Expo.*, 2019, pp. 1210–1215.
[33] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[34] Y. Zhou, J. Peng, and C. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.
[35] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.
[36] Y. Zhang, W. Li, R. Tao, J. Peng, Q. Du, and Z. Cai, "Cross-scene hyperspectral image classification with discriminative cooperative alignment," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: 10.1109/TGRS.2020.3046756.
[37] L. Luo *et al.*, "Discriminative label consistent domain adaptation," 2018, *arXiv:1802.08077*.
[38] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, "Integrating structured biological data by Kernel maximum mean discrepancy," *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, 2006.

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

Her research interests include machine learning and hyperspectral image processing.

Jiangtao Peng (Senior Member, IEEE) received the B.S. degree in information and computing science and 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. His research interests include machine learning and hyperspectral image processing.

Yujie Ning received the B.S. degree in mathematics and applied mathematics from the Faculty of Mathematics and Statistics, Hubei University, Wuhan, China, in 2021, 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 the 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 a Visiting Scholar with the Department of Applied Mathematics, University of Maryland, College Park, MD, USA, to study on the dimensionality reduction of hyperspectral Image. From 2014 to 2016, he was a Postdoc with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, to study intelligent processing in hyperspectral imagery. From 2017 to 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. He is currently a Full Professor with Ningbo University, Ningbo, China. He has authored or coauthored more than 70 journal papers. His current research interests include hyperspectral image processing with manifold learning, anomaly detection, and target recognition of remote sensing imagery using compressive sensing.

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, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013, the Chair for the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014, and 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. She was an Associate Editor for the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *Journal of Applied Remote Sensing*, and *IEEE Signal Processing Letters*. From 2016 to 2020, she was the Editor-in-Chief of the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. She is currently a Member of the IEEE Periodicals Review and Advisory Committee. She is a Fellow of SPIE-International Society for Optics and Photonics.