Adaptive Local Discriminant Analysis and Distribution Matching for Domain Adaptation in Hyperspectral Image Classification

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

Abstract—Multimodally distributed data is very common in remote sensing images, such as hyperspectral images (HSIs). It is important to capture the local manifold structure while preserving the global discriminant information in the multimodal data. In this article, an adaptive local discriminant analysis and distribution matching (ALDADM) method is designed for the domain adaptation (DA) in HSI classification. ALDADM uses adaptive local discriminant analysis to extract features that are discriminative and robust to multimodally distributed data. Meanwhile, considering the spectral properties of HSI, the domain shift is reduced by distribution matching and subspace alignment. During this learning process, ALDADM only selects density peak samples to reduce the influence of interfering samples. Four DA tasks are designing using University of Pavia, Center of Pavia, Yancheng, and Botswana HSI datasets. Compared with existing DA methods, the experimental results demonstrate that our ALDADM offers a superior performance.

Index Terms—Adaptive local discriminant analysis, distribution matching, domain adaptation (DA), hyperspectral images (HSIs).

I. INTRODUCTION

Due to the development of hyperspectral technology, a large number of hyperspectral images (HSIs) can be collected in a short time period. HSI classification has been an important task in real applications. However, labeling HSI is very difficult because it is costly and time-consuming. Therefore, an automatic and efficient method is desired to complete image labeling [1]. An intuitive idea is to train a classifier with already labeled samples and use it directly on new images. However, spectral features may be very different in a new image due to the changed soil properties, environment conditions, and the incident angle of sunlight [2], [3].

In such a situation, domain adaptation (DA) techniques allow us to transfer the knowledge learned from a labeled source domain to a target domain with limited or no labeled samples. This classification problem can be called as cross-domain classification [4]–[6]. In cross-domain classification, the source and target pixels may have large spectral shifts. For example, even if the data comes from the same scene and the same sensor, the distribution of samples can change due to different collection times. When the data comes from different scenes, the pixels of different domains have large spectral shift even if they are the same land cover class. When the sensors are different, they are associated with different feature spaces [7], [8]. Although there exist distribution shift between source and target domains, they are mostly related (e.g., same scene, same sensor, similar materials). Therefore, it is possible and necessary to reduce their distribution differences for appropriate knowledge transfer.

In the past few years, many DA methods have been proposed. Traditional DA methods can be instance-based, classifier-based, or feature-based methods [9]–[11]. Instance-based methods align images by reweighting or picking out important labeled source samples [12]–[14]. Classifier-based methods aim to adapt the classifier trained on the source domain and classify the target domain [15], [16]. Feature-based methods intend to reduce the geometrical or statistical distribution differences between domains in the feature space [17]. The geometrical difference is usually measured as the distance between the subspaces of two domains. Statistical distribution differences represent marginal or/and conditional distribution differences between two domains [18]. Meanwhile, feature-based methods can also be used in conjunction with the first two methods.

Feature-based DA methods are very common. Subspace alignment (SA) learns a transformation matrix that makes the subspaces corresponding to two domains closer together [19], [20]. Correlation alignment (CORAL) learns a linear transformation to transform the source data such that the covariance of transformed source data and target data are close [21]. In
[22], a CORAL-based graph neural network is proposed for DA of HSIs, where a domain-wise and a class-wise CORAL are involved. Geodesic flow kernel (GFK) constructs a geodesic flow between the source and target manifold subspaces to characterize geometric and statistical changes between them [23]. An unsupervised manifold alignment method is proposed for DA in [24]. Transfer component analysis (TCA) maps the source and target domains into a reproducing kernel Hilbert space, where the two domain distributions become similar [25]. Transfer joint matching performs feature matching and source sample weighting across domains in a principled dimensionality reduction process [7], [26]. Based on TCA, joint distribution adaptation (IDA) performs both marginal and conditional distribution adaptation by using pseudolabels of the target domain [27]. Similarly, a discriminative cooperative alignment method is developed to align subspace and distribution of domains [28]. Adaptive local neighbors for transfer discriminative feature learning (ALN-TDFL) method reduces marginal and conditional distribution differences and considers both global discriminative consistency and local information preservation when learning domain-invariant features [29]. Scatter component analysis finds a transformation that can maximize the match between domains to improve the class separability [30].

These traditional DA methods all learn a unified map that projects the source and target domains into a shared subspace. However when the domain shift is too large, there may not exist a common subspace. Therefore, joint geometrical and statistical alignment (JGSA) learns two coupled maps for the source and target domains, considering the discriminative information of source domain and the global information preservation of target domain in addition to distribution matching and SA [31], [32]. However, it ignores the manifold structure of the data. If we force the points of the same class to be closer together, the local data structure can be lost, the transferability of features will be reduced, and it will be difficult to deal with multimodally distributed data. Based on the linear discriminant analysis, locality adaptive discriminant analysis learns transformation matrices to pull similar points together and push dissimilar points farther, thereby capturing the local manifold structure of the data [33]. In addition, many DA strategies involve all source samples and pseudolabeled target samples to learn transformations and iteratively updates pseudolabels based on the previous pseudolabels. The mislabeled samples will disturb the learning process. Discriminant geometrical and statistical alignment method selects density peak landmarks to reduce the interference from the outliers in the source domain and the mispredicted instances in the target domain [34].

To align statistical distribution and subspace between domains and preserve local manifold structure with the use of landmarks, we propose an adaptive local discriminant analysis and distribution matching (ALDADM) for unsupervised DA in this article. It learns two coupled projections for the source and target domains by performing the following steps, i.e., landmark selection, feature alignment, target pseudolabel learning. The density peak based method is first introduced to select landmarks on source and target domains [34], respectively. Then, the features of source and target domains are aligned by simultaneously considering the marginal and conditional distribution alignment, SA, and the local manifold structure and data variance preservation of both domains. In the pseudolabel learning process, graph-based label propagation (GLP) is used to predict labels of the target domain [35]. Overall, the main contributions of this work are summarized as follows.

1) A novel unsupervised DA learning framework is proposed, which well integrates instance-based, feature-based, and classifier-based methods, i.e., instance-based landmark selection, local feature alignment, and pseudolabel learning.

2) A GLP method is used to predict labels of the target domain, and a density peak landmark selection (DPLs) is introduced to filter out more informative samples in source domain and high-confidence pseudolabeled samples in target domain.

3) A local adaptive feature learning module for source and target domains is proposed to handle multimodally distributed data in HSIs, making the learned features discriminative and robust.

The rest of this article is organized as follows. Section II details our proposed ALDADM method. Experimental results and analysis are shown in Section III. Finally, Section IV concludes this article.

II. PROPOSED METHOD

Denote the number of samples in the source and target domains as \( n_s \) and \( n_t \), respectively, and the number of classes as \( C \). Given a labeled source domain \( D_s = \{ (x_i^s, y_i^s) \} \) with \( x_i^s \in \mathbb{R}^d \) and a label set \( L = \{1, 2, \ldots, C\} \) with \( y_i^s \in L \), and an unlabeled target domain \( D_t = \{ x_i^t \} \), \( x_i^t \in \mathbb{R}^d \).

In the DA problem, the marginal or/and conditional distribution between domains are inconsistent. Due to the nature of HSIs, spatial mean filtering is applied to locally preserve the spatial information. In the following, the data samples of two domains are all filtered and denoted as \( X_s \in \mathbb{R}^{d \times n_s}, X_t \in \mathbb{R}^{d \times n_t} \).

Fig. 1 shows the framework of ALDADM. It learns two coupled projection matrices \( A \) and \( B \) for the source and target domains, respectively, by performing landmark selection, feature learning, and target pseudolabel learning.

A. Landmark Selection

Compared to using all instances, selecting the most relevant instances (landmarks) will benefit the subsequent feature learning. With the use of the DPLs method [34], it is likely to select informative instances in \( D_s \) and pseudolabeled instances with high confidence in \( D_t \). DPLs assumes that in the same class, the main instances (cores or landmarks) are clustered together, and some abnormal instances (outliers) are far away from the cluster center.

To implement DPLs, two quantities are computed for each instance \( x \in C_k \), i.e., the intraclass density \( \rho_a(x) \) and the interclass density \( \rho_b(x) \), defined as [34]

\[
\rho_a(x) = \frac{1}{n_k - 1} \sum_{x_i \in C_k, x_i \neq x} K(x, x_i)
\]

(1)
\[ \rho_b(x) = \frac{1}{n_k} \sum_{x_i \in \mathcal{C}_k} K(x, x_i) \]  

(2)

where \( \mathcal{C}_k \) denotes the sample set of the \( k \)th class, \( \mathcal{C}_s \) is the sample set of all other classes except for the \( k \)th class, \( n_k \) and \( n_s \) denote the number of instances in \( \mathcal{C}_k \) and \( \mathcal{C}_s \), respectively, and \( K \) represents the RBF kernel with

\[ K(x, x_i) = \exp \left( -\frac{\|x - x_i\|^2}{2\sigma^2} \right) \]  

(3)

being the similarity between \( x \) and \( x_i \). The closer \( x \) and \( x_i \) are, the larger \( K(x, x_i) \) is. For simplicity, the parameter \( \sigma \) is set to be the median of pairwise distance between source samples or target samples. \( \rho_b(x) \) is called as intraclass density which measures the similarity between \( x \) and other samples in the same class. Similarly, \( \rho_b(x) \) is the interclass density, which measures the similarity between \( x \) and samples in different classes. Finally, a relative density is defined

\[ \Delta \rho(x) = \rho_s(x) - \rho_b(x). \]  

(4)

It is clear that a larger \( \Delta \rho(x) \) means that \( x \) is distributed around the instances with the same label and far away from the instances with different classes. Usually, we calculate the \( \Delta \rho \) of all instances, sort them from large to small, and select the top \( r\% \) instances as DPLs. In the experiments, we compute the relative densities \( \Delta \rho_s(x) \) and \( \Delta \rho_t(x) \) for the source and target domains, respectively. Because the target domain \( D_t \) has no labels, the pseudolabels of target domain is used to compute the relative density \( \Delta \rho_t(x) \).

B. Feature Learning

The domain invariant features are learned by performing distribution and SA with local discriminative information and target variance preservation.

1) Minimizing Distribution Differences: The maximum mean discrepancy (MMD) is used to measure the marginal and conditional probability distribution differences [36]. Considering the large spectral difference between domains in reality, two coupled mappings \( A \) and \( B \) for source and target domains are learned to firstly minimize the marginal distribution difference as [31]

\[ \min_{A,B} \left\| \frac{1}{n_s} \sum_{x_i \in X_s} A^T x_i - \frac{1}{n_t} \sum_{x_j \in X_t} B^T x_j \right\|_F^2. \]  

(5)

For the conditional distribution difference, the following class-wise MMD is computed:

\[ \min_{A,B} \frac{1}{n_s} \sum_{c=1}^C \left( \frac{1}{n_s^{(c)}} \sum_{x_i \in X_s^{(c)}} A^T x_i - \frac{1}{n_t^{(c)}} \sum_{x_j \in X_t^{(c)}} B^T x_j \right)_F^2. \]  

(6)

where \( X_s^{(c)} \) and \( X_t^{(c)} \) are the sample set of the \( c \)th class in the source and target domains, and \( n_s^{(c)} \) and \( n_t^{(c)} \) are the number of samples in these two sets, respectively. Combining the marginal and conditional distribution minimization terms, the final minimization term is

\[ \min_{A,B} \text{Tr} \left( \begin{bmatrix} A^T & B^T \end{bmatrix} \begin{bmatrix} K_s & K_{st} \\ K_{ts} & K_t \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} \right). \]  

(7)

where

\[ K_s = X_s X_s^T, \quad L_s = \frac{1}{n_s} 1_s 1_s^T \]

\[ (L_s^{(c)})_{ij} = \begin{cases} \frac{1}{(n_s^{(c)})^2} & x_i, x_j \in X_s^{(c)} \\ 0 & \text{otherwise} \end{cases} \]  

(8)

\[ K_t = X_t X_t^T, \quad L_t = \frac{1}{n_t} 1_t 1_t^T \]

\[ (L_t^{(c)})_{ij} = \begin{cases} \frac{1}{(n_t^{(c)})^2} & x_i, x_j \in X_t^{(c)} \\ 0 & \text{otherwise} \end{cases} \]  

(9)

Fig. 1. Framework of ALDADM. It mainly contains three parts, i.e., landmark selection, feature learning, and target pseudolabel learning.
\[
K_{st} = X_s \left( L_{st} + \sum_{c=1}^{C} L_{st}^{(c)} \right) X_t^T, \quad L_{st} = -\frac{1}{n_s n_t} 1_s 1_t^T
\]

\[
(L_{st}^{(c)})_{ij} = \begin{cases} \frac{-1}{n_s n_t} & x_i \in X_s^{(c)}, x_j \in X_t^{(c)} \\ 0 & \text{otherwise} \end{cases}
\]  

(10)

\[
K_{ts} = X_t \left( L_{ts} + \sum_{c=1}^{C} L_{ts}^{(c)} \right) X_s^T, \quad L_{ts} = -\frac{1}{n_s n_t} 1_t 1_s^T
\]

\[
(L_{ts}^{(c)})_{ij} = \begin{cases} \frac{-1}{n_s n_t} & x_j \in X_s^{(c)}, x_i \in X_t^{(c)} \\ 0 & \text{otherwise} \end{cases}
\]  

(11)

where \(1_{s/t}\) is the column vector with all ones.

2) Minimizing Subspace Differences: To make the two domains more geometrically aligned, we make the source and target subspaces closer together

\[
\min_{A,B} \| A - B \|_F^2. \tag{12}
\]

3) Maximizing Target Variance: In order to make the projected features retain more original information, we maximize the variance of the target domain as

\[
\max_{B} \text{Tr} \left( B^T S_t B \right) \tag{13}
\]

where \(S_t = X_t H_t X_t^T\) is the variance of target domain, \(H_t = I_t - \frac{1}{n_t} 1_t 1_t^T\) is the centering matrix with \(1_t \in \mathbb{R}^{n_t}\) being the column vector with all ones.

4) Local Discriminative Feature Learning: Multimodally distributed data means that the samples in the same class are distributed in several groups, which is most likely the case in HSIs. Therefore, while maintaining the global relationships of data points, it is equally important to discover the local manifold structure [29]. To address this issue, we propose to incorporate the preservation of local manifold structure into the process of domain-invariant feature learning.

To preserve local manifold structure, a similarity weight matrix for each class is constructed such that the similar points from the same class are as close as possible. For two domains, the formulas are as follows:

\[
L_{\text{same}}^{(s)} = \sum_{c=1}^{C} n_s^{(c)} \sum_{i,j=1}^{n_s^{(c)}} W_{ij}^{(s,c)} || A^T x_i - A^T x_j ||_2^2
\]

\[
= \text{Tr} \left( A^T X_s G_s X_s^T A \right) \tag{14}
\]

\[
L_{\text{same}}^{(t)} = \sum_{c=1}^{C} n_t^{(c)} \sum_{i,j=1}^{n_t^{(c)}} W_{ij}^{(t,c)} || B^T x_i - B^T x_j ||_2^2
\]

\[
= \text{Tr} \left( B^T X_t G_t X_t^T B \right) \tag{15}
\]

where the matrix \(W^{(s/t)} \in \mathbb{R}^{n_{(s/t)} \times n_{(s/t)}}\) is used to capture local relationships between data points, and \(W_{ij}^{(s/t,c)} = 0\) if \(y_i^{(s/t)} \neq y_j^{(s/t)}\).

Define \(d_i = \sum_{j=1}^{n_t} V_{ij}^{(s/t)} D^{(s/t)} = \text{diag}(d_1, \ldots, d_{n_t})\), thus, \(G_{st} = D^{(s/t)} - V^{(s/t)}\), where \(V^{(s/t)}\) is formulated as

\[
V_{ij}^{(s/t)} = \begin{cases} n_{ij}^{(s/t)} W_{ij}^{(s/t,c)}^2 & \text{if } y_i^{(s/t)} = y_j^{(s/t)} \\ 0 & \text{otherwise} \end{cases}
\]

(16)

C. Target Pseudolabel Learning

In the target landmark selection [i.e., the calculate of relative density in (4)], conditional distribution alignment in (6) and local discriminative feature learning in (15), it needs the pseudolabels of target domain. The learning of pseudolabels is critical. Once some instances’ pseudolabels are predicted incorrectly, it will inevitably interfere with subsequent iterative learning process. Different from the most commonly used \(k\)-nearest neighbor classifier, we use the GLP strategy to complete the prediction of target domain [35]. Given a whole data matrix \(X = [X_s, X_t] \in \mathbb{R}^{d \times (n_s + n_t)}\) and a label set \(\mathcal{L} = \{1, 2, \ldots, C\}\). Define the initial \((n_s + n_t) \times C\) label matrix \(Y\) as \(Y_{ij} = 1\) if \(x_i\) is labeled as \(y_i = j\), \(Y_{ij} = 0\) otherwise. The objective of GLP is to learn a \((n_s + n_t) \times C\) label matrix \(F\) to assign the label of \(x_i\) as: \(y_i = \arg \max_{j \in C} F_{ij}\). In detail, it first constructs an affinity matrix \(N\) according to the similarity between data, and then calculates the matrix \(S = E^{-1/2} N E^{-1/2}\), in which \(E = \text{diag}(e_1, \ldots, e_{n_s + n_t})\) is a diagonal matrix and \(e_t = \sum_{j} I_{ij}\). Then, it updates the label matrix according to \(F(t + 1) = \lambda S F(t) + (1 - \lambda) Y\) until convergence, where \(\lambda\) is a parameter in \((0, 1)\), and \(F(0) = Y\). Finally, the label of the target instance \(x'_j\) can be given by

\[
y_{tj} = \arg \max_{r \in \mathcal{L}} F_{rj} \quad \forall r \in \mathcal{L}, j = \{n_s + 1, \ldots, n_s + n_t\}. \tag{17}
\]

D. Objective Function

ALDADM combines (7) and (12)–(15) to obtain the objective function as follows:

\[
\max_{A,B \{\text{Target var}\}} \left\{ \frac{\text{Distribution shift} + \alpha \{\text{Subspace shift} + \beta \{\text{local}\}_{ST}} \right\}
\]

(18)

where \(\alpha\) and \(\beta\) are weighting parameters and \(\{\text{local}\}_{ST}\) is a source and target local discriminative term. The final expression is specifically

\[
\max_{A,B} \frac{\text{Tr} \left( P^T \begin{bmatrix} 0 & 0 \\ 0 & S_t \end{bmatrix} P \right)}{\text{Tr} \left( P^T \begin{bmatrix} K_s + \alpha I + \beta J_s & K_{st} - \alpha I \\ K_{ts} - \alpha I & K_t + (\alpha + \mu) I + \beta J_t \end{bmatrix} P \right)}
\]

(19)

where \(J_s = X_s G_s X_s^T, J_t = X_t G_t X_t^T, P^T = \begin{bmatrix} A^T & B^T \end{bmatrix}\), and \(I \in \mathbb{R}^{d \times d}\) is the identity matrix. As recommended in [31], a parameter \(\mu\) is used to control the scale of \(B\) and set as 0.01 in the experiments.
E. Optimization

An alternative iteration strategy can be used to solve the optimization problem (19). The weight matrix $W^{(s/t)}$ is first initialized. Only the points in the same class $c$ have the weight $1/n_s^{(c)}$ or $1/n_t^{(c)}$, while others are zero. Then, the (19) can be solved by updating $P$ (i.e., $A, B$) and $W^{(s/t)}$ iteratively.

1) Optimizing mappings $P$: Let $M_s = K_s + \alpha I + \beta J_s$ and $M_t = K_t + (\alpha + \mu)I + \beta J_t$. Given $W^{(s/t)}$, (19) can be transformed into

$$\max_P \text{Tr} \left( P^T \begin{bmatrix} 0 & 0 \\ 0 & S_t \end{bmatrix} P \right) = 0$$

s.t. $\text{Tr} \left( P^T \begin{bmatrix} M_s & K_{st} - \alpha I \\ K_{ts} - \alpha I & M_t \end{bmatrix} P \right) = 1.$

(20)

To solve (20), the Lagrange function can be formulated as

$$L = \text{Tr} \left( P^T \begin{bmatrix} 0 & 0 \\ 0 & S_t \end{bmatrix} P \right) - \text{Tr} \left( P^T \begin{bmatrix} M_s & K_{st} - \alpha I \\ K_{ts} - \alpha I & M_t \end{bmatrix} P \right) \Phi$$

(21)

where $\Phi = \text{diag}(\lambda_1, \ldots, \lambda_k)$ are Lagrange multipliers. Setting $\partial L/\partial P = 0$, it can obtain a generalized eigenvector problem

$$\begin{bmatrix} 0 & 0 \\ 0 & S_t \end{bmatrix} P = \begin{bmatrix} M_s & K_{st} - \alpha I \\ K_{ts} - \alpha I & M_t \end{bmatrix} P \Phi.$$  \hspace{1cm} (22)

As a result, $P$ contains the corresponding $k$ eigenvectors of (22).

2) Optimizing weight matrix $W^{(s/t)}$: When $P$ and $W^{(t)}$ are fixed, then (19) becomes

$$\min_{W^{(s)}} \sum_{c=1}^C n_s^{(c)} \sum_{i,j=1} n_s^{(c)} W^{(s,c)}_{ij} \left\| A^T x_i - A^T x_j \right\|_2^2$$

s.t. $W^{(s)} \succeq 0, \quad \sum_j W^{(s)}_{ij} = 1 \quad \forall i.$  \hspace{1cm} (23)

Note that (23) is independent between different $c$ and $i$, so we can solve the following problem individually:

$$\min_{W^{(s,c)}} \sum_{j=1}^{n_s^{(c)}} W^{(s,c)}_{ij} \left\| A^T x_i - A^T x_j \right\|_2^2$$

s.t. $W^{(s,c)}_{ij} \geq 0, \quad \sum_j W^{(s,c)}_{ij} = 1 \quad \forall j.$  \hspace{1cm} (24)

According to (24), the optimal solution is as follows:

$$W^{(s,c)}_{ij} = \frac{1}{\left\| A^T x_i - A^T x_j \right\|_2^2} \left( \sum_{t=1}^{n_s^{(c)}} \frac{1}{\left\| A^T x_i - A^T x_t \right\|_2^2} \right)^{-1}.$$  \hspace{1cm} (25)

Algorithm 1: ALDADM.

Input: The source domain data $D_s = \{x_i, y_i\}_{i=1}^{n_s}$, the target domain data $D_t = \{x_j\}_{j=1}^{n_t}$, the regularization parameters $\alpha$ and $\beta$.

Output: Target labels $Y_t$.

Begin

1: Spatial filtering preprocessing.
2: Predict $Y_t$ by Eq. (17), initialize $W^{(s/t)}$ and compute $G^{(s,t)}$ by Eq. (16).
3: Repeat:
   1) Select the source (target) landmarks based on Eq. (4).
   2) Construct $S_s, K_s, K_t, K_{st}, K_{ts}$ and $L^{(s,t)}_{same}$ using only density peak landmarks.
   3) Solve projection matrix $P$ (i.e., $A, B$) in (22).
   4) Transform the data into subspaces: $Z_s = A^T X_s, \quad Z_t = B^T X_t$.
   5) Update $W^{(s,t)}$ by Eqs. (25) and (26).
6: Update the target pseudo label $Y_t$ based on Eq. (17).

Until convergence

Similarly

$$W^{(t,c)}_{ij} = \frac{1}{\left\| B^T x_i - B^T x_j \right\|_2^2} \left( \sum_{t=1}^{n_t} \frac{1}{\left\| B^T x_i - B^T x_t \right\|_2^2} \right)^{-1}.$$  \hspace{1cm} (26)

By optimizing $P$ and $W^{(s/t)}$ iteratively, ALDADM can continuously reduce the differences between domains, and make the domain-invariant features more discriminative, while being more robust to the multimodally distributed data. The pseudocode of ALDADM is summarized in Algorithm 1.

III. EXPERIMENTAL RESULTS

A. Datasets

University of Pavia (PU) and Center of Pavia (PC): The two scenes were acquired by the ROSIS-03 hyperspectral sensor over the city of Pavia, Italy. The number of spectral bands in the two images is 103 and 102, respectively. They contain 610 $\times$ 340 pixels and 1096 $\times$ 715 pixels. In the experiment, to maintain dimensional consistency, we select the first 102 bands of the PU. Next, we select seven common classes of the two images, and randomly select 800 samples for each class of the corresponding images to form source and target domains, respectively. For this data, we construct the following two DA tasks: 1) PCPU with PC and PU as source and target domains, respectively. 2) PUPC with PU and PC as source and target domains, respectively. The datasets and selected classes are shown in Fig. 2.

Yancheng: The dataset was acquired by the visible-shortwave infrared advanced hyperspectral imager of the China Gaofen-5 satellite over the port of Yancheng City, China, on April 4, 2019. The image scene contains 1175 $\times$ 585 pixels and 267 spectral bands. Similarly, the Yancheng data are also divided into two disjoint regions in the experiments. We name them as source
TABLE I
NUMBER OF SAMPLES IN THE EXPERIMENTS

| PCPU and PUPC | Yancheng | Botswana |
|---------------|----------|----------|
| Class     | Source | Target | Class     | Source | Target | Class     | Source | Target |
| Asphalt  | 800    | 800    | Off shore water | 365    | 940    | Floodplain grasses1 | 88     | 163    |
| Meadows  | 800    | 800    | Aquaculture   | 76     | 152    | Floodplain grasses2 | 136    | 79     |
| Trees    | 800    | 800    | Paddy        | 427    | 405    | Riparian          | 123    | 146    |
| Bare Soil| 800    | 800    | River        | 137    | 80     | Island interior   | 126    | 77     |
| Bitumen  | 800    | 800    | Fallow land  | 53     | 181    | Acacia woodlands  | 210    | 104    |
| Bricks   | 800    | 800    | Dry land     | 82     | 122    | Acacia shrublands | 94     | 154    |
| Shadow   | 800    | 800    |              |        |        |                    |        |        |

Fig. 2. RGB composite image and ground-truth map for (a) and (b) PU and (c) and (d) PC.

Fig. 3. Yancheng data. (a) RGB composite image. (b) Ground-truth map. The red line divides the image into disjoint source (lower part) and target (upper part) regions.

Fig. 4. Botswana data. (a) RGB composite image. (b) Ground-truth map. The red line divides the image into disjoint source (upper part) and target (lower part) regions.

In this article, we focus on the homogeneous DA problem, where the source and target domains are from the same sensor. In the abovementioned four DA tasks in Table I, the source and target domains of PCPU and PUPC tasks are from different scenes although they are acquired by the same sensor. In the Yancheng and Botswana tasks, the source and target domains are from the same scene but disjoint regions.

B. Comparison Methods

For comparison, the following DA methods are considered.

1) No adaptation (NA): Uses the source classification to directly classify target samples.

2) CORAL [21]: Performs the covariance alignment between domains.

3) SA [19]: Aligns the source and target subspaces.

4) GFK [23]: Learns new feature representations through geodesic flows constructed between domains.

5) TCA [25]: Learns transfer components to minimize marginal distribution difference.

6) JDA [27]: Jointly aligns both the marginal and conditional distributions.

and target domains and pick six common classes, as shown in Fig. 3.

Botswana: The data were acquired by the NASA EO-1 satellite over the Okavango Delta, Botswana. The sensor collects data at 30 m pixel resolution. Through a series of processing, the data are remained 145 bands. The image scene has the size of $256 \times 1476$ pixels. In the experiment, the Botswana image is divided into two disjoint regions with similar land cover, which represent the source and target domains, respectively. For the classification task, we selected six classes, as shown in Fig. 4. The details are given in Table I.
TABLE II
CLASSIFICATION ACCURACIES FOR THE PCPU TASK

| Class | 1  | 2  | 3  | 4  | 5  | 6  | 7  | OA  | κ  |
|-------|----|----|----|----|----|----|----|-----|----|
| NA    | 92.3 | 56.3 | 79.6 | 76.6 | 34.9 | 45.3 | 97.4 | 68.9 | 0.637 |
| SA    | 92.6 | 56.5 | 79.6 | 76.4 | 34.5 | 45.5 | 97.4 | 68.9 | 0.638 |
| CORAL | 76.9 | 58.6 | 81.9 | 72.3 | 11.0 | 81.0 | 92.8 | 66.4 | 0.608 |
| GFK   | 92.8 | 56.3 | 79.6 | 76.4 | 34.9 | 45.9 | 97.4 | 69.0 | 0.639 |
| TCA   | 44.6 | 66.4 | 92.0 | 92.4 | 33.4 | 41.6 | 99.0 | 67.4 | 0.620 |
| JDA   | 88.1 | 63.9 | 92.8 | 47.0 | 89.1 | 40.3 | 98.5 | 74.2 | 0.699 |
| JGSA  | 63.3 | 41.9 | 92.9 | 54.6 | 83.0 | 72.0 | 96.4 | 72.0 | 0.673 |
| ALN-TDFL | 83.8 | 56.9 | 91.6 | 50.1 | 92.1 | 53.4 | 98.6 | 75.2 | 0.711 |
| DTJM  | 33.4 | 71.1 | 90.9 | 51.1 | 74.4 | 31.9 | 99.0 | 64.5 | 0.586 |
| ALDADM | 86.5 | 66.1 | 92.3 | 58.4 | 95.6 | 85.6 | 99.8 | 83.5 | 0.807 |

Values in bold indicate the best accuracy among all methods.

7) JGSA [31]: Learns two coupled projections for the source and target domains to achieve geometrical and statistical alignment simultaneously.

8) ALN-TDFL [29]: Considers local manifold structure and global discriminative consistency in the process of feature learning.

9) DTJM [7]: Aligns features by minimizing the empirical MMD and performs sample reweighting on the embedding matrix while maximizing the dependence between the embedding and labels.

10) ALDADM: Picks density peak landmarks while maintaining global information and local manifold structure and domain alignment both statistically and geometrically.

In the abovementioned comparison algorithms, the 1-NN is the basic classifier. For subspace-based methods, the subspace dimension is fixed as $k = 20$. In the proposed method, we set the percentage of density peaks for domains as $r_s = r_t = 0.95$, and fix the number of iterations as $T = 10$. For the regularization parameters, we use the same parameters $\alpha = 1, \beta = 1$ on Botswana and Yangcheng tasks. For Pavia task, due to the large spectral shift between domains, the subspace difference is large and the parameters are set as $\alpha = 20, \beta = 0.1$. The overall accuracy (OA) and the kappa coefficient ($\kappa$) are used to evaluate the classification performance.

C. Experiments

1) Experiments on PCPU Task: The DA task is more difficult when the source domain is PC and the target domain is PU. As can be seen from Table II, the accuracy of the classical DA method is mostly below 70%, and the highest is only about 75%. However, ALDADM learns initial pseudolabels for target domain by GLP, and then selects samples with high confidence to learn domain invariant features, so that the classification accuracy can reach 83.5%, increasing by at least 8%.

Table III provides confusion matrices for NA, JGSA, and ALDADM. In particular, the classification accuracy of the Class 7 (Shadow) achieves good results on all DA methods, almost all above 95%. NA is obviously very poor in Classes 2 (Meadows), 5 (Bitumen), and 6 (Bricks), which misclassifies 40% of the samples in Class 2 into Class 4 and most of the samples in Classes 5 and 6 are misclassified into Class 1. JGSA does a poor job of classifying Classes 2 and 4, they confuse Classes 2 and 4 with less than 60% classification accuracy. ALDADM improves Class 2 to 66%, Class 5 to 95%, and Classes 6 to 85%.

2) Experiments on PUPC Task: Table IV shows the classification performance of several DA methods on the PUPC task. Obviously, the classification accuracy of the general DA method in Class 2 into Class 4 and most of the samples in Classes 5 and 6 are misclassified into Class 1. JGSA does a poor job of classifying Classes 2 and 4, they confuse Classes 2 and 4 with less than 60% classification accuracy. ALDADM improves Class 2 to 66%, Class 5 to 95%, and Class 6 to 85%.
is below 80%, while the accuracy of JGSA and ALN-TDFL can reach about 82% due to the preservation of the global discriminative information. ALDADM performs DPLs, and then utilizes the advantages of global discriminative consistency and local manifold structure simultaneously, making the extracted features more discriminative and robust to multimodally distributed data, and the classification effect is also significantly improved, approaching 88%.

The classification confusion matrix about NA, JGSA, ALDADM is shown in Table V. The NA method performs poorly on Class 6 (Bricks), which misclassifies many samples into Class 1 (Asphalt) because their spectral characteristics are very close. When using JGSA, its classification accuracy on Class 6 is improved by about 40%. Compared with NA and JGSA, our method has been improved on Class 2 and 5–7, and the accuracy is as high as 98% on Classes 5 and 6.

3) Experiments on Yancheng Task: As the source and target domains of Yancheng are adjacent regions, this task is relatively simple. It is obvious from Table VI that the OA of all DA methods can reach over 90%. The proposed ALDADM has the best classification effect, which can reach 99%. However, TCA, JDA, and JGSA completely misclassify Class 6 (Dry land), and basically divided them into Class 5 (Fallow land), as shown in Table VII. In fact, TCA and JDA only consider that the distribution of the two domains is closer, while JGSA further considers global discriminative information preservation. Even so, it is difficult to distinguish between Classes 5 and 6, which are both about land. However, ALDADM also considers the local manifold structure, so as to distinguish these two classes and increase the classification accuracy from 0% to 100%.

4) Experiments on Botswana Task: Table VIII shows the classification effects of different DA algorithms on the Botswana task. Without DA strategy, the accuracy can reach 85% after the data is filtered. If the classical DA algorithm is used, the
Fig. 5. Botswana task. (a) NA (84.8%). (b) SA (85.8%). (c) CORAL (85.3%). (d) GFK (87.4%). (e) TCA (85.2%). (f) JDA (86.2%). (g) JGSA (92.3%). (h) ALN-TDFL (89.8%). (i) DTJM (90.9%). (j) ALDADM (100%).

TABLE VIII
CLASSIFICATION ACCURACIES FOR BOTSWANA TASK

| Class   | 1   | 2   | 3   | 4   | 5   | 6   | OA   | \(\kappa\) |
|---------|-----|-----|-----|-----|-----|-----|------|----------|
| NA      | 93.3 | 100 | 62.3 | 93.5 | 94.2 | 78.6 | 84.8 | 0.815    |
| SA      | 94.5 | 100 | 61.6 | 77.9 | 88.5 | 94.2 | 85.8 | 0.827    |
| CORAL   | 60.1 | 100 | 96.6 | 100 | 86.5 | 85.7 | 85.3 | 0.824    |
| GFK     | 94.5 | 100 | 72.6 | 77.9 | 85.6 | 93.5 | 87.4 | 0.847    |
| TCA     | 59.5 | 100 | 96.6 | 100 | 87.5 | 85.1 | 85.2 | 0.822    |
| JDA     | 62.6 | 100 | 90.4 | 100 | 92.3 | 89.0 | 86.2 | 0.834    |
| JGSA    | 66.9 | 100 | 99.3 | 100 | 99.0 | 100 | 92.3 | 0.906    |
| ALN-TDFL| 63.8 | 100 | 92.5 | 100 | 96.2 | 100 | 89.8 | 0.876    |
| DTJM    | 62.0 | 100 | 100 | 100 | 100 | 100 | 90.9 | 0.890    |
| ALDADM  | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 1.000    |

Values in bold indicate the best accuracy among all methods.

classification accuracy can be increased by 1%–2%. Based on JDA, ALN-TDFL considers the local manifold structure, and its classification accuracy increases by 5%. JGSA can achieve 92% considering the discriminant information preservation and two coupled mappings. Finally, ALDADM learns soft labels through GLP method and selects high confidence samples. By using two mappings, we simultaneously consider local manifold structure and reduce the distribution shift between projected data of two domains. A perfect result is obtained, that is, 100%. The classification map on the target domain of different methods are shown in Fig. 5, where the target region is rotated 90° for displaying. As shown in the central region, many comparison methods misclassify the Class “Floodplain grasses1” in red color to the Class “Floodplain grasses2” in green color because these two classes all belong to the grass and are difficult to be distinguished due to their high spectral similarity. Notwithstanding, our ALDADM can discriminate the subtle spectral differences between them and produce perfect result.

Fig. 6. Landmark visualization on the PCPU task. The source and target domains are shown in the left and right subfigures, where the blue and red colors denote the class “trees” and “bare soil,” respectively, and the solid dots are the landmarks picked by the DPLs.

TABLE IX
ABLATION STUDY OF ALDADM FOR PCPU TASK

|         | JGSA | DPLs | Local DFL | GLP | OA   |
|---------|------|------|-----------|-----|------|
| 1       | ✓    |      |           |     | 72.29|
| 2       | ✓    | ✓    |           |     | 75.91|
| 3       | ✓    |      | ✓         |     | 78.50|
| 4       | ✓    |      |           | ✓   | 81.43|
| 5       | ✓    | ✓    | ✓         |     | 79.84|
| 6       | ✓    | ✓    |           | ✓   | 82.32|
| 7       | ✓    |      | ✓         | ✓   | 83.34|
| 8       | ✓    |      | ✓         | ✓   | 84.66|

D. Ablation Analysis

In order to analyze the effect of each module in ALDADM, we conduct ablation experiments on the PCPU task. We randomly select 800 samples from 7 classes to form the source and target domains, and conduct a total of eight experiments on this task, as shown in Table IX. We regard JGSA as the baseline. Experiment 2 means adding DPLs on the basis of JGSA to test the effect of DPLs. In experiment 3, the source discriminative information preservation of JGSA algorithm is replaced with Local DFL. Experiment 4 represents the pseudolabel learning method from 1-NN to GLP strategy [35].
Comparing experiments 2, 3, and 4 with experiment 1, the addition of each part is effective under the framework of JGSA. When performing DPLs, it can filter out some misclassified target samples, thereby reducing the interference of the subsequent iterative learning process and improving classification accuracy. When the Local DFL is added, the OA is improved by 6% compared to the original JGSA. Then, when using GLP method for pseudolabel learning, it improves the classification performance by 9% over the original JGSA. It means that more target samples get the correct labels.

Experiments 5–7 can show that the effect of adding any two parts together is better than that of a single part, for example, the accuracy of experiment 5 is higher than that of experiments 2 and 3. Because it not only reduces the interference of some misclassified samples, but also considers global information preservation and local manifold structure.

Experiment 8 is ALDADM, which learns the soft labels for the target domain through the GLP strategy, and then uses DPLs to reduce the interference from the outliers in $D_s$ and the mispredicted instances in $D_t$, and reduces domain discrepancy both geometrically and statistically. Therefore, its performance is the best.

E. Landmark Visualization

DPLs are applied to both domains, and the instances with higher intraclass and lower interclass densities are selected. To verify that DPLs in two domains can filter out interfering samples, we selected two classes (i.e., “trees” and “bare soil”) of samples for visualization on the PCPU task to give an intuitive display in Fig. 6. It can be seen from the figure that the DPLs can select landmarks with high density.

F. Feature Visualization

The results of feature visualization for original data, JGSA, and our ALDADM are illustrated in Fig. 7, where four classes (i.e., “Asphalt,” “Meadows,” “Bitumen,” “Bricks”) are selected from the PCPU task and tSNE is used to display features. In the original features, samples from the same class are obviously clustered into several clusters. Taking the yellow “Meadows” class as an example, it has at least three clusters in either source or target domain. This almost coincides with the original sample distribution in the groundtruth map, where the “Meadows” class has distributed mainly in three disjoint regions in PU and many regions in PC. JGSA forces the samples of the same class to be closer to a certain extent. As can be seen in Fig. 7(c), our ALDADM largely solves the multimodally distributed data problem, that is, it makes the samples from the same class gather more closely, and also preserves the local manifold structure of the data.

After all, JGSA only pays attention to the global relationship of the data, but the similarity weight matrix constructed in
the local DFL can relax the global discriminative consistency and capture the local relationship between samples, since only the similar points from the same class are required to drawn closer. Therefore, ALDADM can make the extracted features more discriminative and deal with multimodally distributed data simultaneously.

G. Parameter Analysis

In this section, we investigate the effect of parameters $\alpha$ and $\beta$ on the PCPU task. The OA versus parameters $\alpha$ and $\beta$ is shown in Fig. 8, where the model shows a relative better results over a wide range of parameters. Based on these results, $\alpha = 20$ and $\beta = 0.1$ is chosen for PCPU and PUPC tasks. Due to the existing of large spectral shifts between domains in the PCPU or PUPC task, a relatively larger $\alpha$ and smaller $\beta$ force to decrease the subspace difference at the cost of loosing the requirement on local discriminative preservation as shown in (18).

IV. CONCLUSION

In this article, we propose the ALDADM method for DA in HSIs classification. Its idea is to first perform DPLs on the labeled source samples and the pseudolabeled target samples, thereby reducing the interference of outliers in $D_1$ and mispredicted instances in $D_2$. Then, it learns two mappings to make the mapped subspaces of the source and target domains closer, while dealing with the challenge of multimodally distributed data by respecting the local manifold structure. Extensive experiments on four tasks demonstrate the effectiveness and robustness of ALDADM for cross-domain classification.

In this article, we only focus on the homogeneous DA problem, where the source and target domains are from the same sensor. Notwithstanding, there are still large spectral differences between two domains. The changed soil properties, environment conditions, and the incident angle of sunlight can cause some differences in the spectrum of similar materials in different geographical regions or collection times. In addition, some materials have very similar spectral characteristics, such as Asphalt and Bitumen in the PUPC task, and it is easy to confuse them. Therefore, applying DA methods to HSIs has great difficulty due to spectral differences or distribution shifts. When data comes from different sensors, knowledge transfer and classification is more difficult, and it is prone to bad results. In the future, we will try to extend the homogeneous DA method to multisource DA or heterogeneous DA scenarios.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba for providing the University of Pavia and Center of Pavia datasets.

REFERENCES

[1] 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.

[2] 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.

[3] L. Zhou et al., “Subspace structure regularized nonnegative matrix factorization for hyperspectral unmixing,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 13, pp. 4257–4270, Jul. 2020.

[4] 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.

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

[6] 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, Nov. 2016.

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

[9] Z. Deng, K.-S. Choi, Y. Jiang, and S. Wang, “Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods,” IEEE Trans. Cybern., vol. 44, no. 12, pp. 2585–2599, Dec. 2014.

[10] Z. Cui, W. Li, D. Xu, S. Shan, X. Chen, and X. Li, “Flowing on riemannian manifold: Domain adaptation by shifting covariance,” IEEE Trans. Cybern., vol. 44, no. 12, pp. 2264–2273, Dec. 2014.

[11] L. Zhou, X. Bai, X. Liu, J. Zhou, and E. R. Hancock, “Learning binary code for fast nearest subspace search,” Pattern Recognit., vol. 98, 2020, Art. no. 107040.

[12] V. M. Patel, R. Gopalakrishnan, R. Li, and R. Chellappa, “Visual domain adaptation: A survey of recent advances,” IEEE Signal Process. Mag., vol. 32, no. 3, pp. 53–69, May 2015.

[13] J. Liu, V. Behbood, P. Hao, H. Zuo, S. Xue, and G. Zhang, “Transfer learning using computational intelligence: A survey,” Knowl. Based Syst., vol. 80, pp. 14–23, May 2015.

[14] 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.

[15] D. Tuia, D. Marcos, and G. Camps-Valls, “Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization,” ISPRS J. Photogrammetry Remote Sens., vol. 120, pp. 1–12, Oct. 2016.

[16] L. Sun, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” IEEE Trans. Neural Netw. Learn. Syst., vol. 26, no. 5, pp. 1019–1034, May 2015.

[17] H. Liu, M. Shao, Z. Ding, and Y. Fu, “Structure-preserved unsupervised domain adaptation,” IEEE Trans. Knowl. Data Eng., vol. 31, no. 4, pp. 799–812, Apr. 2019.

[18] L. Li, Z. Wan, and H. He, “Dual alignment for partial domain adaptation,” IEEE Trans. Cybern., vol. 51, no. 7, pp. 3404–3416, Jul. 2021.

[19] 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.

[20] 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.

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

[22] 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. Ear. Observ. Remote Sens., vol. 14, pp. 3170–3184, 2021.

[23] 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. 2006–2013.

[24] 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.
Lin Sun received the B.E. degree in automation from China Jiliang University, Zhejiang, China, in 2020. She is currently working toward the M.E. degree in optical engineering with Yan Shan University, Hebei, China. Her research interest is 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.

Wewei 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 studied with the Department of Applied Mathematics, University of Maryland College Park, as a Visiting Scholar with the famous professor John Benedetto to study on the dimensionality reduction of Hyperspectral Image. From 2014 to 2016, he studied with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, as a Postdoc 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. He is currently a Full Professor with Ningbo University, Zhejiang Province, China. He has authored or coauthored more than 70 journal papers and his current research interests include hyperspectral image processing with manifold learning, anomaly detection and target recognition of remote sensing imagery using compressive sensing.

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. His 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 Shackoul 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 a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). 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. He was 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 currently is a member of IEEE Periodicals Review and Advisory Committee.