Affinity Matrix Learning Via Nonnegative Matrix Factorization for Hyperspectral Imagery Clustering

Yao Qin [ORCID], *Member, IEEE*, Biao Li, Weiping Ni, Sinong Quan [ORCID], Peizhong Wang, and Hui Bian

Abstract—In this article, we integrate the spatial-spectral information of hyperspectral image (HSI) samples into nonnegative matrix factorization (NMF) for affinity matrix learning to address the issue of HSI clustering. This technique consists of three main components: 1) oversegmentation for computing the spectral-spatial affinity matrix; 2) NMF with the guidance of the obtained affinity matrix; and 3) density-based spectral clustering on the final affinity matrix. First, the HSI is oversegmented into superpixels via the entropy rate superpixel algorithm. The spectral-spatial affinity matrix is defined based on the class-consistency assumption of all the HSI samples in each superpixel and the similar HSI samples between adjacent superpixels. Second, to integrate the spectral-spatial information into NMF, the obtained affinity matrix is used to guide the iterative process of NMF. The spectral-spatial affinity matrix is then weighted by the affinity matrix in the obtained low-dimensional subspace to form the final affinity matrix. Third, density-based spectral clustering is applied to the final affinity matrix to obtain clustering maps. Experimental results on three public benchmark HSIs demonstrate that the proposed method is superior to the considered state-of-the-art baseline methods on both the computational cost and clustering accuracy.

Index Terms—Affinity matrix, hyperspectral image (HSI), nonnegative matrix factorization (NMF), oversegmentation, remote sensing, spectral clustering.

I. Introduction

SINCE contiguous spectral information on a wide range of the electromagnetic spectrum can be captured by hyperspectral sensors [1]–[6], hyperspectral images (HSIs) have been employed in various remote sensing applications, such as mineral exploration [7], land-cover classification [8], etc. As the fundamental technique in these applications, HSI classification is usually addressed in the framework of supervised learning, such as support vector machines [9] and deep learning techniques [10]. However, supervised learning always requires labeled samples for training the model. Therefore, when only unlabeled samples

are available, the issue turns to be unsupervised classification, namely clustering. Ideally, clustering can assign similar samples to a group that corresponds to one specific class. However, clustering is always a difficult task for both the machine learning and remote sensing communities because of the nonconvex structure and the intraclass divergence of samples. Furthermore, the large spectral variability caused by multipath scattering, nonhomogeneous mixing of pixel-level spectra, and variations in sun-canopy-sensor geometry still make the HSI clustering a challenging task [11].

Generally, clustering methods proposed by the machine learning community can be divided into the following five categories based on their working mechanisms [12]:
1) *centroid-based*;
2) *density-based*;
3) *biological-based*;
4) *graph-based*; and
5) *deep learning-based* methods.

Conventional methods of centroid and density-based clustering include k-means [13], fuzzy c-means (FCM) [14], k-means++ algorithm [15], fuzzy c-means with spatial constraint (FCM_S) algorithm [16], and fast search and find of density peaks (FSDP) algorithm [17]. Inspired by these typical clustering methods, several HSI clustering methods have been proposed, such as Markov random fields-based clustering algorithm [18], the spectral-spatial diffusion learning (DLSS) method that combines a geometric estimation of class modes with a diffusion-inspired labeling [19]. The biological-based clustering algorithms try to employ the biological models to achieve HSI clustering, such as automatic fuzzy clustering method based on adaptive multiobjective differential evolution [20] and adaptive multiobjective memetic fuzzy clustering algorithm [21]. The graph-based clustering methods are highly popular algorithms, which work by embedding the data in the eigenspaces of the Laplacian matrices and applying k-means on this representation to obtain the clusters [12]. Based on the sparse subspace clustering (SSC) method [22], spatial weighted SSC (SwSSC), SSC incorporating spatial information (SSC_S), spatial sparse subspace clustering (S^4SSC) [23], l_2-norm regularized SSC (L2SSC) [24], and adaptive spatial-spectral neighborhood selection [25] have been proposed for HSI clustering. Recently, a novel total variation (TV) regularized collaborative representation clustering with a locally adaptive dictionary algorithm for HSI has been proposed in [26]. In addition to these methods, deep clustering methods have been recently proposed and the deep embedding of samples is usually achieved via

autoencoder models. Typical deep clustering methods include the deep embedded clustering [27], the information maximizing self-augmented training (IMSAT) techniques [28], deep *k*-means clustering [29], deep self-evolution clustering [30], deep subspace clustering [31], and deep clustering with variational autoencoder [32]. As a pioneer work, the IMSAT method combines fully connected network and regularized information maximization (RIM) that optimizes an intuitive information-theoretic objective function which balances class separation, class balance, and classifier complexity [33]. Please refer to [34] for detailed introduction of deep clustering methods. Inspired by the success of deep clustering, the pioneer work of deep HSI clustering method, namely LSSD, is proposed in [1] and the discriminative deep embedding of HSI samples are learned by training a siamese network based on the (dis)similar sample pairs derived from the set-to-set and sample-to-sample distances. The unsupervised broad learning is first applied for HSI clustering by using a graph regularized sparse autoencoder [35].

Additionally, in order to address the issue of large HSI clustering, several methods based on anchor graph, affinity learning and matrix factorization (MF) are proposed to achieve less computational cost. The fast spectral clustering with anchor graph method first constructs the anchor graph by incorporating the spatial information of HSI and then spectral analysis is efficiently achieved [36]. The scalable graph-based clustering with nonnegative relaxation employs the orthonormal constraint with nonnegative relaxation to build a novel graph-based clustering model and the clustering indicators are directly obtained [37]. The mutual nearest neighbor information is used to merge the segmented results obtained by *k*-means method in [38] and the clustering stages make use of multiple subspaces obtained by deploying a local band selection approach. The heuristic methods, including genetic algorithm, harmony search, and particle swarm optimization are used to explore the effects of the heat kernel parameters aiming to analyze the embedding quality of the graph affinity matrix [39]. The joint spatial-spectral information is effectively incorporated into the proximity graph matrix of which the block diagonal structure is amplified by the "conductivity method" to improve the clustering performance [40]. The optimization of affinity propagation method is introduced to take into account the presence of identical objects in the samples to be partitioned and adapts the preference parameter to each object for HSI clustering [41]. Inspired by the success of MF in hyperspectral unmixing, a robust manifold method consisting of two MF components (RMMF) for HSI clustering is proposed in [42] and the clustering indicators can be directly obtained via the second MF component. The Sketch SSC method is applied for large HSI clustering by extending integrating it with a spatial prior in [43].

The success of the abovementioned methods in HSI clustering can be mainly concluded as the following four aspects.

1) The nonconvex structures of HSI samples are captured by the graph-based method, such as the anchor graph-based methods and SSC-based methods.
2) The spatial information is well exploited in various ways, such as the spatial constraint in FCM_S, TV constraint and the segmented superpixels used in LSSD.

3) With the aid of distance measure of HSI samples derived from spectral-spatial information, the obtained deep discriminative features in LSSD can be easily used to achieve state-of-the-art clustering result.
4) The MF has shown its essential advantage in subspace learning and dimension reduction, thus suitable for HSI clustering if it is well employed.

However, there are several shortcomings of the abovementioned methods that have hinder their clustering performance, such as the large computational costs of the LSSD method and SSC-based methods, the inaccurate similarity based on original spectral features in the RMMF method.

To simultaneously reduce the computational cost and achieve accurate affinity for HSI clustering, we propose a framework to integrate the spectral-spatial information into the nonnegative matrix factorization (NMF) for affinity matrix learning (NMFAML). The merits of the proposed NMFAML method are as follows. First, compared with LSSD and SSC-based methods which involve corresponding iterations of training of the Siamese network and computation of coefficient matrix $\mathbf{C}$, the iterative multiplication operation of the NMF algorithm requires less computational cost. Second, similar to the LSSD method, the proposed NMFAML method can be summarized as affinity measure learning of HSI samples which takes advantage of the merits of both spaital-spectral information and NMF, whereas the original spectral similarity in the RMMF method fails to discover the true affinity between HSI samples.

Specifically, the proposed technique consists of three main components:

1) HSI oversegmentation for the generation of spectral-spatial affinity matrix;
2) generation of NMF-based affinity matrix with the guidance of spectral-spatial affinity matrix; and
3) density-based spectral clustering on the weighted matrix of both spectral-spatial and NMF-based affinity matrix.

To be more specific, oversegmentation is first achieved by applying the entropy rate superpixel (ERS) algorithm and the superpixel number is estimated by detecting the continuous edges of a binary edge map derived from the first principal component of the HSI. Since the superpixel is oversegmented, it is reasonable to regard the samples of each superpixel as belonging to the same class. Meanwhile, it is feasible to believe that the corresponding subsets of two neighboring superpixels with largest similarity may belong to the same class. From this viewpoint, the spectral-spatial affinity matrix is generated in a structured and sparse way. Second, the spectral-spatial affinity matrix is employ as a general graph regularization in the iterative process of NMF. By preserving the graph structure of the spatial-spectral information, the proposed NMF algorithm can have more discriminating power than the typical NMF algorithm [44] and then the related affinity matrix can be easily obtained. Third, the spectral-spatial and NMF-based affinity matrices are weighted and then fed into the density-based spectral clustering algorithm. To summarize, the contribution of the proposed NMFAML method are twofold. First, a novel method of learning accurate affinity matrix of HSI samples is proposed by integrating the spatial-spectral information derived from HSI samples in each
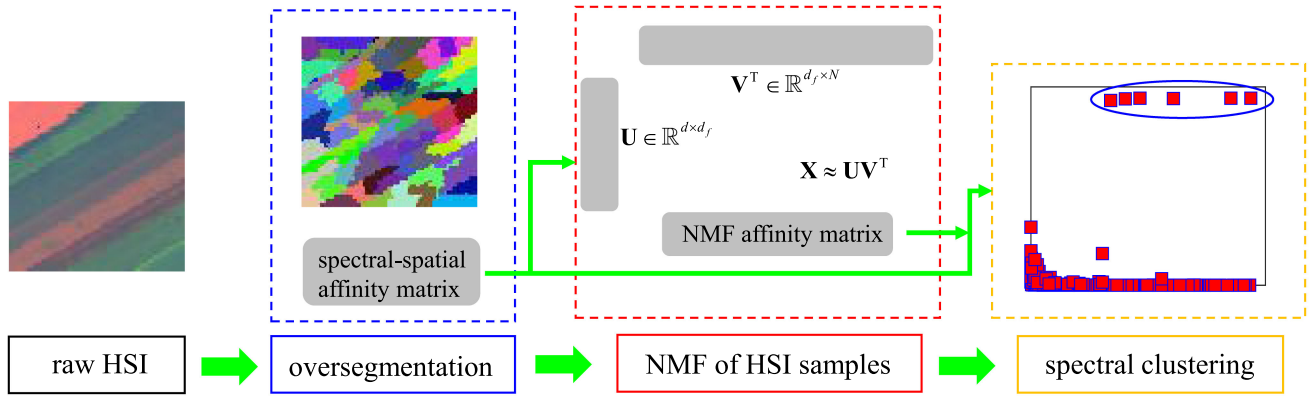
Fig. 1. Illustration of the proposed NMFAML method for HSI clustering. (Best viewed in color).

superpixel and similar HSI samples between spatially adjacent superpixles with the NMF algorithm. Second, comprehensive experiments on three benchmark HSIs have been conducted to demonstrate the superiority of the NMFAML method in terms of computational cost and clustering accuracy.

The rest of the article is organized as follows. The proposed methodology of HSI clustering is presented in Section II and the convergence analysis is given in Section III. Section IV describes the experimental setup and results. Section V summarizes the contributions of our research.

II. PROPOSED METHOD

As shown in Fig. 1, the proposed NMFAML algorithm consists of the following three main steps:

1) HSI oversegmentation;
2) NMF of HSI samples; and
3) density-based spectral clustering.

Specifically, oversegmentation is first employed to generate the spectral-spatial affinity matrix and it can be divided into two components, i.e., estimating number of superpixels and oversegmentation via the ERS algorithm. Second, the obtained spectral-spatial affinity matrix is integrated into the iterative process of NMF algorithm and the NMF-based affinity matrix is then constructed. Third, the spectral-spatial and NMF-based affinity matrices are weighted to form the final affinity matrix to achieve the clustering result via the subsequent FSDP-based spectral clustering. In the following, conventional NMF and FSDP algorithms, the implementation details of the proposed NMFAML method are subsequently described.

A. Conventional NMF Algorithm

For simplicity, the HSI samples are denoted as $\mathbf{X} \in \mathbb{R}^{d \times N}$, where d and N are the dimension and total number of HSI samples, respectively. Assume that $\mathbf{X}$ can be factorized as $\mathbf{UV}^{\mathrm{T}}$, where $\mathbf{U} \in \mathbb{R}^{d \times d_f}$, $\mathbf{V} \in \mathbb{R}^{N \times d_f}$, and d_f is the factorization dimension. Then the NMF loss of $\mathbf{X}$ is written as $||\mathbf{X} - \mathbf{UV}^{\mathrm{T}}||_{\mathrm{F}}^{2}$, where matrices $\mathbf{U} \ 0$ and $\mathbf{V} \ 0$. The corresponding objective function is defined as

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{X} - \mathbf{UV}^{\mathrm{T}}||_{\mathrm{F}}^{2}, \mathbf{U} \ 0, \mathbf{V} \ 0. \tag{1}$$

B. FSDP Algorithm

Given the samples $\mathbf{X} \in \mathbb{R}^{d \times N}$, d_{ij} denotes the Euclidean distance between samples x_i and x_j. For each sample x_i, two quantities are defined, i.e., its local density ρ_i and its distance δ_i from samples of higher density [17]. Specifically, the local density ρ_i of sample x_i is defined as

$$\rho_i = \sum_j \chi(d_{ij} - d_c) \tag{2}$$

where $\chi(y) = 1$ if $y < 0$ and $\chi(y) = 0$ otherwise, d_c is a cutoff distance. The quantity δ_i is computed as the minimum distance from the sample x_i to any other sample with higher density

$$\delta_i = \min_{j:\rho_j > \rho_i} (d_{ij}). \tag{3}$$

For the sample with the highest density, δ_i is defined as $\max_j(d_{ij})$. Since δ_i is much larger than the conventional nearest neighbor distance only for samples that are local or global maximum in the density, the samples with large values of δ_i can be recognized as cluster centers. Once the cluster centers are found, all other samples are assigned to the corresponding clusters as their nearest neighbor of higher density.

C. Oversegmentation

To achieve the oversegmentation of HSI, the principal component analysis (PCA) is first used to decompose the HSI and the 2-D map of first principal component is obtained. Then the 2-D binary edge map is generated by applying the canny detector to the HSI map and the number of continuous edges estimated through the 8-connected component labeling algorithm[1] is considered as the superpixel number N_s. Given the superpixel number N_s and the 2-D map of first three principal components $\mathbf{M}_{3\,pc}$, the oversegmentation is achieved by employing the ERS method[2] [45]. Specifically, the ERS method maps the image to a graph $\mathbf{G} = (\mathbf{P}, \mathbf{E})$, where the vertexes $\mathbf{P}$ and the edge weights $\mathbf{E}$ denote the samples in $\mathbf{M}_{3\,pc}$ and the pairwise similarities

[1][Online]. Available: https://ww2.mathworks.cn/help/images/ref/bwconncomp.html
[2][Online]. Available: https://github.com/mingyuliutw/EntropyRateSuperpixel

between samples, respectively. The objective function includes the entropy rate of the random walk on $\mathbf{G}$ and a balancing term encouraging clusters with similar sizes, reducing the number of unbalanced superpixels. When the optimization process of the ERS method finishes, N_s connected subgraphs corresponding to homogeneous superpixels are generated.

As shown in Fig. 1, since the affinity matrix is used as the guidance of NMF, it is expected to generate accurate affinity matrix. Indeed, with the aid of affinity matrix, the low-dimensional samples obtained by NMF are expected to include the spectral-spatial information, facilitating achieve better final clustering performance. Specifically, the spectral-spatial information is expected to be integrated into the factorized samples as the following objective function:

$$\min_{\mathbf{V}} ||\mathbf{V}^{\mathrm{T}} - \mathbf{V}^{\mathrm{T}}\mathbf{Z}_{sp2}||_{\mathrm{F}}^2 \tag{4}$$

where $\mathbf{Z}_{sp2} \in \mathbb{R}^{N \times N}$ is the affinity matrix related to spectral-spatial information.

In order to generate accurate affinity, we define the matrix $\mathbf{Z}_{sp2}$ based on the following two assumptions of superpixels.

1) Since the superpixels are obtained through oversegmentation, the samples in each superpixel are probable to belong to the same class, which indicates that each sample in the superpixel can be linearly represented by other samples in the same superpixel, i.e.,

$$x_i = \beta_{1i}x_1 + \beta_{2i}x_2 + \cdots + \beta_{N_k i}x_{N_k} \tag{5}$$

where N_k is number of samples in the kth superpixel and $x_i \ (i = 1, \cdots N_k)$ denotes the samples in the superpixel. The corresponding affinity matrix $\mathbf{Z}^{intra}$ can be constructed as

$$\mathbf{Z}_{ji}^{intra} = \begin{cases} \beta_{ji}, & \text{if } x_i \text{ and } x_j \text{ belong to the } k\text{th superpixel} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

For simplicity, samples $x_j \ (j = 1, \ldots, N_k)$ are regarded to contribute equally to the representation of sample x_i. Therefore, the value of β_{ji} herein equals to $\frac{1}{N_k}$.

2) It is obvious that the matrix $\mathbf{Z}^{intra}$ is formed as a block-diagonal structure of the affinities within superpixels, which lacks the affinities between spatially adjacent superpixels. Clearly, if the block-diagonal structure of $\mathbf{Z}^{intra}$ is directly applied into the constraint [see (4)], the matrix $\mathbf{V}$ is probable only with similar affinities within superpixels, instead of the affinities within classes. Therefore, it is essential to take the affinities of samples belonging to different superpixels into consideration to fuse the affinities within superpixels. Indeed, if the samples belonging to the spatially adjacent superpixels have the smallest distances, it is feasible to believe that they belong to the same class. Therefore, the affinities between samples of spatially adjacent superpixels are exploited and the corresponding affinity matrix is denoted as $\mathbf{Z}^{inter}$ for simplicity. Fig. 2 gives an illustration of spatially adjacent superpixels. Specifically, we assume that the kth superpixel $\mathbf{X}_k$ is the centering superpixel and the lth superpixel $\mathbf{X}_l$ is adjacent with $\mathbf{X}_k$. Then the subset of $\mathbf{X}_k$ composing of N_1 samples having the overall smallest distance to $\mathbf{X}_l$ is extracted and each sample x_i^k in the subset is supposed to be represented by the samples $x_j^l \ (j = 1, \ldots, N_2)$

Fig. 2.   Illustration of the spatially adjacent superpixels, where superpixels 1–5 are neighboring with superpixel 0. (Best viewed in color).

Algorithm 1: $\mathbf{Z}^{inter}$ Construction.

Input: HSI superpixels, N_1, N_2, k=1.
Output: $\mathbf{Z}^{inter}$.
1: **Repeat**:
2:    Select the kth superpixel $\mathbf{X}_k$ and l=1.
3:    **Repeat**:
4:       Select the lth adjacent superpixel of $\mathbf{X}_k$ and denote it as $\mathbf{X}_k^l$.
5:       Compute the pairwise distance between $\mathbf{X}_k$ to $\mathbf{X}_k^l$.
6:       Extract the N_1 samples of $\mathbf{X}_k$ having overall the smallest distance to $\mathbf{X}_k^l$.
7:       For all extracted samples x_i, compute the nonnegative affinity coefficients of the N_2 samples in $\mathbf{X}_k^l$ having smallest distances to x_i.
8:       Fill the affinity coefficients into $\mathbf{Z}^{inter}$.
9:       l=$l + 1$.
10:    **Until** all adjacent superpixels of $\mathbf{X}_k$ are selected.
11:    k=$k + 1$.
12: **Until** all superpixels are selected.

---

with the smallest distances in $\mathbf{X}_l$. For better illustration, we denote the assumption as *intersuperpixel affinity*. Different from the assumption of equal contribution of all samples belonging to the same superpixel in the definition of matrix $\mathbf{Z}_{sp2}$, the affinity coefficients are solved as a nonnegative linear least-squares problem as follows:

$$x_i^k = \alpha_{1i}x_1^l + \alpha_{2i}x_2^l + \cdots + \alpha_{N_2 i}x_{N_2}^l, \alpha_{ij} > 0. \tag{7}$$

Therefore, all the affinity coefficients finally form the matrix $\mathbf{Z}^{inter}$, which is defined as

$$\mathbf{Z}_{ji}^{inter} = \begin{cases} \alpha_{ji}, & \text{if } x_i \text{ and } x_j \text{ have intersuperpixel affinity} \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

Algorithm 1 gives the implementation details of matrix $\mathbf{Z}^{inter}$ construction.

With the two class-consistency assumptions of HSI samples within and between superpixels, each sample can be represented by the samples of the same superpixel and samples from adjacent superpixels with the smallest distances. Therefore, in order to better exploit both affinities, the spectral-spatial affinity matrix

is formulated as the weight of $\mathbf{Z}^{intra}$ and $\mathbf{Z}^{inter}$, i.e.,

$$\mathbf{Z}_{sp2} = \lambda_0 \mathbf{Z}^{tra} + (1 - \lambda_0)\mathbf{Z}^{ter} \qquad (9)$$

where $\lambda_0 \in [0, 1]$ is tradeoff coefficient. Note that the rationale of the definition of $\mathbf{Z}_{sp2}$ lies in the representation formulations of (5) and (7). In this way, the affinities of both intrasuperpixels and intersuperpixels are well combined.

D. NMF of HSI Samples

Prior to NMF of HSI samples, the joint spectral-spatial features of HSI are extracted by using the first four principal components. Specifically, the first four components in the local region of 3×3 pixels surrounding each sample are flattened and appended to the raw HSI spectral features to form the joint spectral-spatial features. For better illustration, when the dimension of the raw spectral features is denoted as d, the dimension of the spectral-spatial feature is $d + 36$.

Integrating the constraint of matrix $\mathbf{Z}_{sp2}$ into NMF gives the final optimization problem as follows:

$$\min_{\mathbf{U},\mathbf{V}} ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_{\mathrm{F}}^2 + \lambda_1||\mathbf{V}^{\mathrm{T}} - \mathbf{V}^{\mathrm{T}}\mathbf{Z}_{sp2}||_{\mathrm{F}}^2, \mathbf{U}\ 0, \mathbf{V}\ 0 \quad (10)$$

where λ_1 is a nonnegative tradeoff coefficient for controlling the importance of spectral-spatial affinity. Then two Largrange multipliers are introduced and the final function is defined as

$$\mathcal{L} = \frac{1}{2}||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_{\mathrm{F}}^2 + \frac{\lambda_1}{2}||\mathbf{V}^{\mathrm{T}} - \mathbf{V}^{\mathrm{T}}\mathbf{Z}_{sp2}||_{\mathrm{F}}^2$$
$$+ \mathrm{Tr}(\Phi\mathbf{U}^{\mathrm{T}}) + \mathbf{Tr}(\Psi\mathbf{V}^{\mathrm{T}}). \qquad (11)$$

1) Updating $\mathbf{U}$*:* The partial derivatives of $\mathcal{L}$ with respect to $\mathbf{U}$ is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = -\mathbf{X}\mathbf{V} + \mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V} + \Phi. \qquad (12)$$

Using the KKT conditions $\Phi_{ik}u_{ik}{=}0$, the following are obtained:

$$-(\mathbf{X}\mathbf{V})_{ik}u_{ik} + (\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V})_{ik}u_{ik} = 0. \qquad (13)$$

These equations lead to the following updating rules:

$$u_{ik} \leftarrow \frac{(\mathbf{X}\mathbf{V})_{ik}}{(\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V})_{ik}}u_{ik}. \qquad (14)$$

2) Updating $\mathbf{V}$*:* The partial derivatives of $\mathcal{L}$ with respect to $\mathbf{V}$ is

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}} = -\mathbf{X}^{\mathrm{T}}\mathbf{U} + \mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U} + \Psi + \lambda_1(\mathbf{I} - \mathbf{Z}_{sp2})(\mathbf{I} - \mathbf{Z}_{sp2}^{\mathrm{T}})\mathbf{V}. \qquad (15)$$

By using the KKT conditions $\Psi_{lk}v_{lk}{=}0$, the following are obtained:

$$(-\mathbf{X}^{\mathrm{T}}\mathbf{U} - \lambda_1\mathbf{M}_1\mathbf{V})_{jk}v_{jk} + (\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U} + \lambda_1\mathbf{M}_2\mathbf{V})_{jk}v_{jk} = 0 \qquad (16)$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ denote $\mathbf{Z}_{sp2} + \mathbf{Z}_{sp2}^{\mathrm{T}}$ and $\mathbf{I} + \mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}}$, respectively. These equations lead to the following updating rules:

$$v_{jk} \leftarrow \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{U} + \lambda_1\mathbf{M}_1\mathbf{V})_{jk}}{(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U} + \lambda_1\mathbf{M}_2\mathbf{V})_{jk}}v_{jk}. \qquad (17)$$

Algorithm 2: Proposed NMFAML Method.

Input: HSI samples, cluster number C, parameters N_1 and N_2, parameters λ_0, λ_1 and λ_2.
Output: Clustering map.
1: Estimate the number of superpixels.
2: Oversegment by using ERS method.
3: Compute matrix $\mathbf{Z}^{intra}$ according to (6).
4: Compute matrix $\mathbf{Z}^{inter}$ according to **Algorithm 1**.
5: Compute matrix $\mathbf{Z}_{sp2}$ according to (9).
6: Extract joint spectral-spatial features using PCA.
7: Compute matrices $\mathbf{U}$ and $\mathbf{V}$ according to (14) and (17), respectively.
8: Compute matrix $\mathbf{Z}$ according to (20).
9: Obtain spectral embedding using $\mathbf{Z}$.
10:     Apply the FSDP method to the spectral embedding.

Finally, we normalize $\mathbf{U}$ and $\mathbf{V}$ as follows:[3]

$$u_{ik} \leftarrow \frac{u_{ik}}{\sum_{k=1}^{d_f} u_{ik}} \qquad (18)$$

$$v_{jk} \leftarrow v_{jk}\sum_{k=1}^{d_f} u_{ik}. \qquad (19)$$

E. Density-Based Spectral Clustering

Once the NMF of HSI samples is achieved, each sample x_i is modeled as the low-dimensional sample v_i. Given the spectral-spatial constraint of $\mathbf{Z}_{sp2}$ and the intrinsic merit of NMF algorithm, the low-dimensional matrix $\mathbf{V}$ shows more discriminative power than $\mathbf{X}$. However, since there may exist errors in the affinity matrix $\mathbf{Z}_{sp2}$, such as the nonzero affinity coefficients between samples belonging to different classes, it is difficult to obtain satisfactory performance by directly apply typical clustering method (such as k-means) on matrix $\mathbf{V}$. Therefore, spectral clustering is employed and the final affinity matrix used is defined as follows:

$$\mathbf{Z} = \lambda_2\mathbf{Z}_{sp2} + (1 - \lambda_2)\mathbf{Z}_{nmf} \qquad (20)$$

where $\lambda_2 \in [0, 1]$ is a tradeoff coefficient and the normalized matrix $\mathbf{Z_{nmf}}$ is obtained by applying the K-nearest neighboring algorithm on matrix $\mathbf{V}$ with 0–1 weighting. Specifically, the value of $\mathbf{Z}_{nmf}^{ij}$ is set to 1 only if the sample v_j is the K-nearest neighbor of sample v_i via Euclidean distance and $\mathbf{Z}_{nmf}^{ij} = 0$ otherwise. The value of parameter K is empirically set to 20. When the spectral embedding of $\mathbf{Z}$ is obtained, FSDP[4] method is used to discover the true clustering centers, facilitating achieving accurate clustering maps. The implementation details of the proposed NMFAML method are illustrated in Algorithm 2.

---

[3][Online]. Available: http://www.cad.zju.edu.cn/home/dengcai/Data/GNMF.html
[4][Online]. Available: https://people.sissa.it/laio/Research/Res_clustering.php

III. CONVERGENCE AND ANALYSIS

A. Proof of Convergence

Apparently, since the loss in (10) is nonnegative, we only need to prove that the loss is nonincreasing in the iterative process of NMFAML. Meanwhile, the updating rules of matrix $\mathbf{U}$ is exactly the same as the conventional NMF, keeping the loss nonincreasing when updating $\mathbf{U}$. Thus, we only need to prove the convergence under the updating rule of $\mathbf{V}$. Similar to the convergence proof of graph regularized NMF in [46], the *auxiliary function* approach is employed. In the following, we begin with the introduction of the auxiliary function.

Definition: $\mathcal{A}(v, v')$ is an auxiliary function of $F(v)$ if the following two conditions are satisfied:

$$\mathcal{A}(v, v') \geq F(v), \quad \mathcal{A}(v, v) = F(v). \tag{21}$$

Then we can reach the following lemma: if $\mathcal{A}(v, v')$ is an auxiliary function of $F(v)$, then the value of $F(v)$ is nonincreasing under the following updating rule:

$$v^{(t+1)} = \arg\min_v \mathcal{A}(v, v^{(t)}). \tag{22}$$

Thus, we only need to prove that the updating rule of $\mathbf{V}$ satisfies (22) with a proper $\mathcal{A}(v, v')$. The objective function of NMFAML algorithm is formulated as

$$\mathcal{L} = \frac{1}{2}\|\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}\|_{\mathrm{F}}^2 + \frac{\lambda_1}{2}\|\mathbf{V}^{\mathrm{T}} - \mathbf{V}^{\mathrm{T}}\mathbf{Z}_{sp2}\|_{\mathrm{F}}^2. \tag{23}$$

We then consider the derivatives related to the element v_{ab} as follows:

$$F'_{ab} = [-\mathbf{X}^{\mathrm{T}}\mathbf{U} + \mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U} + \lambda_1(\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})\mathbf{V}]_{ab} \tag{24}$$

$$F''_{ab} = (\mathbf{U}^{\mathrm{T}}\mathbf{U})_{bb} + \lambda_1[(\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})]_{aa} \tag{25}$$

where F'_{ab} and F''_{ab} are the first-order and second-order derivatives, respectively.

The auxiliary function $\mathcal{A}(v, v')$ can be defined as

$$\mathcal{A}(v, v_{ab}^{(t)}) = F_{ab}(v_{ab}^{(t)}) + F'_{ab}(v_{ab}^{(t)})(v - v_{ab}^{(t)})$$
$$+ \frac{(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U})_{ab} + \lambda_1[(\mathbf{I}+\mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}})\mathbf{V}]_{ab}}{2v_{ab}^{(t)}}(v - v_{ab}^{(t)})^2. \tag{26}$$

Proof: Obviously, $\mathcal{A}(v, v) = F_{ab}(v)$. In the following, $\mathcal{A}(v, v') \geq F(v)$ is proved by using the Taylor series expansion of $F_{ab}(v)$, i.e.,

$$F_{ab}(v) = F_{ab}(v_{ab}^{(t)}) + F'_{ab}(v)(v - v_{ab}^{(t)})$$
$$+ \frac{1}{2}\{(\mathbf{U}^{\mathrm{T}}\mathbf{U})_{bb} + [(\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})]_{aa}\}(v-v_{ab}^{(t)})^2. \tag{27}$$

Comparing the $\mathcal{A}(v, v_{ab}^{(t)})$ with the Taylor series expansion of $F_{ab}(v)$, it is clear that we only need to prove the following:

$$\frac{(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U})_{ab} + \lambda_1(\mathbf{I}+\mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}})_{ab}}{v_{ab}^{(t)}} \geq F''_{ab}. \tag{28}$$

We have

$$(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U})_{ab} \geq v_{ab}^{(t)}(\mathbf{U}^{\mathrm{T}}\mathbf{U})_{bb} \tag{29}$$

and

$$[(\mathbf{I} + \mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}})\mathbf{V}]_{ab} \geq [(\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})\mathbf{V}]_{ab}$$
$$= \sum_{j=1}^{N}(\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})_{aj}v_{jb}^{(t)}$$
$$\geq (\mathbf{I}-\mathbf{Z}_{sp2})(\mathbf{I}-\mathbf{Z}_{sp2}^{\mathrm{T}})_{aa}v_{ab}^{(t)}. \tag{30}$$

Therefore, the two conditions of auxiliary function are satisfied and the updating rule of v_{ab} is obtained by solving (22) as follows:

$$v_{ab}^{(t+1)} = \frac{(\mathbf{X}^{\mathrm{T}}\mathbf{U} + \lambda_1\mathbf{M}_1\mathbf{V})_{ab}}{(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U} + \lambda_1\mathbf{M}_2\mathbf{V})_{ab}}v_{ab}^{(t)} \tag{31}$$

where $\mathbf{M}_1$ and $\mathbf{M}_2$ denote $\mathbf{Z}_{sp2} + \mathbf{Z}_{sp2}^{\mathrm{T}}$ and $\mathbf{I} + \mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}}$, respectively.

B. Computational Complexity Analysis

The computational complexity mainly contains three parts: ERS algorithm for oversegmentation, iterative process of NMF, and density-based spectral clustering. First, the complexity of ERS algorithm is averagely $O(N\log(N))$. Second, since the iterative process of the NMFAML method is similar to GNMF, the computational cost is $O(INdd_f)$. Note that the computation of $\mathbf{Z}_{sp2}\mathbf{Z}_{sp2}^{\mathrm{T}}$ is omitted as it is sparse and computed only once. Third, the main computational cost of spectral clustering cost lies in the SVD of $\mathbf{Z}$ with complexity $O(N^2 C)$. Therefore, the total computational cost of the proposed NMFAML method is $O(N\log(N) + INdd_f + N^2 C)$.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, the datasets, the baseline methods, and the evaluation metrics are first introduced. The parameter sensitivity of the proposed NMFAML method analysis is presented. Then, the qualitative and quantitative clustering results are reported to demonstrate the effectiveness of the NMFAML method. All experiments are conducted on a PC with 3.50 GHz Intel i7 CPU and 8.0 GB memory.

A. Experimental Setup

1) Datasets: The following three publicly available benchmark HSIs are employed to validate the robustness of the proposed NMFAML method.

1) The subset of Indian Pines image[5] obtained by the airborne visible infrared imaging spectrometer (AVRIS) over Northwestern Indian Pines contains 200 spectral featurs of 145×145 pixels and four main land-cover classes are considered: *corn_n_t, grass, soybeans_n_t* and *soybeans_m_t*.
2) The Salinas-A image[5] captured AVRIS over Salinas Vally contains 224 bands of 86×83 pixels and 6 different classes represent mostly different types of crops.

[5][Online]. Available: http://www.ehu.eus/ccwintco/index.php/Hyperspectral _Remote_Sensing_Scenes
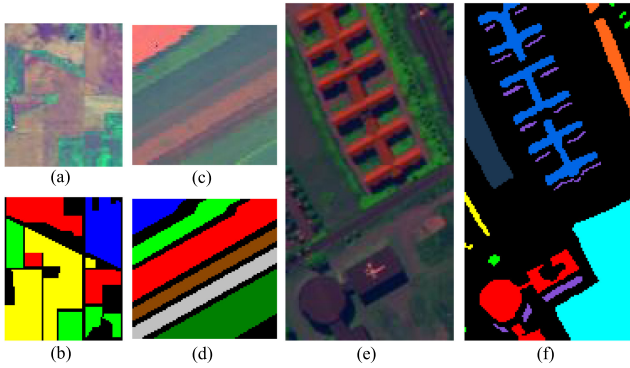
Fig. 3. Datasets used in our experiments. (a) Color composite image and (b) ground truth of the IPS scene. (c) Color composite image and (d) ground truth of the SA scene. (e) Color composite image and (f) ground truth of the PUS scene. (Best viewed in color).

TABLE I
NUMBER OF LABELED SAMPLES FOR IPS, SA, AND PUS DATASETS

IPS		SA		PUS	
Class	GT	Class	GT	Class	GT
Corn_n_t	1005	Brocoli_gw1	391	Asphalt	425
Grass	730	Corn_sgw	1343	Meadows	768
Soybean_n_t	732	Lettuce_r4	616	Tree	63
Soybean_m_t	1924	Lettuce_r5	1525	Metal sheet	1345
		Lettuce_r6	674	Bare soil	2559
		Lettuce_r7	799	Bitumen	860
				Bricks	94
				Shadows	361

3) The subset of Pavia University image[5] collected by the reflective optics spectrographic image system (ROSIS) sensor contains 103 spectral reflectance bands of 200×100 pixels and eight classes are considered herein: *asphalt, meadows, tree, metal sheet, bare soil, bitumen, bricks,* and *shadows*.

For simplicity, we denote the three datasets as IPS, SA, and PUS, respectively. Fig. 3 and Table I give the qualitative and quantitative ground truth (GT) of the three HSIs, respectively.

2) Baseline Methods: Several clustering approaches are employed as baseline methods, including four typical baseline methods (FCM [14], FCM_S [16], FSDP [17], RIM [33]), MF-based method (RMMF [42] and NMF [44]), two SSC-based methods (SSC-S and S4C [23]), DLSS [19], TV [26] and deep learning-based method (LSSD [1]). Note that since it is difficult to reproduce the best results of the baseline methods on different datasets, we then compare the proposed NMFAML method with the corresponding state-of-the-art methods for different datasets, i.e., SSC-S and S4C methods for the IPS scene, DLSS and LSSD methods for the SA scene, TV and LSSD methods for the PUS scene. Meanwhile, the clustering performance of other baseline methods (FCM, FCM_S1, FSDP, RIM, RMMF, and NMF methods) on all the three scenes are reported. Table II illustrates all the methods used in the experiments and their parameter settings.

Regarding the proposed NMFAML method, we fix the parameter setting for all the three scenes to better demonstrate the efficiency. Empirically, the neighbor number in the construction of matrix $\mathbf{Z}_{nmf}$ is set to 20 for the three scenes. Note that five

trials of the clustering have been performed to ensure the stability of the results.

B. Evaluation Metrics

Generally, the overall accuracy (OA), average accuracy (AA), normalized mutual information (NMI) and the Kappa statistics are used to numerically evaluate the accuracy of the clustering. The OA[6] can be computed using the Kuhn–Munkres algorithm [47] and is defined as

$$\text{OA} = \max \frac{\sum_{i=1}^{N} \#\{G_i = \mathbf{map}(C_i)\}}{N} \quad (32)$$

where map denotes the one-to-one permutation between GT labels and clustering labels, G_i and C_i are the GT label and clustering assignment produced using the clustering algorithm for x_i, respectively. In addition, the visualization of the affinity matrices $\mathbf{Z}_{sp2}$ and $\mathbf{Z}$, and the decision graphs of FSDP are employed to validate the effectiveness of the affinity matrices. The computational times of all methods for the three scenes are also reported to describe the computational load.

C. Parameters Sensitivity Analysis

In order to evaluate the proposed method in terms of universal parameter setting, the IPS scene and the corresponding ground truth are used for achieving the optimal parameter setting, whereas the PUS and SA scenes are employed for validation. In detail, the main parameters of the NMFAML method are the three tradeoff coefficients, i.e., λ_0, λ_1, and λ_2.

First, since λ_0 and λ_2 are directly related to the construction of affinity matrix $\mathbf{Z}$, the value of λ_1 is set to 0.5 to discuss the two parameters. In fact, the optimal values of λ_0 and λ_2 can be analyzed as follows.
1) Since the superpixels is achieved by oversegmentation, it is believed that the intrasuperpixel class consistency is relatively more reliable than intersuperpixel class consistency, indicating that the value of λ_0 is expected to be larger than 0.5.
2) Different from the superpixels that only consider the affinity of each sample with the samples in its spatially adjacent area, the low-dimensional samples obtained by NMF discover both local and global affinity of the whole hyperspectral image.

Consequently, compared with the affinity matrix $\mathbf{Z}_{sp2}$, it is expected that more emphasis should be be placed on $\mathbf{Z}_{nmf}$, which means that the optimal value of λ_2 should be smaller than 0.5. Fig. 4(a) illustrates the mean OAs over five trials with respect to different values of λ_0 and λ_2 for the IPS scene. Generally, larger value of λ_2 results in poorer clustering accuracy, indicating that it is difficult to obtain satisfactory performance if only superpixels-related affinity matrix $\mathbf{Z}_{sp2}$ is used. It can be also noticed that the value of λ_0 ranging from 0.6 to 0.8 is relatively preferable for better and sustainable accuracy. Consequently, λ_0 and λ_2 are set to 0.7 and 0.2, respectively. The corresponding

[6][Online]. Available: http://www.cad.zju.edu.cn/home/dengcai/Data/Clustering.html

TABLE II
IMPLEMENTATION DETAILS OF ALL THE METHODS

No.	Methods	Note	Parameter settings
1	FCM[5] [14]	a typical baseline for clustering.	#
2	FCM_S [16]	adopt spatial features in the image segmentation.	window size: 3×3, spatial weight α=0.05.
3	FSDP[4] [17]	classical density-based clustering method.	#
4	RIM[6] [33]	simultaneously cluster the data and train a discriminative classifier.	sweep λ across [0.001,10]
5	RMMF [42]	robust manifold matrix factorization for clustering.	#
6	SSC-S [23]	sparse subspace clustering incorporating spatial information.	*
7	S4C [23]	sparse subspace clustering with spectral–spatial structural information.	*
8	DLSS[7] [19]	diffusion inspired labeling incorporating spectral-spatial information.	*
9	TV [26]	regularized collaborative representation with a locally adaptive dictionary.	*
10	LSSD [1]	learn deep embedding based on two distance measures.	*
11	NMF [44]	spectral clustering on the affinity matrix obtained by typical NMF.	#
12	**NMFAML**	non-negative matrix factorization for affinity matrix learning.	λ_0=0.7, λ_1=0.6, λ_2=0.2, N_1=50 and N_2=100.

[#]The corresponding source codes are used.
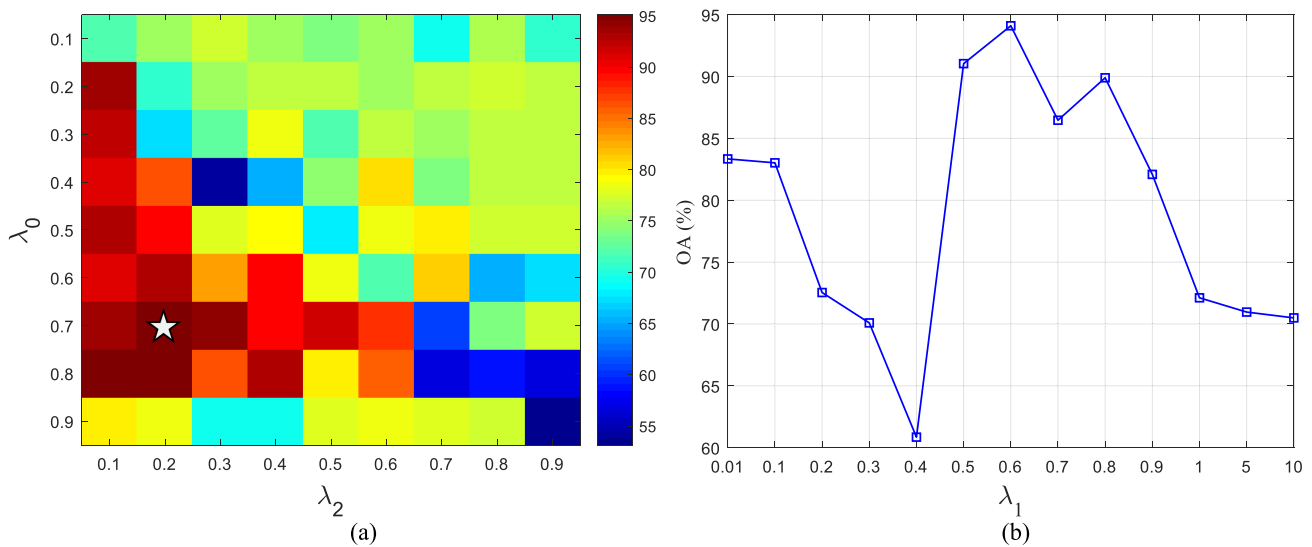[*]The results in the corresponding literature are reported.

Fig. 4. Mean OAs obtained by the proposed NMFAML method on the IPS scene using different values of (a) λ_0 and λ_2, (b) λ_1. (Best viewed in color).

weights of matrices $\mathbf{Z}^{tra}$, $\mathbf{Z}^{ter}$, and $\mathbf{Z}_{nmf}$ are 14%, 6%, and 80%, respectively.

Second, the value of λ_1 ranging from 0.01 to 10 are considered with λ_0 and λ_2 fixed as 0.7 and 0.2, respectively. Fig. 4(b) shows the corresponding clustering accuracy over five trials and the maximum value of mean OA reaches more than 94%. Therefore, the value of λ_1 is set to 0.6.

Furthermore, as shown in Fig. 5, we have elaborated the effect of the number of principal components when obtaining spectral-spatial features on clustering. One can note that the clustering accuracies of the SA and PUS scenes are more stable than that of the IPS scene. However, the empirical setting that four PCs are used can achieve acceptable clustering accuracies for the three scenes.

D. IPS Scene

The clustering maps of IPS scene obtained by the different methods are shown in Fig. 6. One can observe that there is some noise in the clustering map obtained by the proposed
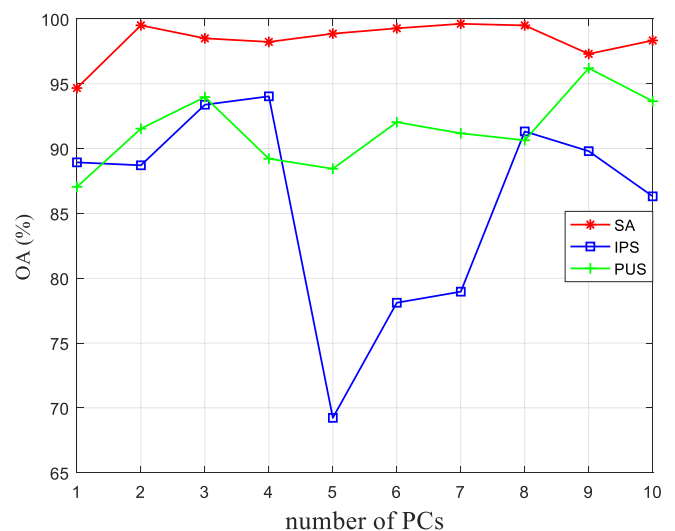

Fig. 5. Mean OAs obtained by the proposed NMFAML method on all the three scenes using different number of PCs. (Best viewed in color).
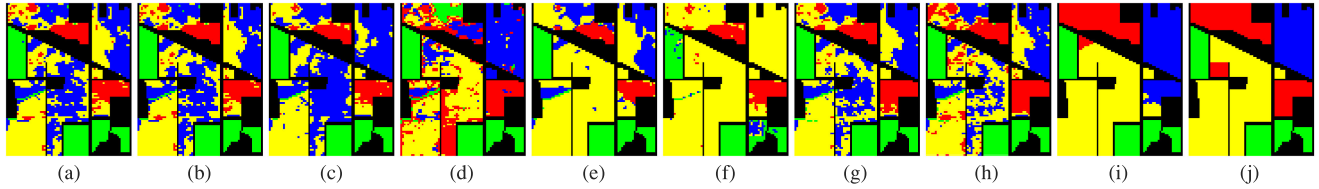
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j)

Fig. 6.　Clustering maps of the different methods on the IPS scene. (a) FCM (OA = 62.24%), (b) FCM_S (OA = 62.40%), (c) FSDP (OA = 61.15%), (d) RIM (OA = 71.24%), (e) NMF (OA = 74.17%), (f) RMMF (OA = 67.48%), (g) SSCS (OA = 68.12%), (h) S4C (OA = 70.08%), (i) NMFAML (OA = 97.02%), and (j) ground truth. (Best viewed in color).
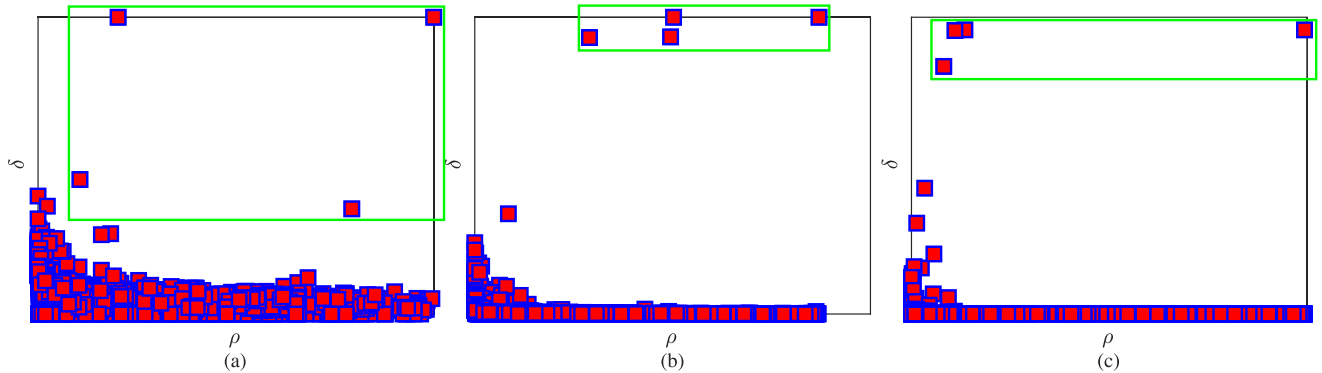

(a) (b) (c)

Fig. 7.　Decision maps of the different methods on the IPS scene. (a) FSDP, (b) NMF, and (c) NMFAML. The clustering centers are marked by the green rectangle. (Best viewed in color).

TABLE III
MEAN CLUSTERING ACCURACIES FOR THE IPS SCENE. THE BEST TWO RESULTS FOR EACH ROW ARE REPORTED IN BOLD AND UNDERLINED, RESPECTIVELY. THE PROPOSED METHOD OUTPERFORMS ALL BASELINE METHODS

Class	FCM	FCM_S	FSDP	RIM	NMF	RMMF	SSCS	S4C	**NMFAML**
Corn_n_t	44.78	45.77	39.70	52.70	35.32	41.99	58.05	61.00	**75.42**
Grass	99.86	**100.00**	100.00	99.89	100.00	87.40	100.00	100.00	100.00
Soybean_n_t	60.79	59.84	82.51	82.90	47.19	0.00	68.85	65.30	**100.00**
Soybean_m_t	57.64	57.80	49.48	52.12	93.56	**98.91**	64.76	65.28	95.15
OA (%)	62.24	62.40	61.15	65.32	73.57	67.48	68.12	70.08	**92.25**
AA (%)	71.06	70.79	74.38	71.27	80.37	63.97	\	\	93.91
Kappa (%)	47.60	47.78	47.49	52.94	60.10	48.16	55.45	58.25	**88.85**
NMI (%)	43.09	43.22	46.16	44.79	47.25	41.51	\	\	81.99

　\　The results in the corresponding literature are not reported.

NMFAML method [see Fig. 6(i)], which may be caused by the strategy of exploiting affinity between superpixels. However, the clustering map of the NMFAML method shows less scattered noise than other baseline methods, facilitating better overall clustering performance with OA of 97.02%. Meanwhile, the NMF method shows better accuracy than other baselines, validating that the effectiveness of the combination of MF and spectral clustering techniques. In addition, the decision graphs in the process of density-based clustering of FSDP, NMF, and NMFAML methods are shown in Fig. 7. Clearly, the clustering centers obtained by the NMF and NMFAML methods have larger densities (large values of δ) than the FSDP method. Therefore, it can be deduced that the low-dimensional representation **V** is more discriminative than original HSI features.

Table III reports the mean quantitative evaluation of the clustering results over 5 trials on the IPS scene. It is clear that the FCM, FCM_S, and FSDP methods are inferior to other methods resulting in lowest mean OAs of 62.24%, 62.40%, and 61.15%, respectively. Among all the baseline methods, the NMF and S4C methods obtain the best clustering performances (mean OAs of 73.57% and 70.08%, respectively) by exploiting MF and the spatial contextual information, respectively. However, the proposed NMFAML method achieves a significantly higher OA (92.25%) with main improvements lie in the "Corn_n_t" and "Soybean_n_t" classes, with class accuracy increments of more than 14.42% and 17.10%, respectively.

E. PUS Scene

The clustering maps of the PUS scene obtained by all the methods are shown in Fig. 8. Clearly, the proposed NMFAML method outperforms all the other methods with large margin of
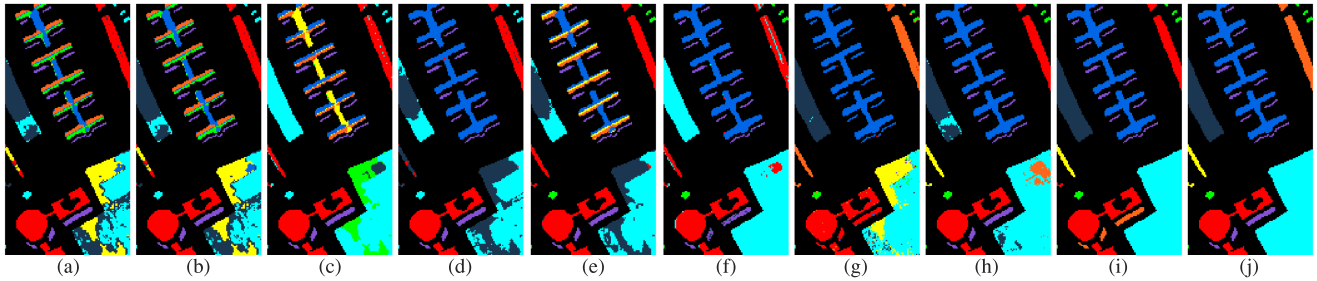
Fig. 8. Clustering maps obtained via different methods on the PUS scene. (a) FCM (OA=54.10%), (b) FCM_S (OA=54.06%), (c) FSDP (OA=54.21%), (d) RIM (OA=68.94%), (e) NMF (OA=62.92%), (f) RMMF (OA=78.20%), (g) TV (OA=79.67%), (h) LSSD (OA=87.40%), (i) NMFAML (OA=96.26%), and (j) ground truth. (Best viewed in color).
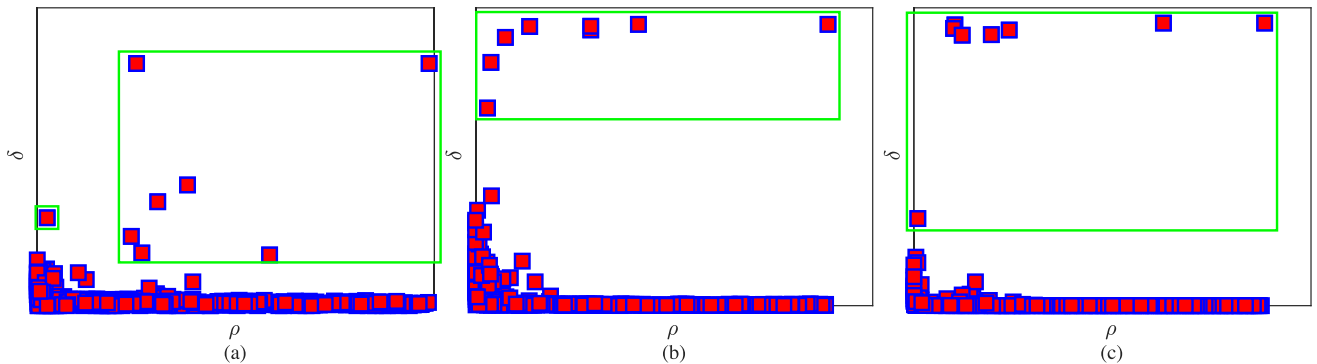

Fig. 9. Decision maps of the different methods on the PUS scene. (a) FSDP, (b) NMF, and (c) NMFAML. The clustering centers are marked by the green rectangle. (Best viewed in color).

TABLE IV
MEAN CLUSTERING ACCURACIES FOR THE PUS SCENE. THE BEST TWO RESULTS FOR EACH ROW ARE REPORTED IN BOLD AND UNDERLINED, RESPECTIVELY. THE PROPOSED METHOD OUTPERFORMS ALL BASELINE METHODS

Class	FCM	FCM_S	FSDP	RIM	NMF	RMMF	TV	LSSD	**NMFAML**
Asphalt	0.00	0.00	0.00	0.00	0.00	0.00	**98.82**	0.00	0.00
Meadows	68.75	<u>82.03</u>	0.00	57.34	46.09	0.00	**99.61**	36.74	80.00
Tree	0.00	0.00	0.00	0.00	**100.00**	<u>98.41</u>	66.67	**100.00**	80.32
Metal Sheet	36.30	37.34	37.57	92.02	54.49	99.39	**100.00**	**100.00**	<u>99.95</u>
Bare soil	47.71	43.06	69.52	58.05	64.25	<u>97.46</u>	68.58	94.83	**100.00**
Bitumen	**100.00**	**100.00**	**100.00**	<u>99.93</u>	99.44	98.60	97.33	**100.00**	**100.00**
Bricks	34.04	42.55	0.00	0.00	0.00	0.00	0.00	**97.87**	<u>52.98</u>
Shadows	**100.00**	**100.00**	**100.00**	<u>99.50</u>	<u>99.50</u>	91.14	0.28	**100.00**	58.78
OA (%)	53.99	54.06	54.21	67.57	61.94	78.20	79.67	<u>83.79</u>	**87.83**
AA (%)	48.11	49.22	41.37	46.56	59.77	54.51	63.72	**80.80**	<u>74.90</u>
Kappa (%)	44.45	45.31	41.51	58.76	51.97	69.74	74.59	<u>78.16</u>	**83.58**
NMI (%)	61.41	60.94	68.34	71.51	69.23	74.55	78.14	<u>80.91</u>	**86.50**

mean OA of 96.26%, whereas the largest mean OA obtained by the baseline methods is 87.40% (achieved by the LSSD method). Furthermore, scattered noise can hardly be found in the clustering map of NMFAML method, whereas FCM, FCM_S, FSDP, and NMF methods produce much noise with lowest mean OA of 54.10%, 54.06%, 54.21%, and 62.92%, respectively. In addition, the corresponding decision graphs in the process of

density-based clustering of FSDP, NMF, and NMFAML methods are shown in Fig. 9. Obviously, the NMF and NMFAML methods are superior to the FSDP method with clustering centers having larger densities (large values of δ).

Table IV reports the quantitative clustering performance obtained by all the methods over five trials on the PUS scene. It is clear that the proposed NMFAML method is superior to
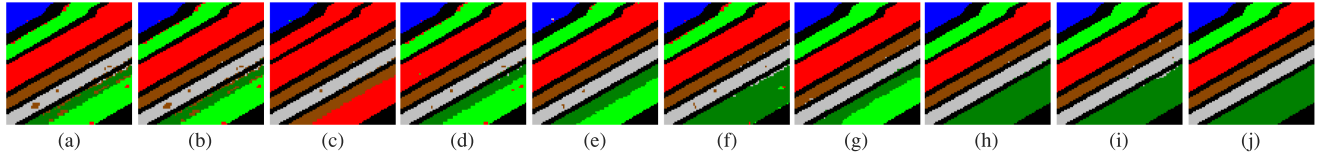
Fig. 10. Clustering maps of the different methods on the SA scene. (a) FCM (OA=81.53%), (b) FCM_S (OA=81.68%), (c) FSDP (OA=63.00%), (d) RIM (OA=83.75%), (e) NMF (OA=86.31%), (f) RMMF (OA=98.20%), (g) DLSS (OA=84.76%), (h) LSSD (OA=99.89%), (i) NMFAML (OA= 99.57%), and (j) ground truth. (Best viewed in color).
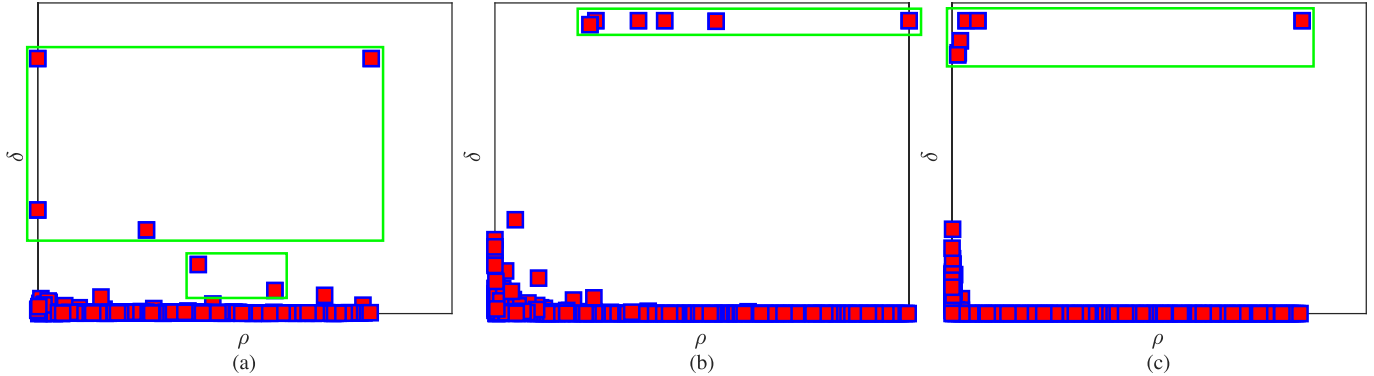

Fig. 11. Decision maps of the different methods on the SA scene. (a) FSDP, (b) NMF, and (c) NMFAML. The clustering centers are marked by the green rectangle. (Best viewed in color).

TABLE V
MEAN CLUSTERING ACCURACIES FOR THE SA SCENE. THE BEST TWO RESULTS FOR EACH ROW ARE REPORTED IN BOLD AND UNDERLINED

Class	FCM	FCM_S	FSDP	RIM	NMF	RMMF	DLSS	LSSD	**NMFAML**
Brocoli_gw1	<u>99.74</u>	<u>99.74</u>	99.49	79.74	99.34	<u>99.74</u>	**100.00**	**100.00**	**100.00**
Corn_sgw	31.57	32.27	0.00	43.77	44.63	96.72	40.43	**100.00**	<u>98.85</u>
Lettuce_r4	95.78	95.29	0.00	96.72	<u>97.99</u>	94.16	**100.00**	**100.00**	**100.00**
Lettuce_r5	**100.00**	**100.00**	**100.00**	79.44	<u>99.99</u>	**100.00**	99.87	**100.00**	**100.00**
Lettuce_r6	**99.85**	**99.85**	99.70	99.73	<u>99.82</u>	99.26	99.11	**99.85**	99.55
Lettuce_r7	94.87	94.99	98.00	97.17	<u>99.20</u>	98.75	99.12	98.77	**99.35**
OA (%)	81.53	81.66	63.00	77.70	85.67	98.20	84.76	**99.80**	<u>99.56</u>
AA (%)	86.65	86.77	51.10	81.71	90.43	98.34	90.18	**99.86**	<u>99.55</u>
Kappa (%)	77.49	77.65	52.95	72.86	82.47	97.75	81.39	**99.75**	<u>99.45</u>
NMI (%)	81.04	81.21	56.33	80.13	88.27	94.76	88.24	**99.28**	<u>98.48</u>

other methods with largest mean OA of 87.83%. Among all the baseline methods, the recently published TV and LSSD methods achieves the best two clustering performances (mean OAs of 79.67% and 83.79%, respectively) by exploiting the spatial information based on local dictionary and deep learning technique, respectively.

F. Salinas-A Scene

The examples of the clustering maps obtained by the different methods on the SA scene are shown in Fig. 10. One can note that comparable clustering accuracies are obtained by the RMMF, LSSD and NMFAML methods with OA of 98.20%, 99.89%,

and 99.57%, respectively. In detail, compared with the clustering map of LSSD, the clustering map of NMFAML has noise on the edge of different classes, which leads to a slight reduction in clustering accuracy. Meanwhile, although the NMF method achieve better performance than the DLSS method, they both fails in the clustering of the "Corn_sgw" class [see the deep green color in Fig. 10(e) and (g)]. Fig. 11 gives the density-based decision graphs of FSDP, NMF and NMFAML methods. It can easily seen that the low dimensional representation of HSI samples obtained by NMFAML method is more distinguishable with clustering centers of apparently large densities.

To further illustrate the clustering performance, Table V reports the mean quantitative evaluation of the clustering
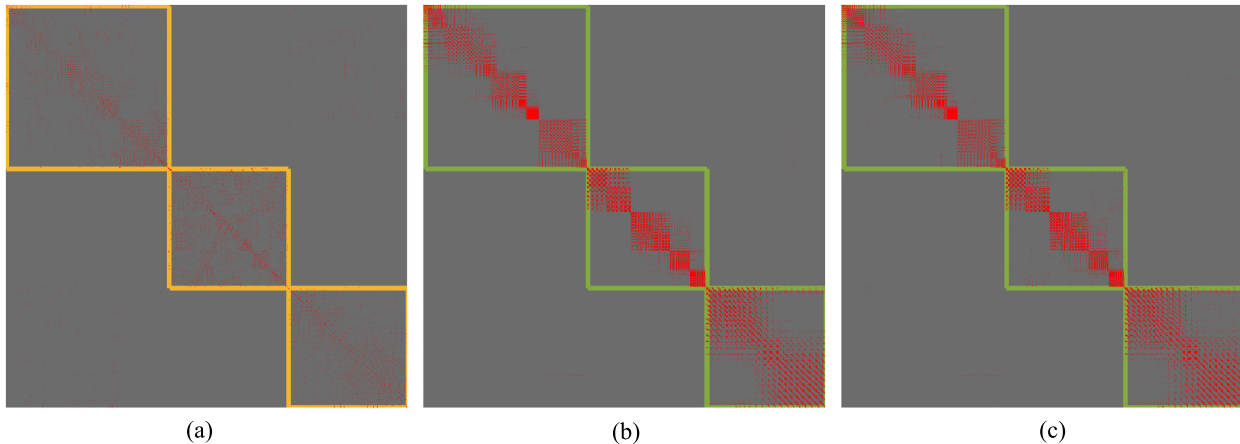
Fig. 12. Illustration of affinity matrices obtained by the NMF and NMFAML methods on the three classes of the IPS scene. (a) Matrix $\mathbf{Z}$ of NMF, (b) matrix $\mathbf{Z}_{sp2}$ of NMFAML, and (c) matrix $\mathbf{Z}$ of NMFAML.

TABLE VI
COMPUTATIONAL TIMES OF DIFFERENT METHODS ON THE THREE SCENES (IN SECONDS)

Dataset	FCM	FCM_S	FSDP	RIM	NMF	RMMF	SSCS/TV/DLSS	S4C/LSSD	**NMFAML**
IPS	1.56	4.17	9.56	39.46	10.95	0.99	1032.70 (SSCS)	1567.90 (S4C)	73.30
SA	3.12	5.32	14.73	58.55	19.03	1.51	8.03 (DLSS)	578.32 (LSSD)	103.28
PUS	1.80	3.92	12.59	86.63	21.42	7.93	6933.00 (TV)	674.58 (LSSD)	142.66

performance on the SA scene over five trials. One can note that the FSDP and RIM methods perform worse than all the other methods with mean OAs of 63.00% and 76.01%, respectively. The FCM, FCM_S, NMF, and DLSS methods result in OAs of 81.53%, 81.66%, 85.67%, and 84.76%, respectively. Compared with all the other baseline methods, the RMMF, LSSD, and NMFAML methods result in the highest mean OAs (98.20%, 99.80%, and 99.56%) with large improvements on the "Corn_sgw" class.

G. Affinity Matrix Visualization and Computational Cost Analysis

For better illustration of the rationale of the proposed NM-FAML method, Fig. 12 gives the visualization of the affinity matrices in the NMF and NMFAML methods on the "Corn_n_t," "Grass," and "Soybean_n_t" classes of the IPS scene. Both the spectral-spatial affinity matrix $\mathbf{Z}_{sp2}$ and the final affinity matrix $\mathbf{Z}$ of the NMFAML method are employed. Note that the elements of the three matrices are arranged so that the square areas marked by the orange and green lines give the intraclass affinities. In fact, the nonzero elements of affinity matrix is expected to be regularized in the intraclass area for capturing the accuracy affinity of HSI samples. The following two observations can be easily drawn. First, there exist less nonzero elements in the affinity matrix of NMF method, indicating that it is hardly to achieve satisfactory clustering performance by exploiting the affinity matrix. Second, since the affinity matrix of NMFAML method is derived from superpixels, its major nonzero elements shows the pattern of diagonal block and the interblock nonzero

elements in each class share a small proportion. These interblock elements are critical to assemble the HSI samples of same class belonging to superpixels.

The computational times for different methods on the three scenes are reported in Table VI. In general, three typical methods (FCM, FCM_S, and FSDP) take the least time among all methods, whereas the representation matrix-based methods (SSCS, S4C, and TV) require much larger times due to the computation of representation matrices. It is interesting to note that the RMMF method also requires much lower times than the NMF and NMFAML methods as it regularizes the low-dimensional to be orthogonal instead of nonnegative. Compared with the state-of-the-art baselines, the proposed NMFAML method is superior in term of less computation costs and better clustering performance.

V. CONCLUSION

In order to address the task of HSI clustering, this article has employed superpixels and the NMF algorithm to compute the affinity matrix between HSI samples, which is finally used for density-based clustering to obtain clustering maps. As the key of spectral clustering, the affinity matrix related to spectral-spatial information is derived from the intrasuperpixels and intersuperpixels affinity of HSI samples. In order to further capture the rich and accurate affinity between HSI samples, the obtained affinity matrix is used to guide the iterative process of the NMF algorithm. The final affinity matrix is defined as the weight of the spectral-spatial affinity matrix and the affinity matrix of low-dimensional representation of HSI samples obtained by

NMF. The experiments have been conducted on three public benchmark HSIs, i.e., subsets of the Indian Pines and Pavia University datasets, Salinas-A. The NMFAML method outperforms the state-of-the-art clustering methods for all the three HSIs on both clustering accuracy and computational load.

Acknowledgment

References

[1] Y. Qin, L. Bruzzone, and B. Li, "Learning discriminative embedding for hyperspectral image clustering based on set-to-set and sample-to-sample distances," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 473–485, Jan. 2020.

[2] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[3] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[4] 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.

[5] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[6] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.

[7] K. Tiwari, M. Arora, and D. Singh, "An assessment of independent component analysis for detection of military targets from hyperspectral images," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 13, no. 5, pp. 730–740, 2011.

[8] Y. Qin, B. Li, and Y. Ye, "Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification," *IEEE Tran. Geoscience Remote Sensing*, vol. 57, no. 6, pp. 3952–3966, 2019.

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

[10] L. Mou, L. Bruzzone, and X. X. Zhu, "Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery," *IEEE Trans. Geoscience Remote Sensing*, vol. 57, no. 2, pp. 924–935, 2019.

[11] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.

[12] U. Shaham, K. Stanton, H. Li, B. Nadler, R. Basri, and Y. Kluger, "SpectralNet: Spectral clustering using deep neural networks," in *Proc. Int. conf. learn. Representations*, 2018.

[13] J. Anderson and J. Bodie, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 21, pp. 379–387, 1975.

[14] W. Pedrycz, "Fuzzy sets in pattern recognition: Methodology and methods," *Pattern Recognit.*, vol. 23, no. 1/2, pp. 121–146, 1990.

[15] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proc. 18th Annu. ACM-SIAM Symp. Discrete Algorithms*, 2007, pp. 1027–1035.

[16] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern., Part B (Cybern.)*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[17] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[18] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[19] J. M. Murphy and M. Maggioni, "Unsupervised clustering and active learning of hyperspectral images with nonlinear diffusion," *IEEE Trans. Geosci. Remote Sens. Mag.*, vol. 57, no. 3, pp. 1829–1845, Mar. 2019.

[20] Y. Zhong, S. Zhang, and L. Zhang, "Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2290–2301, Oct. 2013.

[21] A. Ma, Y. Zhong, and L. Zhang, "Adaptive multiobjective memetic fuzzy clustering algorithm for remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4202–4217, Aug. 2015.

[22] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[23] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[24] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.

[25] M. B. Salem, K. S. Ettabaa, and M. S. Bouhlel, "An unsupervised classification approach for hyperspectral images based adaptive spatial and spectral neighborhood selection and graph clustering," *Procedia Comput. Sci.*, vol. 126, pp. 743–750, 2018.

[26] H. Zhai, H. Zhang, L. Zhang, and P. Li, "Total variation regularized collaborative representation clustering with a locally adaptive dictionary for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 166–180, Jan. 2019.

[27] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 478–487.

[28] W. Hu, T. Miyato, S. Tokui, E. Matsumoto, and M. Sugiyama, "Learning discrete representations via information maximizing self-augmented training," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1558–1567.

[29] M. M. Fard, T. Thonet, and E. Gaussier, "Deep k-means: Jointly clustering with k-means and learning representations," *Pattern Recognit. Lett.*, vol. 138, pp. 185–192, 2020.

[30] J. Chang, G. Meng, L. Wang, S. Xiang, and C. Pan, "Deep self-evolution clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 809–823, Apr. 2020.

[31] X. Peng, J. Feng, J. T. Zhou, Y. Lei, and S. Yan, "Deep subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 12, pp. 5509–5521, 2020.

[32] K. Lim, X. Jiang, and C. Yi, "Deep clustering with variational autoencoder," *IEEE Signal Proc. Lett.*, vol. 27, pp. 231–235, 2020.

[33] A. Krause, P. Perona, and R. G. Gomes, "Discriminative clustering by regularized information maximization," in *Proc. Adv. Neural Inf. Proc. Syst.*, 2010, pp. 775–783.

[34] E. Min, X. Guo, Q. Liu, G. Zhang, J. Cui, and J. Long, "A survey of clustering with deep learning: From the perspective of network architecture," *IEEE Access*, vol. 6, pp. 39 501–39 514, 2018.

[35] Y. Kong, Y. Cheng, C. L. P. Chen, and X. Wang, "Hyperspectral image clustering based on unsupervised broad learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1741–1745, Nov. 2019.

[36] R. Wang, F. Nie, and W. Yu, "Fast spectral clustering with anchor graph for large hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2003–2007, Nov. 2017.

[37] R. Wang, F. Nie, Z. Wang, F. He, and X. Li, "Scalable graph-based clustering with nonnegative relaxation for large hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7352–7364, Oct. 2019.

[38] A. Mehta and O. Dikshit, "Segmentation-based projected clustering of hyperspectral images using mutual nearest neighbour," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5237–5244, Dec. 2017.

[39] G. Takn and O. Ceylan, "An adaptive affinity matrix optimization for locality preserving projection via heuristic methods for hyperspectral image analysis," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4690–4697, Dec. 2019.

[40] L. Fan and D. W. Messinger, "Joint spatial–spectral hyperspectral image clustering using block-diagonal amplified affinity matrix," *Proc. SPIE*, vol. 57, 2018, Art. no. 033107.

[41] J. Alameddine, K. Chehdi, and C. Cariou, "Optimization of unsupervised affinity propagation clustering method," *Proc. SPIE*, vol. 11155, pp. 115–124, 2019. [Online]. Available: https://doi.org/10.1117/12.2533164

[42] L. Zhang, L. Zhang, B. Du, J. You, and D. Tao, "Hyperspectral image unsupervised classification by robust manifold matrix factorization," *Inf. Sci.*, vol. 485, pp. 154–169, 2019.

[43] S. Huang, H. Zhang, Q. Du, and A. Pižurica, "Sketch-based subspace clustering of hyperspectral images," *Remote Sens.*, vol. 12, no. 5, 2020, Art. no. 775.

[44] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Int. Conf. Neural Inf. Proc. Syst.*, 2000, pp. 535–541.

[45] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. Comput. Vis. Pattern Recognit. IEEE Conf.*, 2011, pp. 2097–2104.

[46] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[47] J. Munkres, "Algorithms for the assignment and transportation problems," *J. Soc. Ind. Appl. Math.*, vol. 5, no. 1, pp. 32–38, 1957.

Sinong Quan received the B.S. degree in mechanical and electronic engineering from the South China University of Technology, Guangzhou, China, in 2013, and the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2015 and 2019, respectively.

He is currently a Lecturer with the School of Electronic Science, National University of Defense Technology. His current research interests include polarimetric anti-interference and target detection, pattern recognition, synthetic aperture radar/polarimetric synthetic aperture radar image processing, and machine learning.

Dr. Quan was the recipient of the Outstanding Graduate Award from the National University of Defense Technology, in 2015.
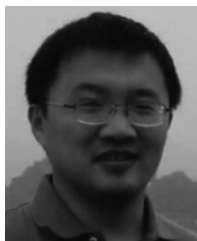
Yao Qin (Member, IEEE) received the B.S degree in information engineering from Shanghai Jiaotong University, Shanghai, China, in 2013, the M.S. and Ph.D. degrees in information and communication engineering from the National University of Defense Technology, Changsha.

He was a visiting Ph.D. with the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, Trento, Italy. He has been a Research Assistant with Northwest Institute of Nuclear Technology, Xi'an, China, since 2020. His research interests include infrared small target detection, hyperspectral image classification and clustering, and domain adaptation.

Peizhong Wang He was born in China in 1981. He received the B.S. degree in science and technology of remote sensing from Chang'an University, Xi'an, China.

He has been a Research Associate with the Northwest Institute of Nuclear Technology, Xi'an, China, since 2017. His research interest includes hyperspectral image processing and remote sensing.

Biao Li received the Ph.D. degree in information and communication engineering from the College of Electronic Science, National University of Defense Technology, Changsha, China, in 1998.
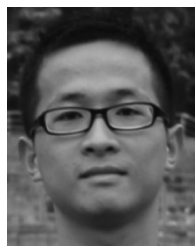
He is currently a Full Professor with the National University of Defense Technology. His current research interests include signal processing and infrared image processing.

Hui Bian was born in 1971. He received the B.S. degree in test engineering from the Rocket Force University of Engineering, Xi'an, China, in 1994, and the M.S. degree in signal and information processing from Tsinghua University, Beijing, China, in 2006.

He is an Associate Research Fellow with the Northwest Institute of Nuclear Technology, Xi'an, China. His research interests include image fusion, target detection, and pattern recognition.

Weiping Ni was born in China in 1980. He received the B.S. degree in electrical science and technology from the University of Science and Technology of China, Hefei, China, in 2004, the M.S. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2006, and the Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2016.

He has been a Research Associate with the Northwest Institute of Nuclear Technology, Xi'an, China, since 2014. His research interest includes remote sensing image processing, automatic target recognition, and computer vision.