

# Spectral–Spatial Feature Extraction for Hyperspectral Anomaly Detection

Jie Lei<sup>ID</sup>, Weiying Xie<sup>ID</sup>, *Member, IEEE*, Jian Yang, Yunsong Li, and Chein-I Chang<sup>ID</sup>, *Fellow, IEEE*

**Abstract**—Hyperspectral anomaly detection faces various levels of difficulty due to the high dimensionality of hyperspectral images (HSIs), redundant information, noisy bands, and the limited capability of utilizing spectral–spatial information. In this paper, we address these problems and propose a novel approach, called spectral–spatial feature extraction (SSFE), which is based on two main aspects. In the spectral domain, we assume that the anomalous pixels are rarely present and all (or most) of the samples around the anomalies belong to background (BKG). Using this fact, we introduce a suppression function to construct a discriminative feature space and utilize a deep brief network to learn spectral representation and abstraction automatically that are used as inputs to the Mahalanobis distance (MD)-based detector. In the spatial domain, the anomalies appear as a small area grouped by pixels with high correlation among them compared to BKG. Therefore, the objects appearing as a small area are extracted based on attribute filtering, and a guided filter is further employed for local smoothness. More specifically, we extract spatial features of anomalies only from one single band obtained by fusing all bands in the visible wavelength range. Finally, we detect anomalies by jointly considering the spectral and spatial detection results. Several experiments are performed, which show that our proposed method outperforms the state-of-the-art methods.

**Index Terms**—Anomaly detection, deep brief network, feature extraction, hyperspectral image (HSI), interference suppression.

## I. INTRODUCTION

BENEFITTED from the combination of spectral and spatial information, hyperspectral image (HSI) can distinguish targets of interest from background (BKG) based on spectral signatures and spatial characteristics. In particular, hyperspectral anomaly detection has attracted increasing attention since it is significant in military and civilian applications [1], [2]. In theory, anomalies are different from normal majority classes or objects, and they are outliers [3]. Anomaly detection is a fundamental task in hyperspectral data exploitation that aims at unveiling targets or objects whose spectral signatures are distinct from their surrounding data samples. In practice, it is a challenge to do anomaly detection because there is no prior knowledge of the anomalies or BKG and the BKG is usually complex. Besides, the HSIs are contaminated with noise that also increases the difficulty of detection. In general, the anomalies exhibit two characteristics compared to BKG, which can be used to distinguish from their surroundings: 1) anomalies have different spectra from BKG and 2) anomalous objects in real remote sensing HSIs appear as relatively small area embedded in the local smooth BKG.

The traditional methods mainly focus on anomalies with different spectral signatures. As an important milestone of anomaly detection, the global Reed–Xiaoli (RX) (GRX) algorithm proposed by Reed and Yu [4] is based on a hypothesis testing in which the conditional probability density functions under the two hypotheses (without and with anomaly) are assumed to be Gaussian. Then, the anomalies are detected by calculating the Mahalanobis distance (MD) between the pixel to be tested and the BKG. Many extensions based on RX have been investigated for hyperspectral anomaly detection [5]–[10]. Among them, the local RX (LRX) detector is a typically evolved version of RX, which estimates BKG using local statistics (i.e., mean and covariance matrix) [9]. Obviously, the Gaussian distribution assumption of BKG in these methods is not fully reasonable due to the complexity of real remote sensing HSIs [11]–[14]. To address this issue, a kernel RX detector [15], a low-rank-based detector [16], [17], and a sparse representation-based detector [18], [19] have been proposed. The kernel methods model the non-Gaussian distributions of BKG in a higher dimensional feature space. The low-rank-based detector is designed based on the assumption that BKG is low rank and anomalies are sparse. In sparse representation, collaborative-representation-based detector (CRD) [19] exploits the concept that each

Manuscript received November 26, 2018; revised March 20, 2019 and May 8, 2019; accepted May 17, 2019. Date of publication June 12, 2019; date of current version September 25, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61801359, Grant 61571345, Grant 91538101, Grant 61501346, Grant 61502367, and Grant 61701360, in part by the 111 Project under Grant B08038, in part by the Fundamental Research Funds for the Central Universities under Grant JB180104, in part by Yangtze River Scholar Bonus Schemes under Grant CJT160102, in part by the Ten Thousand Talent Program, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JQ153, Grant 2016JQ6023, and Grant 2016JQ6018, in part by the General Financial Grant from the China Postdoctoral Science Foundation under Grant 2017M620440 and in part by the Science and Technology on Electro-Optic Control Laboratory and Aeronautical Science Foundation of China under Grant 6142504190206. (Corresponding authors: Weiying Xie; Yunsong Li.)

J. Lei is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China, and also with the Science and Technology on Electro-Optic Control Laboratory, Luoyang 471000, China (e-mail: jielei@mail.xidian.edu.cn).

W. Xie, J. Yang, and Y. Li are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China (e-mail: wxyxie@xidian.edu.cn; ysl@mail.xidian.edu.cn).

C.-I. Chang is with the Center of Hyperspectral Imaging in Remote Sensing, Information and Technology College, Dalian Maritime University, Dalian 116026, China, also with the Department of Computer Science and Information Engineering, National Yunlin University of Science and Technology, Douliu 64002, Taiwan, also with the Remote Sensing Signal and Image Processing Laboratory, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore, MD 21250 USA, and also with the Department of Computer Science and Information Management, Providence University, Taichung 02912, Taiwan (e-mail: cchang@umbc.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TGRS.2019.2918387

0196-2892 © 2019 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See [http://www.ieee.org/publications\\_standards/publications/rights/index.html](http://www.ieee.org/publications_standards/publications/rights/index.html) for more information.

pixel in BKG can be approximately represented by its spatial neighborhoods, while anomalies cannot, and the collaboration of representation is reinforced by the  $\ell_2$ -norm minimization of the representation weight vector. In addition, Li *et al.* [20] proposed a transferred deep convolutional neural network-based strategy for anomaly detection (CNND) in HSI. In the CNND, a reference data set was employed to train the CNN with pixel pairs generated from labeled samples. The disadvantages are that the training process may cause time-consuming and the training samples are limited for anomaly detection.

Besides the previous methods based on spectral signatures, spatial feature extraction methods are recently proposed for anomaly detection [21], [22], which depend on the assumption that anomalies in real remote sensing HSIs appear as small area objects compared to BKG.

In reality, if an anomaly is detected from only one aspect (spectral or spatial), the detection result will be unsatisfactory. Li *et al.* [23] linearly combined the spectral information based on the structured sparse representation model and the spatial information based on the intracluster reconstruction model. Zhao and Zhang [24] exploited the stacked autoencoder to extract spatial-spectral feature matrices for anomaly detection. These recent works proved that using both spatial and spectral information is more useful than using only spectral information. Obviously, the key to success for these methods is an appropriate feature extraction method for effective BKG suppression in both spectral and spatial domains.

The principal component analysis (PCA) and the independent component analysis (ICA) are popular feature extraction methods in the spectral domain, which project HSIs into another feature space to carry on a dimensionality reduction for feature extraction [25]–[27]. These methods may suppress BKG and thus increase the separability of anomaly and BKG. However, the anomaly appeared as single pixel like noise may be eliminated in the significant PCs. The disadvantage of ICA-based methods is that they just independently deal with each pixel without considering the spatial information.

Recent studies have highlighted deep neural networks (DNNs), which can extract multiple layers of representation and abstraction to help model and generalize from complex data sets. Thus, DNNs are believed to yield higher detection accuracy than those traditional, shallow feature extractors [28]–[30]. Unfortunately, there are no training samples that can be provided for hyperspectral anomaly detection. Deep belief network (DBNs) are promising DNNs, which can extract high-level features in an unsupervised manner [31]–[34]. Ma *et al.* [35] considered the reconstruction errors between the input and the output of the DBN autoencoder as anomaly score directly, but they only used the spectral features of HSIs without including the spatial features.

In this paper, a spectral-spatial feature extraction-based hyperspectral anomaly detection method, named SSFE, is proposed to address the above problems. In the spectral domain, the proposed SSFE method is effectively performed based on a simple assumption that the anomalous pixels are rarely present and all (or most) of the samples surrounding the anomalies belong to BKG. Based on this assumption, we can construct a suppression function to increase the separability between anomalies and BKG, then the reconstructed data are

input into DBN with the aim at extracting a reduced number of features to which both anomaly and BKG can be well characterized. To obtain an initial detection map based on the spectral features extracted by DBN, the MD method is used. Because the spatial information can also benefit anomaly detection performance, we further detect anomalies based on the spatial features. In the spatial domain, the proposed SSFE assumes that the anomalies in real remote sensing HSIs appear as a small size of pixels compared to BKG, and it employs the spatial correlation among adjacent pixels. Such priors can help to capture the underlying spatial characteristics of anomalies. Due to the fact that spatial features can be captured by human visible wavelengths approximately ranging from 400 to 760 nm and it is time-consuming to extract spatial features from high-dimensional HSIs, the bands in the visible light wavelength range are further fused. Since the closing and opening operations in the attribute filter can remove the dark and bright connected components in small areas [36], [37], a simple differential operation is then introduced to preserve connected components with small areas. According to the local smoothness prior, we make use of a guided filter (GF) [38] to obtain an initial detection map based on its good edge preserving capability. The final detection map is a linear combination obtained from both spectral and spatial domains. Extensive experiments have been conducted on real-world data sets to verify the superior performance of the proposed SSFE method.

The rest of this paper is organized into five sections. Section II is a brief description of DBN. In Section III, we detail our proposed SSFE-based hyperspectral anomaly detection method. Experiments and parameter settings are performed in Section VI. Section V draws some conclusions.

## II. RELATED WORK

DBN can effectively produce different layers of nonlinear latent features in an unsupervised and layer-by-layer manner, where restricted Boltzmann machines (RBM) is commonly used as a layerwise training model [33]. This network can be employed as a tool for feature extraction and dimensionality reduction. We first review the main component in the construction of a DBN, i.e., the RBM.

RBM is a layerwise network, which consists of visible and hidden units. A symmetrical weight matrix  $\mathbf{W}$  can connect the visible units  $\mathbf{v}$  and the hidden units  $\mathbf{h}$ . It is noted that there is no connection between two visible units or between two hidden units. The joint distribution of the visible and hidden units can be given by

$$p(\mathbf{v}, \mathbf{h}; \theta) = \frac{e^{-E(\mathbf{v}, \mathbf{h}; \theta)}}{Z(\theta)} \quad (1)$$

$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} E(\mathbf{v}, \mathbf{h}; \theta) \quad (2)$$

where  $Z(\theta)$  is a normalization constant defined as the sum of  $E(\mathbf{v}, \mathbf{h}; \theta)$  over all possible values of the visible units  $\mathbf{v}$  and the hidden units  $\mathbf{h}$ .  $E(\mathbf{v}, \mathbf{h}; \theta)$  is the energy function of a joint distribution given by

$$\begin{aligned} E(\mathbf{v}, \mathbf{h}; \theta) &= -\sum_{i=1}^K a_i v_i - \sum_{j=1}^J b_j h_j - \sum_{i=1}^K \sum_{j=1}^J W_{ij} v_i h_j \\ &= -\mathbf{a}^T \mathbf{v} - \mathbf{b}^T \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h} \end{aligned} \quad (3)$$

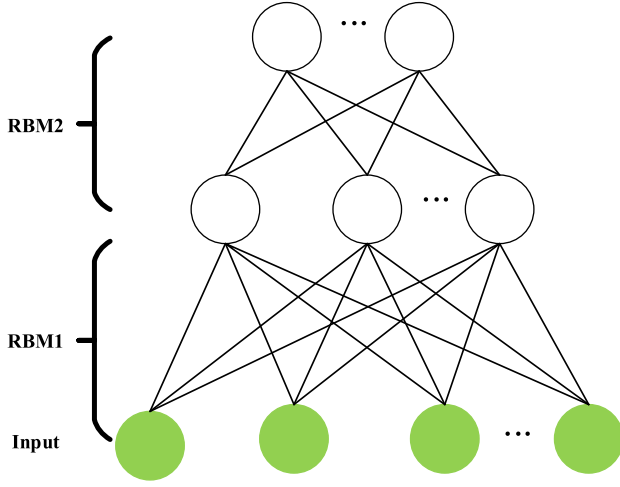


Fig. 1. Illustration of a two-layer DBN.

where  $\theta = \{a_i, b_j, W_{ij}\}$ ;  $W_{ij}$  denotes the weight between the visible unit  $i$  and the hidden unit  $j$ ;  $a_i$  and  $b_j$  represent the bias terms of the visible and hidden layers, respectively; and  $K$  and  $J$  are the total numbers of units in the visible and hidden layers, respectively.

It is clear from (3) that the hidden units are independent with the given visible units, and so are the visible units. The conditional distributions of the input data and the hidden unit are specified by the following logistic functions:

$$p(v_i = 1 | \mathbf{h}, \theta) = g\left(a_i + \sum_{i=1}^K W_{ij} v_i\right) \quad (4)$$

$$p(h_j = 1 | \mathbf{v}, \theta) = g\left(b_j + \sum_{j=1}^J W_{ij} h_j\right) \quad (5)$$

where  $g(\cdot)$  represents the logistic function given by

$$g(x) = \frac{1}{1 + e^{-x}}. \quad (6)$$

The weight matrix  $\mathbf{W}$  can be optimized by the contrastive divergence (CD) [39]. It is clear that RBM can reconstruct the input data by using the information in the hidden units. The information in the hidden units is the learned features from the input data. The reconstruction performance depends on whether the hidden units retain enough information about the input data.

Learning one layer at a time is relatively limited. The hidden units of RBM can be fed into another RBM to model the significant dependences between the hidden units of the earlier RBM. More specifically, stacking RBMs one atop another can progressively capture the hierarchical representation of the input data, which is called DBN. In this case, DBN can be employed as a means for dimensionality reduction on high-dimensional data. Fig. 1 shows an implementation of a two-layer DBN.

In this paper, DBN is used to extract discriminative features in a spectral domain from an HSI. Because of deep non-linear transformation, DBN can effectively extract high-level

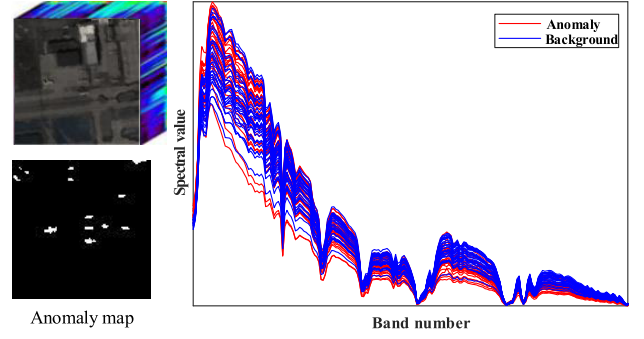


Fig. 2. Illustration of spatial and spectral information in an HSI.

representation of spectra to further capture the underlying characteristics in the spectral domain.

### III. PROPOSED METHOD

Given an HSI, anomalies are considered as interesting targets which are quite different from their surrounding pixels by exhibiting two characteristics: significantly different spectral signatures from their neighboring BKG or the global BKG and their presence in small area compared to the local homogeneous BKG.

However, in practical applications, the targets usually have spectral signatures similar to spectral signatures of their surrounding BKG or the global BKG, as shown in Fig. 2. Furthermore, anomalies have poor structural information from the perspective of the spatial domain. Especially, an anomaly that appears as a single pixel may not contain structural information in the spatial context. In reality, if an anomaly is detected from only one aspect (spectral or spatial), the detection result will be unsatisfactory. Therefore, we employ a joint spectral and spatial approach by linearly combining two detection maps as

$$D = \omega_1 D_{\text{spectral}} + \omega_2 D_{\text{spatial}} \quad (7)$$

where  $\omega_1$  and  $\omega_2$  are the balance parameters. Without any prior knowledge, we do not know whether the anomalies in the data sets are present in the form of pixels or an area with structural information. Therefore, we set  $\omega_1 = \omega_2 = 0.5$  inspired by [23], which means that the spectral features and spatial features have the same contribution to anomaly detection. Besides, these default parameter settings provide stable and acceptable performance for all data sets tested in our experiments. If the anomalies in the data set have structural and area information, the weight of spatial features, i.e.,  $\omega_2$ , should be adjusted to a larger value. On the contrary, the weight of spectral features, i.e.,  $\omega_1$ , should be adjusted to a larger value. In other words, these two set parameters complement each other in terms of spectral and spatial features. The schematic of the proposed SSFE is given in Fig. 3. We first illustrate a detection process based on spectral feature extraction.

Formally, let  $\mathbf{Y} \in \mathbb{R}^{L \times mn}$  denote an HSI with  $L$  spectral bands and  $mn$  pixels in the spatial domain.  $\mathbf{Y}$  may be interpreted as  $mn$  vectors with  $L$  dimensions, i.e.,  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{mn}]$ . It can also be interpreted as a collection of  $L$  2-D images of size  $m \times n$ , i.e.,  $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_L]$ .

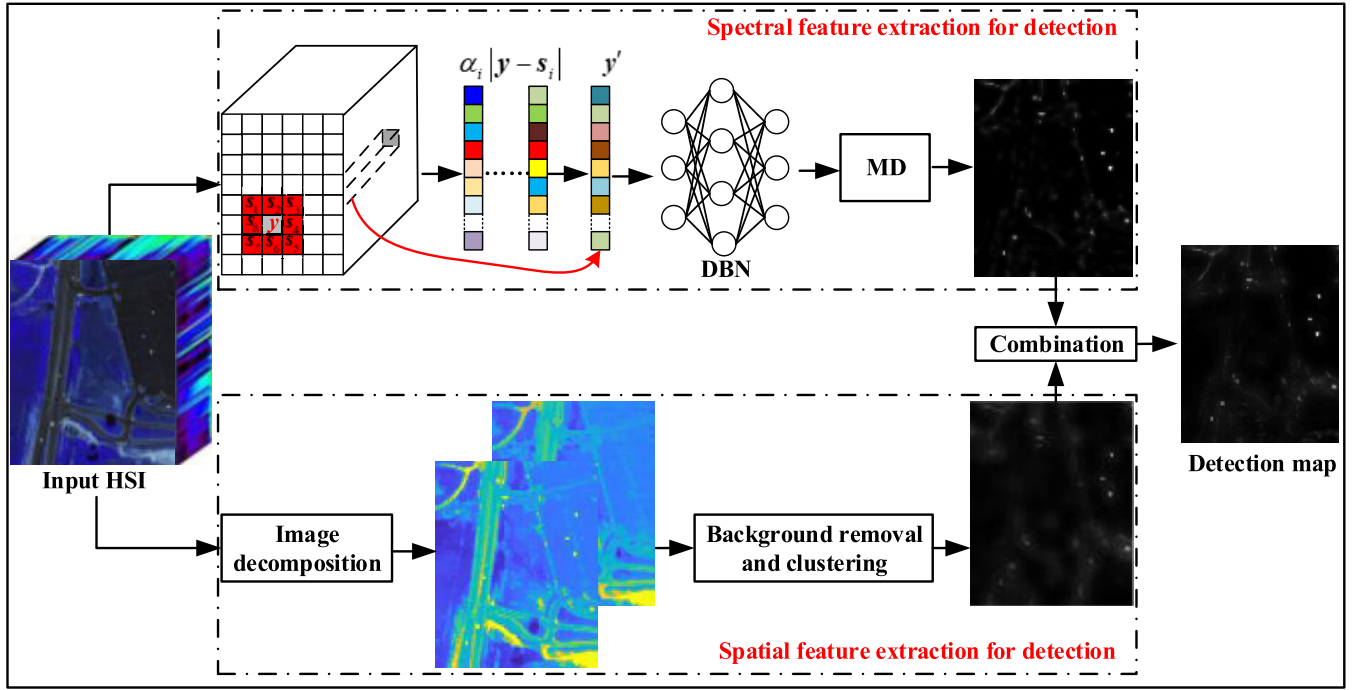


Fig. 3. Overview of the proposed spectral-spatial anomaly detection (SSFE) algorithm.

#### A. Spectral Feature Extraction for Anomaly Detection

This section focuses on anomalies corresponding to a few pixels even a single pixel. As mentioned above, the spectral feature extraction for anomaly detection is based on the assumption that the anomalous pixels are rarely present and all (or most) of the samples around the anomalies belong to BKG. Based on this fact, we introduce a suppression function to construct a discriminative feature space. The current vector  $\mathbf{y} \in \mathbb{R}^{L \times 1}$  serves as a center, and there are eight samples around  $\mathbf{y}$  represented by  $[s_1, s_2, \dots, s_8]$ , as shown in a  $3 \times 3$  window in Fig. 3. Then, the absolute value of the difference vector between the current vector and its surroundings is calculated, i.e.,  $[|\mathbf{y} - s_1|, |\mathbf{y} - s_2|, \dots, |\mathbf{y} - s_8|]$ . This represents the absolute distance between the current pixel and its surrounding pixels. In order to further increase this distance, we introduce a nonnegative parameter which depends on the absolute distance, defined as

$$\alpha_i = 1 - e^{-10|\mathbf{y} - s_i|}, \quad i = 1, 2, \dots, 8. \quad (8)$$

Fig. 4 shows the shape of  $1 - e^{-10|t|}$ . It can be observed that when  $|t|$  is relatively large, the value of  $1 - e^{-10|t|}$  is approximate to 1. When  $|t|$  is small, the value of  $1 - e^{-10|t|}$  decreases rapidly.

Then, each distance  $|\mathbf{y} - s_i|$  is reconstructed by multiplying  $\alpha_i$ , i.e.,  $\alpha_i |\mathbf{y} - s_i|$ . In this case, if  $|\mathbf{y} - s_i|$  is large, the reconstructed distance is still large. Otherwise, the reconstructed distance will be suppressed. Finally, the current vector  $\mathbf{y}$  is replaced by the relatively distinctive vector as follows:

$$\mathbf{y}' = \frac{1}{8} \sum_{i=1}^8 \alpha_i |\mathbf{y} - s_i| = \frac{1}{8} \sum_{i=1}^8 (1 - e^{-10|\mathbf{y} - s_i|}) |\mathbf{y} - s_i|. \quad (9)$$

There are four situations in the constructed discriminative feature space as follows.

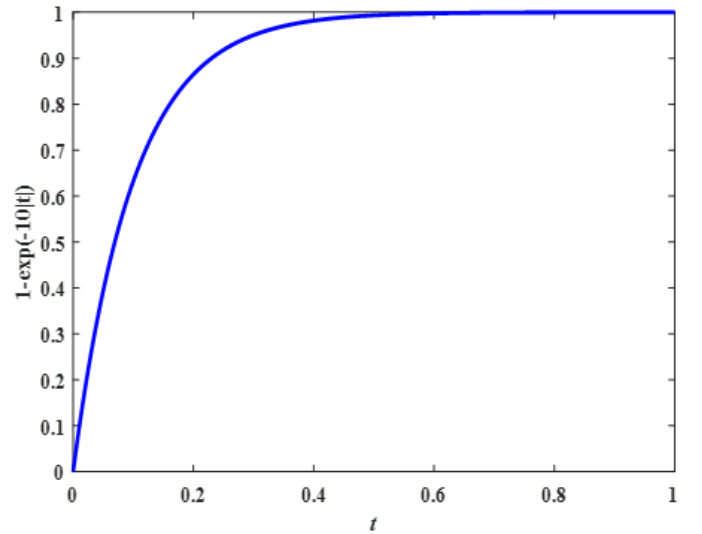


Fig. 4. Shape of function  $1 - e^{-10|t|}$ .

- 1) If the current pixel  $\mathbf{y}$  belongs to BKG and its surrounding pixels all belong to BKG too, the constructed vector  $\mathbf{y}'$  is close to 0.
- 2) If the current pixel  $\mathbf{y}$  belongs to BKG but one or two pixels of its surroundings are anomalies, the constructed vector  $\mathbf{y}'$  is still small.
- 3) If the current pixel  $\mathbf{y}$  is an anomaly and its surrounding pixels all belong to BKG, the constructed vector  $\mathbf{y}'$  is relatively large.
- 4) If the current pixel  $\mathbf{y}$  is an anomaly and one or two pixels of its surroundings are anomalies too, the constructed vector  $\mathbf{y}'$  is still relatively large.

All these situations are based on the assumption that the anomalous pixels are rarely present and all (or most) of



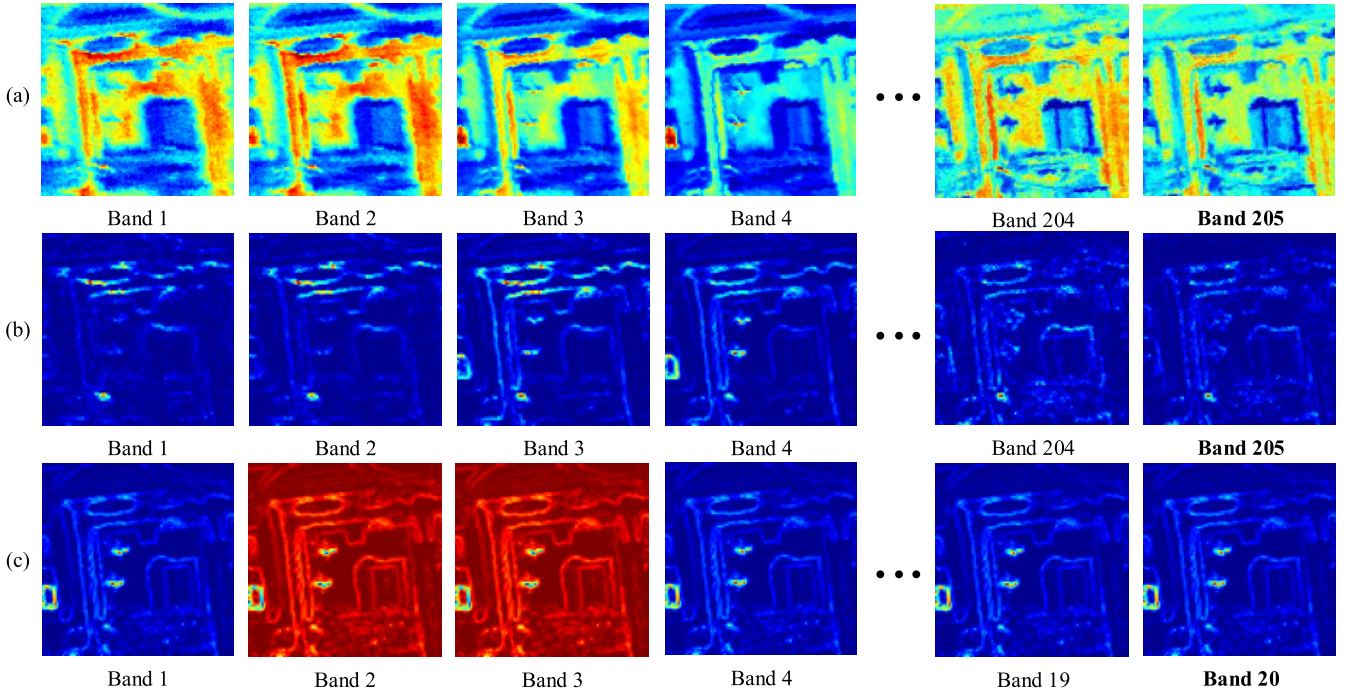


Fig. 5. Intermediate results of the spectral domain in visual. (a) Bands in the original HSI  $\mathbf{Y}$ . (b) Bands in the suppressed HSI  $\mathbf{Y}'$ . (c) Bands in the extracted feature matrix by DBN. It is noted that there are 205 spectral bands in  $\mathbf{Y}$  and  $\mathbf{Y}'$  and only 20 spectral bands in the extracted feature matrix.

the samples around anomalies are assumed to belong to BKG. In the situations of (1) and (2), the parameter  $\alpha_i$  can enforce  $|\mathbf{y} - \mathbf{s}_i|$  to close to 0. In the situations of (3) and (4), the parameter  $\alpha_i$  is relatively near to 1, and thus  $|\mathbf{y} - \mathbf{s}_i|$  is still large. Therefore, the separability between anomalies and BKG is further increased. The previous operation is conducted on all the pixels of  $\mathbf{Y}$  to produce a new data matrix  $\mathbf{Y}' = [\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_{mn}]$ . It is apparent that the suppressed data set is a relatively discriminative data space. Therefore, the reconstructed data space has positive influence on the performance of anomaly detection. Then, we utilize DBN to extract deep and underlying features of spectral signatures from  $\mathbf{Y}' = [\mathbf{y}'_1, \mathbf{y}'_2, \dots, \mathbf{y}'_{mn}]$ .

The traditional feature extraction methods, such as PCA and ICA, cannot extract deep features. Here, we employ DBN to extract deep discriminative features which can better describe anomalies without any prior knowledge. In detail, we use DBN as a nonlinear feature extraction algorithm to transform the high-dimensional data  $\mathbf{Y}'$  into a low-dimensional set of features. This process can be considered as dimensionality reduction. To overcome the overfitting problem, we employ a DBN with only two hidden layers. As mentioned above, the input data are the 1-D vector from  $\mathbf{Y}'$  and there are  $mn$  vectors. Let  $\mathbf{h}^{(1)}$  denote the first hidden layer and  $\mathbf{h}^{(2)}$  denote the second hidden layer. Each hidden layer can get higher order features from the lower layer through learning. The first layer is learned for extracting refined features from the input vector  $\mathbf{y}'$  and the extracted features are preserved in  $\mathbf{h}^{(1)}$  as follows:

$$\mathbf{h}^{(1)} = g \left( a_i + \sum_{i=1}^K W_{ij} y'_i \right) \quad (10)$$

where  $W_{ij}$  denotes the weight between the input data (visible unit  $i$ ) and hidden unit  $j$ ,  $a_i$  is the bias term of the input data, and  $K$  is the total number of units in the visible layer. Then, the features in  $\mathbf{h}^{(1)}$  are the visible data of the second layer and the extracted abstract features are kept in  $\mathbf{h}^{(2)}$ , defined as

$$\mathbf{h}^{(2)} = g \left( b_j + \sum_{j=1}^J W_{jc} h_j^{(1)} \right) \quad (11)$$

where  $W_{jc}$  denotes the weight between the first hidden unit  $j$  and the second hidden unit  $c$ ,  $b_j$  represents the bias term of the first hidden layer, and  $J$  is the total number of units in the first hidden layer. Fig. 5 shows the intermediate results. It can be observed that the anomalies are not so obvious in the original HSI  $\mathbf{Y}$ , which means that the spectral response values of anomalies are smaller than that of BKG. The suppressed data  $\mathbf{Y}'$  can enhance the spectral response values of anomalies in some bands, but there still exists useless bands. The DBN is conducted on  $\mathbf{Y}'$  to extract effective features that can directly enhance the spectral response values of anomalies. Here, the features are extracted based on a two-layer DBN. It is noted that the number of spectral bands is decreased from 205 to 20. Therefore, DBN can reduce dimensionality while extracting effective feature matrix.

In the following experiments, we employ a two-layer DBN with  $C$  units in the second hidden layer. Therefore, the final extracted feature matrix via DBN can be represented by  $\mathbf{h}^{(2)} = [\mathbf{h}_1^{(2)}, \mathbf{h}_2^{(2)}, \dots, \mathbf{h}_{mn}^{(2)}]$ , which has  $mn$  vectors with  $C$  dimensions. It means that the HSI has reduced dimensions from  $L$  to  $C$  by DBN. Then, we can utilize an MD method to detect anomalies in  $\mathbf{h}^{(2)}$ . The detection result can be represented as

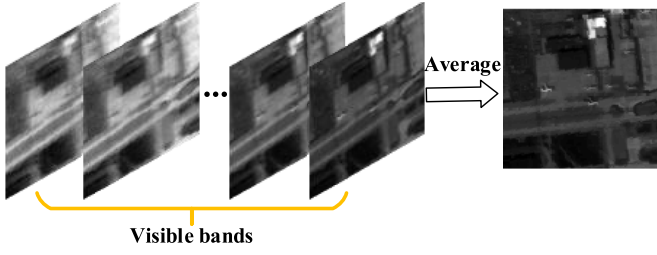


Fig. 6. Visible bands' fusion result by the averaging method.

follows:

$$D_{\text{spectral}} = (\mathbf{h}_i^{(2)} - \mu)^T \Gamma^{-1} (\mathbf{h}_i^{(2)} - \mu), \quad i = 1, \dots, mn \quad (12)$$

where  $\mathbf{h}_i^{(2)} \in \mathbb{R}^{C \times 1}$  is the  $i$ th vector in  $\mathbf{h}^{(2)}$ , and  $\mu \in \mathbb{R}^{C \times 1}$  and  $\Gamma \in \mathbb{R}^{C \times C}$  are the mean and the covariance matrix of  $\mathbf{h}^{(2)}$ .

### B. Spatial Feature Extraction for Anomaly Detection

The previous anomaly detection was based on the spectral feature extraction method which focused on anomalies appeared as a few pixels even a single pixel. Different from the previous method, this section assumes that anomalies in real remote sensing HSIs generally appear as a small number of pixels with high correlation among them compared with BKG. In addition, it focuses on anomalies with structural and area information in the spatial domain. It is time-consuming to detect spatial information of anomalies in high-dimensional HSIs. Besides, human visible wavelength is approximately ranging from 400 to 760 nm. Therefore, we extract spatial features of anomalies only in the visible wavelength range.

In the visible wavelength range, the adjacent bands are highly correlated and contain redundant information [40]. First, the bands in the visible wavelength range are fused together to find a new image for three reasons: 1) to remove the noisy pixels; 2) to overcome redundant information; and 3) to extract useful spatial information. The most commonly used and simple image fusion method is the averaging method, that is

$$\mathbf{S} = \frac{\sum_{l=1}^{L_v} \mathbf{Y}_l}{L_v} \quad (13)$$

where  $\mathbf{S} \in \mathbb{R}^{m \times n}$  is the average band and  $L_v$  refers to the total number of bands in the visible wavelength range. Fig. 6 shows the fusion results of the visible wavelength bands by the averaging method, which proves that this step can remove noise and redundant information to a certain degree.

Subsequently, the spatial features of anomalies are separated from BKG by performing a morphological attribute filter on  $\mathbf{S}$ . The morphological attribute filter proposed in [36] is an adaptive filter, which can effectively describe the spatial structure of HSIs by setting different attribute values to preserve or remove the connected components. Due to its advantages of size feature, it has been widely used in HSI classification [37]. The morphological attribute filter is mainly based on attribute openings and closings. We test on a large number of HSIs to study the influence of attribute filtering on HSIs. Two examples

are shown in Fig. 7. The spatial image is referred to as  $\mathbf{S}$  that is obtained by averaging the visible bands. It can be observed that the closing and opening operations in an attribute filter are able to remove the dark and bright connected components in small area, respectively. For anomaly detection, we want to preserve the objects in small areas. Therefore, the differential map is obtained as follows:

$$A = |\mathbf{S} - A^o(\mathbf{S})| + |\mathbf{S} - A^c(\mathbf{S})| \quad (14)$$

where  $A^o(\mathbf{S})$  and  $A^c(\mathbf{S})$  represent the attribute openings and closings extracted from  $\mathbf{S}$ , respectively. The term  $|\mathbf{S} - A^o(\mathbf{S})|$  preserves the bright connected components in small area, and  $|\mathbf{S} - A^c(\mathbf{S})|$  preserves the dark connected components in small area. It can be seen from the differential image in Fig. 7 that (14) can effectively preserve objects in small areas.

As previously mentioned, this section focuses on anomalies in small areas compared with BKG as well as spatial correlation among adjacent pixels. Here, we use a GF to consider the prior knowledge of local smoothness in the differential map due to good performance and fast speed of GF. The final spatial detection result is given by

$$D_{\text{spatial}} = \bar{a}_i A_i + \bar{b}_i \quad (15)$$

where  $(\bar{a}, \bar{b})$  are the average values of linear coefficients in different local windows. The linear coefficients can be solved by minimizing the squared difference between the input and the filtering output in a local window. For the proposed SSFE in the spatial domain, the filter size and blur degree of GF are set to 3 and 0.5 for all test data sets, respectively, learned from experience.

Kang *et al.* [22] employed PCA to preserve three principal components (PCs) and then utilized a morphological attribute filter to extract spatial features on these three PCs, while we conduct a morphological attribute filter only on one image obtained by fusing the visible bands. Besides, we utilize a GF rather than a domain transform recursive filter (DTRF) because the GF is famous for its outstanding performance in terms of speed and edge preservation.

Once the initial maps are obtained based on spectral and spatial features [i.e., (12) and (15)], the final detection map is calculated by (7).

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, the proposed SSFE method is evaluated on three real hyperspectral data sets captured at different scenes, which are listed as follows.

### A. Data Sets

1) *Airport-Beach-Urban Data Set*: The Airport-Beach-Urban (ABU) data set was made available online.<sup>1</sup> The detailed acquisition procedure has been described in [22]. These sample images contain  $100 \times 100$  or  $150 \times 150$  pixels with the corresponding references. The airplane and urban scenes were captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, and most of the beach

<sup>1</sup><http://xudongkang.weebly.com/>

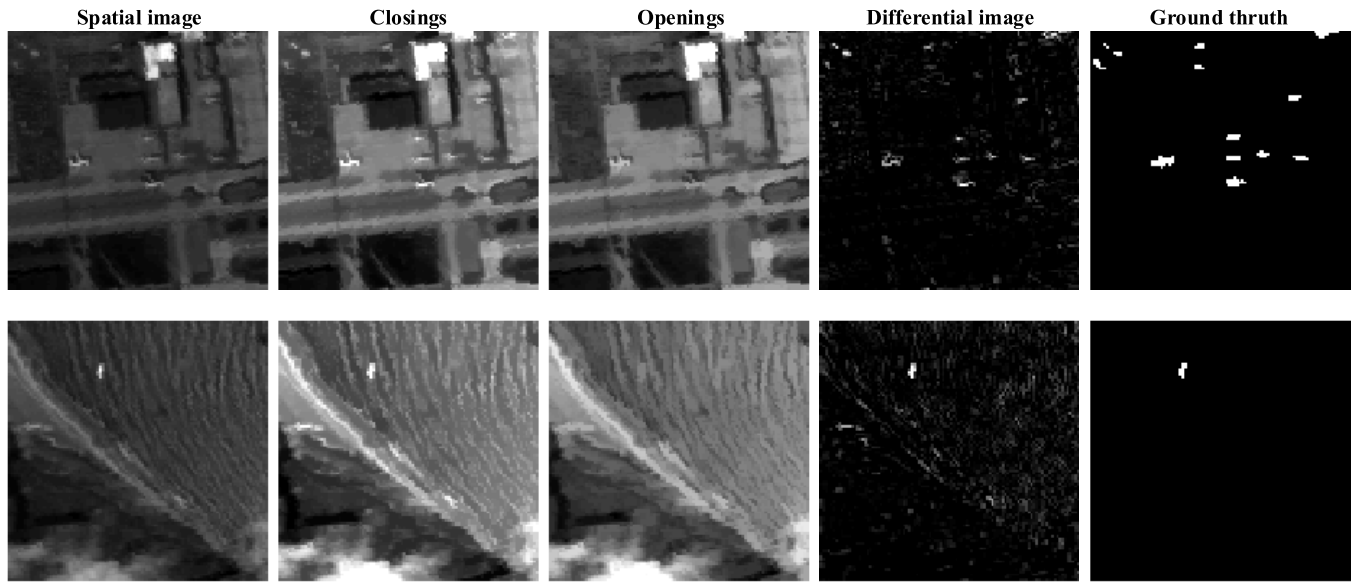


Fig. 7. Intermediate results of the spatial domain in visual. Note that the spatial image refers to  $S$  that is obtained by the average of the visible bands.

scenes are also captured by the AVIRIS sensor. One of the beach scenes was captured by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor. The reference maps are manually labeled by the Environment for Visualizing Images (ENVI) software [22].

2) *San Diego Data Set*: Another real HSI is widely used in related publications, which is also collected by the AVIRIS sensor. The scene is part of the San Diego airport area, CA, USA. This image contains  $100 \times 100$  pixels in each band and each pixel contains 189 dimensions, in which the noisy bands have been removed. All the compared algorithms are used to detect three aircrafts in this data set.

3) *HYDICE Data Set*: This data set was acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor over an urban area, CA, USA. A subimage of size  $80 \times 100$  is widely used in many publications. Among them, 21 pixels are considered as anomalies, which are mainly cars. Each pixel contains 175 bands after removing the noisy and water vapor absorption bands.

### B. Compared Methods

To verify the effectiveness of the proposed method, typical methods, the RX detector and the LRX detector are employed for comparison purpose. We also compare with the state-of-the-art algorithm called CRD. For the LRX detector, the window size is a significant parameter. There is big difference in detection performance with different window sizes. In our experiments, the optimal window size of LRX for each data set is selected according to the detection performance. For the CRD method, the optimal regularization parameter is set to  $10^{-6}$  according to the original literature [19].

The performance of the proposed SSFE method is evaluated by the most widely used metric, the receiver operating characteristic (ROC), and the area under the curve (AUC) metric [41]. The ROC curve, with its origins in communication

theory, describes detection performance with the true positive rate (TPR) plotted as a function of the false positive rate (FPR) at each threshold. TPR defines the number of correct positive detections occur among all positive samples and FPR defines the number of incorrect positive detections occur among all negative samples available. This curve characterizes the tradeoff between the TPR and the FPR [42], [43]. Specifically, the AUC metric depends only on the order of the pixels rather than absolute detection values. The ideal value of AUC is 1.

### C. Parameter Settings

In the spectral feature extraction procedure by DBN, we have several parameters that need to be adjusted: the number of hidden layers, the number of hidden units and the batch size. The number of hidden layers and units can determine the quality of learned features, which play significant roles in the performance of anomaly detection. In the experiments, we find that the batch size also has an influence on the detection performance in the spectral domain. Concretely, we evaluate the impact of different parameter settings on the postprocessing, i.e., AUC, with the others fixed. It is observed from Fig. 8(a) that the performance of the proposed method in the spectral domain decreases when the number of hidden layers is larger than 2 for all test data sets. Moreover, the performance is degraded when the number of hidden layers is smaller than 2. These experiments show that the detection performance can achieve the best results when employing two hidden layers. In general, a deep architecture can make full use of HSI characteristics. However, as the training samples are limited for HSIs, the overfitting problem often occurs if the network is too deep that may lead to unsatisfactory detection performance, which had also been shown in [31] and [34]. Thus, the number of hidden layers is set as 2 for all HSIs. In the first hidden layer, we set the number of hidden units as 70 for all the data sets. The number of hidden units for



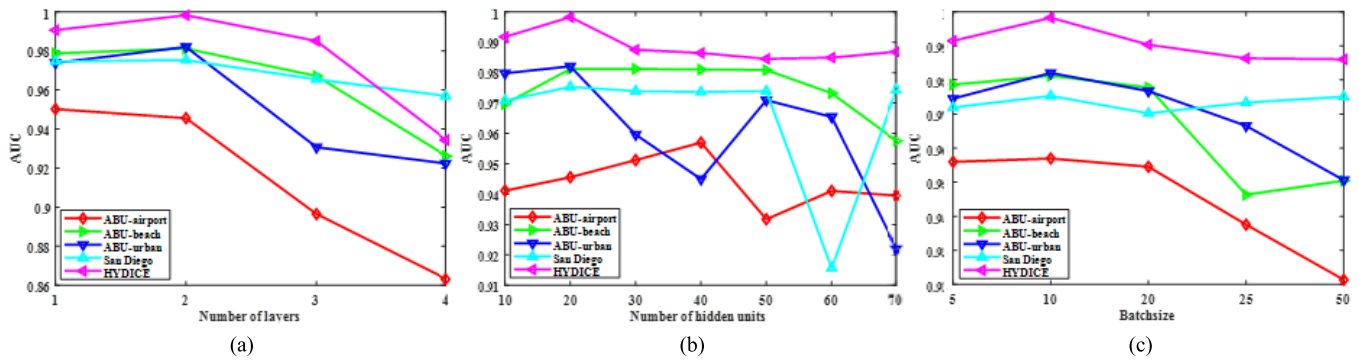


Fig. 8. Parameter sensitivity analysis of the influence on different data sets. (a) Number of hidden layers. (b) Number of hidden units. (c) Batch size.

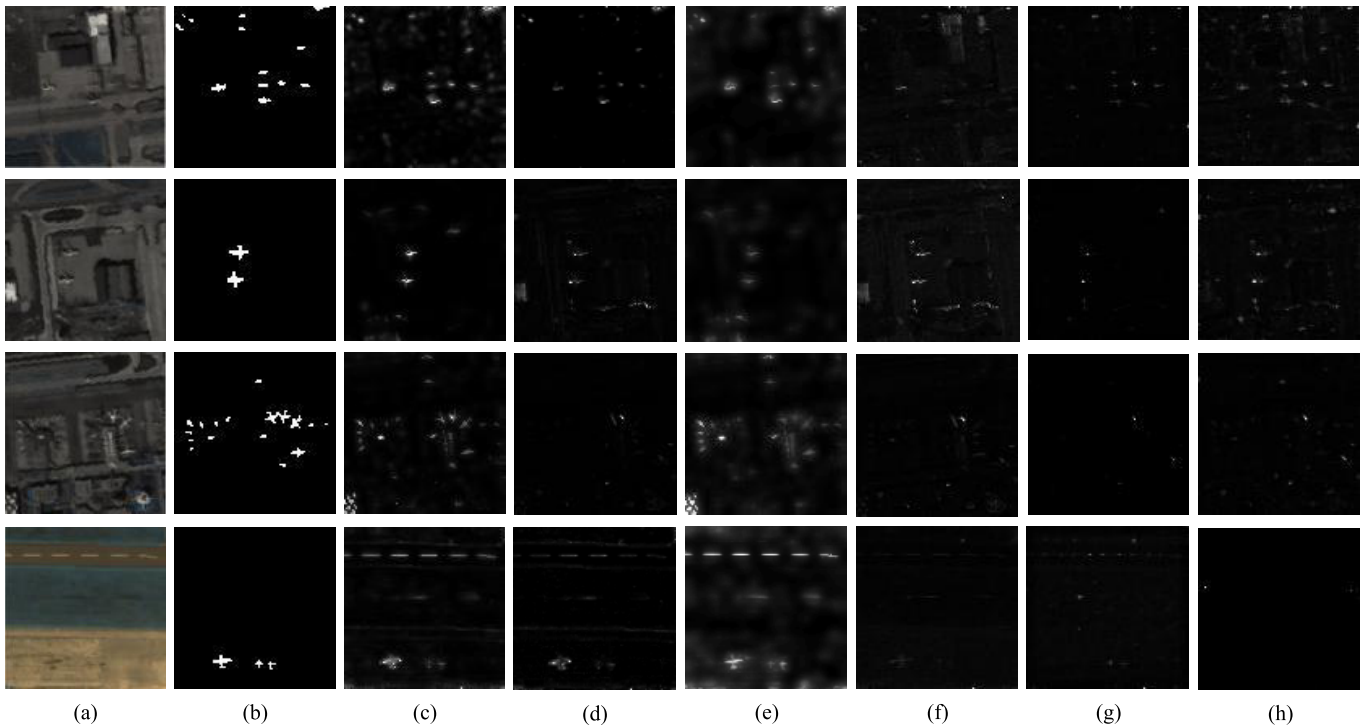


Fig. 9. Detection maps of the compared methods. (a) Color composites of airport scene. (b) Reference map. (c) SSFE. (d) SSFE-Spe. (e) SSFE-Spa. (f) RX. (g) LRX. (h) CRD.

the second hidden layer should be optimized under the precondition of the aforementioned parameter. We systematically vary the parameter setting each by one and report the AUC on the final version. Fig. 8(b) studies the effect of the proposed method in the spectral domain with different numbers of hidden units. We can see that the best AUC is achieved when the number of hidden units is 20 for most of HSIs except ABU-airport for which the number of hidden units is 40. Next, the effect of batch size on detection performance is evaluated with the fixed hidden layers and units. It can be seen from Fig. 8(c) that, when the batch size is ten, the AUC achieves the best value for all data sets. With a smaller or larger value, the results are not ideal.

#### D. Detection Results

For the ABU data set, the reference map and detection results in visual are shown in Figs. 9–11. The proposed

SSFE method causes the smallest visual difference between the reference and detection maps compared with RX, LRX, and CRD methods. More specifically, SSFE can detect the anomalies of different sizes, for example, the second and fourth rows in Fig. 9. The airport objects of different sizes in these HSIs can be well detected by the SSFE method. However, the RX, LRX, and CRD methods cannot detect the anomalies of different sizes. Besides, SSFE is applicable to different scenes, including airport, beach, and urban scenes, which is its another major advantage.

The corresponding AUC values are reported in Tables I–III. The AUC values of the proposed SSFE method are approximate to the ideal value 1 for all the scenes. The average AUC value is 0.98774 for airport scene and it is much larger than the RX, LRX, and CRD algorithms. For the beach and urban scenes, the proposed SSFE method can still achieve the best AUC values.



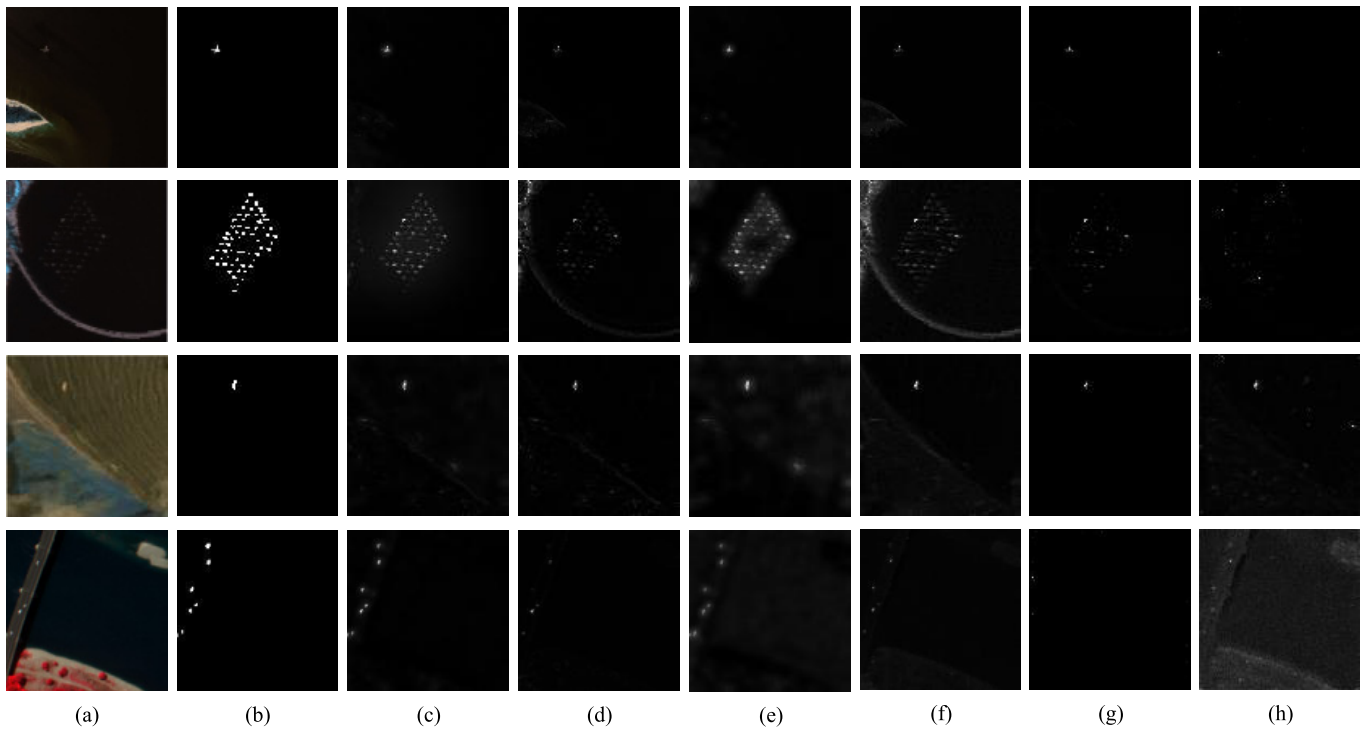


Fig. 10. Detection maps of the compared methods. (a) Color composites of beach scene. (b) Reference map. (c) SSFE. (d) SSFE-Spe. (e) SSFE-Spa. (f) RX. (g) LRX. (h) CRD.

TABLE I  
AUC VALUES OF THE COMPARED METHODS  
ON THE ABU-AIRPORT SCENE

HSIs	SSFE	SSFE-Spe	SSFE-Spa	RX	LRX	CRD
airport-1	<b>0.98709</b>	0.94268	0.98688	0.82259	0.96630	0.96071
airport-2	<b>0.98848</b>	0.96203	0.98564	0.84059	0.94912	0.92071
airport-3	<b>0.98458</b>	0.94207	0.98368	0.92899	0.93464	0.94722
airport-4	<b>0.99083</b>	0.98112	0.98714	0.95206	0.88730	0.82168
Average	<b>0.98774</b>	0.95697	0.98583	0.88606	0.93434	0.91258

TABLE II  
AUC VALUES OF THE COMPARED METHODS ON THE ABU-BEACH SCENE

HSIs	SSFE	SSFE-Spe	SSFE-Spa	RX	LRX	CRD
beach-1	<b>0.99992</b>	0.99602	0.99991	0.98281	0.99777	0.96720
beach-2	<b>0.99769</b>	0.93499	0.99764	0.91060	0.97532	0.92214
beach-3	<b>0.99998</b>	0.99987	0.99989	0.99985	0.99997	0.99914
beach-4	<b>0.99874</b>	0.99404	0.99859	0.95335	0.95200	0.90093
Average	<b>0.99908</b>	0.98123	0.99901	0.96165	0.98127	0.93009

In order to validate that both the spectral and spatial components in the proposed SSFE method have positive influence on anomaly detection, we illustrate the corresponding detection maps and AUC values of these two parts. In the experiments, the spectral component in the SSFE method is called SSFE-Spe and the spatial component is called SSFE-Spa. It can be observed from Figs. 9–11 that SSFE-Spe can detect anomalies appeared as single or several pixels, which complies with our assumption that the anomalous pixels are rarely

TABLE III  
AUC VALUES OF THE COMPARED METHODS ON THE ABU-URBAN SCENE

HSIs	SSFE	SSFE-Spe	SSFE-Spa	RX	LRX	CRD
urban-1	<b>0.99925</b>	0.99342	0.99865	0.99064	0.99671	0.99279
urban-2	<b>0.99896</b>	0.99645	0.99769	0.99463	0.90233	0.93042
urban-3	<b>0.99890</b>	0.97317	0.99876	0.95119	0.96032	0.93970
urban-4	<b>0.99726</b>	0.98379	0.99671	0.98170	0.91912	0.95163
urban-5	<b>0.99418</b>	0.96353	0.99340	0.96932	0.93212	0.93045
Average	<b>0.99771</b>	0.98207	0.99704	0.97750	0.94212	0.94899

present and all (or most) of the samples around the anomaly belong to BKG in Section III-C. We can also observe from Figs. 9–11 that SSFE-Spa can detect anomalies with structural information, which depends on the fact that the adjacent pixels have high correlations with each other. These two priors can fulfill a reasonable spectral-spatial constraint, and as a result, the detection performance can be improved. The AUC values of SSFE-Spe and SSFE-Spa are all larger than that of the RX, LRX, and CRD methods, which illustrate that both spectral and spatial components proposed in SSFE have superior performance to detect anomalies.

For San Diego and HYDICE data sets, the detection results of the compared methods are shown in Figs. 12 and 13 and the corresponding AUC values are listed in Tables IV and V. It can be observed that the detection maps obtained by the proposed SSFE method are closer to the reference maps than those obtained by the other methods. Especially, the detection map of San Diego data set contains less BKG compared with the RX method, which illustrates that the proposed method can

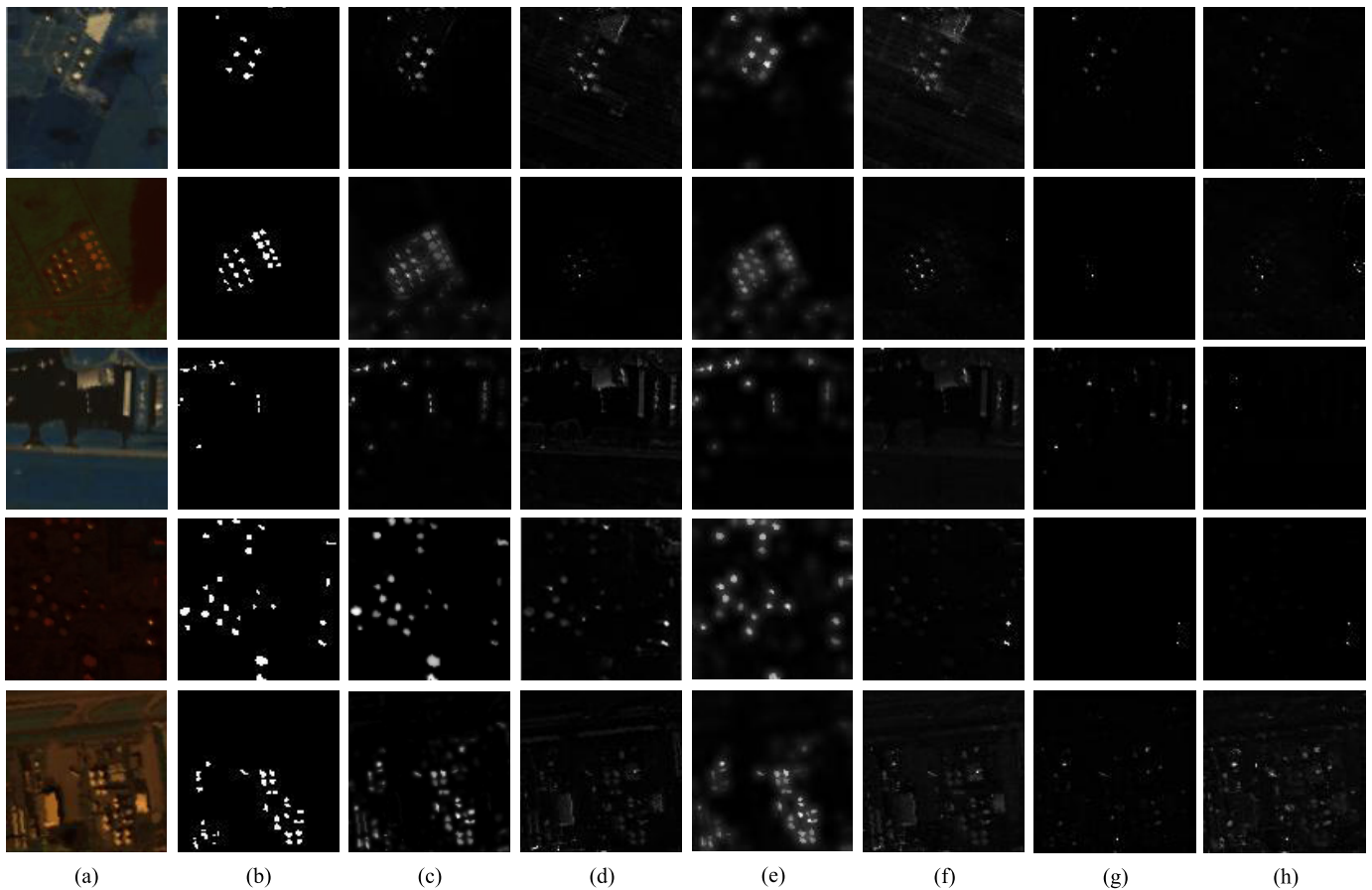


Fig. 11. Detection maps of the compared methods. (a) Color composites of urban scene. (b) Reference map. (c) SSFE. (d) SSFE-Spe. (e) SSFE-Spa. (f) RX. (g) LRX. (h) CRD.

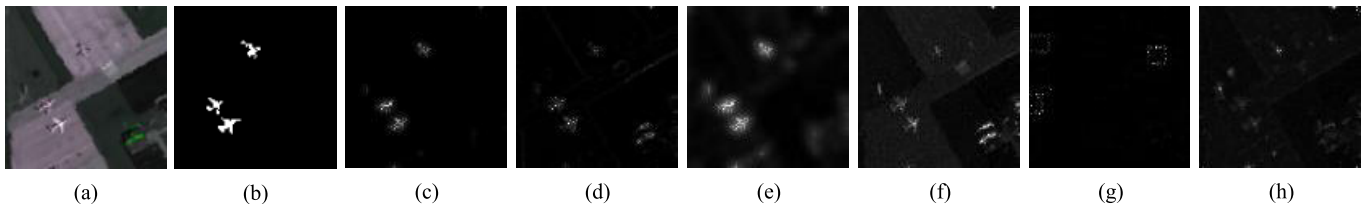


Fig. 12. Detection maps of the compared methods. (a) Color composites of San Diego image. (b) Reference map. (c) SSFE. (d) SSFE-Spe. (e) SSFE-Spa. (f) RX. (g) LRX. (h) CRD.

TABLE IV

AUC VALUES OF THE COMPARED METHODS ON THE SAN DIEGO IMAGE

HSIs	SSFE	SSFE-Spe	SSFE-Spa	RX	LRX	CRD
San Diego	<b>0.99469</b>	0.97532	0.99449	0.95151	0.86330	0.96410

TABLE V

AUC VALUES OF THE COMPARED METHODS ON THE HYDICE DATA

HSIs	SSFE	SSFE-Spe	SSFE-Spa	RX	LRX	CRD
HYDICE	<b>0.99858</b>	0.99828	0.99810	0.97637	0.93407	0.96206

effectively suppress BKG. In addition, the AUC values listed in Tables IV and V also demonstrate the superior performance of the proposed SSFE method. Especially, the detection results based on spectral and spatial feature matrices have larger AUC values than the other compared methods. The complementary of spectral and spatial domains can further improve the detection performance.

With the best parameters, Fig. 14 shows the ROC performance of different anomaly detectors for San Diego and

HYDICE data. From the results, it is obvious that the proposed SSFE is always superior to RX, LRX, PCAF, and CRD, because the curves of the detection results remain over those compared methods. It confirms that our method has promising results for these two scenes. In general, the detection results are consistent with the AUC values reported in Tables IV and V.

It can be observed from Tables I–V, except for the HYDICE data set, SSFE-Spa has higher accuracy than SSFE-Spe for

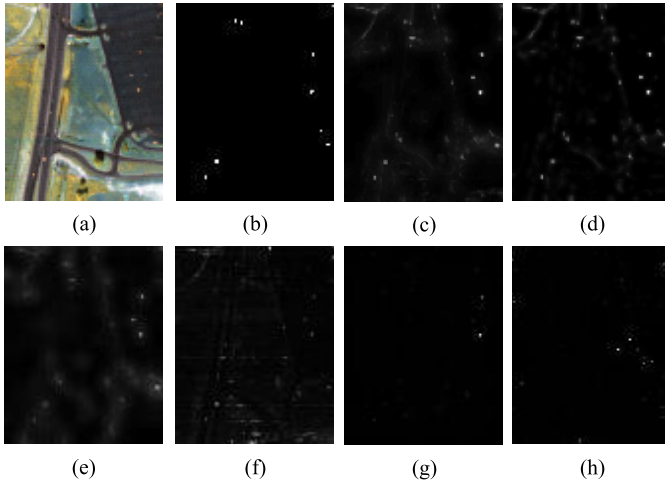


Fig. 13. Detection maps of the compared methods. (a) Color composites of HYDICE. (b) Reference map. (c) SSFE. (d) SSFE-Spe. (e) SSFE-Spa. (f) RX. (g) LRX. (h) CRD.

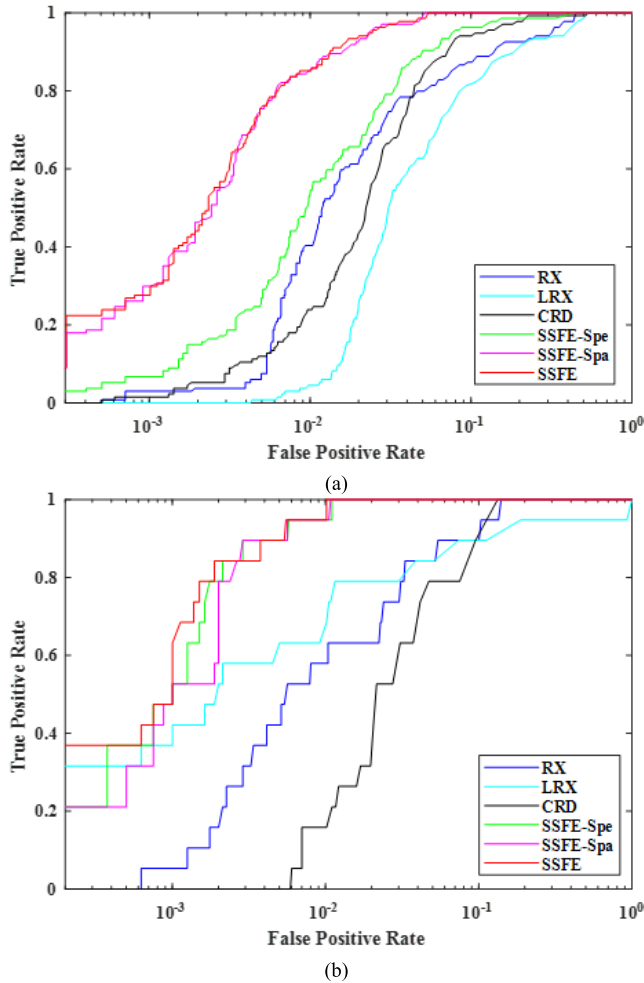


Fig. 14. ROC curves of the compared methods on (a) San Diego and (b) HYDICE.

all the other data sets, because all the data sets except the HYDICE data set contain the structural and area information of the anomalies. From the current results, SSFE-Spe is more

suitable for the data sets containing point anomalies, which is in line with our assumption in the spectral feature extraction part that the anomalous pixels are rarely present and all (or most) of the samples around anomalies belong to BKG.

## V. CONCLUSION

This paper develops an approach to hyperspectral anomaly detection, which depends on both spectral and spatial feature extraction aspects. In the proposed method, we first extract deep spectral features via DBN from a suppressed data space. Then, the MD-based detector is utilized to detect anomalies appeared as a single pixel or a few pixels. Additionally, in order to extract anomalies appeared as small area and consider local smoothness prior, we introduce a simple differential operation to remove BKG and employ the GF to preserve the edge information of anomalies. The final detection map is achieved by combining the spectral and spatial results. Comparative analyses validate that the proposed SSFE method presents superior performance on several hyperspectral data sets captured from different scenes. In the future, we intend to optimize DBN to further reduce the false-alarm rate of the proposed SSFE method.

## REFERENCES

- [1] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [2] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Norwell, MA, USA: Kluwer, 2003.
- [3] C.-I. Chang, *Real-Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection*. New York, NY, USA: Springer, 2016.
- [4] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [5] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.
- [6] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [7] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.
- [8] S. Matteoli, T. Veracini, M. Diani, and G. Corsini, "A locally adaptive background density estimator: An evolution for RX-based anomaly detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 323–327, Jan. 2014.
- [9] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [10] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [11] Y. Yuan, D. Ma, and Q. Wang, "Hyperspectral anomaly detection by graph pixel selection," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3123–3134, Dec. 2016.
- [12] W. Xie, Y. Shi, Y. Li, X. Jia, and J. Lei, "High-quality spectral-spatial reconstruction using saliency detection and deep feature enhancement," *Pattern Recognit.*, vol. 88, pp. 139–152, Apr. 2019.
- [13] B. Du and L. Zhang, "A discriminative metric learning based anomaly detection method," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6844–6857, Nov. 2014.



- [14] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 2919–2930, May 2018.
- [15] H. Goldberg, H. Kwon, and N. M. Nasrabadi, "Kernel eigenspace separation transform for subspace anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 581–585, Oct. 2007.
- [16] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [17] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, p. 289, Mar. 2016.
- [18] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.
- [19] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [20] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.
- [21] W. Xie, T. Jiang, Y. Li, X. Jia, and J. Lei, "Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, to be published. doi: [10.1109/TGRS.2018.2890212](https://doi.org/10.1109/TGRS.2018.2890212).
- [22] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [23] F. Li, X. Zhang, L. Zhang, D. Jiang, and Y. Zhang, "Exploiting structured sparsity for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4050–4064, Jul. 2018.
- [24] C. Zhao and L. Zhang, "Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Infr. Phys. Technol.*, vol. 92, pp. 166–176, Aug. 2018.
- [25] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.
- [26] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [27] W. Liao, A. Pizurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
- [28] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1581–1586, May 2019. doi: [10.1109/TNNLS.2018.2868836](https://doi.org/10.1109/TNNLS.2018.2868836).
- [29] Q. Wang, Z. Qin, F. Nie, and X. Li, "Spectral embedded adaptive neighbors clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1265–1271, Apr. 2019. doi: [10.1109/TNNLS.2018.2861209](https://doi.org/10.1109/TNNLS.2018.2861209).
- [30] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognit.*, vol. 63, pp. 371–383, Mar. 2017.
- [31] J. Li, B. Xi, Y. Li, Q. Du, and K. Wang, "Hyperspectral classification based on texture feature enhancement and deep belief networks," *Remote Sens.*, vol. 10, no. 3, p. 396, Mar. 2018.
- [32] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 1805–1812.
- [33] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [34] Y. Chen, X. Zhao, and X. Jia, "Spectral-spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.
- [35] N. Ma, Y. Peng, S. Wang, and P. H. W. Leong, "An unsupervised deep hyperspectral anomaly detector," *Sensors*, vol. 18, no. 3, p. 693, Feb. 2018.
- [36] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vis. Image Understand.*, vol. 64, no. 3, pp. 377–389, Nov. 1996.
- [37] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral-spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.
- [38] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
- [39] G. E. Hinton, "Training products of experts by minimizing contrastive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1711–1800, Aug. 2002.
- [40] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, to be published. doi: [10.1109/TGRS.2019.2904108](https://doi.org/10.1109/TGRS.2019.2904108).
- [41] C. Ferri, J. Hernández-Orallo, and P. Flach, "A coherent interpretation of AUC as a measure of aggregated classification performance," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 657–664.
- [42] C.-I. Chang, "Multiparameter receiver operating characteristic analysis for signal detection and classification," *IEEE Sensors J.*, vol. 10, no. 3, pp. 423–442, Mar. 2010.
- [43] C.-I. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ, USA: Wiley, 2013, ch. 3.



**Jie Lei** received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2006 and 2010, respectively.

He has been a Visiting Scholar with the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA, from 2014 to 2015. He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University and also a member of the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks. His research interests include image and video processing, computer vision, and customized computing for big-data applications.



**Weiying Xie** received the B.S. degree in electronic information science and technology from the University of Jinan, Jinan, China, in 2011, the M.S. degree in communication and information systems from Lanzhou University, Lanzhou, China, in 2014, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2017.

She is currently a Post-Doctoral Researcher and a Teacher with Xidian University. Her research interests include neural networks, deep learning, hyperspectral image processing, and high-performance computing.



**Jian Yang** received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2018, where he is currently pursuing the M.S. degree with the Image Coding and Processing Center, State Key Laboratory of Integrated Service Network.

His research interests include hyperspectral image, anomaly detection, and unsupervised learning.



**Yunsong Li** received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 1999 and 2002, respectively.

He joined the School of Telecommunications Engineering, Xidian University, in 1999, where he is currently a Professor. He is also the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Service Networks. His research interests include image and video processing and high-performance computing.



**Chein-I Chang** (S'81–M'87–SM'92–F'10–LF'17) received the B.S. degree in mathematics from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree in mathematics from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, the M.A. degree in mathematics from Stony Brook University, Stony Brook, NY, USA, in 1977, the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, in 1980 and 1982, respectively, and the Ph.D. degree in electrical engineering from the University of Maryland at College Park, College Park, MD, USA, in 1987.

He was a Visiting Research Specialist with the Institute of Information Engineering, National Cheng Kung University, Tainan, Taiwan, from 1994 to 1995, and a Distinguished Visiting Fellow/Fellow Professor from 2009 to 2010 both of which were sponsored by the National Science Council, Taiwan. He was the Distinguished Lecturer Chair with National Chung Hsing University, Taichung, Taiwan, from 2005 to 2006, sponsored by the Ministry of Education, Taiwan. He has been a Distinguished Chair Professor with National Chung Hsing University since 2014 and a Chair Professor with Providence University, Taichung, since 2012. He has been with the University of Maryland, Baltimore County, Baltimore, MD, USA, since 1987, where he

is currently a Professor with the Department of Computer Science and Electrical Engineering. He is holding the Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing, Dalian Maritime University, Dalian, China. He is also holding the Hua Shan Scholar Chair Professorship with Xidian University, Xi'an, China. He has seven patents on hyperspectral image processing. He has authored four books, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic, 2003), *Hyperspectral Data Processing: Algorithm Design and Analysis* (Wiley, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). He has edited two books, *Recent Advances in Hyperspectral Signal and Image Processing* in 2006 and *Hyperspectral Data Exploitation: Theory and Applications* (Wiley, 2007) and co-edited with A. Plaza a book *High Performance Computing in Remote Sensing* (CRC, 2007). His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang is a fellow of SPIE. He received the National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, Maryland. He was a Plenary Speaker for the Society for Photo-Optical Instrumentation Engineers (SPIE) Optics + Applications, Remote Sensing Symposium, in 2009. He was a Keynote Speaker with the User Conference of Hyperspectral Imaging 2010, Industrial Technology Research Institute, Hsinchu; the 2009 Annual Meeting of the Radiological Society of the Republic of China, 2009, Taichung; the 2008 International Symposium on Spectral Sensing Research in 2008; and the Conference on Computer Vision, Graphics, and Image Processing 2003, Kimen 2013, Nantou, Taiwan. He was an Associate Editor in the area of hyperspectral signal processing of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2001 to 2007. He was a Guest Editor of a special issue for the *Journal of High Speed Networks on Telemedicine and Applications* in 2000 and a Co-Guest Editor of another special issue on Broadband Multimedia Sensor Networks in Healthcare Applications of the *Journal of High Speed Networks on Telemedicine and Applications* in 2007. He is a Co-Guest Editor of special issues on High Performance Computing of Hyperspectral Imaging of the *International Journal of High Performance Computing Applications* in 2007, on Signal Processing and System Design in Health Care Applications of the *EURASIP Journal on Advances in Signal Processing* in 2009, and on Multispectral, Hyperspectral, and Polarimetric Imaging Technology of the *Journal of Sensors*. He is currently an Associate Editor of *Artificial Intelligence Research* and also on the editorial boards of the *Journal of High Speed Networks*, *Recent Patents on Mechanical Engineering*, the *International Journal of Computational Sciences and Engineering*, the *Journal of Robotics and Artificial Intelligence*, and the *Open Remote Sensing Journal*.