# Reweighted Nuclear Norm and Total Variation Regularization With Sparse Dictionary Construction for Hyperspectral Anomaly Detection

Xiaoyi Wang, Liguo Wang, Jiawen Wang, Kaipeng Sun, and Qunming Wang

***Abstract*—Hyperspectral anomaly detection is an important technique in the field of remote sensing image processing. Over the last few years, low rank and sparse matrix decomposition (LRSMD) has played an increasingly significant role in hyperspectral anomaly detection. The detection performance of the LRSMD-based anomaly detectors is primarily determined by prior constraints and the background dictionary construction method. To increase the detection accuracy, we proposed the reWeighted Nuclear Norm and total variation regularization with Sparse Dictionary construction for hyperspectral Anomaly Detection (WNNSDAD), which incorporated reweighted nuclear norm and total variation regularizations as the prior constraints into the LRSMD model, and constructed a sparse background dictionary without the need of clustering. Compared to the standard nuclear norm, the reweighted nuclear norm helped to overcome the challenge of an unbalanced penalty for a singular value and ensure a more effective low rank approximation. Simultaneously, total variation regularization was introduced as a piecewise smoothing constraint, which helped to maintain the spatial correlation of the hyperspectral image. Additionally, we proposed a background dictionary construction method, by which a relatively complete background dictionary could be obtained without clustering, and the background part could be represented more reliably. The experiments on seven real-world hyperspectral datasets show that in comparison to eight state-of-the-art anomaly detection methods, the proposed WNNSDAD method demonstrated greater accuracy.**

***Index Terms*—Hyperspectral anomaly detection, low rank and sparse matrix decomposition (LRSMD), reweighted nuclear norm, sparse background dictionary, total variation (TV).**

## I. Introduction

A HYPERSPECTRAL image (HSI) comprises three-dimensional cube data. Two of these dimensions represent spatial position; and the other, the reflectance of ground objects in different bands. In comparison to multispectral images, HSIs have hundreds of continuous spectra, which could provide abundant spectral information. Based on variations in spectral characteristics, HSIs can be used to recognize ground objects more reliably. Therefore, HSIs are suitable for many remote sensing image processing applications, such as spectral unmixing [1]–[3], classification [4]–[6], change detection [7], [8], and target detection [9]–[11].

Target detection primarily aims to distinguish target pixels from the background pixels according to the spectral characteristics. The target pixels possess two characteristics: first, as compared to surrounding background pixels, significant spectral differences exist; second, the volume of targets is small, with a lower probability of occurrence. Target detection is divisible into supervised and unsupervised versions based on whether there exists prior spectral information on the target pixels. The unsupervised version is alternately known as anomaly detection. As there is no requirement for prior spectral information on anomalous target pixels, anomaly detection is more efficient in applications. Therefore, hyperspectral anomaly detection has been widely used in environmental monitoring [12], search and rescue [13], military reconnaissance [14], and other fields.

### A. Development of Hyperspectral Anomaly Detection Methods

Over the past decades, internal and external researchers have increasingly focused on hyperspectral anomaly detection. Presently, anomaly detection methods could be mainly divided into two groups: statistics-based and representation-based groups. The Reed-Xiaoli (RX) method [15] was first proposed in numerous statistics-based anomaly detection methods. The RX method assumes that the background pixels of HSIs obey Gaussian distribution, and estimates a statistical model by calculating the mean and covariance of pixel spectral vectors. The anomalies are detected by calculating the Mahalanobis distance between tested pixels and the statistical model. Based on the criteria for selecting global or local pixels to characterize the statistical model, the RX methods are divisible into two versions: the global RX (GRX) and local RX (LRX) methods [16]. In recent years, a number of enhanced RX-based methods have been proposed.

Manuscript received November 24, 2021; revised January 13, 2022; accepted January 25, 2022. Date of publication February 1, 2022; date of current version February 24, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42171345, Grant 41971297, and Grant 62071084 and in part by the Tongji University under Grant 02502350047. *(Corresponding author: Liguo Wang.)*

Xiaoyi Wang is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: wangxyi11@126.com).

Liguo Wang is with the College of Information and Communication Engineering, Dalian Minzu University, Dalian 116600, China, and also with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: wangliguo@hrbeu.edu.cn).

Jiawen Wang and Kaipeng Sun are with the Shanghai Institute of Satellite Engineering, Shanghai 201109, China.

Qunming Wang is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: wqm11111@126.com).

Digital Object Identifier 10.1109/JSTARS.2022.3147195

To ensure that the background pixels in each class are more likely to obey a single Gaussian distribution, the cluster-based RX method [17] divides all pixels into several classes. The kernel RX (KRX) method [18] transfers an inseparable HSI into a high-dimensional feature space, to facilitate separation of the background and anomalies. The KRX method was developed to describe a more complex data distribution through nonlinear mapping. However, the KRX method is computationally more complex, and moreover, the selection of the optimal kernel parameters is challenging. To mitigate the influence of anomalies and noise in statistical model estimation, the weighted RX method [19] assigns larger weights to the background and smaller weights to anomalies. These enhanced RX-based methods overcome the disadvantages of the original RX method to a certain extent. However, satisfying the assumption of a Gaussian distribution is challenging for real-world HSIs. Therefore, the detection performance of statistics-based methods is practically often.

Representation-based anomaly detection methods have garnered increasing attention, as there is no need to make any specific assumptions about background distribution. The basic principle is that background could be represented linearly by a background dictionary, whereas anomalies could not. Yuan *et al.* [20] proposed a hyperspectral anomaly detection method based on local sparsity divergence without any distribution hypothesis, which utilizes a sliding dual-window strategy to construct local spectral and spatial dictionaries and determines the anomalies based on the difference in distribution of the sparse coefficients of pixels. Similarly, Li and Du [21] proposed a collaborative representation-based detector (CRD), which uses a background dictionary composed of pixels between dual-windows to collaboratively represent the center tested pixel, and moreover, uses $\ell_2 - norm$ minimization to enhance the collaborative ability between dictionary atoms. To reduce the computational complexity of the CRD, Ma *et al.* [22] proposed a collaborative representation hyperspectral anomaly detection algorithm, which was a fast recursive algorithm, wherein two elementary transformation matrices were constructed based on the position of pixels, and furthermore, a recursive updating method was derived by using matrix inversion lemma, which increases the detection speed of the CRD to a certain extent. Su *et al.* [14] further proposed an anomaly detection method by combining low rank representation (LRR) and CR theories, and meanwhile, introduced the global low rank and local collaborative properties to constrain the representation coefficient matrix.

Low rank and sparse matrix decomposition (LRSMD) [23] is also an important branch of representation-based anomaly detection methods. The primary concept behind the LRSMD-based methods is that the background of HSIs could be represented with few pixels, which is also known as low rank property. Conversely, the anomaly portion is simultaneously constrained by column sparsity, due to the global sparsity of anomaly targets in HSIs. Therefore, the original HSIs could be decomposed into background and anomaly components by low rank property and sparsity. The robust PCA (RPCA) hypothesis [24] was also applied to hyperspectral anomaly detection. RPCA is a special case of LRSMD. It does not need to construct background dictionary, but directly imposes low rank constraints on the original HSI. Zhang *et al.* [25] proposed an LRSMD-based method, which explored prior low rank information on the background to compute the background statistic model and utilized the Mahalanobis distance to detect anomalies. The low rank and sparse representation (LRASR) [26] method combines the LRR and SR theories for hyperspectral anomaly detection. To maintain the local geometric structure and spatial relationship of HSIs, Cheng and Wang [27] introduced graph and total variation (TV) regularizations into the LRSMD model and proposed the graph and total variation regularized low rank representation (GTVLRR) model. Subsequently, Cheng and Wang [28] proposed an anomaly detection method based on a union-dictionary model, wherein a union-dictionary was used to represent background and anomalies, which not only helped to better separate the background and anomalies but also achieved the purpose of noise reduction. Fu *et al.* [29] proposed a hyperspectral anomaly detection method via deep plug-and-play denoising convolutional neural network (CNN) regularization. Instead of exerting a handcrafted regularization, a CNN-based denoiser network was used to constrain the background representation coefficient matrix. Additionally, deep learning and tensor theory have recently emerged as research hotspots in hyperspectral anomaly detection. Li *et al.* [30] presented a prior-based tensor approximation method that divides HSI into background tensor and anomaly tensor. To identify hyperspectral anomalies, Zhang and Cheng [31] employ a tensor-based transferred CNN. Xie *et al.* [32] initially integrated autoencoder and adversarial network learning potential background distribution, and then utilized Mahalanobis distance as detection method, which is named SBEM in brief. Subsequently, Li *et al.* [33] proposed a sparse coding-inspired generative adversarial network for hyperspectral anomaly detection, which utilizes a cascaded autoencoder as the generator.

### *B. Motivation of the Proposed Reweighted Nuclear Norm and Total Variation Regularization With Sparse Dictionary Construction for Hyperspectral Anomaly Detection (WNNSDAD)*

The effectiveness of LRSMD-based methods mainly depends on the design of prior constraints and the background dictionary construction methods. As mentioned above, the backgrounds of HSIs have high spatial correlation (i.e., low rank). Low rank constraints help to effectively maintain the low-dimensional structure of a background. A low rank is an NP-hard problem; hence, a nuclear norm is generally used as a convex relaxation optimization objective considering the low rank. The nuclear norm expresses the sum of matrix singular values, which has been used widely in many fields, such as matrix completion [34] and background modeling [35]. In LRSMD-based hyperspectral anomaly detection methods, the nuclear norm is used as the regularization for the representation coefficient matrix of a background dictionary. Normally, the main properties of a matrix are determined by the larger singular values.

However, the standard nuclear norm exhibits an unbalanced penalty for various singular values, that is, it has the same penalty for all singular values. Using the standard nuclear norm as a low rank constraint, the background matrix finds it challenging to approximate the original background in an iterative updation process, which indirectly affects the separation efficiency between the background and anomalies.

Additionally, the dictionary construction method directly affects the ability of dictionary representation. Most of the existing background dictionary construction methods are based on an unsupervised clustering scheme. For example, in [27], an HSI is first clustered by *k*-means. Thereafter, the RX detector is used for each class, and some pixels with smaller detection values in each class are selected as the background dictionary atoms. This is a mainstream method of dictionary construction, and its derivative methods are presently widely used. However, *k*-means has several limitations. In particular, the number of clusters needs to be given or estimated in advance, but it is practically quite challenging. Additionally, *k*-means is highly sensitive to the data type. If the distribution of each class in an HSI is unbalanced, it could negatively impact the cluster result, introducing include noise and anomalies into background clusters. In [28], the density peak-based (DPC) clustering method was used to construct a background dictionary. In comparison to *k*-means, the DPC clustering method could deal with clusters of various shapes and sizes; hence, it is more robust to noise and anomalies. However, the DPC clustering method faces the issue of the curse of dimensionality for HSIs. Moreover, similar to that for *k*-means, the clustering result is affected by the setting of cluster centers. Additionally, when the number of a certain background material is small, it could be limited as noise or anomaly, which could lead to incomplete background dictionary (i.e., some background pixels cannot be expressed), resulting in a decrease in detection accuracy.

Considering the above challenges, in this article, we proposed a novel hyperspectral anomaly detection method, which incorporates reweighted nuclear norm and TV regularizations as the prior constraints into an LRSMD model. TV regularization and reweighted nuclear norm could characterize the spatial smoothness and low rank property, which have been confirmed in the remote sensing community [36], [37]. Moreover, a sparse background dictionary is constructed without any clustering techniques. The proposed method is herein abbreviated as WNNSDAD. The proposed method has the following main contributions.

1) The reweighted nuclear norm regularization is used as a low rank constraint for the representation coefficient matrix of a background dictionary, which helps to ensure that the reconstructed background part is closer to the real background. Furthermore, the TV regularization is introduced as the piecewise smoothness constraint of the background dictionary representation coefficient matrix to take advantage of the spatial correlation between adjacent pixels.
2) A nonclustering method is proposed to construct a sparse background dictionary. By combining anomaly detection with sparse representation theories, an over-complete background dictionary is constructed without clustering techniques. It is not only unaffected by the cluster parameters but also effectively retains the characteristic spectrum of a background, especially for small background materials.

The rest of this article is organized as follows. The proposed WNNSDAD and sparse dictionary construction strategy are presented in Section II. Section III provides experimental results based on seven real-world HSI datasets. In Section IV, the proposed method and experimental results are further discussed. Finally, Section V concludes this article.

## II. Methods

We defined a three-dimensional HSI as $\mathbb{Y} \in \mathbb{R}^{H\times W\times D}$, where *H* represents a row, *W* a column, and *D* a spectral band. Here, the LRSMD-based method was primarily introduced. We transformed the original three-dimensional HSI into a two-dimensional matrix. Let $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_l] \in \mathbf{R}^{D\times L}$ be a transformed two-dimensional HSI, where $L = H \times W$ represents the total number of pixels, and $\mathbf{y}_l$ is the *l*th pixel with *D* dimensions. A two-dimensional HSI can be decomposed into two components: background $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_l] \in \mathbf{R}^{D\times L}$ and anomalies $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_l] \in \mathbf{R}^{D\times L}$, i.e., $\mathbf{Y} = \mathbf{B} + \mathbf{S}$.

### A. Theoretical Basis of LRSMD

LRSMD is used to consider the product of background dictionary $\mathbf{A}$ and representation coefficient matrix $\mathbf{X}$ to replace the background component $\mathbf{B}$. It is considered that the representation coefficient matrix and the anomalies have low rank and sparse properties, respectively. The optimization objective function of LRSMD is expressed as follows:

$$\min_{\mathbf{X},\mathbf{S}} \|\mathbf{X}\|_* + \lambda\|\mathbf{S}\|_{2,1}, \text{ s.t. } \mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{S} \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k] \in \mathbf{R}^{D\times K}$ is the background dictionary and $\mathbf{a}_k$ is the *k*th dictionary atom; $\|\cdot\|_*$ expresses the nuclear norm of the matrix; $\|\cdot\|_{2,1}$ is $\ell_{2,1} - norm$ and represents the sum of $\ell_2 - norm$ in each row of matrix $\mathbf{X}$; and $\lambda$ is a tradeoff parameter. By optimizing (1) to obtain $\mathbf{X}$ and $\mathbf{S}$, $\mathbf{S}$ can be used to detect anomalies. In particular, the anomaly probability $Q(\mathbf{y}_i) = \|\mathbf{s}_i\|_2$ of each pixel could be obtained by calculating $\ell_2 - norm$ for each column of $\mathbf{S}$. Given a threshold $\eta$, if $Q(\mathbf{y}_i) > \eta$, the pixel is considered to be an anomaly; otherwise, it is a background.

### B. Reweighted Nuclear Norm and TV Regularization

*1) Reweighted Nuclear Norm:* The low rank constraint of a reweighted nuclear norm on $\mathbf{X}$ could be realized by the singular value decomposition (SVD) method. In particular, $\mathbf{X}$ could be decomposed into $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are identity orthogonal matrices, i.e., $\mathbf{U}\mathbf{U}^T = \mathbf{I}$ and $\mathbf{V}\mathbf{V}^T = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix. $\mathbf{U}$ and $\mathbf{V}$ are called left and right singular matrices, respectively. $\mathbf{\Sigma}$ is a diagonal matrix, and we defined $\Sigma_{ii}$ as the *i*th diagonal element. Varying from the standard nuclear norm, a weighting factor is introduced into the iterative optimization, giving a larger weight to a smaller singular value and a smaller weight to a larger singular value. In particular, the weight factor was set to $w_i = (\frac{1}{\Sigma_{ii}} + \varepsilon)$, where $\varepsilon$ is a small

positive parameter. The reweighted soft-threshold shrinkage was expressed as

$$G_w(\Sigma_{ii}) = \max(\Sigma_{ii} - \rho \times w_i, 0) \tag{2}$$

where $\rho$ is a tradeoff parameter. It could be seen that a smaller singular value corresponds to a smaller weight and tends to be zero after the subsequent iteration. The solution of the objective function is as follows:

$$\mathbf{X} = \mathbf{U}G_w(\Sigma_{ii})\mathbf{V}. \tag{3}$$

*2) TV Regularization:* The adjacent pixels $\mathbf{y}_i$ and $\mathbf{y}_j$ were considered to have similar spectral characteristics; hence, the corresponding representation coefficients $\mathbf{x}_i$ and $\mathbf{x}_j$ were also similar. TV regularization is often regarded as the piecewise smoothness constraint, which is defined as an $\ell_{1,1} - norm$ of difference considering the neighbor pixels by

$$\|\mathbf{HX}\|_{1,1} = \left\| \begin{bmatrix} \mathbf{H}_h\mathbf{X} \\ \mathbf{H}_v\mathbf{X} \end{bmatrix} \right\|_{1,1} \tag{4}$$

where $\mathbf{H}_h$ denotes the linear operator, by calculating the difference of representation coefficients between each pixel and its horizontal neighboring pixel, and $\mathbf{H}_v$ denotes the difference between each pixel and its vertical neighboring pixel. The $\ell_{1,1} - norm$ is defined as the sum of the $\ell_1 - norm$ of each column in a matrix.

### C. Proposed WNNSDAD Method

The optimization objective function of the proposed WNNSDAD could be expressed as

$$\begin{aligned} \min_{\mathbf{X},\mathbf{S}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX} - \mathbf{S}\|_F^2 + \|\mathbf{X}\|_{w*} + \lambda\|\mathbf{HX}\|_{1,1} + \beta\|\mathbf{S}\|_{2,1} \\ \text{s.t.} \quad \mathbf{Y} = \mathbf{AX} + \mathbf{S} \end{aligned} \tag{5}$$

where $\|\cdot\|_{w*}$ expresses the reweighted nuclear norm regularization, and is used to constrain $\mathbf{X}$, $\ell_{2,1} - norm$ is the sparse regularization term of anomalies, and $\lambda$ and $\beta$ are tradeoff parameters.

As the objective optimization function was inseparable, auxiliary variables were introduced to represent the optimization problem as follows:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{V}_1,\mathbf{V}_2,\mathbf{S}} \frac{1}{2}\|\mathbf{Y} - \mathbf{AX} - \mathbf{S}\|_F^2 + \|\mathbf{V}_1\|_{w,*} + \lambda\|\mathbf{V}_3\|_{1,1} \\ + \beta\|\mathbf{S}\|_{2,1} \\ \text{s.t.} \quad \mathbf{V}_1 = \mathbf{X}, \mathbf{V}_2 = \mathbf{X}, \mathbf{V}_3 = \mathbf{HV}_2. \end{aligned} \tag{6}$$

Equation (6) is separable, and the optimization problem could be efficiently solved by the alternating direction method of multipliers (ADMM) [38]. The augmented Lagrange function is expressed as follows:

$$\begin{aligned} &L\{\mathbf{X}, \mathbf{S}, \mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mu\} \\ &= \frac{1}{2}\|\mathbf{Y} - \mathbf{AX} - \mathbf{S}\|_F^2 + \|\mathbf{V}_1\|_{w,*} + \lambda\|\mathbf{V}_3\|_{1,1} \\ &\quad + \beta\|\mathbf{S}\|_{2,1} + \frac{\mu}{2}\|\mathbf{V}_1 - \mathbf{X} - \mathbf{D}_1\|_F^2 \\ &\quad + \frac{\mu}{2}\|\mathbf{V}_2 - \mathbf{X} - \mathbf{D}_2\|_F^2 + \frac{\mu}{2}\|\mathbf{V}_3 - \mathbf{HV}_2 - \mathbf{D}_3\|_F^2 \end{aligned} \tag{7}$$

where $\mu$ is a penalty parameter, and $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ are Lagrange multipliers. The minimization of (7) could be achieved by iteratively optimizing the objective function for one variable and simultaneously fixing other variables. As the augmented Lagrangian function is convex, the global optimality of the solution can be guaranteed. This problem could be divided into the following subproblems.

1) Update $\mathbf{X}$: By fixing other variables, the optimization problem of $\mathbf{X}$ could be expressed as

$$\begin{aligned} \mathbf{X}^{t+1} = \arg\min_{\mathbf{X}^t} &\frac{1}{2}\left\|\mathbf{Y} - \mathbf{AX}^t - \mathbf{S}^t\right\|_F^2 \\ &+ \frac{\mu}{2}\left\|\mathbf{V}_1{}^t - \mathbf{X}^t - \mathbf{D}_1{}^t\right\|_F^2 \\ &+ \frac{\mu}{2}\left\|\mathbf{V}_2{}^t - \mathbf{X}^t - \mathbf{D}_2{}^t\right\|_F^2 \end{aligned} \tag{8}$$

where $t$ is the iteration index. A closed-form solution could be obtained by deriving (8) for $\mathbf{X}^t$ and setting the equation to equal zero

$$\begin{aligned} \mathbf{X}^{t+1} &= (\mathbf{A}^T\mathbf{A} + 2\mu\mathbf{I})^{-1} \\ &\quad \left[\mathbf{A}^T(\mathbf{Y} - \mathbf{S}^t) + \mu(\mathbf{V}_1{}^t - \mathbf{D}_1{}^t) + \mu(\mathbf{V}_2{}^t - \mathbf{D}_2{}^t)\right]. \end{aligned} \tag{9}$$

2) Update $\mathbf{V}_1$: The objective function is written as follows:

$$\mathbf{V}_1{}^{t+1} = \arg\min_{\mathbf{V}_1{}^t} \left\|\mathbf{V}_1{}^t\right\|_{w,*} + \frac{\mu}{2}\left\|\mathbf{V}_1{}^t - \mathbf{X}^{t+1} - \mathbf{D}_1{}^t\right\|_F^2. \tag{10}$$

Equation (10) could be updated by the SVD method

$$\mathbf{V}_1{}^{t+1} = \Theta_{w,1/\mu^t}(\mathbf{X}^{t+1} + \mathbf{D}_1{}^t). \tag{11}$$

3) Update $\mathbf{V}_2$: The optimization problem of $\mathbf{V}_2$ could be expressed as follows:

$$\begin{aligned} \mathbf{V}_2{}^{t+1} = \arg\min_{\mathbf{V}_2{}^t} &\frac{\mu}{2}\left\|\mathbf{V}_2{}^t - \mathbf{X}^{t+1} - \mathbf{D}_2{}^t\right\|_F^2 \\ &+ \frac{\mu}{2}\left\|\mathbf{V}_3{}^t - \mathbf{HV}_2{}^t - \mathbf{D}_3{}^t\right\|_F^2. \end{aligned} \tag{12}$$

The closed-form solution could be written as follows:

$$\mathbf{V}_2{}^{t+1} = (\mathbf{H}^T\mathbf{H} + \mathbf{I})^{-1}(\mathbf{X}^{t+1} + \mathbf{D}_2{}^t + \mathbf{H}^T(\mathbf{V}_3{}^t - \mathbf{D}_3{}^t)). \tag{13}$$

4) Update $\mathbf{V}_3$: $\mathbf{V}_3$ could be updated through the following objective function:

$$\begin{aligned} \mathbf{V}_3{}^{t+1} = \arg\min_{\mathbf{V}_3{}^t} &\lambda\left\|\mathbf{V}_3{}^t\right\|_{1,1} \\ &+ \frac{\mu}{2}\left\|\mathbf{V}_3{}^t - \mathbf{HV}_2{}^{t+1} - \mathbf{D}_3{}^t\right\|_F^2. \end{aligned} \tag{14}$$

Furthermore, the soft threshold method was used to solve the problem

$$\mathbf{V}_3{}^{t+1} = \text{soft}(\mathbf{HV}_2{}^{t+1} + \mathbf{D}_3{}^t, \frac{\mu^t}{\lambda}). \tag{15}$$

5) Update $\mathbf{S}$: The optimization problem was as follows:

$$\mathbf{S}^{t+1} = \arg\min_{\mathbf{S}} \left\|\mathbf{S}^t\right\|_{2,1} + \frac{1}{2}\left\|\mathbf{Y} - \mathbf{AX}^{t+1} - \mathbf{S}^t\right\|_F^2. \tag{16}$$

**Algorithm 1:** Optimization of the Proposed WNNSDAD.

**Input**: Two-dimensional HSI $\mathbf{Y}$; background dictionary $\mathbf{A}$; parameters $\lambda$, $\beta$, $\mu$, $\mu_{\max}$, and $\rho$; the maximum iteration index $t_{\max}$; and residual threshold $\varepsilon$.
**Output**: Background representation coefficient $\mathbf{X}^*$ and abnormal part $\mathbf{S}^*$.
**Initialization**: Set $\mathbf{X}$, $\mathbf{S}$, $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$, $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ to zero matrices.
**while** converge = false, **do**
  Update $\mathbf{X}$ by (9).
  Update $\mathbf{V}_1$ by (11).
  Update $\mathbf{V}_2$ by (13).
  Update $\mathbf{V}_3$ by (15).
  Update $\mathbf{S}$ by (17).
    Update $\mathbf{D}_1$, $\mathbf{D}_2$ and $\mathbf{D}_3$ by (18).
  Update $\mu$ by (19).
  **if** (20) is satisfied or $t = t_{\max}$, converge = true.
  **end if**
**end while**

Equation (16) could be solved by the $\ell_{2,1} - norm$ minimization operator $\Omega$ [33]. Therefore, we have

$$\mathbf{S}^{t+1} = \Omega_{\beta/\mu^t}(\mathbf{Y} - \mathbf{A}\mathbf{X}^{t+1}). \tag{17}$$

6) Update Lagrange multipliers $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$

$$\begin{cases} {\mathbf{D}_1}^{t+1} = {\mathbf{D}_1}^t - (\mathbf{V}_1^{t+1} - \mathbf{X}^{t+1}) \\ {\mathbf{D}_2}^{t+1} = {\mathbf{D}_2}^t - (\mathbf{V}_2^{t+1} - \mathbf{X}^{t+1}) \\ {\mathbf{D}_3}^{t+1} = {\mathbf{D}_3}^t - (\mathbf{V}_3^{t+1} - \mathbf{H}{\mathbf{V}_2}^{t+1}) \end{cases}. \tag{18}$$

7) Update penalty parameter $\mu$

$$\mu^{t+1} = \min(\rho\mu^t, \mu_{\max}) \tag{19}$$

where $\rho$ is the growth index of the penalty parameter and $\mu_{\max}$ is the maximum penalty parameter.

The iteration process stops until the maximum iteration index is reached or if it meets the following requirements:

$$\begin{cases} \|\mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{S}\|_F \le \varepsilon \\ \|\mathbf{X} - \mathbf{V}_1\|_F \le \varepsilon \\ \|\mathbf{X} - \mathbf{V}_2\|_F \le \varepsilon \end{cases} \tag{20}$$

where $\varepsilon$ is a residual threshold. The process for solving the objective function (5) is summarized in Algorithm 1.

### *D. Proposed Sparse Background Dictionary Construction Method*

Here, we proposed a sparse background dictionary construction method, which combined anomaly detection with sparse representation theory. Varying from the previous cluster-based dictionary construction methods, we could obtain a relatively complete background dictionary without unsupervised clustering. The construction of a sparse background dictionary could be regarded as a three-stage strategy.

1) First, the RX method was used to detect the original HSI. By calculating the Mahalanobis distance between each tested pixel and the corresponding background mean vector, we obtained an initial detection map, which reflected the anomaly intensity of each pixel.
2) Then, we used an adaptive threshold method inspired from [39] to screen out potential background pixels. The threshold $\delta$ was set as follows:

$$\delta = \varphi(E + (M - E)\sqrt{\frac{E}{M}}) \tag{21}$$

where $E$ and $M$ are the mean and maximum values of the preliminary detection map, respectively. To select more pure background pixels, we add an adjustment parameter $\varphi$ less than 1. We considered the pixels with detection values lower than the threshold to be closer to the background; hence, the pixels with a value greater than the threshold were removed, the remaining pixels constituted the background sample set. We define the background sample set as $\mathbf{P}$.
3) Lastly, to avoid the of nonconvergence or over fitting in the process of the training dictionary, we randomly sorted the samples in $\mathbf{P}$. Subsequently, the K-SVD algorithm [40] and sparse coding technique orthogonal matching pursuit (OMP) [41] was applied to train the sparse background dictionary. Here, we needed to set a dictionary size. Note that to construct an over-complete dictionary, the dictionary size must be larger than the dimension of the HSI. Empirically, this parameter is usually set with an integer power of 2, such as 128, 256, and 512, among others. Further details on K-SVD and OMP could be found in [40] and [41].

As the benchmark method of anomaly detection, the RX method is advantageous considering its simple form and fast operation speed; also, it requires no parameter adjustment. Although limited by the Gaussian distribution hypothesis, it could achieve the purpose of preliminary detection. Using the RX method as an initial anomaly detection method is a common process for constructing a background dictionary, such as in [27] and [29]. Varying from the cluster-based background dictionary construction methods, we utilized a sparse representation scheme to train a background dictionary. Through the sparse representation technique, we could obtain a complete dictionary containing a characteristic background spectrum, especially for small background materials. The proposed sparse dictionary construction method is summarized in Algorithm 2; moreover, Fig. 1 shows the flowchart of the whole algorithm.

## III. EXPERIMENTS

### *A. HSI Datasets and Experimental Setup*

Seven real-world HSI datasets were used in the experiments. Table I lists some important image characteristics of the seven HSI datasets. Note that AVIRIS and ROSIS-03 are abbreviations of the airborne visible/infrared imaging spectrometer and reflective optics spectrographic imaging system, respectively. False-color images and corresponding ground reference maps of the seven datasets are shown in Fig. 2.

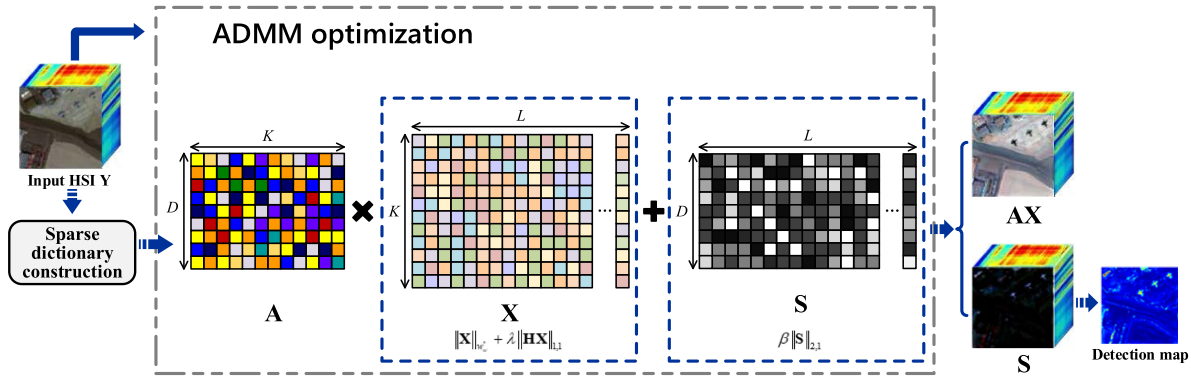



Fig. 1. Flowchart of the proposed method for hyperspectral anomaly detection.

**Algorithm 2:** Sparse Background Dictionary Construction.

1. **Input**: Two-dimensional HSI $\mathbf{Y}$, dictionary size $K$, the maximum iteration index $t_{\max}$, maximal sparsity $\alpha$ (set to 4 in this article), and termination threshold $\xi$.
2. **Output**: Background dictionary $\mathbf{A}= [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_k]$.
3. Compute the mean vector and covariance matrix of $\mathbf{Y}$.
4. Compute Mahalanobis distance to the mean vector between each tested pixel.
6. Equation (21) is used to calculate the threshold $\delta$; furthermore, the pixel with intensity values smaller than the threshold are regarded as pure background pixels to form the background sample $\mathbf{P}$.
7. Initialize the background dictionary $\hat{\mathbf{A}}$, and randomly select $K$ pixels from to $\mathbf{P}$ form dictionary atoms.
8. Utilizing K-SVD and OMP algorithms to update background dictionary while $t_{\max}$ is reached or the difference between two consecutive realizations is less than $\xi$.

To demonstrate the effectiveness of the proposed WNNSDAD, seven state-of-the-art anomaly detection methods, including GRX, LRX, RPCA, CRD, LRASR, LSMAD, GTVLRR, and SBEM, were selected for comparison. The rest of this section comprises four parts. Section III-B provides the experimental results of WNNSDAD and the seven comparative methods for seven real-world HSI datasets. Section III-C describes the parameter analysis of WNNSDAD. Section III-D demonstrates the effectiveness of the sparse background dictionary construction method.

### B. Detection Performance

Here, two-dimensional and three-dimensional receiver operating characteristic (ROC) curves [42], [43] and the area under the curve (AUC) [44] were used as the evaluation indexes for the anomaly detection. Table II lists the parameter setting of different methods for the seven datasets. The notations in the table are represented as: $w_{\text{in}}$ and $w_{\text{out}}$ are the window sizes of the dual-window in LRX and CRD, and $\lambda$, $\beta$, and $\gamma$ are the predefined tradeoff parameters. Additionally, the learning rate of autoencoder and adversarial network in SBEM were set to 10e–3 and 10e–4 based on the suggestion in [32].

TABLE I
SOME IMAGE CHARACTERISTICS OF THE SEVEN HSIS

| Datasets | Location | Sensor | Spatial Size | Number of Bands |
|---|---|---|---|---|
| Dataset I | San diego | AVIRIS | 100×100 | 189 |
| Dataset II | San diego | AVIRIS | 100×100 | 189 |
| Dataset III | Los angeles | AVIRIS | 100×100 | 205 |
| Dataset IV | Gulfport | ROSIS-03 | 100×100 | 191 |
| Dataset V | Pavia | AVIRIS | 150×150 | 102 |
| Dataset VI | Los angeles | AVIRIS | 100×100 | 205 |
| Dataset VII | Shanghai | GF-5 | 119×109 | 150 |

Fig. 3 shows the detection maps of the nine methods for the seven HSI datasets. Through visual inspection, it could be evidently seen that the detection maps of WNNSDAD were more satisfactory than those of the other methods. For example, for dataset I, GRX, LRX, and RPCA basically failed to separate the background and anomalies; consequently, it was challenging difficult to locate the anomaly aircraft from the detection maps. Although CRD, LRASR, LRSMD, GTVLRR, and SBEM could help detect the position of anomaly aircraft, they could not completely retain the shape and edge contours of anomaly aircraft. However, the detection map for WNNSDAD clearly separates the background and anomalies; this was visible from the brightness: the brightness of the background approaches 0, whereas that of the anomaly approached 1. Additionally, WNNSDAD effectively retains the shape of the anomaly aircraft. As another example, considering dataset IV, although each method could detect anomalies, the detection map of the proposed method was more accurate. In particular, the complete shape of the anomaly vehicles could not be detected by GRX, LRX, RPCA, LRASR, LRSMD, and SBEM. For CRD and GTVLRR, although the shape of the anomaly vehicles was detected, the background below the image could not be suppressed effectively. The detection maps of the remaining datasets were similar to the former two, i.e., WNNSDAD facilitated more satisfactory visual inspection as compared to the other seven methods.

The two-dimensional and three-dimensional ROC curves for the nine methods are shown in Figs. 4 and 5, respectively.

TABLE II
PARAMETER SETTING OF DIFFERENT METHODS FOR THE SEVEN DATASETS

| | RPCA | LRX | | CRD | | | LRASR | | LSMAD | GTVLRR | | | WNNSDAD | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\lambda$ | $w_{in}$ | $w_{out}$ | $\lambda$ | $w_{in}$ | $w_{out}$ | $\lambda$ | $\beta$ | Rank value | $\lambda$ | $\beta$ | $\gamma$ | $\lambda$ | $\beta$ | Dictionary size |
| Dataset I | 10e-2 | 3 | 23 | 10e-6 | 17 | 21 | 10e-2 | 1 | 3 | 0.5 | 0.2 | 0.05 | 1 | 1 | 256 |
| Dataset II | 10e-2 | 9 | 13 | 10e-6 | 19 | 21 | 10e-2 | 1 | 2 | 0.5 | 0.2 | 0.05 | 0.08 | 0.05 | 256 |
| Dataset III | 10e-2 | 5 | 7 | 10e-6 | 11 | 13 | 10e-2 | 1 | 19 | 0.5 | 0.2 | 0.05 | 0.2 | 0.2 | 256 |
| Dataset IV | 10e-2 | 9 | 13 | 10e-6 | 17 | 19 | 10e-2 | 1 | 20 | 0.5 | 0.2 | 0.05 | 0.02 | 0.2 | 256 |
| Dataset V | 10e-2 | 15 | 17 | 10e-6 | 7 | 9 | 10e-2 | 1 | 18 | 0.5 | 0.2 | 0.05 | 0.005 | 0.002 | 256 |
| Dataset VI | 10e-2 | 21 | 23 | 10e-6 | 3 | 13 | 10e-2 | 1 | 20 | 0.5 | 0.2 | 0.05 | 0.2 | 0.2 | 1024 |
| Dataset VII | 10e-2 | 11 | 3 | 10e-6 | 11 | 13 | 10e-2 | 1 | 3 | 0.5 | 0.2 | 0.05 | 0.2 | 0.2 | 256 |

TABLE III
AUC VALUES OF THE NINE METHODS FOR THE SEVEN DATASETS

| | GRX | LRX | RPCA | CRD | LRASR | LSMAD | GTVLRR | SBEM | WNNSDAD |
|---|---|---|---|---|---|---|---|---|---|
| Dataset I | 0.8885 | 0.7590 | 0.9165 | 0.9865 | 0.9891 | 0.9905 | 0.9800 | 0.9845 | **0.9949** |
| Dataset II | 0.9403 | 0.9417 | 0.9294 | 0.9664 | 0.8353 | 0.9710 | 0.9311 | 0.9856 | **0.9879** |
| Dataset III | 0.8221 | 0.8845 | 0.8089 | 0.9555 | 0.7955 | 0.8420 | 0.9007 | 0.9402 | **0.9752** |
| Dataset IV | 0.9526 | 0.9273 | 0.9628 | 0.9767 | 0.9813 | 0.9899 | 0.9865 | 0.9872 | **0.9955** |
| Dataset V | 0.9539 | 0.9084 | 0.9599 | 0.9505 | 0.9506 | 0.9733 | 0.9801 | 0.9856 | **0.9958** |
| Dataset VI | 0.9692 | 0.9151 | 0.9658 | 0.9523 | 0.9085 | 0.9610 | 0.9337 | 0.9664 | **0.9804** |
| Dataset VII | 0.9811 | 0.9851 | 0.9897 | 0.9924 | 0.8204 | 0.9929 | 0.9967 | 0.9846 | **0.9993** |

From Fig. 4, it can be seen that the two-dimensional ROC curves of WNNSDAD are located in the upper-left corner for all seven datasets, suggesting that the proposed method had a more accurate detection performance and lower false alarm rate. Additionally, based on the two-dimensional ROC curves, CRD, GTVLRR, and SBEM exhibited a relatively stable performance. As shown in Fig. 4(c), when the false alarm rate was lower than 0.1, the detection rate of CRD, GTVLRR, and SBEM approached that of the proposed method. However, when the detection rate approached 1, CRD, GTVLRR, and SBEM exhibited a larger false alarm rate. Using Mahalanobis distance to create detection map, SBEM enhances the difference between anomalies and background by autoencoder, and the results of most datasets produced by SBEM are more satisfactory than GRX and LRX. Fig. 5 also demonstrates that the performance of WNNSDAD was relatively more satisfactory than the other methods. The AUC values of the eight methods are listed in Table III, where the most accurate result for each dataset is highlighted in bold. Being consistent with the ROC curves, the AUC values of the proposed method were the largest, for all seven datasets. Considering the detection results for dataset I as an example, the AUC value for WNNSDAD was 0.9949; those for GRX, LRX, RPCA, CRD, LRASR, LSMAD, GTVLRR, and SBEM were 0.8885, 0.7590, 0.9165, 0.9865, 0.9891, 0.9905, 0.9800, and 0.9845, respectively. Similarly, for datasets II–IV, which showed both aircraft as anomalies, the proposed method still demonstrated reliable scores. For datasets V–VII, the AUC values for WNNSDAD were 0.9955, 0.9804, and 0.9993, respectively, and the AUC values for the other methods were obviously smaller.

We utilized a box-whisker plot to further demonstrate the separation ability between the background and anomalies. As shown in Fig. 6, we utilized purple and blue boxes to represent the distributions of the background and anomalies, respectively. In general, if there existed a larger distance and smaller overlap between a background box and an anomaly box, this method was considered to demonstrate good separation ability for the background and anomalies. A similar conclusion could be drawn that WNNSDAD could achieve the most satisfactory separation ability between the background and anomalies. Based on the proposed method, it could be seen from Fig. 6 that a background box and an anomaly box had the smallest overlap and the largest distance for all datasets. This indicated that WNNSDAD could be suitably applied across various datasets. Additionally, GTVLRR demonstrated a separation ability more stable, but still inferior to that of WNNSDAD.

### C. Parameter Analysis

In the proposed method, three parameters needed to be tuned: the tradeoff parameters $\lambda$ and $\beta$, and the background dictionary size $K$. $\lambda$ and $\beta$ control piecewise smoothness and sparse property, respectively. To validate the robustness of the proposed method considering variations in the tradeoff parameters, $\lambda$ and $\beta$ were set in the candidate pool {0.002, 0.005, 0.02, 0.05, 0.08, 0.2, 0.4, 0.6, 0.8, and 1}, and the dictionary size for

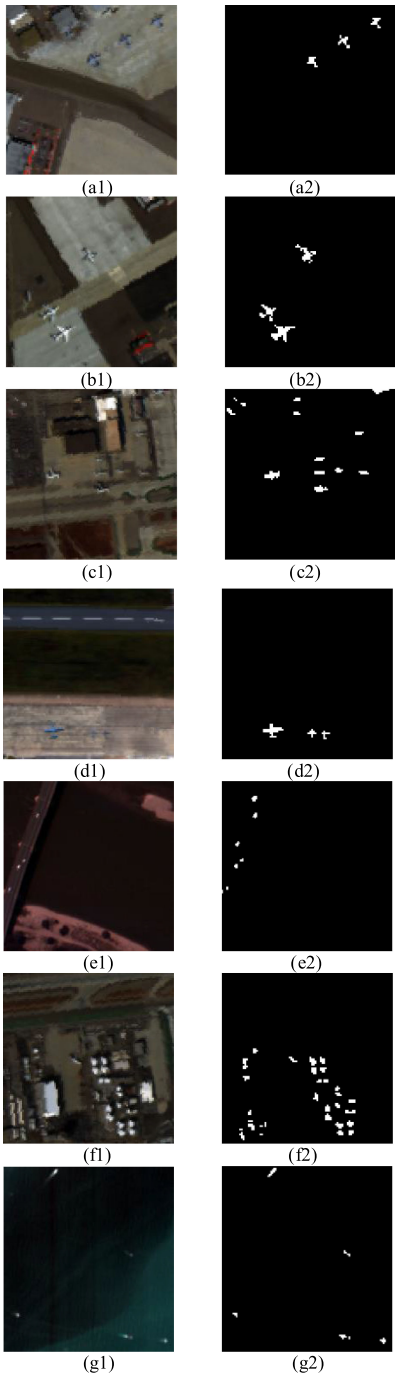



Fig. 2. Seven real-world HSI datasets used in the experiments. (a1)–(g1) are false-color images [bands 37, 18, and 8 as RGB for (a1)–(f1), and bands 70, 22, and 17 as RGB for (g1)] and (a2)–(g2) are the corresponding ground-reference maps of the seven datasets. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII.

the seven datasets was generally 256. Fig. 7 shows the AUC values with a change in $\lambda$ and $\beta$ for the seven datasets. With a change of parameters, the AUC values of each dataset were evidently relatively stable, and the fluctuation was not apparent. For example, for datasets II, V, and VII, with a change in tradeoff parameters, the fluctuation of the AUC values was less than 0.05. For dataset I, if $\lambda$ or $\beta$ approached 1, the AUC values were all above 0.99. However, when the two parameters lay between 0.002 and 0.2, the AUC values decreased up to even 0.94. For datasets III and VI, when the tradeoff parameters in range lay between 0.005 and 0.05, the detection performance was satisfactory; however, the AUC values decreased noticeably under other parameters.

As a parameter of the proposed method, dictionary size also played a crucial role in the experimental results. In [45], it was pointed out that larger dictionaries possessed more expressive power. Here, by fixing the other two parameters $\lambda$ and $\beta$ in the optimal values, we empirically trained three dictionaries with sizes of 256, 512, and 1024 for the seven datasets. Fig. 8 shows the AUC values of the three dictionary sizes for the seven datasets, and Table IV lists the corresponding running-time. We found that a larger dictionary size does not necessarily produce more accurate experimental results; also, the process would take longer. For example, for dataset III, when the dictionary sizes were 256, 512, and 1024, the corresponding AUC values were 0.9752, 0.9691, and 0.9718, respectively. Although the difference in the three AUC values was small, the running-time of the larger dictionary size was obviously longer. The possible reason behind this is that the spectral bands of HSIs are highly correlated, i.e., severe information redundancy exists between spectral bands. Using a larger dictionary size could affect the subsequent iterative update process and reduce the efficiency in obtaining an optimal solution. In particular, for dataset VII, for a dictionary size of 1024, the corresponding operation time was up to 1111 s.

### D. Performance of Sparse Dictionary

To validate the effectiveness of our proposed sparse dictionary construction method, we compared it with the other two cluster-based dictionary construction methods mentioned in [27] and [28]; for convenience, we called them as the *k*-means-based and DPC-based dictionary construction methods, respectively. We used the three dictionary construction methods combined with the proposed algorithm, taking the seven datasets as examples; furthermore, Table V lists the AUC values. In the *k*-means-based method, the *k* value and the number for each cluster were set to 15 and 20 for the seven datasets, and the clusters with less than 20 pixels were not considered. In the DPC-based methods, the predefined distance threshold $\eta$ and $k$ value were 0.1 and 6, respectively. As in Table V, the largest value was highlighted in bold. In comparison to the other two methods, the proposed dictionary construction method demonstrated more reliable performance. In particular, for dataset III, the AUC values for the *k*-means-based, DPC-based, and the proposed dictionary construction methods were 0.9105, 0.7946, and 0.9752, respectively. This primarily occurred as the clustering-based methods are often affected by parameter setting, which is generally subjective. Additionally, some small background materials are easily excluded as noise or anomalies, and they lose their spectral characteristics. This could negatively impact the background dictionary representation ability. The proposed sparse dictionary construction method was affected by the clustering parameter setting; moreover, the characteristic spectrum was effectively preserved by the sparse coding technique. Small background

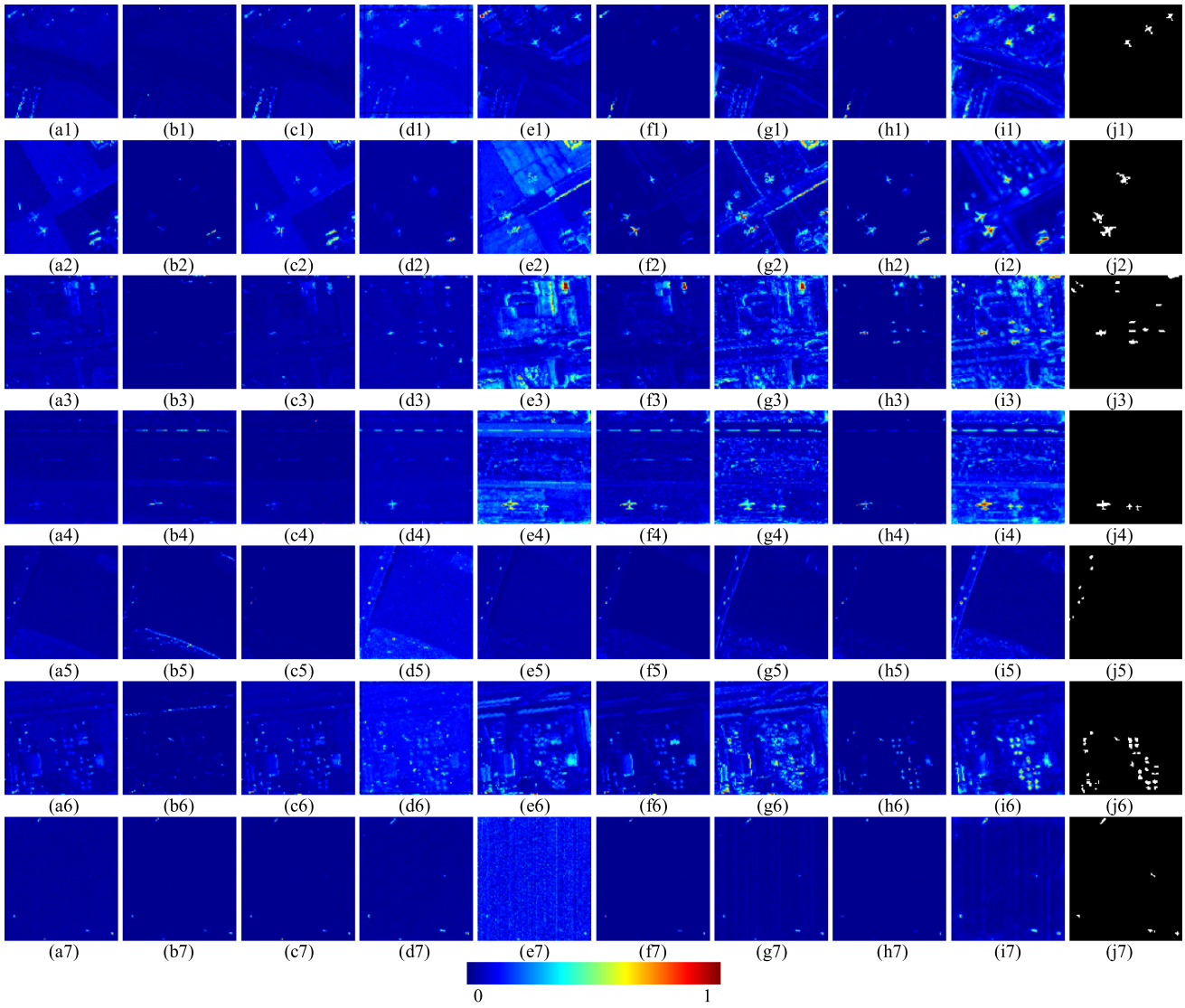



Fig. 3. Detection maps based on nine methods for the seven real-world HSI datasets. (a)–(i) are detection maps (from the left to the right) for GRX, LRX, RPCA, CRD, LRASR, LSMAD, GTVLRR, SBEM, and WNNSDAD, respectively, and (j) is the corresponding ground-reference map. Lines 1–7 represent dataset I, dataset II, dataset III, dataset IV, dataset V, dataset VI, and dataset VII, respectively.

TABLE IV
RUNNING-TIME OF DIFFERENT BACKGROUND DICTIONARY SIZES FOR THE SEVEN DATASETS (IN SECONDS)

| Dictionary size | Dataset I | Dataset II | Dataset III | Dataset IV | Dataset V | Dataset VI | Dataset VII |
|---|---|---|---|---|---|---|---|
| 256 | 99 | 115 | 103 | 111 | 154 | 92 | 246 |
| 512 | 233 | 242 | 223 | 246 | 310 | 264 | 456 |
| 1024 | 477 | 507 | 407 | 519 | 756 | 437 | 1111 |

TABLE V
AUC VALUES OF DIFFERENT BACKGROUND DICTIONARY CONSTRUCTION METHODS

| | Dataset I | Dataset II | Dataset III | Dataset IV | Dataset V | Dataset VI | Dataset VII |
|---|---|---|---|---|---|---|---|
| $k$-means-based dictionary | 0.9900 | 0.9453 | 0.9105 | 0.9525 | 0.9898 | 0.8438 | 0.9898 |
| DPC-based dictionary | 0.9929 | 0.9625 | 0.8276 | 0.8824 | 0.9934 | 0.8914 | 0.9978 |
| Sparse dictionary | **0.9949** | **0.9879** | **0.9752** | **0.9955** | **0.9958** | **0.9804** | **0.9993** |

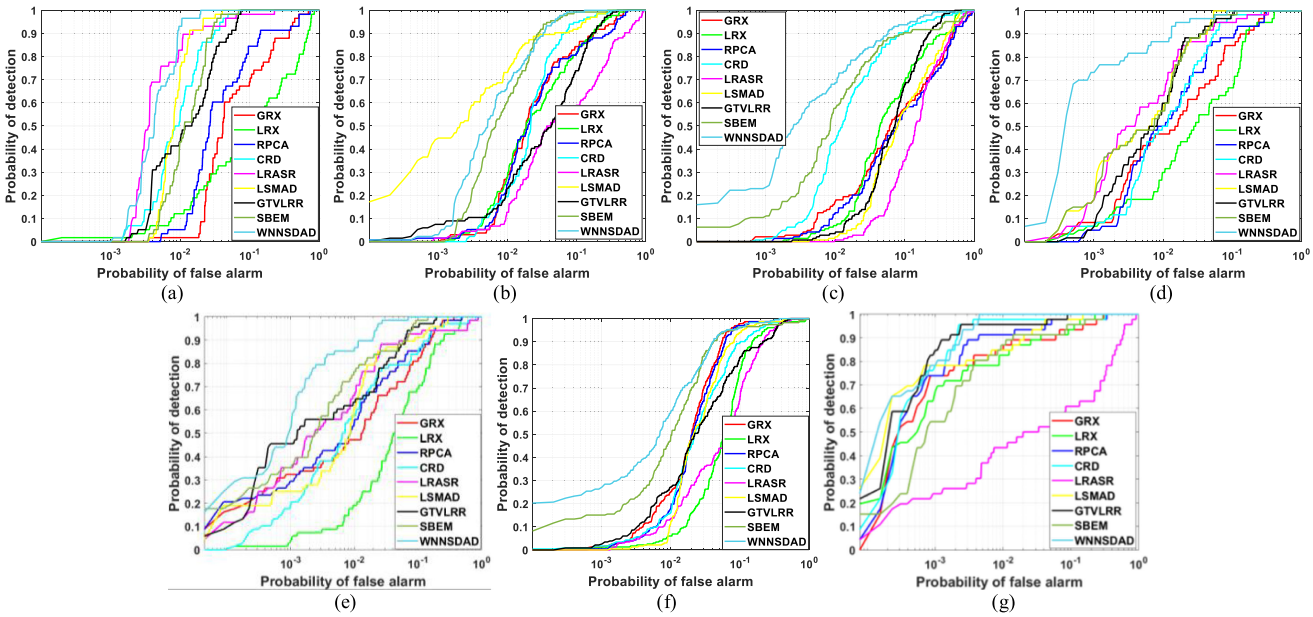



Fig. 4. Two-dimensional ROC curves of the nine anomaly detectors for the seven real-world HSI datasets. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII.

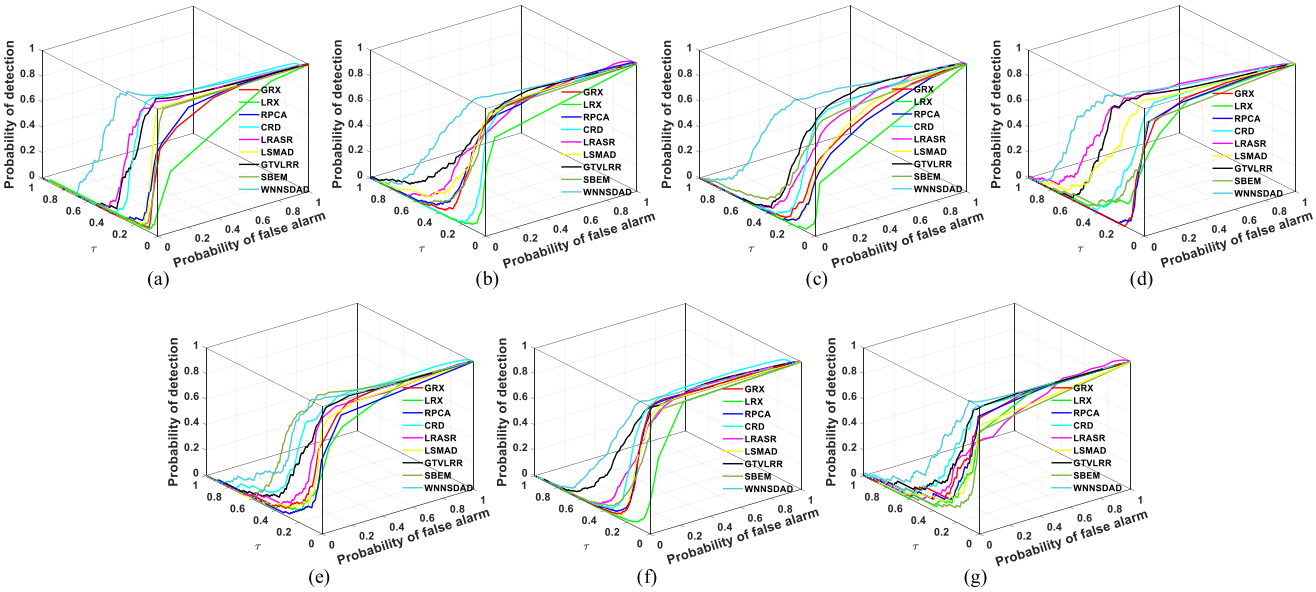



Fig. 5. Three-dimensional ROC curves of the nine anomaly detectors for the seven real-world HSI datasets. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII.

materials could also be well-represented; therefore, the background and anomalies could be separated effectively. Fig. 9 shows the false-images of the separated background part through different background dictionary construction methods. Ideally, we expected the separated background part to be consistent with the background part in the original HSI; also, the lower the brightness of the anomaly targets, the better. From Fig. 9, we can see that the background false-color images based on the proposed background construction method were closer to the real background part. For dataset III considering the *k*-means-based and DPC-based dictionaries, the background false-color images exhibited more noise and color distortion. However, the sparse dictionary provided a smooth and clear background. For datasets IV and VI, the brightness of the anomaly targets was close to black based on the sparse dictionary, but was almost the same as the original HSI considering the *k*-means-based and DPC-based dictionaries. This is because the *k*-means-based and DPC-based dictionaries were incomplete, which affected the expression ability of the background part.

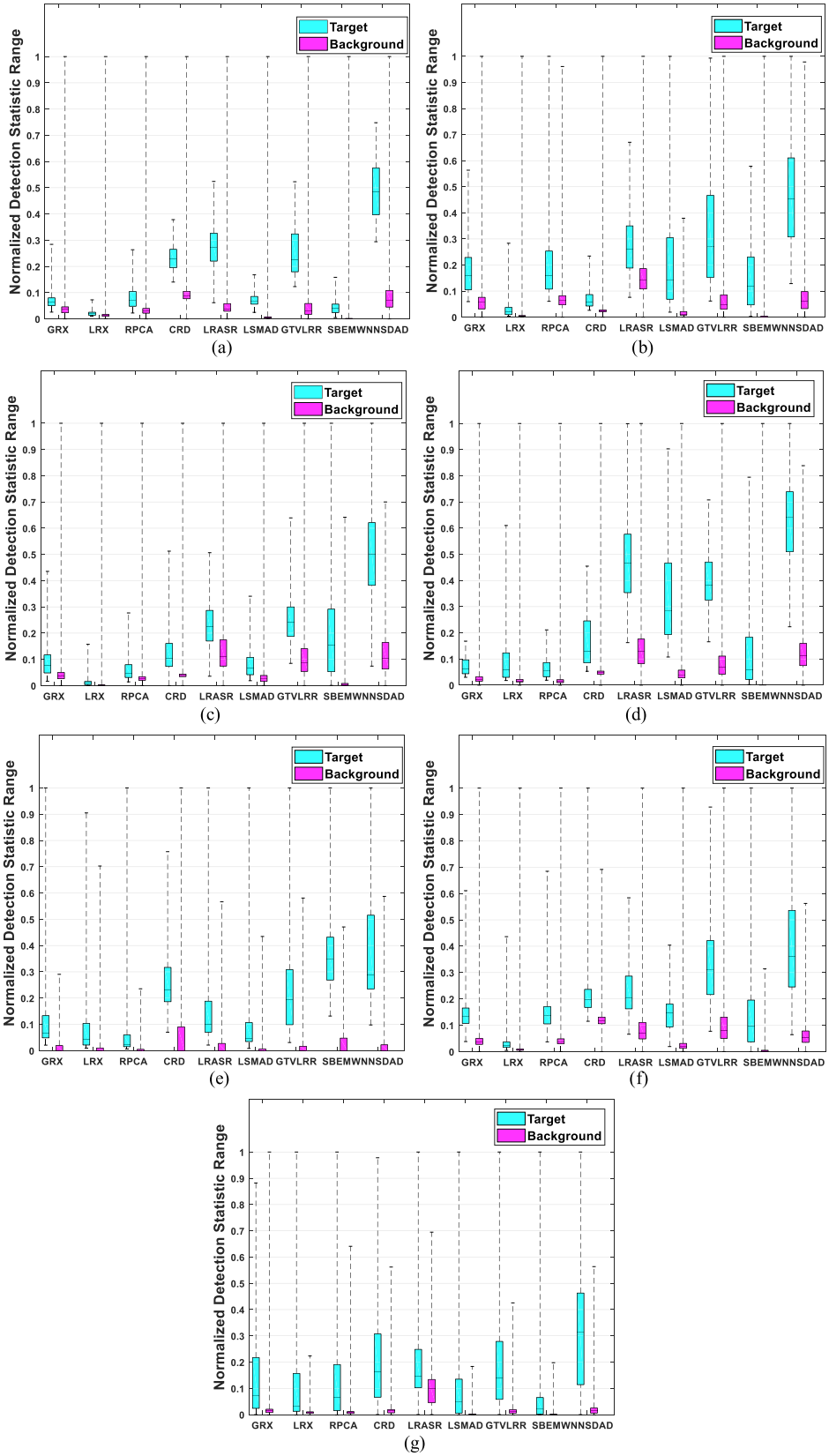



Fig. 6. Box-whisker plots for the nine anomaly detection methods for the seven HSI datasets. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII.

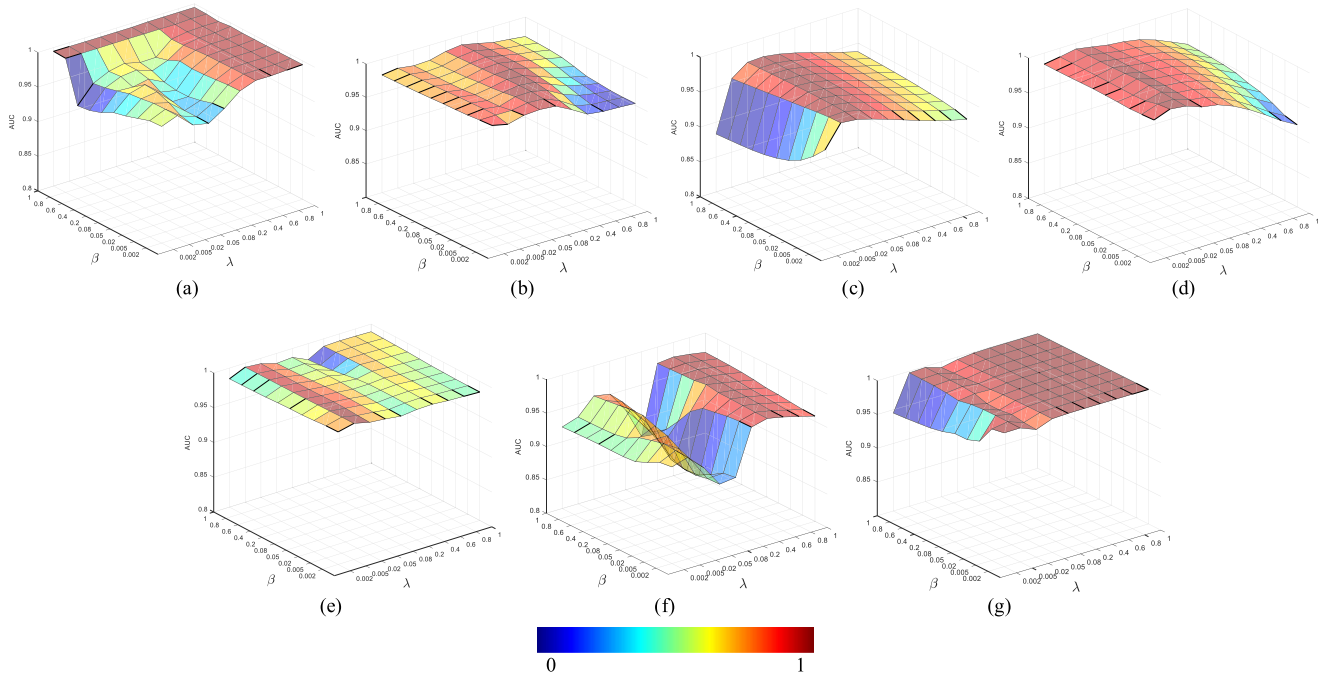



Fig. 7. AUC values of the WNNSDAD with different values of λ and $\beta$. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII.

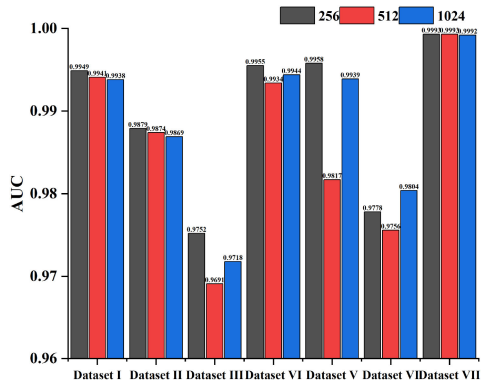



Fig. 8. AUC values of various background dictionary sizes for the seven datasets.

### *E. Performance of Reweighted Nuclear Norm and TV Regularization*

In this section, we first examined the influence of the reweighted nuclear norm and the standard nuclear norm. Taking dataset IV as an example, under the same parameter settings, the reweighted nuclear norm and the standard nuclear norm produce different results. As shown in Fig. 10, the contour of abnormal aircraft in Fig. 10(a1) is clearer, whereas the contour in Fig. 10(b1) is blurred, especially for the first and second aircraft, see the parts marked with white oval. In addition, it can be seen from the detection maps that the reweighted nuclear norm leads to smaller false alarm rate, as shown in the white oval in Fig. 10(a2) and (b2). The main reason is that the reweighted nuclear norm introduces a weight factor into the iterative optimization, which imposes a smaller penalty on the larger singular value and a larger penalty on the smaller singular value. It can effectively approach the original background part and more reliably separate the background and anomalies.

To check the benefit of the TV regularization, we removed the constraint of TV regularization on the representation coefficient of the background dictionary. The optimization objective function without TV regularization is

$$\min_{\mathbf{X},\mathbf{V}_1} \|\mathbf{Y}-\mathbf{A}\mathbf{X}-\mathbf{S}\|_F^2 + \|\mathbf{V}_1\|_{w,*} + \beta\|\mathbf{S}\|_{2,1} \quad \text{s.t.} \quad \mathbf{V}_1 = \mathbf{X}. \tag{22}$$

With the number of the background dictionary atoms and the tradeoff parameter $\beta$ remain unchanged, we took datasets I and II as examples. The false-color images of the separated background parts are shown in Fig. 11. It can be seen that compared with Fig. 11(a2) and (b2), Fig. 11(a1) and (b1) shows more satisfactory performance in separating the background part.

## IV. Discussion

From Fig. 9 and Table IV, we saw how the proposed dictionary construction method overcame the influence of cluster parameter settings, subjectively to a certain extent. Important characteristic spectral information was effectively retained through the sparse coding technique; moreover, the background representation ability was enhanced noticeably. The sparse dictionary construction method could be applied for hyperspectral target detection. As there are some known target spectra, we could design a union-dictionary model similar to [28]. Additionally, the sparse

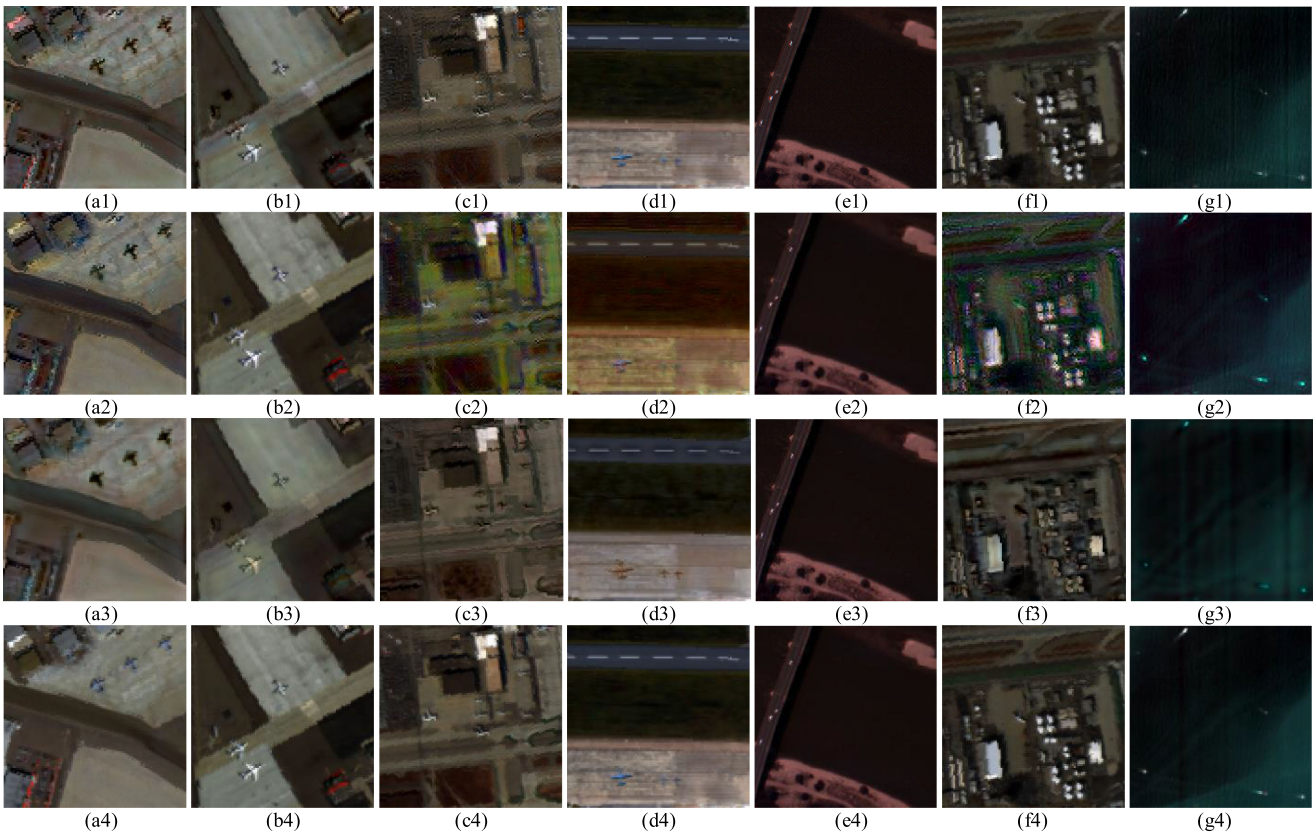



Fig. 9. False-color images [bands 37, 18, and 8 as RGB for (a)–(f), and bands 70, 22, and 17 as RGB for (g)] of the separated background part based on different background dictionary construction methods. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V. (f) Dataset VI. (g) Dataset VII. Lines 1, 2, and 3 express the *k*-means-based, DPC-based, and sparse dictionaries, respectively. As a reference, Line 4 represents the false-color images of the original HSI.

dictionary construction method could also be applied in other HSI processing fields, which have a similar working principle with LRSMD, such as HSI noise reduction and HSI unmixing, among others.

The effectiveness of WNNSDAD in detecting anomalies and in separating background and anomalies was illustrated by a number of experiments considering real-world datasets. However, from Figs. 3 and 5, WNNSDAD was seen to be inadequate in suppressing false alarms, especially for the San Diego and Los Angeles Airport datasets. This primarily occurred because in comparison to the backgrounds of other datasets, the backgrounds of datasets I and III were more complex; moreover, the shape of the anomaly targets was irregular, which could have contributed to challenges in detection. However, in the final analysis, we needed to essentially address the problem of a large false alarm rate for complex hyperspectral datasets, which generally exists in most hyperspectral target and anomaly detection methods.

Recently, deep learning methods have garnered significant attention, and a considerable volume of algorithms based on deep learning for HSI processing have been proposed [46]–[48]. Xie *et al.* [32] and Li *et al.* [33] utilized autoencoders and cascaded autoencoders, respectively, to increase the difference between background and anomalies. To a certain extent, SBEM helped overcome the difficulty in network training without any prior abnormal information. However, it also faced limitations in learning-based methods, for example, the cumbersome hyperparameter tuning process and the network structure were more dependent on data types. Additionally, training an adversarial network is more challenging, which limits the practical application of this type of method. Going ahead, we aim to design a semisupervised/unsupervised network structure for hyperspectral anomaly detection while researching learning-based methods. Furthermore, we will mainly research the impact of data type in tuning hyperparameters, to reduce the operation time utilized by parameter tuning in learning-based methods.

Through numerous comparative experiments, both representation-based and LRSMD-based models needed their tradeoff parameters to be tuned multiple times, or needed to be based on empiricism. Although WNNSDAD was very robust, considering the change in tradeoff parameters (see Fig. 7), it still faced challenges in parameter adjustment in practical application. As shown in Fig. 7, for most datasets, $\lambda$ tends to play a more important role. Moreover, for the datasets with more complex background distribution, the change of $\beta$ also affects the AUC value, as shown in the results for datasets I and VI. If a strategy could be designed to adaptively estimate the parameter range based on various data types, the efficiency of the algorithm could be greatly improved. We will focus on this issue in future research. Additionally, by analyzing the tradeoff parameters, $\lambda$ was easily found to play a greater role in detection performance than $\beta$. This indicated

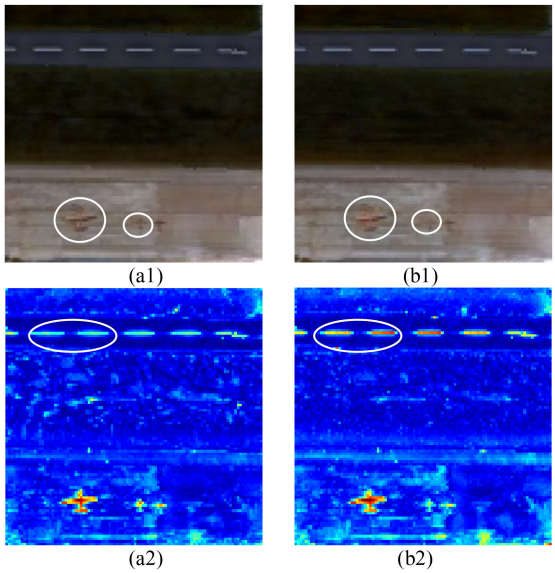



Fig. 10. Results of different nuclear norm solutions. Lines 1 and 2 express the false-color images (bands 37, 18, and 8 as RGB) of the separated background part and final detection maps. (a) Reweighted nuclear norm. (b) Standard nuclear norm.

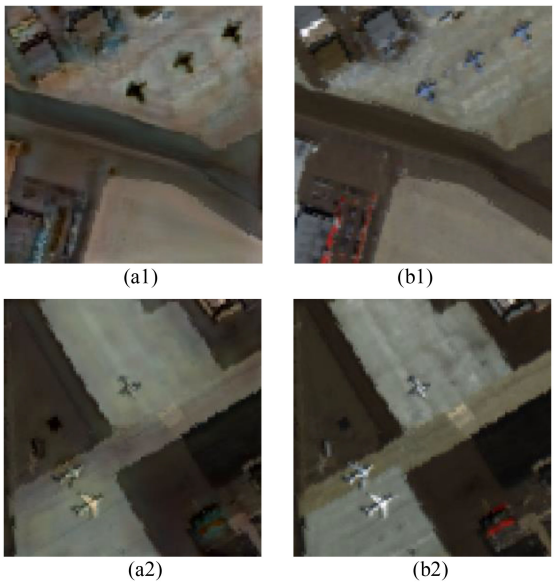



Fig. 11. False-color images (bands 37, 18, and 8 as RGB) of the separated background part based on TV regularization and without TV regularization. Lines 1 and 2 are results for Datasets I and II, respectively. (a) With TV regularization. (b) Without TV regularization.

the importance of spatial correlation in HSIs. Therefore, fully exploring the spatial information in HSIs is also important toward improving detection performance. Additionally, we found that the background dictionary size has a small influence on the AUC value, but a larger background dictionary will cost more optimization time. Hence, in practical applications, a smaller dictionary size tends to be a preferable choice.

In general, for most datasets, the anomalies have the property of global sparsity and low probability of occurrence. We found that the accuracies of WNNSDAD for datasets III and VI are not as great as that for the other datasets. It is seen that the anomalies in dataset VI are densely distributed and with a larger probability of occurrence. Therefore, we believe that WNNSDAD is more suitable for the datasets with less anomalies and sparse distribution, such as datasets IV and VII in this article. This is because we only consider the sparsity of anomalies, an assumption similar to other LRSMD-based methods.

The RX method was utilized to choose training samples for background in our proposed background dictionary construction method since it is simple and fast to implement. Nevertheless, there is still uncertainty in the training samples extracted by RX. If the abnormal pixel is incorrectly selected as the background dictionary atom, it will not only pollute the background dictionary but will also have a negative impact on the separation between background and anomaly. To construct a pure background dictionary, we still need to investigate how to extract training samples for background more accurately in future research. For example, deep-learning-based methods may be used to explore the characteristics of data distribution, and the extracted features may be used as training samples instead of original ones.

## V. Conclusion

The WNNSDAD method was developed for hyperspectral anomaly detection in this article, which integrated reweighted nuclear norm and TV regularizations as prior constraints into the well-known LRSMD model. The reweighted nuclear norm precludes an unbalanced penalty for a singular value of a matrix. Therefore, the low rank property of the background was effectively preserved in an iterative process, so that the background could be better represented. To retain the spatial correlation of the HSIs, TV regularization was applied to construct a piecewise smoothing constraint. Additionally, we integrated anomaly detection with the sparse representation technique and proposed a sparse dictionary construction method. It helped to overcome the problem that the cluster-based background dictionary construction methods depended on the subjectivity setting of clustering parameters. Moreover, the sparse representation technique could help to effectively preserve the characteristic background spectrum, especially for small background materials. A complete background dictionary is an important condition for representing the background part, so as to better separate the background and anomalies. In future research, we will concentrate on schemes for more reliable dictionary construction and extension of the method for more complex distributions of the background and anomalies.

## References

[1] H. Li, R. Feng, L. Wang, Y. Zhong, and L. Zhang, "Superpixel-based reweighted low rank and total variation sparse unmixing for hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 629–647, Jan. 2021.

[2] X. R. Li, R. S. Huang, and L. Zhao, "Correntropy-based spatial-spectral robust sparsity-regularized hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1453–1471, Feb. 2021.

[3] J. T. Peng, Y. C. Zhou, W. W. Sun, Q. Du, and L. K. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1501–1515, Feb. 2021.

[4] Z. Wu, J. Sun, Y. Zhang, Y. Zhu, and Z. Wei, "Scheduling-guided automatic processing of massive hyperspectral image classification on cloud computing architectures," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3588–3601, Jul. 2021.

[5] S. Abbas *et al.*, "Characterizing and classifying urban tree species using bi-monthly terrestrial hyperspectral images in Hong Kong," *ISPRS J. Photogramm. Remote Sens.*, vol. 177, pp. 204–216, 2021.

[6] M. E. Paoletti, J. M. Haut, X. Tao, J. Plaza, and A. Plaza, "FLOP-Reduction through memory allocations within CNN for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5938–5952, Jul. 2021.

[7] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 137–150, 2018.

[8] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4913–4928, Jul. 2021.

[9] Y. Shi, J. Li, Y. Zheng, B. Xi, and Y. Li, "Hyperspectral target detection with RoI feature transformation and multiscale spectral attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5071–5084, Jun. 2021.

[10] X. Song, L. Zou, and L. Wu, "Detection of subpixel targets on hyperspectral remote sensing imagery based on background endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 2365–2377, Mar. 2021.

[11] F. Vincent and O. Besson, "One-step generalized likelihood ratio test for subpixel target detection in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4479–4489, Jul. 2020.

[12] R. Feng, H. Li, L. Wang, Y. Zhong, L. Zhang, and T. Zeng, "Local spatial constraint and total variation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2021, Art. no. 5512216, doi: 10.1109/TGRS.2021.3098814.

[13] W. Y. Xie *et al.*, "Unsupervised spectral mapping and feature selection for hyperspectral anomaly detection," *Neural Netw.*, vol. 132, pp. 144–154, 2020.

[14] H. J. Su, Z. Y. Wu, A. X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 195–211, 2020.

[15] 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.

[16] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5801–5820, Oct. 2016.

[17] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.

[18] H. Kwon and N. 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.

[19] 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.

[20] Z. Yuan, H. Sun, K. Ji, Z. Li, and H. Zou, "Local sparsity divergence for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1697–1701, Oct. 2014.

[21] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[22] N. Ma, Y. Peng, and S. Wang, "A fast recursive collaboration representation anomaly detector for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 588–592, Apr. 2019.

[23] L. Du, Z. Wu, Y. Xu, W. Liu, and Z. Wei, "Kernel low rank representation for hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2016, pp. 477–480.

[24] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, "Low rank decomposition-based anomaly detection," *Proc. SPIE*, vol. 8743, 2013, Art. no. 87430N.

[25] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[26] 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.

[27] T. Cheng and B. Wang, "Graph and total variation regularized low rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jun. 2020.

[28] T. Cheng and B. Wang, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1472–1486, Feb. 2021.

[29] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.

[30] L. Li, W. Li, Y. Qu, C. H. Zhao, R. Tao, and D. Qian, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2020.3038659.

[31] L. Zhang and B. Cheng, "Transferred CNN based on tensor for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2115–2119, Dec. 2020.

[32] W. Xie, B. Liu, Y. Li, J. Lei, and Q. Du, "Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5416–5427, Aug. 2020.

[33] Y. Li, T. Jiang, W. Xie, J. Lei, and Q. Du, "Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2021, Art. no. 5512811, doi: 10.1109/TGRS.2021.3102048.

[34] X. Jia, X. Feng, W. Wang, and L. Zhang, "Generalized unitarily invariant gauge regularization for fast low rank matrix recovery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 4, pp. 1627–1641, Apr. 2021.

[35] S. Javed, A. Mahmood, T. Bouwmans, and S. K. Jung, "Spatiotemporal low rank modeling for complex scene background initialization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1315–1329, Jun. 2018.

[36] Z. Wu *et al.*, "Total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *J. Electron. Imag.*, vol. 25, no. 1, 2016, Art. no. 013037.

[37] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3909–3921, Jul. 2017.

[38] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[39] Z. Li, G. Kuang, W. Yu, and Q. Xue, "Algorithm on small target detection based on principal component of hyperspectral imagery," *J. Infrared Millimeter Waves*, vol. 23, no. 4, pp. 286–290, 2004.

[40] M. Aharon, "An algorithm for designing over-complete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[41] G. Davis, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, no. 1, pp. 57–98, 1997.

[42] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Feb. 2008.

[43] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.

[44] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, Apr. 2011.

[45] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[46] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jul. 2021.

[47] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 30, pp. 2810–2825, Feb. 2021.

[48] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 535–549, Jan. 2021.

**Xiaoyi Wang** received the M.A. degree from Heilongjiang University, Harbin, China, in 2018. She is currently working toward the Ph.D. degree with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China.

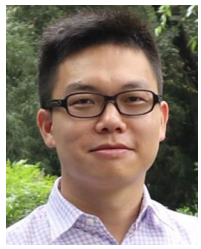
Her research interests include remote sensing image fusion and machine learning.

**Liguo Wang** received the M.A. degree in 2002 and Ph.D. degree in signal and information processing in 2005 from Harbin Institute of Technology, Harbin, China.

He is currently a Professor with the College of Information and Communication Engineering, Dalian Minzu University, Dalian, China, and an Adjunct Professor with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China. His research interests include remote sensing image processing and machine learning. He has authored/coauthored three books, 25 patents, and more than 170 papers in journals and conference proceedings.

**Jiawen Wang's**, photograph and biography not available at the time of publication.

**Kaipeng Sun's**, photograph and biography not available at the time of publication.



**Qunming Wang** received the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. From 2017 to 2018, he was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K. His three-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored 70 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang is a member the Editorial Board of *Remote Sensing of Environment*, and is an Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering and Remote Sensing*. He was an Associate Editor for *Computers and Geosciences* (2017−2020).