

Dual Collaborative Constraints Regularized Low-Rank and Sparse Representation via Robust Dictionaries Construction for Hyperspectral Anomaly Detection

Sheng Lin , Min Zhang , Xi Cheng, Kexue Zhou , Shaobo Zhao, and Hai Wang 

Abstract—The low rank and sparse representation (LRSR) technique has attracted increasing attention for hyperspectral anomaly detection (HAD). Although a large quantity of research based on LRSR for HAD is proposed, the detection performance is still limited, due to the unsatisfactory dictionary construction and insufficient consideration of global and local characteristics. To tackle the above-mentioned concern, a novel HAD method, termed dual collaborative constraints regularized low-rank and sparse representation via robust dictionaries construction, is proposed in this article. Concretely, a robust dictionary construction strategy, which thoroughly excavates the potential of the density estimation model and local outlier factor, is proposed to yield pure and representative dictionary atoms. To fully exploit the global and local characteristics of hyperspectral images, dual collaborative constraints corresponding to the background and anomaly components are imposed on the LRSR model. Notably, two weighted matrices are further exerted on the representation coefficients to improve the effect of collaborative constraints, considering the fact that the surrounding pixels similar to the testing pixel should be given a large weight, otherwise the weight is expected to be small. In this way, the background and anomaly components can be well modeled. Additionally, a nonlinear transformation operation, which combines the output of the density estimation model and local outlier factor with the detection result derived from the LRSR model, is developed to suppress the background. The experiments conducted on one simulated dataset and three real datasets demonstrate the superiority of the proposed method compared with the four typical methods and four state-of-the-art methods.

Index Terms—Anomaly detection, dual collaborative constraints, hyperspectral images (HSIs), low-rank and sparse representation (LRSR), robust dictionaries construction.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) have been widely used in the field of band selection [1], [2], [3], image classification [4], [5], [6], hyperspectral pansharpening [7], [8],

Manuscript received 25 May 2022; revised 12 August 2022 and 14 September 2022; accepted 11 October 2022. Date of publication 14 October 2022; date of current version 16 February 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 12003018, in part by the Fundamental Research Funds for the Central Universities under Grant XJS191305, and in part by the China Postdoctoral Science Foundation under Grant 2018M633471. (Corresponding author: Min Zhang.)

The authors are with the School of Aerospace Science and Technology, Xidian University, Xi'an 710071, China (e-mail: linsheng@stu.xidian.edu.cn; minzhang@xidian.edu.cn; cheng_xi@stu.xidian.edu.cn; kxzhou@stu.xidian.edu.cn; shaobozhao@stu.xidian.edu.cn; wanghai@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3214508

[9], hyperspectral unmixing [10], [11], anomaly detection [12], [13], [14], [15], and target detection [16], [17], [18], owing to the rich spectral information. Hyperspectral anomaly detection (HAD), which aims to search for spectral signatures deviating from the background, has become a research hot topic due to its widespread application in military defense, maritime rescue, and mineral exploration. Compared with hyperspectral target detection, HAD brings more challenges because of the absence of prior information about the anomaly.

In recent years, many researchers proposed various methods to detect anomalies in the HSIs. The most typical method is the RX detector [19], which is proposed by Reed and Yu. Also, it is regarded as the milestone of the statistic theory-based methods for HAD. The RX detector holds that the background conforms to the multivariate Gaussian distribution, and the mean value vector and covariance matrix of all pixels belonging to the HSI are estimated to calculate the Mahalanobis distance of each testing pixel. Different from the RX detector, which utilizes the global statistic characteristics, the Local RX [20] employs dual windows to acquire the local statistic characteristics for a better background model. Nevertheless, two-aspect challenges, which consist of the improper multivariate Gaussian assumption of background and the inaccurate background model in the interference of anomaly, exist in the aforementioned methods for the real hyperspectral scenes. To resolve the above-mentioned issues, various extensions of RX detector, containing subspace RX [21], kernel RX [22], random-selection-based anomaly detector [23], multiwindow RX [24], [25], and fractional Fourier entropy RX [12], are gradually proposed to promote the detection accuracy. However, these methods can hardly accurately model the background, with the interference of the anomalies and noise.

To model the background as accurately as possible, the representation-based methods, which can well represent the background pixels and poorly express the anomalous pixels, are widely used for HAD. The residual of the representation-based methods can be employed to measure the anomalies. In general, the larger the residual of a testing pixel, the higher the possibility of a testing pixel being an anomaly. The representation-based methods can be separated into three categories: sparse representation-based methods, collaborative representation (CR)-based methods, and low-rank representation-based methods, considering the different regularization constraints imposed

on the representation models. The sparse representation assumes that the background pixels can be linearly expressed by the atoms belonging to the overcomplete dictionary, whereas the representation effect for the anomalous pixels is terrible. The background joint sparse representation detector, which is a typical representative to explore the sparse model, models the background by selecting the representative atoms in the local regions [26]. Additionally, constructing a reliable overcomplete dictionary is the other key research direction for the sparse representation-based methods. The discriminative feature learning with multiple-dictionary sparse representation detector, proposed by Ma et al. [27], is an effective strategy to construct a reliable overcomplete dictionary for HAD. However, due to the lack of prior information, it is hard to completely prevent interference from anomalies, constructing an accurate overcomplete dictionary.

The classical CR detector (CRD), proposed by Li and Du [28], holds that background pixels can be linearly represented by the pixels within the dual windows, whereas the anomalous pixels could not. Unfortunately, the pixels located in the dual windows are vulnerable to being contaminated by the anomalous pixels, resulting in unsatisfactory detection performance. To alleviate this problem, a large number of variants, including dual CR [29], self-weighted CR [30], and outlier removal-based CR [31], are proposed to enhance detection performance. With regard to the CR-based methods, the local characteristics are considered in modeling the background, whereas the global attributes fail to be exploited, leading to limited detection performance.

In recent years, the low-rank representation-based methods play a vital role in HSI processing, such as denoising [32], [33], clustering [34], [35], classification [36], [37], and anomaly detection [38], [39]. Different from the CR-based methods, the low-rank representation-based methods assume that the background has a low-rank attribute and the anomalies are treated as sparse. The low-rank and sparse matrix decomposition (LSMAD) method proposed by Zhang et al. [40] is an early attempt to model the background with the low-rank property of background and sparsity of anomalies. Further, the low-rank and sparse decomposition method with a mixture of Gaussian (LSDM-MoG) [41] is proposed to better characterize data distribution. Furthermore, the background component is converted into the product of the background dictionary and the representation coefficient matrices, which has attracted increasing attention due to the great potential in the construction of the background dictionary. The low-rank and sparse representation (LRASR) method [42] can be regarded as a successful exploration for expressing the background part with the background dictionary and representation coefficient matrices. However, the purity of the background dictionary cannot be guaranteed for LRASR. To tackle this issue, Jiang et al. [43] proposed a low-rank embedded network to detect anomalies under the condition of the sample-free. Xie et al. [44] proposed a weakly supervised low-rank representation method for constructing a representative and discriminative background dictionary. To further effectively detect anomalies and lower the influence of noise, Huyan et al. [45] proposed a novel dictionary construction strategy, termed potential anomaly and background dictionary construction (PAB-DC), to describe the anomaly and

background characteristics. Notably, the above low-rank representation-based methods only consider the global information, and the local attributes are ignored. To fully consider the local geometrical structure information and global property, Cheng and Wang [46] imposed the graph and total variation regularized terms on the low-rank representation model. In addition, Su et al. [47] proposed a low-rank and CRD (LRCRD), considering both the global attribute and local property. Although the global attribute and local property are considered in LRCRD, the collaborative constraint related to the anomaly component in the dual dictionaries-based low-rank and sparse representation model fails to be considered. Clearly, the anomalous pixels in HSIs can be represented by the potential anomaly dictionary, indicating that CR is suitable for the anomaly component. Besides, the purity and representativeness of dictionaries are vital to detection performance.

To achieve this goal, a novel HAD method, termed as dual collaborative constraints regularized low-rank and sparse representation via robust dictionaries construction, is proposed in this article, as illustrated in Fig. 1. First, a robust dictionaries construction strategy is proposed, jointly leveraging the density maps generated by density estimation model and local outlier factor (LOF). Dual collaborative constraints are then imposed on the low-rank and sparse representation model, in the comprehensive consideration of the global property and local attribute of the HSI. Furthermore, two weighted matrices are exerted on the representation coefficients considering the fact that different atoms should have different importance. Finally, a nonlinear transformation operation is introduced to perform the background suppression, expanding the difference between the background and anomaly. Compared with the existing HAD methods, the main contributions of our method are as follows.

- 1) A novel dual dictionaries construction strategy based on two-stream density maps is proposed in this article. With the dictionary construction strategy, the pure and representative pixels can be selected to act as the dictionary atoms.
- 2) The dual collaborative constraints, which consider both background and anomaly components, are imposed into the low-rank and sparse representation model. In this way, global and local information can be fully exploited.
- 3) To carry out the background suppression, a nonlinear transformation operation, in consideration of the two-stream density maps, is developed to exert on the detection result generated by the low-rank and sparse representation model.

The remainder of this article is organized as follows. Section II presents the related work. The proposed method is presented in Section III. Section IV introduces the experiments. Finally, the conclusions are outlined in Section V.

II. RELATED WORK

This section is to introduce the classical models utilized in this article. The detailed description of these classical models, consisting of low rank representation (LRR) model and CR model, is as follows.

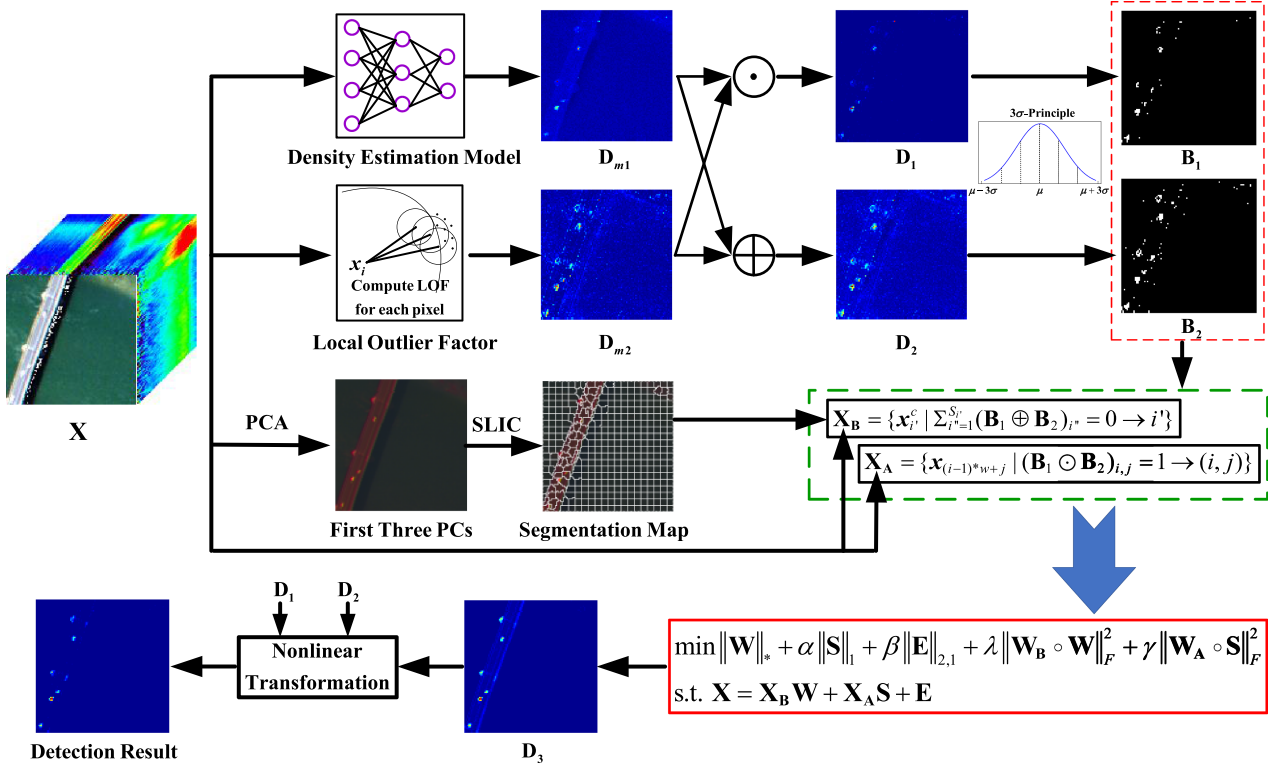


Fig. 1. Schematic of the proposed method. D_{m1} and D_{m2} separately denote the density maps generated by the density estimation model and local outlier factor. D_1 , D_2 , and D_3 represent the background suppression map, anomalies enhancement map, and initial detection result, respectively. B_1 and B_2 indicate binary maps generated from D_1 and D_2 .

A. Low-Rank and Sparse Representation Model

In HSIs, the background is regarded as low rank, whereas the anomalies are sparse [42]. Inevitably, noise with stochastic or deterministic characteristics should be considered due to its corruption for HSIs [40]. Based on the above-mentioned analysis, the HSI \mathbf{Y} can be decomposed into three components, i.e., the low-rank component \mathbf{B} , sparse component \mathbf{A} , and noise component \mathbf{N} , which are expressed as follows:

$$\mathbf{Y} = \mathbf{B} + \mathbf{A} + \mathbf{N}. \quad (1)$$

Considering the fact that there is a strong correlation between background pixels [42] in HSIs, indicating that each of them can be linearly represented by some representative background pixels. Similarly, each anomalous pixel can be expressed by the potential prior of anomalies in HSIs [45]. In this way, the (1) can be converted into the following equation:

$$\mathbf{Y} = \mathbf{D}_B \mathbf{W}_Y + \mathbf{D}_A \mathbf{S}_Y + \mathbf{N} \quad (2)$$

where $\mathbf{D}_B \mathbf{W}_Y$ and $\mathbf{D}_A \mathbf{S}_Y$ separately symbolize the low-rank term and sparse term. \mathbf{D}_B and \mathbf{D}_A stand for the background and potential anomaly dictionaries, respectively. \mathbf{W}_Y and \mathbf{S}_Y refer to the representation coefficient matrices corresponding to the low-rank term and sparse term, respectively.

To model the problem mentioned above, an optimization objective function is expressed as follows:

$$\begin{aligned} \min_{\mathbf{W}_Y, \mathbf{S}_Y, \mathbf{N}} & \|\mathbf{W}_Y\|_* + \nu_1 \|\mathbf{S}_Y\|_1 + \nu_2 \|\mathbf{N}\|_{2,1} \\ \text{s.t. } & \mathbf{Y} = \mathbf{D}_B \mathbf{W}_Y + \mathbf{D}_A \mathbf{S}_Y + \mathbf{N} \end{aligned} \quad (3)$$

where $\|\cdot\|_*$, $\|\cdot\|_1$, and $\|\cdot\|_{2,1}$ separately indicate the nuclear norm, l_1 norm, and $l_{2,1}$ norm. $\nu_1 > 0$ and $\nu_2 > 0$ are the regularization coefficients to balance the contribution of each term.

B. Collaborative Representation Model

The core idea of CR model is that the background pixels can be represented by the surrounding pixels, whereas the anomalous pixels cannot [28]. Let \mathbf{x} denote the testing pixel, \mathbf{D}_S represent the background dictionary, and φ indicate the weight vector. The goal of the CR is to determine the weight vector φ by minimizing both $\|\mathbf{x} - \mathbf{D}_S \varphi\|_2^2$ and $\|\varphi\|_2^2$. As a result, the objective function can be described as

$$\arg \min_{\varphi} \|\mathbf{x} - \mathbf{D}_S \varphi\|_2^2 + v \|\varphi\|_2^2 \quad (4)$$

where v indicates the regularization parameter. Generally speaking, the surrounding pixels that are similar to the testing pixel should be given a large weight, whereas the surrounding pixels which are different from the testing pixel should have a smaller weight. Accordingly, the distance-weighted-based optimization

problem becomes

$$\arg \min_{\varphi} \|\mathbf{x} - \mathbf{D}_S \varphi\|_2^2 + v \|\mathbf{\Gamma}_x \cdot \varphi\|_2^2 \quad (5)$$

where $\mathbf{\Gamma}_x = \text{diag}(\|\mathbf{x} - \mathbf{d}_1\|_2, \|\mathbf{x} - \mathbf{d}_2\|_2, \dots, \|\mathbf{x} - \mathbf{d}_s\|_2)$ indicates the diagonal regularization matrix, each item of which denotes the diagonal entry, and $\text{diag}(\cdot)$ represents the operation to construct a diagonal matrix. $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_s$ are the columns of \mathbf{D}_S .

III. PROPOSED METHOD

Generally speaking, the 3-D HSI cubic can be converted into a 2-D matrix denoted as $\mathbf{X} \in \mathbb{R}^{d \times N}$, in which d and N signify the number of spectral bands and pixels in an HSI, respectively. For an HSI, the background and anomalies are separately deemed as low rank and sparse [42]. To distinguish the background and anomaly, the dual dictionaries-based representation method is adopted, which can be expressed as $\mathbf{X} = \mathbf{X}_B \mathbf{W} + \mathbf{X}_A \mathbf{S} + \mathbf{E}$, where $\mathbf{X}_B \mathbf{W} \in \mathbb{R}^{d \times N}$, $\mathbf{X}_A \mathbf{S} \in \mathbb{R}^{d \times N}$, and $\mathbf{E} \in \mathbb{R}^{d \times N}$ separately symbolize the low-rank term, sparse term, and noise term. $\mathbf{X}_B \in \mathbb{R}^{d \times m}$ and $\mathbf{X}_A \in \mathbb{R}^{d \times n}$ separately stand for the background and potential anomaly dictionaries, where m and n separately represent the number of atoms in background dictionary and potential anomaly dictionary. $\mathbf{W} \in \mathbb{R}^{m \times N}$ and $\mathbf{S} \in \mathbb{R}^{n \times N}$ refer to the representation coefficient matrices corresponding to the low-rank term and sparse term, respectively.

In an HSI, the CR theory is widely used to characterize the background [28], [30]. With this situation, to exploit the global low-rank and local collaborative attributes, Su et al. [47] regularized the low-rank representation model with l_2 norm of each column of the representation coefficient matrix corresponding to the low-rank component. Motivated by this, dual collaborative constraints, which are composed of the l_2 norm of each column of representation coefficient matrices belonging to the background and anomaly components, are imposed on the low-rank and sparse representation model to act as the regularization terms. In this case, the optimization objective function in the consideration of dual collaborative constraints can be reformulated as

$$\begin{aligned} & \min_{\mathbf{W}, \mathbf{S}, \mathbf{E}} \|\mathbf{W}\|_* + \alpha \|\mathbf{S}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda \sum_{i=1}^N \|\mathbf{w}_i\|_2^2 \\ & + \gamma \sum_{i=1}^N \|\mathbf{s}_i\|_2^2 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}_B \mathbf{W} + \mathbf{X}_A \mathbf{S} + \mathbf{E} \end{aligned} \quad (6)$$

where $\alpha > 0$, $\beta > 0$, $\lambda > 0$, and $\gamma > 0$ stand for the regularization coefficients to tradeoff each term. \mathbf{w}_i and \mathbf{s}_i refer to the i th column of \mathbf{W} and \mathbf{S} , respectively. Note that $\sum_i \|\mathbf{w}_i\|_2^2 = \|\mathbf{W}\|_F^2$ and $\sum_i \|\mathbf{s}_i\|_2^2 = \|\mathbf{S}\|_F^2$ in (6). Therefore, the optimization objective function can be converted into

$$\begin{aligned} & \min_{\mathbf{W}, \mathbf{S}, \mathbf{E}} \|\mathbf{W}\|_* + \alpha \|\mathbf{S}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda \|\mathbf{W}\|_F^2 + \gamma \|\mathbf{S}\|_F^2 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}_B \mathbf{W} + \mathbf{X}_A \mathbf{S} + \mathbf{E}. \end{aligned} \quad (7)$$

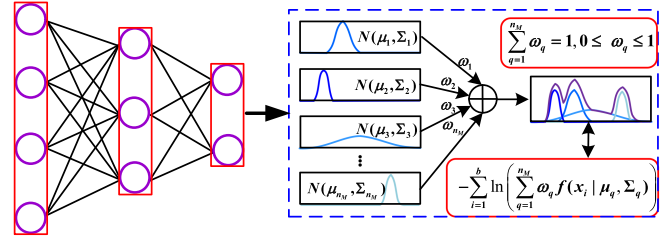


Fig. 2. Schematic of the density estimation model. $N(\mu_q, \Sigma_q)$ and ω_q separately indicate the q th Gaussian component and corresponding weight.

In general, the CR-based methods assume that the surrounding pixels that are similar to the center pixel should be given a large weight, whereas the surrounding pixels which are different from the center pixel should have a smaller weight [28]. To this end, two weighted matrices are imposed on the background and anomaly representation coefficients to adjust the contribution of each atom in the dictionary. Therefore, (7) can be written as

$$\begin{aligned} & \min_{\mathbf{W}, \mathbf{S}, \mathbf{E}} \|\mathbf{W}\|_* + \alpha \|\mathbf{S}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda \|\mathbf{W}_B \circ \mathbf{W}\|_F^2 \\ & + \gamma \|\mathbf{W}_A \circ \mathbf{S}\|_F^2 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}_B \mathbf{W} + \mathbf{X}_A \mathbf{S} + \mathbf{E} \end{aligned} \quad (8)$$

where $\mathbf{W}_B = \{w_{B(i,j)}\}_{i=1, j=1}^{i=m, j=N}$ and $\mathbf{W}_A = \{w_{A(i,j)}\}_{i=1, j=1}^{i=n, j=N}$ represent the weighted matrices corresponding to background and anomaly terms, respectively, in which $w_{*(i,j)}$ indicates the Euclidean distance between the i th dictionary atom and the j th pixel of HSI. \circ denotes the Hadamard product.

A. Density Estimation via Gaussian Mixture Components

In the existing HAD task, the HSIs are generally considered to follow a single multivariate Gaussian distribution [19], [22]. Apparently, this assumption is not applicable to the real scene due to the highly complicated characteristics of the background of HSIs [43]. As a result, we consider the mixture of Gaussian distribution to characterize the HSIs. There are twofold effects for Gaussian mixture model (GMM), which can be not only for density estimation but also for clustering. A detailed description of GMM-based density estimation is as follows, and the GMM-based clustering is recounted in Section III-C.

To implement GMM-based density estimation, a density estimation model, which contains n_H fully connected layers, is established, as illustrated in Fig. 2. Clearly, there is no direct relation between the density estimation model and GMM. To cope with this concern, the output of the density estimation model is set to the number of Gaussian mixture components. In this way, the density estimation model can be closely combined with GMM.

Let $f(x)$ represent the probability density function (pdf), which can be expressed as

$$f(x) = \frac{1}{\sqrt{2\pi} |\Sigma|} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (9)$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and covariance matrix, respectively. $|\cdot|$ indicates the determinant of a matrix. Clearly, $f(\mathbf{x})$ can be determined by $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Therefore, we termed it as $f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$. In this case, the pdf of GMM can be defined as

$$f_M(\mathbf{x}) = \sum_{q=1}^{n_M} \omega_q f(\mathbf{x}|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \quad (10)$$

where $\omega_q > 0$ denotes the coefficient to balance the Gaussian components, and the sum of all coefficients is 1. $\boldsymbol{\mu}_q$ and $\boldsymbol{\Sigma}_q$ represent the mean vector and covariance matrix corresponding to the q th Gaussian mixture component, respectively. n_M signifies the number of Gaussian mixture components. It is worth noting that ω_q , $\boldsymbol{\mu}_q$, and $\boldsymbol{\Sigma}_q$ are unknown, indicating that these parameters should be estimated. Generally speaking, the parameters of pdf can be estimated by the maximum-likelihood estimation, which can be formulated as follows:

$$\begin{aligned} \max_{\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\Sigma}} \sum_{i=1}^N \ln \left(\sum_{q=1}^{n_M} \omega_q f(\mathbf{x}_i|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \right) \\ \text{s.t. } \begin{cases} 0 \leq f(\mathbf{x}_i|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \leq 1 \\ \sum_{q=1}^{n_M} \omega_q = 1, \omega_q \geq 0 \end{cases} \end{aligned} \quad (11)$$

To train the density estimation model, the maximum-likelihood function of GMM is employed to act as the loss function. However, the maximum-likelihood function is to calculate the maximum value, whereas the loss function of the density estimation model should be minimized. To this end, the inverse of the maximum-likelihood function of GMM is adopted to act as the loss function to learn the density estimation model. The detailed loss function can be formulated as

$$l = - \sum_{i=1}^b \ln \left(\sum_{q=1}^{n_M} \omega_q f(\mathbf{x}_i|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \right) \quad (12)$$

$$\omega_q = \sum_{i=1}^b \frac{o_{iq}}{b} \quad (13)$$

$$\boldsymbol{\mu}_q = \frac{\sum_{i=1}^b o_{iq} \mathbf{x}_i}{\sum_{i=1}^b o_{iq}} \quad (14)$$

$$\boldsymbol{\Sigma}_q = \frac{\sum_{i=1}^b o_{iq} (\mathbf{x}_i - \boldsymbol{\mu}_q)(\mathbf{x}_i - \boldsymbol{\mu}_q)^T}{\sum_{i=1}^b o_{iq}} \quad (15)$$

where $o_i = \text{softmax}(N_{\text{dem}}(\mathbf{x}_i; \theta_{\text{dem}}))$ signifies the output of the density estimation model corresponding to the i th training sample, in which N_{dem} represents the mapping operation of the density estimation model, and θ_{dem} symbolizes the parameters of the density estimation model. b indicates the batch size of training samples. Once the training process finished, the HSI can be fed into the density estimation model to obtain the output, which can be employed to calculate the parameters of GMM by (13)–(15). With this condition, we can acquire the density map through utilizing pdf under the estimated parameters. Nevertheless, there is low density value for the position corresponding to anomalous pixels, which is contrary to our original intension. To tackle this concern, the negative log pdf is employed to act as the indicator

for obtaining the density map, which can be formulated as

$$\begin{aligned} d_{m1}^i &= -\log \left(\sum_{q=1}^{n_M} \omega_q f(\mathbf{x}_i|\boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q) \right) \\ &= -\log \left(\sum_{q=1}^{n_M} \frac{\omega_q}{\sqrt{2\pi} |\boldsymbol{\Sigma}_q|} e^{-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_q)^T \boldsymbol{\Sigma}_q^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_q)} \right) \end{aligned} \quad (16)$$

where d_{m1}^i denotes the density value of the i th pixel in HSI. When all pixels of an HSI are considered, a 2-D density map \mathbf{D}_{m1} can be obtained.

B. Density Evaluation Based on Local Outlier Factor

LOF [48], which is first proposed for outlier detection, is an effective strategy to evaluate the anomalous degree of each sample. Clearly, this technique can serve as a performance indicator to determine the probability that each sample belongs to an anomaly. In addition, the LOF can also be regarded as a kind of density evaluation without any distribution assumptions. Based on these aspects, the LOF can be employed to evaluate the density of each pixel in HSI. Taking the testing pixel \mathbf{x}_t in HSI as an example, the detailed description of LOF can be expressed as follows.

- 1) Compute the Euclidean distance of testing pixel \mathbf{x}_t and other pixels and define the k th nearest pixel of \mathbf{x}_t as k -distance.
- 2) Set the collection to the pixels whose distance from \mathbf{x}_t is less than k as $N_k(\mathbf{x}_t)$.
- 3) Define the reachability distance of \mathbf{x}_t to a pixel \mathbf{x}_t^o in $N_k(\mathbf{x}_t)$ as

$$\text{reach} - \text{dist}_k(\mathbf{x}_t, \mathbf{x}_t^o) = \max \{k\text{-distance}(\mathbf{x}_t^o), d(\mathbf{x}_t, \mathbf{x}_t^o)\} \quad (17)$$

where $d(\mathbf{x}_t, \mathbf{x}_t^o)$ denotes the Euclidean distance of \mathbf{x}_t and \mathbf{x}_t^o .

- 4) Calculate the local reachability density of \mathbf{x}_t (i.e., $\text{lrd}_k(\mathbf{x}_t)$) in $N_k(\mathbf{x}_t)$, which can be defined as

$$\text{lrd}_k(\mathbf{x}_t) = \left(\frac{\sum_{\mathbf{x}_t^o \in N_k(\mathbf{x}_t)} \text{reach} - \text{dist}_k(\mathbf{x}_t, \mathbf{x}_t^o)}{|N_k(\mathbf{x}_t)|} \right)^{-1} \quad (18)$$

where $|N_k(\mathbf{x}_t)|$ represents the number of pixels in $N_k(\mathbf{x}_t)$.

- 5) Compute the LOF of \mathbf{x}_t with the following formula:

$$\text{LOF}(\mathbf{x}_t) = \frac{\sum_{\mathbf{x}_t^o \in N_k(\mathbf{x}_t)} \text{lrd}_k(\mathbf{x}_t^o)}{|N_k(\mathbf{x}_t)|} \cdot \frac{1}{\text{lrd}_k(\mathbf{x}_t)} \quad (19)$$

In the above case, the pixel whose LOF value approaches 1 is more likely to be the same class as k -nearest neighbors. In addition, the pixel whose LOF value is much greater than 1 will be reckoned as an anomaly, owing to the sparse characteristics of the anomalies in HSI. Through the above way, the anomaly degree of each pixel in HSI can be calculated. Eventually, the 2-D density map \mathbf{D}_{m2} is formed.

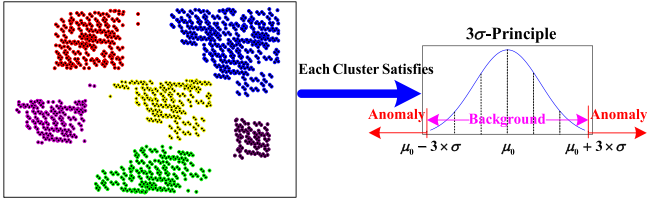


Fig. 3. Example to illustrate the process of anomaly judgment. Here, the points with the same color represent the same cluster, and each cluster obeys the Gaussian distribution.

C. Dual Collaborative Constraints Regularized Low-Rank and Sparse Representation

1) *Dual Dictionaries Construction*: In the field of HAD, the GMM-based density estimation approach or LOF-based density evaluation method is generally used to detect anomalies [49], [50], whereas there is a limited capability for each of them to measure the anomalous degree of testing pixels in HSIs. As a result, we jointly consider both GMM and LOF to estimate the anomalous degree of each pixel. Concretely, the elementwise product of \mathbf{D}_{m1} and \mathbf{D}_{m2} is employed to search for the most representative anomalies. In this way, the anomaly degree of anomalous pixels will be highlighted, whereas that of background pixels will be suppressed. Similarly, the addition operation pixel by pixel of \mathbf{D}_{m1} and \mathbf{D}_{m2} is executed to reinforce the potential anomalies. The above description can be expressed as

$$\mathbf{D}_1 = \mathbf{D}_{m1} \odot \mathbf{D}_{m2} \quad (20)$$

$$\mathbf{D}_2 = \mathbf{D}_{m1} \oplus \mathbf{D}_{m2}. \quad (21)$$

As stated in Section III-A, the Gaussian mixture distribution can be used to cluster the HSIs, which can be expressed as

$$c_i = \operatorname{argmax}(\mathbf{o}_{i1}, \dots, \mathbf{o}_{in_M}) \quad (22)$$

where $c_i \in \{1, 2, 3, \dots, n_M\}$ indicates the cluster to which the i th sample belongs to. $\operatorname{argmax}(\cdot)$ refers to the index of the maximum value. \mathbf{o}_{ik} stands for the k th output of the density estimation model for the i th sample. With this circumstance, each cluster can be deemed to obey a Gaussian distribution. In terms of the Gaussian distribution, normal samples fall into the interval of 3σ principle, i.e., $(\mu_0 - 3 \times \sigma, \mu_0 + 3 \times \sigma)$, whereas the outlier samples are outside of the aforementioned range, where μ_0 and σ are the mean value and standard deviation of the samples, respectively. For better comprehension, an example is exhibited in Fig. 3. Based on the above-mentioned analysis, the output can be represented by “0” and “1,” corresponding to the normal samples (i.e., background) and outlier samples (i.e., anomaly), respectively. Nevertheless, it is hard to directly adopt the intensity values of pixels in an HSI to calculate μ_0 and σ , due to the fact that there is a wealth of bands for each pixel in an HSI, so we utilize the value of \mathbf{D}_1 and \mathbf{D}_2 to act as the intensity value of the HSI. Through the aforementioned way, two binary maps, i.e., \mathbf{B}_1 and \mathbf{B}_2 , are obtained.

To construct the background and potential anomaly dictionaries as accurately as possible, \mathbf{B}_1 and \mathbf{B}_2 are jointly combined

to assist in selecting the atoms of background and potential anomaly dictionaries. Similar to \mathbf{D}_1 and \mathbf{D}_2 , the elementwise product of \mathbf{B}_1 and \mathbf{B}_2 acts as the indicator to sift the potential anomaly atoms, and the sum of \mathbf{B}_1 and \mathbf{B}_2 is used to be the director to choose the background atoms. Notably, to reduce the computation complexity, the principal component analysis is first utilized to reduce the dimensionality of the HSI, and then the simple linear iterative clustering (SLIC) [51] is performed on the first three principal components to obtain the superpixels. Finally, the background atoms can be acquired through the superpixels and the sum of \mathbf{B}_1 and \mathbf{B}_2 , whereas the potential anomaly atoms can be obtained via elementwise product of \mathbf{B}_1 and \mathbf{B}_2 . The details can be formulated as

$$\mathbf{X}_A = \{\mathbf{x}_{(i-1)*w+j} | (\mathbf{B}_1 \odot \mathbf{B}_2)_{i,j} = 1 \rightarrow (i, j)\} \quad (23)$$

$$\mathbf{X}_B = \{\mathbf{x}_i^c | \sum_{i'=1}^{S_{i'}} (\mathbf{B}_1 \oplus \mathbf{B}_2)_{i'} = 0 \rightarrow i'\} \quad (24)$$

where \mathbf{X}_A and \mathbf{X}_B represent the collection of potential anomaly pixels and background pixels, respectively. $i \in [1, h]$ and $j \in [1, w]$ are separately the indexes corresponding to the row and column, in which h and w refer to the height and width of an HSI, respectively. $i' \in [1, n_s]$ denotes the index of superpixels, where n_s represents the number of predefined superpixels. $S_{i'}$ indicates the number of pixels in the i' th superpixel. $\sum_{i'=1}^{S_{i'}} (\cdot)_{i'}$ signifies the sum of response value of all pixels belonging to the i' th superpixel. \mathbf{x}_i^c refers to the centroid of the i' th superpixel.

2) *Low-Rank and Sparse Representation*: As stated in (8), our goal is to solve the optimization problem in the consideration of the dual collaborative constraints, with the constructed background dictionary \mathbf{X}_B and potential anomaly dictionary \mathbf{X}_A . To tackle the aforementioned low-rank and sparse representation problem, four auxiliary variables \mathbf{P} , \mathbf{Q} , \mathbf{R} , and \mathbf{V} are introduced to make the objective function separable. Consequently, the problem (8) can be converted into the following form:

$$\begin{aligned} & \min_{\mathbf{W}, \mathbf{S}, \mathbf{E}, \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{V}} \|\mathbf{P}\|_* + \alpha \|\mathbf{Q}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda \|\mathbf{W}_B \circ \mathbf{R}\|_F^2 \\ & + \gamma \|\mathbf{W}_A \circ \mathbf{V}\|_F^2 \\ & \text{s.t. } \mathbf{X} = \mathbf{X}_B \mathbf{W} + \mathbf{X}_A \mathbf{S} + \mathbf{E}, \mathbf{W} = \mathbf{P}, \mathbf{S} = \mathbf{Q}, \mathbf{W} = \mathbf{R}, \mathbf{S} = \mathbf{V}. \end{aligned} \quad (25)$$

The augmented Lagrange function of problem (25) is

$$\begin{aligned} l = & \|\mathbf{P}\|_* + \alpha \|\mathbf{Q}\|_1 + \beta \|\mathbf{E}\|_{2,1} + \lambda \|\mathbf{W}_B \circ \mathbf{R}\|_F^2 \\ & + \gamma \|\mathbf{W}_A \circ \mathbf{V}\|_F^2 + \operatorname{tr}[\mathbf{Y}_1^T (\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E})] \\ & + \operatorname{tr}[\mathbf{Y}_2^T (\mathbf{W} - \mathbf{P})] + \operatorname{tr}[\mathbf{Y}_3^T (\mathbf{S} - \mathbf{Q})] + \operatorname{tr}[\mathbf{Y}_4^T (\mathbf{W} - \mathbf{R})] \\ & + \operatorname{tr}[\mathbf{Y}_5^T (\mathbf{S} - \mathbf{V})] + \frac{\eta}{2} (\|\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E}\|_F^2 \\ & + \|\mathbf{W} - \mathbf{P}\|_F^2 + \|\mathbf{S} - \mathbf{Q}\|_F^2 + \|\mathbf{W} - \mathbf{R}\|_F^2 + \|\mathbf{S} - \mathbf{V}\|_F^2) \end{aligned} \quad (26)$$

where \mathbf{Y}_1 , \mathbf{Y}_2 , \mathbf{Y}_3 , \mathbf{Y}_4 , and \mathbf{Y}_5 are the Lagrange multipliers and η is a penalty parameter. $\operatorname{tr}[\cdot]$ represents the trace of the matrix. The problem (26) can be solved with the alternating direction method of multipliers (ADMM) method [52], of which the core idea of this is to fix other variables while updating one variable.

- 1) Update \mathbf{P} while fixing other variables. The objective function can be derived as

$$\min_{\mathbf{P}} \|\mathbf{P}\|_* + \frac{\eta}{2} \left\| \mathbf{P} - \left(\mathbf{W} + \frac{\mathbf{Y}_2}{\eta} \right) \right\|_F^2. \quad (27)$$

- 2) Update \mathbf{Q} while fixing other variables. The objective function can be derived as

$$\min_{\mathbf{Q}} \alpha \|\mathbf{Q}\|_1 + \frac{\eta}{2} \left\| \mathbf{Q} - \left(\mathbf{S} + \frac{\mathbf{Y}_3}{\eta} \right) \right\|_F^2. \quad (28)$$

- 3) Update \mathbf{E} while fixing other variables. The objective function can be derived as

$$\min_{\mathbf{E}} \beta \|\mathbf{E}\|_{2,1} + \frac{\eta}{2} \left\| \mathbf{E} - \left(\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} + \frac{\mathbf{Y}_1}{\eta} \right) \right\|_F^2. \quad (29)$$

- 4) Update \mathbf{R} while fixing other variables. The objective function can be derived as

$$\min_{\mathbf{R}} \lambda \|\mathbf{W}_B \circ \mathbf{R}\|_F^2 + \text{tr}[\mathbf{Y}_4^T (\mathbf{W} - \mathbf{R})] + \frac{\eta}{2} \|\mathbf{W} - \mathbf{R}\|_F^2 \quad (30)$$

$$\Rightarrow \mathbf{R} = \left(\frac{1}{2\lambda \mathbf{W}_B \circ \mathbf{W}_B + \eta} \right) \circ (\mathbf{Y}_4 + \eta \mathbf{W}). \quad (31)$$

- 5) Update \mathbf{V} while fixing other variables. The objective function can be derived as

$$\min_{\mathbf{V}} \gamma \|\mathbf{W}_A \circ \mathbf{V}\|_F^2 + \text{tr}[\mathbf{Y}_5^T (\mathbf{S} - \mathbf{V})] + \frac{\eta}{2} \|\mathbf{S} - \mathbf{V}\|_F^2 \quad (32)$$

$$\Rightarrow \mathbf{V} = \left(\frac{1}{2\gamma \mathbf{W}_A \circ \mathbf{W}_A + \eta} \right) \circ (\mathbf{Y}_5 + \eta \mathbf{S}). \quad (33)$$

- 6) Update \mathbf{W} while fixing other variables. The objective function can be derived as

$$\begin{aligned} \min_{\mathbf{W}} \text{tr}[\mathbf{Y}_1^T (\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E})] + \text{tr}[\mathbf{Y}_2^T (\mathbf{W} - \mathbf{P})] \\ + \text{tr}[\mathbf{Y}_4^T (\mathbf{W} - \mathbf{R})] + \frac{\eta}{2} (\|\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E}\|_F^2 \\ + \|\mathbf{W} - \mathbf{P}\|_F^2 + \|\mathbf{W} - \mathbf{R}\|_F^2) \end{aligned} \quad (34)$$

$$\Rightarrow \mathbf{W} = (\mathbf{X}_B^T \mathbf{X}_B + 2\mathbf{I})^{-1} [\mathbf{X}_B^T (\mathbf{X} - \mathbf{X}_A \mathbf{S} - \mathbf{E} + \mathbf{Y}_1/\eta) + (\mathbf{P} - \mathbf{Y}_2/\eta) + (\mathbf{R} - \mathbf{Y}_4/\eta)]. \quad (35)$$

- 7) Update \mathbf{S} while fixing other variables. The objective function can be derived as

$$\begin{aligned} \min_{\mathbf{S}} \text{tr}[\mathbf{Y}_1^T (\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E})] + \text{tr}[\mathbf{Y}_3^T (\mathbf{S} - \mathbf{Q})] \\ + \text{tr}[\mathbf{Y}_5^T (\mathbf{S} - \mathbf{V})] + \frac{\eta}{2} (\|\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E}\|_F^2 \\ + \|\mathbf{S} - \mathbf{Q}\|_F^2 + \|\mathbf{S} - \mathbf{V}\|_F^2) \end{aligned} \quad (36)$$

$$\Rightarrow \mathbf{S} = (\mathbf{X}_A^T \mathbf{X}_A + 2\mathbf{I})^{-1} [\mathbf{X}_A^T (\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{E} + \mathbf{Y}_1/\eta) + (\mathbf{Q} - \mathbf{Y}_3/\eta) + (\mathbf{V} - \mathbf{Y}_5/\eta)]. \quad (37)$$

The solution of the optimization process is outlined in Algorithm 1. Note that Φ , Ω , and Θ used in Algorithm 1 separately

Algorithm 1: Solve (26) by ADMM.

Input: HSI $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{i=N} \in \mathbb{R}^{d \times N}$, regularization parameter α, β, λ and γ .

Initialize: $\mathbf{W} = \mathbf{S} = \mathbf{P} = \mathbf{Q} = \mathbf{R} = \mathbf{V} = \mathbf{0}$, $\mathbf{E} = \mathbf{0}$, $\mathbf{Y}_1 = \mathbf{Y}_2 = \mathbf{Y}_3 = \mathbf{Y}_4 = \mathbf{Y}_5 = \mathbf{0}$, $\eta = 10^{-6}$, $\eta_{\max} = 10^{10}$, $\rho = 1.2$, $\varepsilon = 10^{-6}$.

- 1: While $\max(\|\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E}\|_F, \|\mathbf{W} - \mathbf{P}\|_F, \|\mathbf{S} - \mathbf{Q}\|_F, \|\mathbf{W} - \mathbf{R}\|_F, \|\mathbf{S} - \mathbf{V}\|_F) \geq \varepsilon$ do
- 2: Update \mathbf{P} while fixing other variables by (27):
 $\mathbf{P} \leftarrow \Phi_{1/\eta}(\mathbf{W} + \mathbf{Y}_2/\eta)$
- 3: Update \mathbf{Q} while fixing other variables by (28):
 $\mathbf{Q} \leftarrow \Omega_{\alpha/\eta}(\mathbf{S} + \mathbf{Y}_3/\eta)$
- 4: Update \mathbf{E} while fixing other variables by (29):
 $\mathbf{E} \leftarrow \Theta_{\beta/\eta}(\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} + \mathbf{Y}_1/\eta)$
- 5: Update \mathbf{R} while fixing other variables by (31):
 $\mathbf{R} \leftarrow [1/(2\lambda \mathbf{W}_B \circ \mathbf{W}_B + \eta)] \circ (\mathbf{Y}_4 + \eta \mathbf{W})$
- 6: Update \mathbf{V} while fixing other variables by (33):
 $\mathbf{V} \leftarrow [1/(2\gamma \mathbf{W}_A \circ \mathbf{W}_A + \eta)] \circ (\mathbf{Y}_5 + \eta \mathbf{S})$
- 7: Update \mathbf{W} while fixing other variables by (35):
 $\mathbf{W} \leftarrow (\mathbf{X}_B^T \mathbf{X}_B + 2\mathbf{I})^{-1} [\mathbf{X}_B^T (\mathbf{X} - \mathbf{X}_A \mathbf{S} - \mathbf{E} + \mathbf{Y}_1/\eta) + (\mathbf{P} - \mathbf{Y}_2/\eta) + (\mathbf{R} - \mathbf{Y}_4/\eta)]$
- 8: Update \mathbf{S} while fixing other variables by (37):
 $\mathbf{S} \leftarrow (\mathbf{X}_A^T \mathbf{X}_A + 2\mathbf{I})^{-1} [\mathbf{X}_A^T (\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{E} + \mathbf{Y}_1/\eta) + (\mathbf{Q} - \mathbf{Y}_3/\eta) + (\mathbf{V} - \mathbf{Y}_5/\eta)]$
- 9: Update the five Lagrange multipliers:
 $\mathbf{Y}_1 \leftarrow \mathbf{Y}_1 + \eta(\mathbf{X} - \mathbf{X}_B \mathbf{W} - \mathbf{X}_A \mathbf{S} - \mathbf{E})$
 $\mathbf{Y}_2 \leftarrow \mathbf{Y}_2 + \eta(\mathbf{W} - \mathbf{P})$
 $\mathbf{Y}_3 \leftarrow \mathbf{Y}_3 + \eta(\mathbf{S} - \mathbf{Q})$
 $\mathbf{Y}_4 \leftarrow \mathbf{Y}_4 + \eta(\mathbf{W} - \mathbf{R})$
 $\mathbf{Y}_5 \leftarrow \mathbf{Y}_5 + \eta(\mathbf{S} - \mathbf{V})$
- 10: Update the balance parameter η :
 $\eta \leftarrow \min(\rho\eta, \eta_{\max})$
- 11: End While

Output: $\mathbf{W}, \mathbf{S}, \mathbf{E}$

refer to the singular value thresholding [53], soft thresholding [54], and $l_{2,1}$ norm minimization operator [55].

Once the optimization process is terminated, the detection result \mathbf{D}_3 can be determined by the sparse component $\mathbf{X}_A \mathbf{S}$ as

$$T(\mathbf{x}_i) = \left\| [\mathbf{X}_A \mathbf{S}]_{:,i} \right\|_2 = \sqrt{\sum_{j=1}^d ([\mathbf{X}_A \mathbf{S}]_{j,i})^2} \quad (38)$$

where $[\cdot]_{:,i}$ denotes the i th column of a matrix.

To suppress the background interference in the detection result, a nonlinear transformation, which considers both \mathbf{D}_1 and \mathbf{D}_2 , is performed on \mathbf{D}_3 for better detection performance. The nonlinear transformation can be formulated as

$$\mathbf{D}_F = \mathbf{D}_3 (1 - e^{-\tau_1 \mathbf{D}_1}) (1 - e^{-\tau_2 \mathbf{D}_2}). \quad (39)$$

where \mathbf{D}_F denotes the final detection result. $\tau_1 > 0$ and $\tau_2 > 0$ indicate the parameters to adjust the shape of nonlinear transformation.

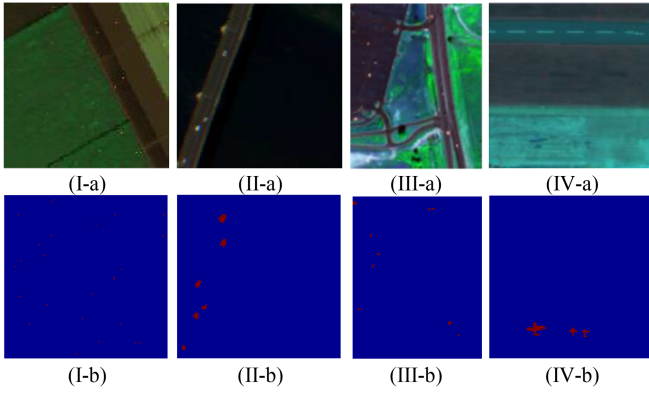


Fig. 4. Visualization of the experimental datasets. (a) Pseudocolor image. (b) Reference map. (I) Salinas. (II) Pavia. (III) HYDICE. (IV) Gulfport.

IV. EXPERIMENTS

In this section, we conducted a flurry of experiments to validate the effectiveness and superiority of the proposed method.

A. Datasets and Evaluation Metrics

1) *Data Description*: One simulated dataset and three real datasets, acquired by three different sensors, are used to evaluate the performance of the proposed method. The anomalies existed in the datasets appear in different forms, and the details are described in the following.

Salinas Dataset [56]: The first dataset, generated by exerting the target implantation method [57] on a real dataset captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensors, covering the area of Salinas Valley, CA, USA, is the simulated dataset. There are 204 spectral bands in total for the dataset, and each of them contains 120×120 pixels. The spatial resolution of the dataset is 3.7 m. The pseudocolor image and corresponding reference map are displayed in Fig. 4(I-a) and (I-b), respectively.

Pavia Dataset [40]: The second dataset, collected by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor, covers the area of the center of Pavia city located in northern Italy. The number of spectral bands is 102 for the dataset, ranging from 0.43 to $0.86 \mu\text{m}$. There are 100×100 pixels and a 1.3-m spatial resolution for the dataset. The pseudocolor image and corresponding reference map are separately shown in Fig. 4(II-a) and (II-b), respectively.

HYDICE Dataset [58]: The third dataset was obtained by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor, covering an urban area of CA, USA. There are 80×100 spatial pixels and 175 spectral bands ranging from 0.4 to $2.5 \mu\text{m}$ for the dataset, and its spatial resolution is 1 m. The pseudocolor image and corresponding reference map are exhibited in Fig. 4(III-a) and (III-b), respectively.

Gulfport Dataset [47]: The fourth dataset with a size of $100 \times 100 \times 191$ was acquired by the AVIRIS sensor at Gulfport, with a spatial resolution of 3.4 m. The pseudocolor image and corresponding reference map are separately illustrated in Fig. 4(IV-a) and (IV-b).

2) *Evaluation Metrics*: To evaluate the performance of the proposed method, three commonly utilized evaluation metrics, containing the receiver operating characteristic (ROC) curve [59], the area under the curve (AUC) [60], and the separability map [61], are employed. The ROC curve depicts the relationship between the true positive rate (P_D) and false positive rate (P_F) at different thresholds. The commonly used ROC curves are (P_D, P_F) and (P_F, τ) , which are separately employed to characterize the detection rate and false alarm rate. For the ROC curve of (P_D, P_F) , the higher the curve, the better the performance. Inversely, the low ROC curve of (P_F, τ) means a low false alarm rate, indicating superior performance. Correspondingly, two kinds of AUC scores are adopted, which can be expressed as AUC score of (P_D, P_F) and AUC score of (P_F, τ) , respectively. A detector whose AUC score of (P_D, P_F) closely approaches to 1 is viewed as an excellent one. Conversely, the AUC score of (P_F, τ) for a superior detector should be close to 0. The separability map, which is generated from the boxplot, characterizes the ability of the detector to extract anomalies from background. The better the detector, the larger the separation distance between the background and anomalies.

B. Experimental Setup

1) *Implementation Details*: The running environment of our method is Python 3.6.13, Pytorch-GPU 1.8.2, and CUDA 10.2, which are executed on an Intel Core i7-9700T CPU 2.00GHz machine with 16GB of RAM and a GeForce RTX 2080Ti graphics card with 11GB memory. Other compared methods are carried out on a running environment with MATLAB R2017b except for PAB-DC, which is conducted on a running environment with Python 3.6.13. We optimize the density estimation model via the Adagrad [62] optimizer with a learning rate of 0.0001, and the number of iterations is 1000. The number of hidden nodes is empirically set to 128. The batch size of the density estimation model is set to N , which is the number of total pixels in HSI. For the proposed method, we empirically set the parameters to fixed values, which are expressed as $n_H = 1$, $n_M = 8$, $k = 20$, $n_s = 400$, $\alpha = 0.1$, $\beta = 0.1$, $\lambda = 0.1$, $\gamma = 0.1$, $\tau_1 = 1$, and $\tau_2 = 1$, respectively. In our experiments, the above parameters are configured by default, and the optimal detection results can be obtained by tuning these parameters by the user.

2) *Compared Methods*: To validate the superiority of the proposed method, eight widely cited methods, consisting of RX [19], CRD [28], LRASR [42], LSMAD [40], MLW_LRRSTO [63], PAB-DC [45], LSDM-MoG [41], and KIFD [58], are utilized for comparison. Among them, RX, CRD, LRASR, and LSMAD are typical methods, and the last four methods are proposed recently.

C. Parameter Sensitivity Analysis

This section aims to analyze the performance of the proposed method with different parameters. There are ten parameters to be considered, which are separately the number of hidden layers n_H , the number of Gaussian mixture components n_M , the number of k -nearest neighbors k , the number of superpixels n_s , the

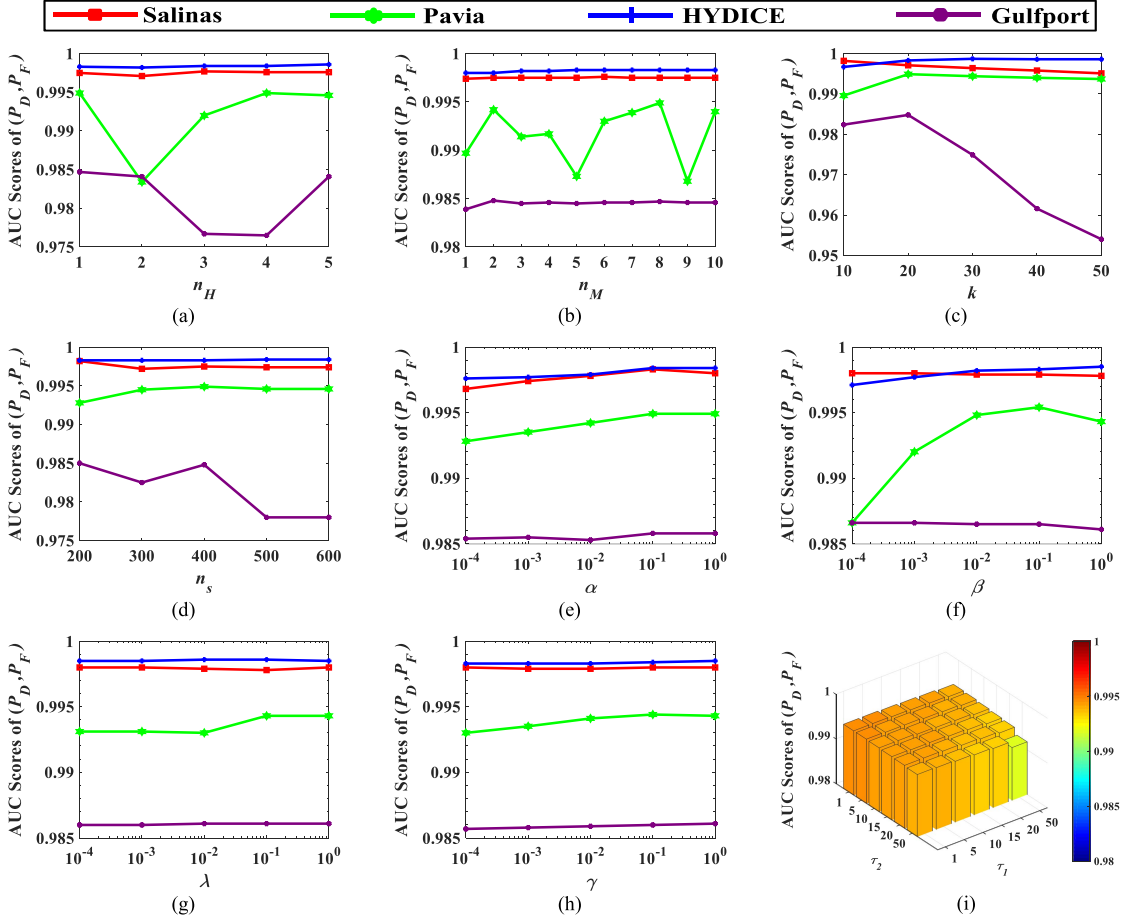


Fig. 5. Parameter analysis on the experimental datasets. (a) Number of hidden layers. (b) Number of Gaussian mixture components. (c) Number of k -nearest neighbors. (d) Number of superpixels. (e) α . (f) β . (g) λ . (h) γ . (i) Effect of τ_1 and τ_2 on Pavia dataset.

regularization coefficients α , β , λ , γ , and transformation coefficients τ_1 and τ_2 . In our experiments, we fix the other parameters when a certain parameter is analyzed, and the other parameters are set to the preconfigured values, as stated in Section IV-B1. The details of parameter analysis are illustrated in Fig. 5, in which the abscissa represents the range of the parameters and the ordinate indicates the AUC scores of (P_D, P_F) .

1) *Number of Hidden Layers n_H* : To analyze the influence of the number of hidden layers in the density estimation model on the detection performance, a group of numbers, ranging from 1 to 5 at an interval of 1, are set, as displayed in Fig. 5(a). By observing Fig. 5(a), we can notice that the proposed method performs well when n_H is equivalent to 1 or 5 on all experimental datasets. Clearly, the computation burden of the density estimation model with five hidden layers is higher than that of the density estimation model with one hidden layer. Therefore, it is considered to be very wise to set the number of hidden layers of the density estimation model as 1.

2) *Number of Gaussian Mixture Components n_M* : Fig. 5(b) shows the effect of n_M with different values on the detection performance. Concretely, n_M spans from 1 to 10 with an interval of 1. For the Gaussian mixture components, a proper parameter n_M is helpful to characterize the background of HSI and isolate

the anomalies. Through Fig. 5(b), we can find that the proposed method has an obvious fluctuation with the change of n_M for Pavia dataset, while it keeps stable on the other datasets. Notably, the proposed method can achieve excellent results when $n_M = 8$ for all experimental datasets.

3) *Number of K-Nearest Neighbors k* : For the number of k -nearest neighbors k , if k is relatively small, the central region of the anomalies occupying a large area will reside in a low-density area. Inversely, the computation burden is quite high when k is particularly large. Therefore, the value of k is set as $\{10, 20, 30, 40, 50\}$, as illustrated in Fig. 5(c). From Fig. 5(c), we can see that the detector can achieve satisfactory detection performance when $k = 20$.

4) *Number of Superpixels n_s* : The effect of the number of superpixels on the detection performance is analyzed, as exhibited in Fig. 5(d). For the detector, fewer superpixels mean that fewer pixels participate in the construction of the background dictionary, resulting in low detection performance. Contrarily, the more superpixels will lead to heavy calculation consumption. To this end, we set n_s as $\{200, 300, 400, 500, 600\}$ to tradeoff the detection performance and calculation burden. By observing Fig. 5(d), it can be found that the detection performance of the proposed method maintains stable with the variation of n_s for

TABLE I
OPTIMAL PARAMETERS SETTINGS FOR EXPERIMENTAL DATASETS

Parameters	n_H	n_M	k	n_s	α	β	λ	γ	τ_1	τ_2
Salinas	1	8	20	400	0.1	0.1	0.1	0.1	1	1
Pavia	1	8	20	400	0.1	0.1	0.1	0.1	1	5
HYDICE	1	8	20	400	0.1	0.1	0.1	0.1	1	1
Gulfport	1	8	20	400	0.1	0.1	0.1	0.1	5	1

the HYDICE dataset. Different from the above, there is a certain degree of change for the remaining three datasets, especially for the Gulfport dataset. Evidently, it is intuitively plausible to achieve the excellent detection performance when n_s is equal to 400.

5) *Regularization Coefficients α , β , λ , and γ* : For the regularization coefficients, all of them are configured as $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^0\}$, and the corresponding detection results under different values are displayed in Fig. 5(e)–(h), respectively. As a whole, the detector has an obvious change in detection performance for Pavia dataset compared with that of other datasets over all regularization coefficients. Through Fig. 5(e)–(h), we can find that the detector can obtain gratifying detection performance when the overall regularization coefficients are equivalent to 10^{-1} , 10^{-1} , 10^{-1} , and 10^{-1} , respectively.

6) *Transformation Coefficients τ_1 and τ_2* : With regard to the transformation coefficients, both of them are configured as $\{1, 5, 10, 15, 20, 50\}$, considering the fact that the high response values in (38) are expected to be reinforced, while the corresponding low response values should be suppressed. Taking the Pavia dataset as an example, the detection results plotted in a 3-D manner are visualized in Fig. 5(i). By observing

Fig. 5(i), it can be found that the detection performance remains stable and at a high level. Notably, the detection performance achieves the optimal result when $\tau_1 = 1$ and $\tau_2 = 5$. To obtain the optimal performance, a great quantity of experiments is executed on all experimental datasets, and the corresponding parameter settings are listed in Table I.

D. Component Analysis

1) *Effectiveness Evaluation of Dual Dictionaries Construction Strategy*: To validate the effect of dual dictionaries construction strategy based on two-stream density maps, the proposed method is compared with the typical dual dictionaries-based method (i.e., PAB-DC), as displayed in Fig. 6. Here, the postprocessing operation, i.e., the nonlinear transformation in (39), is ignored in the proposed method for a fair comparison. By observing Fig. 6, we can notice that the detection performance of the proposed method outperforms that of the typical method on all experimental datasets. Concretely, the detection performance of the proposed method is much higher than that of the typical method on Salinas, Pavia, and Gulfport datasets. For the HYDICE dataset, the improvement is limited for the proposed method relative to PAB-DC, owing to the fact that the detection performance of both detectors is quite close to 1. Furthermore to clearly show the superiority of the proposed method, the background and potential anomaly dictionaries belonging to the proposed method and PAB-DC are visualized in

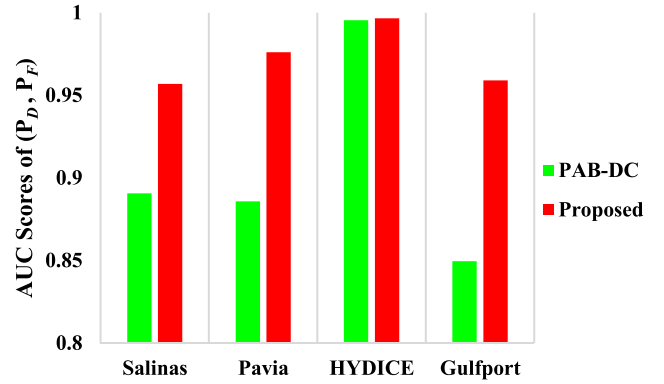


Fig. 6. 2-D histogram of detection performance with different dictionaries construction.

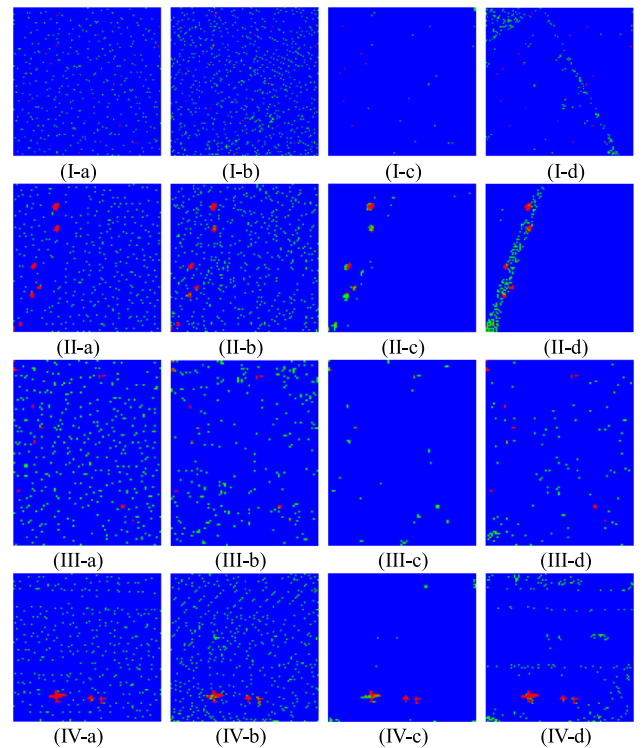


Fig. 7. Visualization of the selected atoms in background dictionary and potential anomaly dictionary for the proposed method and PAB-DC on experimental datasets. (I) Salinas. (II) Pavia. (III) HYDICE. (IV) Gulfport. (a) Selected atoms in background dictionary for the proposed method. (b) Selected atoms in background dictionary for PAB-DC. (c) Selected atoms in potential anomaly dictionary for the proposed method. (d) Selected atoms in potential anomaly dictionary for PAB-DC. The red areas refer to the anomalies existed in the reference map, and the green dots represent the selected atoms in the background and potential anomaly dictionaries.

Fig. 7. Compared with the background dictionary constructed in PAB-DC, that constructed in the proposed method is hardly contaminated by the anomalous pixels. Similarly, the interference degree of background pixels to potential anomaly dictionaries in the proposed method is much lower than that of PAB-DC. This means that the proposed method can acquire more pure and representative dictionary atoms relative to those of PAB-DC, achieving more excellent detection performance.

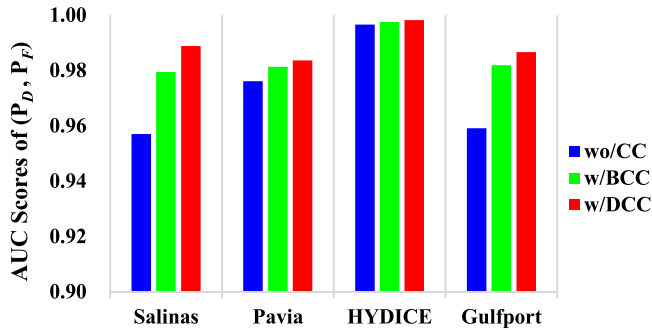


Fig. 8. 2-D histogram of detection performance with or without collaborative constraints. Note that “wo/CC” represents the proposed method without collaborative constraints, “w/BCC” indicates the proposed method with the background collaborative constraint, and “w/DCC” signifies the proposed method with dual collaborative constraints.

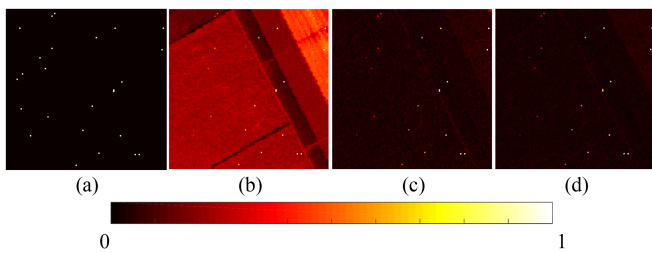


Fig. 9. Visualization of the detection results under different constraint terms for the Salinas dataset. (a) Reference map. (b) Proposed method without collaborative constraints. (c) Proposed method with background collaborative constraint. (d) Proposed method with dual collaborative constraints.

2) *Effectiveness Evaluation of Dual Collaborative Constraints*: To verify the effect of the dual collaborative constraints imposed on the low-rank and sparse representation model, three aspects, containing the proposed method without collaborative constraints, the proposed method with background collaborative constraint, and the proposed method with dual collaborative constraints, are considered, as plotted in Fig. 8. Similar to Section IV-D1, the nonlinear transformation is also omitted for fully excavating the effect of the dual collaborative constraints. As we can see in Fig. 8, the detection performance of the proposed method far exceeds that of other aspects of Salinas, Pavia, and Gulfport datasets. Clearly, the improvement with regard to the detection performance is not obvious for the HYDICE dataset, which is mainly because the detection performance of the proposed method is nearly close to the limitation. Taking the Salinas dataset for example, the detection maps separately corresponding to the three comparative components are visualized, as illustrated in Fig. 9. Through Fig. 9, it can be found that the detection performance of the proposed method with dual collaborative constraints is remarkably higher than those of other types, indicating the effectiveness of the dual collaborative constraints.

E. Detection Performance

We employ detection maps and three evaluation metrics (i.e., ROC curve of (P_D, P_F) and ROC curve of (P_F, τ) , AUC

score of (P_D, P_F) and AUC score of (P_F, τ) , and separability maps) to assess the performance of the proposed method. Fig. 10 displays the detection maps of the proposed method and compared methods. The ROC curves of (P_D, P_F) and (P_F, τ) with respect to the proposed method and compared methods are separately shown in Figs. 11 and 12. Correspondingly, the AUC scores of (P_D, P_F) and (P_F, τ) are exhibited in Tables II and III, respectively. In addition, the separability maps between the proposed method and compared methods are illustrated in Fig. 13. It is worth noting that the bold numbers and underlined numbers indicate the best result and worst result, respectively.

For the Salinas dataset, there is the highest AUC score of (P_D, P_F) (i.e., 0.9982) and the lowest AUC score of (P_F, τ) (i.e., 0.0003) for the proposed method relative to the compared methods. The KIFD achieves the first rank over AUC score of (P_D, P_F) among compared methods, whereas superiority over the AUC score of (P_F, τ) is not conspicuous enough for KIFD. Notably, the LSDM-MoG has the lowest AUC score of (P_D, P_F) , which is much lower than the proposed method, indicating that the localization capability of LSDM-MoG is insufficient for the simulated dataset. In terms of PAB-DC, AUC score of (P_F, τ) is the highest among all methods, which means a very high false alarm rate, and the localization ability is unsatisfactory. Similarly, the effect of RX and LRASR is poor for locating anomalies. Additionally, the CRD, LSMAD, and MLW_LRRSTO can identify anomalies with a high detection rate and a low false alarm rate.

With regard to Pavia dataset, as illustrated in Fig. 10(II-i), we notice that the proposed method can identify the total anomalies, preserving low false alarms. The RX, LRASR, LSMAD, and MLW_LRRSTO are second to the proposed method in terms of the AUC score of (P_D, P_F) , which are 0.9887, 0.9889, 0.9842, and 0.9886, listed in Table II, respectively. Though the anomalies are well located for CRD, the relatively high false alarm rate results in lower detection accuracy. Notably, PAB-DC and KIFD fail to identify the anomalies comprehensively due to the very high false alarm rate, which are up to 0.3009 and 0.2360, respectively.

For HYDICE dataset, the proposed method still outperforms compared methods both on AUC score of (P_D, P_F) and AUC score of (P_F, τ) , especially for LRASR. The CRD, MLW_LRRSTO, PAB-DC, and KIFD are slightly lower than the proposed method over the AUC score of (P_D, P_F) , which are separately 0.9951, 0.9960, 0.9955, and 0.9966. At the same time, the false alarm rate of CRD, MLW_LRRSTO, and PAB-DC is satisfactory relative to KIFD. Besides, it is acceptable in the consideration of both AUC score of (P_D, P_F) and AUC score of (P_F, τ) for RX and LSMAD.

The detection map of the proposed method on Gulfport dataset is displayed in Fig. 10(IV-i). By observing Fig. 10(IV-i), we can notice that the anomalies (i.e., three airplanes) are identified to a large extent. Though two small airplanes are not very clear from a visual point of view, the detection accuracy (i.e., AUC score of (P_D, P_F)) is very high in comparison with other methods due to the low false alarm rate. In terms of visual effects, the anomalies are submerged into the background for LRASR, PAB-DC, and KIFD owing to the strong false alarm rate. Remarkably, the AUC

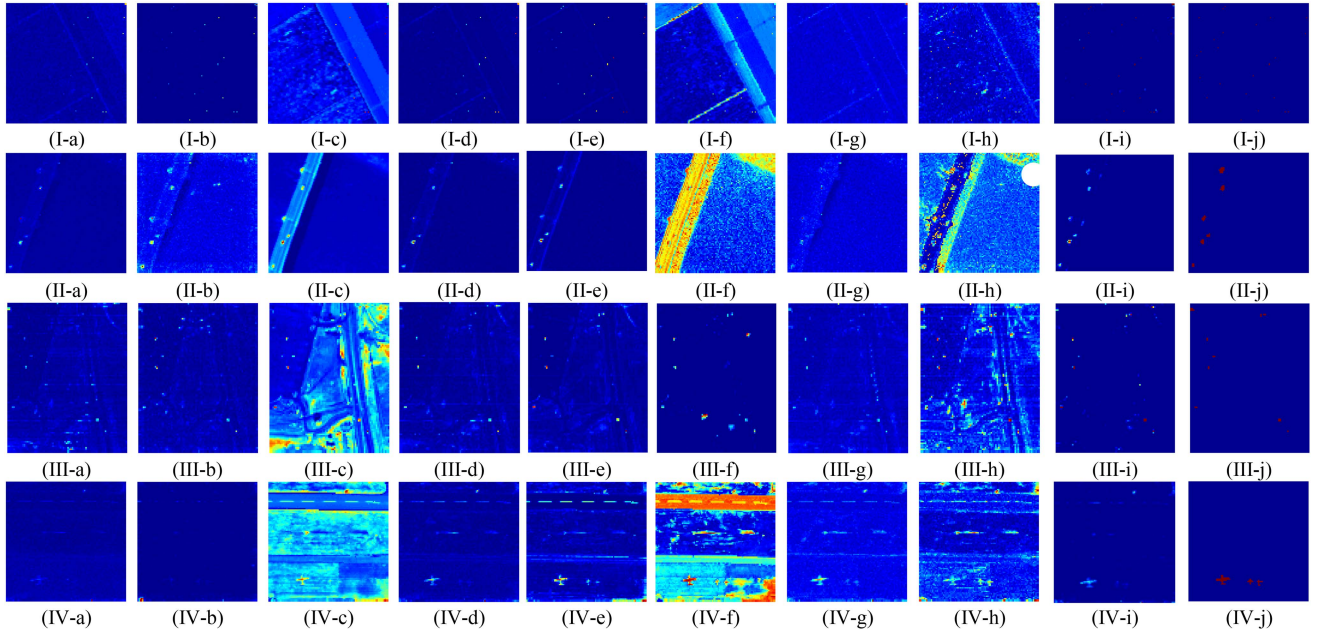


Fig. 10. Detection maps of different methods. (I) Salinas. (II) Pavia. (III) HYDICE. (IV) Gulfport. (a) RX. (b) CRD. (c) LRASR. (d) LSMAD. (e) MLW_LRRSTO. (f) PAB-DC. (g) LSDM-MoG. (h) KIFD. (i) Proposed method. (j) Reference map.

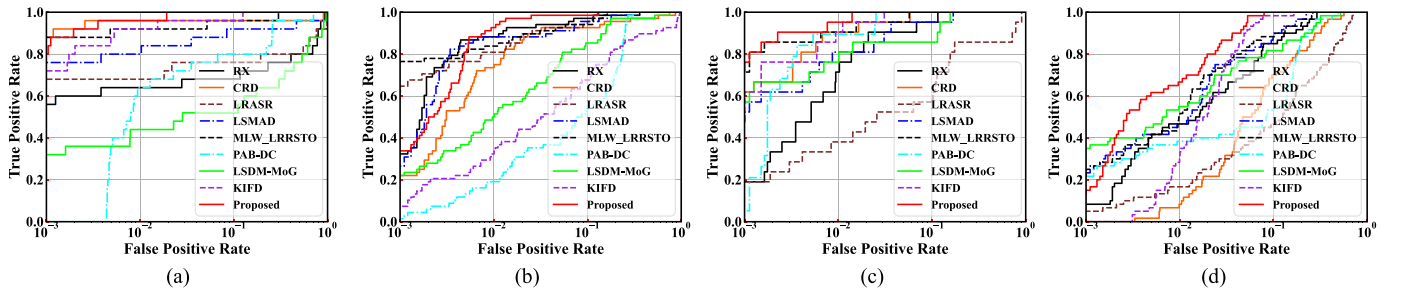


Fig. 11. ROC curves of (P_D, P_F) over compared methods on (a) Salinas, (b) Pavia, (c) HYDICE, and (d) Gulfport.

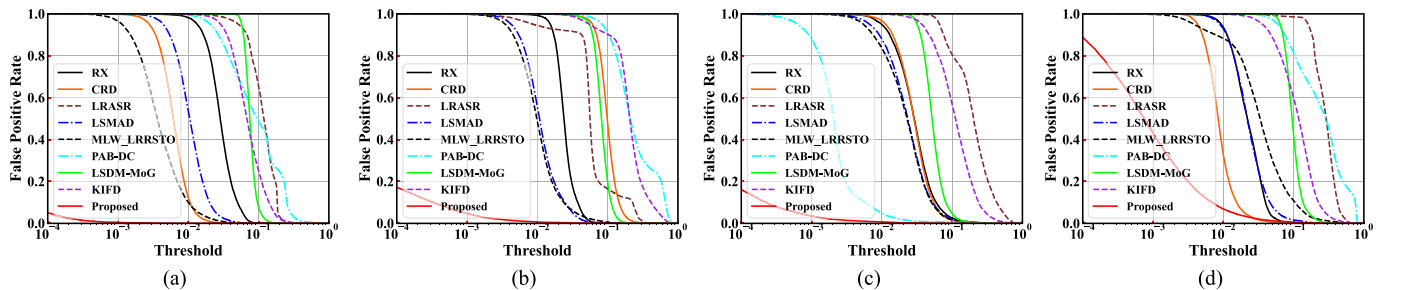


Fig. 12. ROC curves of (P_F, τ) over compared methods on (a) Salinas, (b) Pavia, (c) HYDICE, and (d) Gulfport.

TABLE II
AUC SCORES OF (P_D, P_F) FOR COMPARED METHODS ON EXPERIMENTAL DATASETS

	RX [19]	CRD [28]	LRASR [42]	LSMAD [40]	MLW_LRRSTO [63]	PAB-DC [45]	LSDM-MoG [41]	KIFD [58]	Proposed
Salinas	0.8073	0.9635	0.8332	0.9375	0.9857	0.9228	<u>0.7558</u>	0.9935	0.9982
Pavia	0.9887	0.9650	0.9889	0.9842	0.9886	0.8858	0.9482	<u>0.8589</u>	0.9949
HYDICE	0.9857	0.9951	<u>0.8402</u>	0.9861	0.9960	0.9955	0.9792	0.9966	0.9986
Gulfport	0.9526	0.8921	<u>0.8374</u>	0.9665	0.9665	0.8976	0.9436	0.9750	0.9879
Average	0.9336	0.9539	<u>0.8749</u>	0.9686	0.9846	0.9254	0.9067	0.9560	0.9949

TABLE III
AUC SCORES OF (P_F, τ) FOR COMPARED METHODS ON EXPERIMENTAL DATASETS

	RX [19]	CRD [28]	LRASR [42]	LSMAD [40]	MLW	LRRSTO [63]	PAB-DC [45]	LSDM-MoG [41]	KIFD [58]	Proposed
Salinas	0.0314	0.0069	0.1227	0.0123		0.0053	<u>0.1247</u>	0.0779	0.0803	0.0003
Pavia	0.0259	0.1046	0.0803	0.0135		0.0134	<u>0.3009</u>	0.0847	0.2360	0.0004
HYDICE	0.0351	0.0333	<u>0.2319</u>	0.0291		0.0277	0.0052	0.0570	0.1335	0.0004
Gulfport	0.0247	0.0108	<u>0.2957</u>	0.0275		0.0500	0.3252	0.1021	0.1337	0.0041
Average	0.0293	0.0389	0.1827	0.0206		0.0241	<u>0.1890</u>	0.0804	0.1459	0.0013

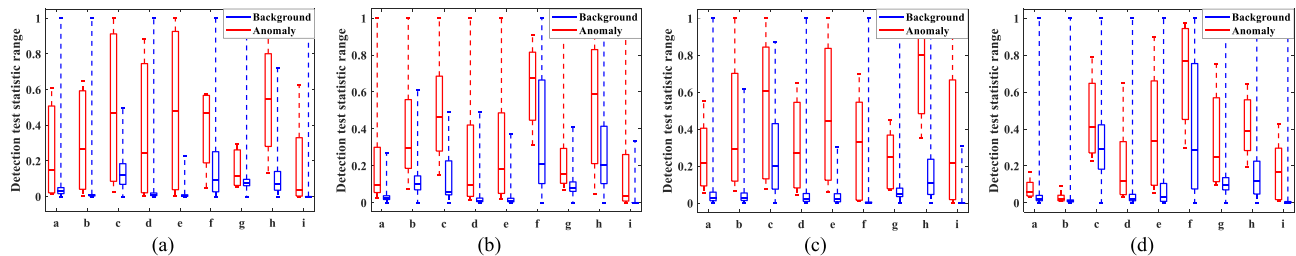


Fig. 13. Separability maps over compared methods on (a) Salinas, (b) Pavia, (c) HYDICE, and (d) Gulfport. Here, a-“RX,” b-“CRD,” c-“LRASR,” d-“LSMAD,” e-“MLW_LRRSTO,” f-“PAB-DC,” g-“LSDM-MoG,” h-“KIFD,” and i-“Proposed.”

scores of (P_D, P_F) on LRASR and PAB-DC are much lower than that of KIFD, which is slightly lower than the proposed method. The LSMAD, MLW_LRRSTO, and LSDM-MoG can well detect the anomalies, while the high false alarm rate leads to a low detection rate. In addition, the RX and CRD achieve encouraging false alarm rates, whereas the detection rate of CRD is poor relative to other methods except for LRASR.

To prove the superiority of the proposed method, both ROC curves (i.e., ROC curve of (P_D, P_F) and ROC curve of (P_F, τ)) are utilized to judge the performance, as illustrated in Figs. 11 and 12, respectively. By observing the ROC curves of (P_D, P_F) in Fig. 11, we can find that the ROC curve of (P_D, P_F) of the proposed method is higher than that of the compared methods on HYDICE dataset, indicating that the detection performance of the proposed method is optimal on HYDICE dataset. Additionally, the ROC curves of (P_D, P_F) of the proposed method lie nearer the upper left corner relative to the compared methods on Salinas, Pavia, and Gulfport datasets when the false positive rates are greater than 10^{-2} , which means that the proposed method performs well on these three datasets when the false positive rates are in the range of $[10^{-2}, 10^0]$. Although the performance of the ROC curves of (P_D, P_F) of the proposed method is not optimal on the aforementioned three datasets when the false positive rates are in the range of $[10^{-3}, 10^{-2}]$, the area under the ROC curves of (P_D, P_F) listed in Table II is optimal. In summary, as long as the overall performance is excellent, the local behavior is slightly suboptimal and can be acceptable. With regard to the ROC curve of (P_F, τ) , all compared methods are strongly higher than the proposed method, meaning that the false alarm rate of the proposed method is completely lower than that of the others. Notably, both ROC curve of (P_D, P_F) and ROC curve of (P_F, τ) are displayed in a logarithmic-scale way for better visualization.

The separability maps on experimental datasets are further adopted to evaluate the effect of extracting anomalies from

background, as shown in Fig. 13. In Fig. 13, two-column boxplots corresponding to each detector separately represent the anomaly (in red) and background (in blue). The bottom bound and upper bound of the box are separately 10% and 90% of the statistical interval of the detection results, respectively, and the line in the middle of the box is the median of the statistical interval. The lines of top and bottom for each column are the extremums, which are separately the maximum and minimum values of the statistical interval, while the 0%~10% and 90%~100% intervals are represented by dotted lines. The interval between the red box and the blue box indicates the separability degree of the anomalies and background. The height of the blue box represents the suppression effect for the background. For Salinas dataset, there are overlaps between the red box and blue box to some degree for RX, LRASR, PAB-DC, and LSDM-MoG, indicating that their separability degree is limited. The rest of compared methods and the proposed method perform well in separating the background and anomaly, whereas the background suppression capability of the proposed method is superior to the rest of compared methods, especially for KIFD. Similarly, for Pavia dataset, we can notice that the overlaps with regard to the red box and blue box are obvious for PAB-DC, LSDM-MoG, and KIFD, which means that the separability performance is unsatisfactory. Although the separability of the other compared methods is slightly better than the proposed method, the background suppression effect of the proposed method evidently outperforms all compared methods. Likewise, for HYDICE dataset, we can find that there is a nice tradeoff between the separability (i.e., the ability to extract the anomaly from background) and background suppression in comparison with the compared methods. It is worth noting that the proposed method still performs well for Gulfport dataset, while the separability is poor for most compared methods except for KIFD. In summary, on experimental datasets, the proposed method has the best background suppression capability and acceptable

TABLE IV
AVERAGE RUNNING TIME (IN SECONDS) OF COMPARED METHODS

	RX [19]	CRD [28]	LRASR [42]	LSMAD [40]	MLW_LRRSTO [63]	PAB-DC [45]	LSDM-MoG [41]	KIFD [58]	Proposed
Salinas	0.55	8.08	18.32	17.65	69.20	<u>532.61</u>	14.33	70.71	280.61
Pavia	0.29	9.48	39.16	13.93	37.82	<u>352.12</u>	8.57	56.37	102.45
HYDICE	0.23	2.62	29.84	8.25	37.50	<u>257.15</u>	6.61	46.42	96.86
Gulfport	0.32	7.17	14.06	11.40	56.65	<u>366.59</u>	9.37	60.03	190.37
Average	0.35	6.84	25.35	12.81	50.29	<u>377.12</u>	9.72	58.38	167.57

separability between the anomaly and background relative to the compared methods.

Furthermore, the running time of all methods is listed in Table IV to evaluate the computational burden. Notably, the bold numbers and underlined numbers represent the best result and worst result, respectively. The RX achieves the first-rank real-time performance in comparison with other methods. Although the proposed method is more time-consuming relative to most methods, the running time of the proposed method is less than that of PAB-DC, which is the typical dual dictionaries-based low-rank and sparse representation method, indicating that the proposed methods can be acceptable in the consideration of the detection performance and running time.

V. CONCLUSION

In this article, a novel HAD method, which imposes dual collaborative constraints on low-rank and sparse representation models, in conjunction with the robust dictionary construction, is proposed. Concretely, to thoroughly employ the global and local information in the HSI, dual collaborative constraints act as the regularization terms to model the background and anomaly components. Considering the fact that surrounding pixels with high similarity relative to the pixel under test should be given high weight, two weighted coefficient matrices corresponding to background and anomaly components are further introduced. To acquire better detection performance, a robust dictionary construction strategy, which fully exploits the density maps generated by the density estimation model and LOF, is proposed, yielding a pure and representative background and potential anomaly dictionaries. To validate the effectiveness and superiority of the proposed method, the experiments are carried out on four datasets consisting of one simulated dataset and three real datasets with different sensors. The effectiveness of the proposed method is verified by the component analysis of experimental datasets. Similarly, the superiority of the proposed method is also proved by comparing it with the four typical methods and four state-of-the-art methods. In future work, we will think about how to reduce the complexity of the proposed model, reducing the running time.

ACKNOWLEDGMENT

The authors would like to thank the researchers who provide the datasets and codes. The authors would also acknowledge the handling editor and the anonymous reviewers for their critical and constructive comments in dealing with this manuscript.

REFERENCES

- [1] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.
- [2] Q. Wang, F. Zhang, and X. Li, "Hyperspectral band selection via optimal neighborhood reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8465–8476, Dec. 2020.
- [3] J. Feng et al., "Deep reinforcement learning for semisupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501719.
- [4] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2623–2634, Jun. 2018.
- [5] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.
- [6] Y. Ding, X. Zhao, Z. Zhang, W. Cai, and N. Yang, "Multiscale graph sample and aggregate network with context-aware learning for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4561–4572, 2021.
- [7] L. He, J. Zhu, J. Li, D. Meng, J. Chanussot, and A. Plaza, "Spectral-fidelity convolutional neural networks for hyperspectral pansharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5898–5914, 2020.
- [8] L. He, J. Zhu, J. Li, A. Plaza, J. Chanussot, and B. Li, "HyperPNN: Hyperspectral pansharpening via spectrally predictive convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3092–3100, Aug. 2019.
- [9] K. Yan, M. Zhou, L. Liu, C. Xie, and D. Hong, "When pansharpening meets graph convolution network and knowledge distillation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5408915.
- [10] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial pre-processing for hyperspectral unmixing," *Remote Sens. Environ.*, vol. 204, pp. 333–346, Jan. 2018.
- [11] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravorty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.
- [12] R. Tao, X. Zhao, W. Li, H.-C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.
- [13] X. Ma, X. Zhang, X. Tang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection based on low-rank representation with data-driven projection and dictionary construction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2226–2239, 2020.
- [14] S. Yuan, L. Shi, B. Yao, F. Li, and Y. Du, "A hyperspectral anomaly detection algorithm using sub-features grouping and binary accumulation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6007505.
- [15] S. Lin, M. Zhang, X. Cheng, L. Wang, M. Xu, and H. Wang, "Hyperspectral anomaly detection via dual dictionaries construction guided by two-stage complementary decision," *Remote Sens.*, vol. 14, no. 8, Apr. 2022, Art. no. 1784.
- [16] D. Zhu, B. Du, and L. Zhang, "Target dictionary construction-based sparse representation hyperspectral target detection methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1254–1264, Apr. 2019.
- [17] C.-I. Chang, H. Cao, and M. Song, "Orthogonal subspace projection target detector for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4915–4932, 2021.
- [18] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object detection in remote sensing images based on improved bounding box regression and multi-level features fusion," *Remote Sens.*, vol. 12, no. 1, Jan. 2020, Art. no. 143.

- [19] 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.
- [20] J. M. Molero, E. M. Garzon, I. Garcia, 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.
- [21] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, 2004, vol. 3, pp. 1818–1824.
- [22] K. Heesung 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.
- [23] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
- [24] W. Liu and C. Chang, "Multiple-window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 644–658, Apr. 2013.
- [25] W. Li and Q. Du, "Decision fusion for dual-window-based hyperspectral anomaly detector," *J. Appl. Remote Sens.*, vol. 9, no. 1, pp. 97297–97297, Jan. 2015.
- [26] 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.
- [27] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, May 2018, Art. no. 745.
- [28] 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.
- [29] G. Zhang, N. Li, B. Tu, Z. Liao, and Y. Peng, "Hyperspectral anomaly detection via dual collaborative representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4881–4894, 2020.
- [30] R. Wang, H. Hu, F. He, F. Nie, S. Cai, and Z. Ming, "Self-weighted collaborative representation for hyperspectral anomaly detection," *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107718.
- [31] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.
- [32] H. Zhang, J. Cai, W. He, H. Shen, and L. Zhang, "Double low-rank matrix decomposition for hyperspectral image denoising and destriping," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502619.
- [33] L. Sun, T. Zhan, Z. Wu, L. Xiao, and B. Jeon, "Hyperspectral mixed denoising via spectral difference-induced total variation and low-rank approximation," *Remote Sens.*, vol. 10, no. 12, 2018, Art. no. 1956.
- [34] J. Xu, J. E. Fowler, and L. Xiao, "Hypergraph-regularized low-rank subspace clustering using superpixels for unsupervised spatial-spectral hyperspectral classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 5, pp. 871–875, May 2021.
- [35] J. Xu, L. Xiao, and J. Yang, "Unified low-rank subspace clustering with dynamic hypergraph for hyperspectral image," *Remote Sens.*, vol. 13, no. 7, 2021, Art. no. 1372.
- [36] Z. He, J. Li, K. Liu, L. Liu, and H. Tao, "Kernel low-rank multitask learning in variational mode decomposition domain for multi-/hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4193–4208, Jul. 2018.
- [37] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.
- [38] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Deep low-rank prior for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527017.
- [39] S. Wang, X. Wang, Y. Zhong, and L. Zhang, "Hyperspectral anomaly detection via locally enhanced low-rank prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6995–7009, Oct. 2020.
- [40] 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.
- [41] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2020.
- [42] 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.
- [43] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "LREN: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4139–4146.
- [44] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3889–3900, Aug. 2021.
- [45] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [46] T. Cheng and B. J. 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, Jan. 2020.
- [47] H. Su, Z. 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, Nov. 2020.
- [48] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2000, pp. 93–104.
- [49] S. Chang, B. Du, and L. Zhang, "BASO: A background-anomaly component projection and separation optimized filter for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3747–3761, Jul. 2018.
- [50] K. Jiang et al., "E2E-LIAD: End-to-end local invariant autoencoding density estimation model for anomaly target detection in hyperspectral image," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11385–11396, Nov. 2022, doi: [10.1109/TCYB.2021.3079247](https://doi.org/10.1109/TCYB.2021.3079247).
- [51] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [52] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, Apr. 1992.
- [53] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
- [54] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*.
- [55] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
- [56] 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, 2022, Art. no. 5512216.
- [57] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [58] S. Li, K. Zhang, P. Duan, and X. Kang, "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 319–329, Jan. 2020.
- [59] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.
- [60] P. A. Flach, J. Hernández-Orallo, and C. F. Ramirez, "A coherent interpretation of AUC as a measure of aggregated classification performance," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 657–664.
- [61] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [62] A. Lydia and S. Francis, "Adagrad—An optimizer for stochastic gradient descent," *Int. J. Inf. Comput. Sci.*, vol. 6, no. 5, pp. 566–568, May 2019.
- [63] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1578.



Sheng Lin received the B.E. degree in electrical engineering and automation and the M.S. degree in electrical engineering from Zhengzhou University of Light Industry, Zhengzhou, China, in 2017 and 2020, respectively. He is currently working toward the Ph.D. degree in instrument science and technology with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include hyperspectral anomaly detection and object detection.



Kexue Zhou received the B.E. degree in measurement and control technology and instruments in 2018 from Xidian University, Xi'an, China, where he is currently working toward the Ph.D. degree in instrument science and technology with the School of Aerospace Science and Technology.

His research interests include computer vision and machine learning, especially for object detection in remote sensing.



Min Zhang received the B.E. and Ph.D. degrees in measurement and control from Xidian University, Xi'an, China, in 2010 and 2017, respectively.

She is currently an Associate Professor with the School of Aerospace Science and Technology, Xidian University. Her current research interests include time and frequency measurements, digital signal processing, object detection, and hyperspectral images processing.



Shaobo Zhao received the M.S. degree in aerospace engineering in 2019 from the School of Aerospace Science and Technology, Xidian University, Xi'an, China, where he is currently working toward the Ph.D. degree in instrument science and technology.

His research interests include machine learning, enhancement, and restoration.



Xi Cheng received the M.S. degree in control theory and control engineering from Zhengzhou University of Light Industry, Zhengzhou, China, in 2020. He is currently working toward the Ph.D. degree in instrument science and technology with the School of Aerospace Science and Technology, Xidian University, Xi'an, China.

His research interests include hyperspectral anomaly detection and saliency object detection.



Hai Wang received the M.S. degree in communication and information system and the Ph.D. degree in instrument science and technology from the Xidian University, Xi'an, China, in 2003 and 2007, respectively.

He is currently a Professor with the School of Aerospace Science and Technology, Xidian University, Xi'an, China. His current research interests include time and frequency measurements, frequency and phase measurements methods based on field-programmable gates array and system-on-chip, electronic instrument design, time-frequency signal processing, object detection, and hyperspectral image processing.