Feedback Band Group and Variation Low-Rank Sparse Model for Hyperspectral Image Anomaly Detection

Lan Li, Qiang Zhang, *Member, IEEE*, Meiping Song, *Member, IEEE*, and Chein-I Chang, *Life Fellow, IEEE*

Abstract— For scenes with complex backgrounds and weak anomalies, how to effectively distinguish anomaly targets from the background is the key to perform hyperspectral image (HSI) anomaly detection (AD). Data decomposition-based methods have been widely studied due to their potential in separating background and anomaly components. However, due to its unclean background extraction and sensitivity to noise, it has an adverse effect on the detection of anomaly targets. In addition, a large amount of spectral data can lead to an increase in computation during data decomposition. To address this issue, we propose an AD method based on a feedback band group and variation low-rank (LR) sparse model (FBGVLRS-AD). First, we employ a uniform band selection (BS) strategy to partition spectral bands and perform data decomposition on the selected band group, to separate LR and sparse components. This decomposition on the band group can reduce computational time and mitigate the interference from spectral variability. Second, to preserve the integrity of abnormal target spectra during the background extraction process, the $L_{2,1}$ norm is employed for joint correlated total variation (TV) to extract the desired anomalous targets. Then, utilizing the detection information from the existing band groups, a feedback-driven iterative framework has been designed to consider the consistency and complementarity in AD across band groups. This framework facilitates the extraction of sparse components in the subsequent band groups and reinforces the anomalous elements. Iteratively addressing these subproblems on band groups helps prevent the loss of useful spectral information, maintaining sufficient anomaly information while reducing interference from redundant information and spectral variations. Finally, the proposed FBGVLR-AD is optimally solved by the augmented Lagrange multiplier (ALM) method. A comparison with state-of-the-art anomaly detectors on multiple data validates the competitiveness of the proposed method for AD tasks.

Index Terms— Anomaly detection (AD), correlated total variation (CTV), group sparsity, hyperspectral image (HSI) denoising.

I. INTRODUCTION

THE rapid development of high spectral resolution in hyperspectral images (HSIs) has increasingly positioned it as a vital component in Earth observation missions. It has a wide range of applications in areas such as land cover classification [1], [2], target detection [3], and super-resolution [4]. Anomaly detection (AD) has the advantage of being able to detect potential targets without any prior knowledge, and it has highly practical applications in military reconnaissance, mineral exploration, and food quality detection [5], [6]. Given the absence of prior knowledge about anomalies, research on AD primarily focuses on distinguishing anomaly targets from the background by analyzing the disparities in spectral characteristics between the anomaly targets and the background [7]. From the perspective of the core principles of AD methods, AD techniques for HS data can be categorized into model-driven-based methods [8], [9] and data-driven-based methods [10], [11].

The strong feature expression capability of data-driven methods has shown promising results in many fields [12], [13], [14]. In HSI AD, due to the limited number of pixels occupied by anomalous targets, they cannot dominate the convergence of the reconstruction model. This results in poor reconstruction of anomalous targets, whereas background samples can be reconstructed effectively. In other words, the background typically exhibits smaller reconstruction errors, whereas anomalies tend to manifest larger ones. Following this idea, many deep-learning (DL) AD methods have been developed. For example, Lei et al. [15] introduced an anomaly detector relying on spectral learning and discriminative reconstruction. This method constrains the model to exclusively reconstruct the background by acquiring the spectral information of the input data. To further exploit the spatial information of HSI, robust graph autoencoders (RGAEs) [16] introduced graph regularization based on superpixel segmentation, aiming to achieve the preservation of geometric structures and local consistency in HSI. DL methods possess powerful learning capabilities; however, due to the absence of any prior knowledge in AD, its training samples may be prone to errors. In cases where the learning samples are incorrect, it can lead to erroneous learning directions. The lack of prior knowledge hinders DL methods from fully leveraging their advantages [17]. To better separate background component and anomalous target,

an unsupervised discriminative reconstruction constrained generative adversarial network was introduced [18]. This network compels the model to focus on the superior reconstruction of the background. In addition, some studies [19], [20] utilized weak supervision and semi-supervision to select pure background samples for training. They leverage well-trained networks to achieve promising detection results. However, these methods require training specific detectors for each HSI. Furthermore, while autoencoder (AE) network-based AD methods exhibit excellent performance, their effectiveness is often hampered by a lack of prior information. To enhance the reconstruction of background features, guided autoencoder (GAED) [21] reduced the reconstruction of anomalous targets by incorporating guiding modules into the network. In a similar vein, another method [22] utilized a fully convolutional AE for background reconstruction. This method employed an adaptive weighted loss function to suppress abnormal reconstruction, with the advantage of dynamically updating weight coefficients based on reconstruction errors. This eliminates the need for manual parameter tuning. However, it is worth noting that retraining becomes necessary when applying this method to different testing scenarios. Recently, some data-driven studies incorporating low-rank (LR) priors were proposed [11], [23], [24]. For example, Wang et al. [11] introduced a method based on deep LR priors, achieving promising detection results. In this method, the model-driven LR prior and the data-driven fully convolutional AE are jointly modeled as an energy minimization problem.

Model-driven-based methods rely on mathematical descriptions of models and typically exhibit strong predictive capabilities. The classical Reed-Xiaoli (RX) algorithm, considered one of the representative algorithms of the statistical AD baseline, was proposed by Reed and Xu [25]. This detector assumes that the probability density function of the background obeys a multivariate Gaussian distribution and does not require a priori knowledge of the target spectral. Instead, it only computes the global mean and background covariance matrix to evaluate anomaly pixels. Subsequently, some variations of the RX algorithm have been reported for AD, including improved KRX [26], local RX [27], and subspace RX [28], among others. Chang and Chiang [29] simplified the operation of RXD without affecting the detection efficiency and proposed **R**-based correlation matrix AD (RAD), which has become one of the more widely used AD algorithms.

In addition to RX and its variants, various AD algorithms based on data decomposition have been proposed [30], [31]. In recent years, owing to the high correlation in HSI, the LR prior has been widely considered in various tasks [32], such as super-resolution [33], object detection [34], tensor completion [35], and denoising [36]. Similarly, in the task of HSI AD, considering that background information is included in the lowest rank part of HS pixels, Guo et al. [30] developed a hierarchical mutually incoherent induced discrimination (AHMID) learning strategy to model background, anomalies, and noise. Li et al. [37] proposed utilizing tensor decomposition instead of matrix decomposition. The HSI is decomposed into background tensor and anomaly tensor. Anomalies are characterized with the help of spatial group

sparse prior on the anomaly tensor (PTA). In [38], the fractional-order Fourier transform combined with LR sparse matrix factorization was proposed for AD. It is the result of the integration of feature extraction and background purification (IFEBP). Chang et al. [39] jointly employed LR matrices and sparse matrices to achieve background suppression and target detection within the different subspaces they form. They proposed orthogonal subspace projection-based Go-Decomposition (OSP-GoDec) for AD, and this method has shown promising results. Chen et al. [40] developed an AD method called component decomposition. In [40], HS data were represented as a linear orthogonal decomposition of three components: principal component, independent component, and noise component, and performs well for HS data AD.

In the real world, the acquired HS data are usually disturbed by noise due to the limitations of the imaging system and device conditions [41], [42]. Spectral signature corrupted by noise can also affect the improvement of subsequent AD performance [43], [44]. Therefore, there are studies focused on addressing HS data contaminated by noise. They perform denoising operations on noise-contaminated HS data and then perform AD on the recovered data to reduce the impact of noise on anomaly targets. For example, in [45], an AD method was proposed for handling noise-contaminated HS data using spectral unmixing-based dictionary construction and LR matrix decomposition. Zhuang et al. [46] introduced an AD algorithm for noise-contaminated HS data. They utilize LR representation to promote self-similarity among nonlocal similar patches and, through collaborative sparsity, preserve rare pixels to achieve denoising and detection in HS data. Taking into account spatial factors and noise interference, Feng et al. [47] combined superpixel segmentation with total variation (TV) model to better separate the anomaly pixels from HS data by maintaining the local structure of the background pixels. To provide a more comprehensive characterization of the rich spatial and spectral features in HS data, Wang et al. [48] proposed an LR matrix decomposition method with Schatten 1/2 quasi-norm and noise removal, for AD in HS data. However, more regularization implies that a greater number of parameters need to be adjusted to balance the importance of each term. Recently, the correlated TV (CTV) [49] was proposed for HSI restoration tasks. This method utilized nuclear norm to encode strong correlations among gradient images, aiming to recover the LR and local smooth components of the data. It is evident that this primarily considers components occupying the major part of the HSI and does not give more attention to anomaly targets that distinguished from noise. Abnormal targets are typically composed of a few pixels known as "rare pixels," especially weak targets. However, LR approximations cannot effectively approximate the spectral of these rare pixels, and even the spectral of abnormal targets may be corrupted by denoising [46]. These abnormal spectral features consist of rare pixels that might be crucial targets in subsequent applications, such as AD tasks. Unlike the HSI restoration task [36], [41], AD tasks pay more attention to regions or pixels that differ from the image background mode. These pixels typically occur with a relatively low probability and are composed of a smaller

number of pixels. Considering that abnormal targets exhibit a highly sparse distribution in the spatial dimension but still maintain density in the spectral dimension. To enable the model to more fully retain information about abnormal pixels, we employ a mixed norm that satisfies column-wise sparsity to characterize the sparse component $\mathbf{S}$ (which is not well represented in CTV). This is advantageous for preserving the anomaly targets of interest and protecting them from being corrupted during denoising processing.

Furthermore, due to the information-rich nature of HS data, simultaneously decomposing the acquired full-band HS data not only increases computational time but can also lead to significant mutual "interference," thereby affecting the efficiency and accuracy of AD tasks [50]. In order to effectively improve the computational accuracy of AD and reduce the computation, Wang et al. [51] utilized principal component analysis (PCA) to perform preprocessing operations on the observed HS data and proposed a PCA-based tensor LR and sparse representation model. This type of preprocessing by feature extraction can effectively reduce the computation of the model and retain a certain amount of information for subsequent tasks. However, PCA is more concerned with the retention of LR information, so the resulting bands are not conducive to anomaly processing. In addition, AD based on band selection (BS) has been successively proposed to reduce the redundancy interference and computationally heavy problem of HS data. For instance, Shang et al. [52] introduced a residual-driven BS method to enhance the capability of a band subset for anomaly recognition, to cope with the challenge of lacking prior information about anomalies. Xie et al. [53] proposed a spectral–spatial AD method based on a specific subset of bands by utilizing the underlying physical features favoring the AD of HS data to constrain the unsupervised network. It selects a band subset in an unsupervised manner and combines the suppression background of simultaneous iterations of spatial and spectral detectors. However, AD is unsupervised, it is challenging to provide a guiding criterion for BS in such cases. Therefore, how to reduce the computational time of HS data decomposition without losing anomaly information is an issue that needs to be further explored.

In this study, we propose an AD method based on feedback band group and variation LR sparse model (FBGVLRS-AD). The proposed FBGVLRS-AD method is developed from data decomposition and can be effective in significantly reducing computational time for decomposition. First, in contrast to previous AD methods based on the combination of LR and TV models as two separate regularization terms, we utilize the nuclear norm on the gradient map to simultaneously encode LR and local smoothness properties, thus avoiding the need for additional parameter tuning. Furthermore, taking into account that anomaly targets spectral is distinguished from random distributed sparse noise, it exhibits sparsity in the spatial domain but remains dense in the spectral domain. Therefore, we constrain them by jointly considering the spectral–spatial sparsity, to preserve the spectral characteristics of abnormal target pixels during the background extraction process. Second, directly decomposing full-band HS data imposes a heavier computational burden. Unlike the preprocessing methods of PCA and

BS, we do not selectively retain or discard any information or bands. Instead, we use BS [uniform BS (UBS)] to partition the computation of large-scale full-band data into small-scale computations on multiple band groups using uniform UBS and then iteratively process these subproblems. This approach enables the preservation of sufficient information for subsequent processing, mitigating limitations on AD performance resulting from the loss of valuable spectral information due to a lack of prior knowledge. In addition, decomposing for the band groups also reduces the computational load. Furthermore, high band redundancy can increase the difficulty of distinguishing between $\mathbf{N}$ and $\mathbf{S}$ components, thereby affecting the accuracy of $\mathbf{S}$ extraction. This division process helps reduce interference from nonanomalous components. Third, these band groups contain abundant information, comprising interesting anomalous targets and background components that require suppression. In HS data, information occupying large areas is typically deemed nonanomalous targets and exhibits piecewise smooth characteristics. The CTV model can enhance the continuity of these regions, facilitating the separation of nonanomalous components, and thereby preserving the anomaly targets of interest. Fourth, there is a degree of consistency and complementarity in the AD performance among different band groups. We have designed an iterative framework that prompts us to utilize existing knowledge of the detection for band groups to guide and facilitate the separation of the sparse component $\mathbf{S}$ for subsequent band groups, thereby improving the performance of AD. Compared to existing data decomposition-based AD methods, the contributions of our presented method can be summarized as follows.

1) This article introduces a band group-based iterative feedback framework by leveraging the consistency and complementarity of AD results across various band groups. In this framework, the computation of large-scale full-band data is divided into multiple small-scale computations on band groups. The extracted information about anomalous targets serves as feedback to guide the decomposition of subsequent band groups, aiming to better preserve and highlight anomalous targets. The proposed framework accelerates detection speed while avoiding issues related to the loss of anomaly information caused by spectral BS.

2) The CTV model is employed in data decomposition AD tasks, incorporating a combined $L_{2,1}$ norm. The CTV model considers both the low rankness and local smoothness of nonanomalous components in HS data, eliminating the need for adjusting multiple parameters. The $L_{2,1}$ norm is employed to preserve the spectral integrity of anomalous targets during the background extraction process.

3) An iterative optimization algorithm based on augmented Lagrange multiplier (ALM) is designed to derive and solve the proposed FBGVLRS-AD method. Experimental results on multiple real datasets demonstrate the effectiveness of the proposed method, particularly for weak anomalous target features.

The remaining sections of this article are organized as follows. Section II provides an overview of existing methods,

including LR and sparse model, and TV model-based AD. Section III offers a detailed introduction to the proposed AD method based on feedback band group and variation LR sparse model. Section IV covers the experimental results and performance analysis. Finally, the conclusions drawn from the study and the outline of future work are presented in Section V.

II. RELATION METHOD

A. LRaSMD for AD

$\mathcal{X} \in \mathbb{R}^{h \times w \times b}$ is a 3-D HS data cube with a spatial size of $h \times w$ and comprising b bands. Here, we unfold $\mathcal{X}$ into $\mathbf{X} \in \mathbb{R}^{b \times n}$, where $n = h \times w$ represents the total number of pixels in the HS data. Background pixels in HS data exhibit a high correlation and can be linearly represented by other background pixels. This can be expressed through the following matrix factorization model:

$$\mathbf{X} = \mathbf{L} + \mathbf{S}. \tag{1}$$

Equation (1) implies that the original data $\mathbf{X}$ can be regarded as a combination of an LR matrix $\mathbf{L}$ and a sparse matrix $\mathbf{S}$. Equation (1) has numerous feasible solutions, and the common way is to impose constraints on the background component $\mathbf{L}$ and sparse component $\mathbf{S}$. Building upon the high correlation of HS data, a common way is to apply an LR constraint to the background component $\mathbf{L}$, which requires suppression. The anomalous targets that need to be detected typically reside in $\mathbf{S}$, and spatially, they often exhibit characteristics of sparse distribution. Therefore, (1) can be extended to the minimization form in the following equation:

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \beta \|\mathbf{S}\|_0$$
$$\text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S} \tag{2}$$

Considering that the anomalous signatures are different from noise, they are still dense in spectral dimension. This implies that $\mathbf{S}$ is column-wise sparse. Based on the above analysis, the following models can be derived:

$$\min_{\mathbf{L},\mathbf{S}} \operatorname{rank}(\mathbf{L}) + \beta \|\mathbf{S}\|_{2,1}$$
$$\text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S}. \tag{3}$$

In this context, β is a tradeoff coefficient greater than 0, intended to balance the importance of background and sparse components. rank(.) and $\|\cdot\|_{2,1}$ denote the rank function and a mixed norm that promotes column sparsity in the matrix $\mathbf{S}$ [51], respectively. The definition of $\|\cdot\|_{2,1}$ is as follows:

$$\|\mathbf{S}\|_{2,1} = \sum_{j=1}^{rc} \sqrt{\sum_{i=1}^{b} \mathbf{S}_{i,j}^2}. \tag{4}$$

However, (3) is nonconvex and NP-hard. A commonly employed approach is to replace the nonconvex rank function with a convex function approximating it, known as the nuclear norm. The equation is then reformulated as follows:

$$\min_{\mathbf{L},\mathbf{S}} \|\mathbf{L}\|_* + \beta \|\mathbf{S}\|_{2,1}$$
$$\text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{S} \tag{5}$$

where $\|\cdot\|_*$ represents the matrix nuclear norm, solved using the singular value thresholding method.

B. TV-Based for AD

Owing to the piecewise smooth characteristics of the TV model, it can effectively portray the structural features of HS data. The TV model has been employed in AD methods that take spatial relationships into account, thereby enhancing the extraction of background component

$$\min_{\mathbf{Z},\mathbf{S}} \frac{1}{2} \left\| (\mathbf{Z} - \hat{Z})\mathbf{W} \right\|_F^2 + \lambda_1 \text{TV}(\mathbf{Z}) + \lambda_2 \|\mathbf{S}\|_{2,1}$$
$$\text{s.t. } \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{S} \tag{6}$$

where $\hat{Z}$ is the guide matrix containing local spatial information, $\mathbf{W}$ is a weighted diagonal matrix based on anomaly probability, $\mathbf{D}\mathbf{Z}$ represents the background component, and $\mathbf{D}$ and $\mathbf{Z}$ denote the pretrained dictionary and its corresponding representation coefficients, respectively. In (6), the TV term is used to enhance the correlation between pixels in order to approximate the background component. Simultaneously, it also has a noise-suppressing effect. Furthermore, due to the weaker correlation between anomalies and surrounding pixels, utilizing the TV model can effectively isolate anomalous pixels, which is advantageous for background extraction.

In addition, considering the potential influence of Gaussian noise contamination or model errors on HS data [54], the model in (1) can be further extended as

$$\mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N} \tag{7}$$

where $\mathbf{N} \in \mathbb{R}^{b \times n}$ is additive Gaussian noise or modeling error.

III. FEEDBACK BAND GROUP AND VARIATION LR SPARSE AD

In this article, we achieve this by partitioning the full-band data into multiple band groups and performing a data decomposing on band groups, which can efficiently reduce the computational burden of the decomposition process and avoid the increase in residual information after extracting the LR part, caused by the spectral variability of full-band data, especially when the rank is low. The increase in residual information leads to a greater challenge in distinguishing between the $\mathbf{N}$ and $\mathbf{S}$. In addition, by jointly considering the CTV LR property and sparse norm, we effectively separate the nonanomalous LR component and retain the anomaly targets of interest. Finally, leveraging obtained anomaly knowledge in a feedback loop facilitates subsequent decomposition processes, achieving rapid decomposition while enhancing the performance of AD task. Fig. 1 shows the schematic of the proposed FBGVLRS-AD method. The FBGVLR-AD algorithm consists of three components: 1) data decomposition on a band group aims to reduce the computation time for data decomposition while minimizing interference from redundant information and spectral variations; 2) local smoothing and LR, and jointly spectral–spatial sparsity constraints, are employed to extract interesting anomalous targets on band groups; and 3) iterative feedback mechanism, utilizing an
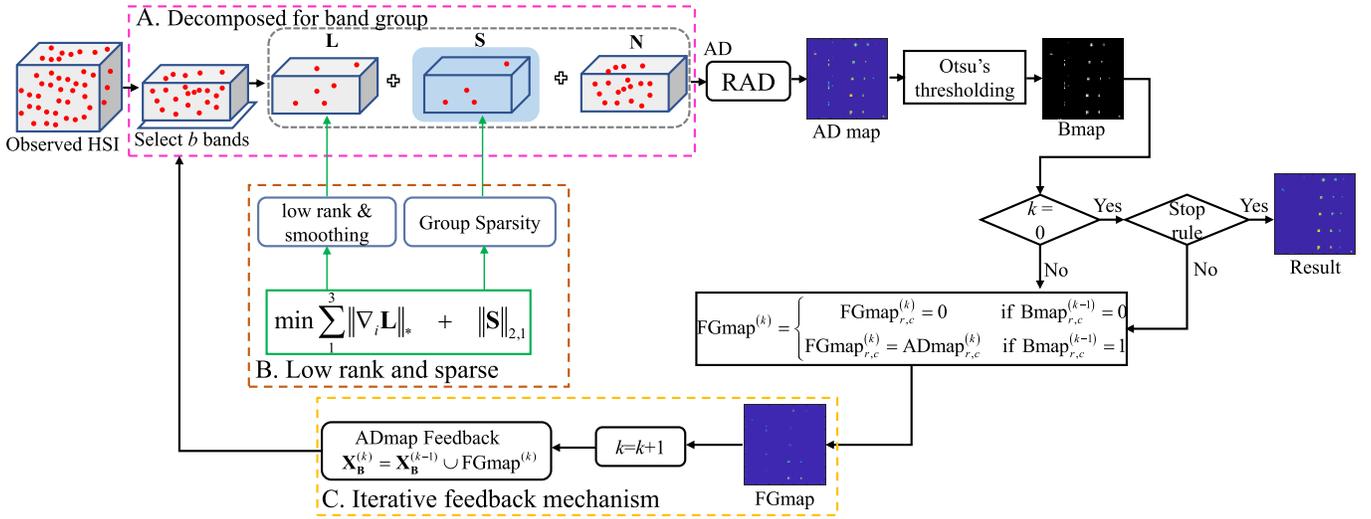
Fig. 1. Flowchart of the proposed FBGVLRS for HSI AD. (a) Decomposed for band group. (b) LR and sparse. (c) Iterative feedback mechanism.

existing anomaly target detection information-guided subsequent band group decomposition. A more detailed description will be provided in the following.

A. LRaSMD for Band Group

UBS is one of the simplest but quite usable BS methods, which does not require any prior knowledge or BS criteria. Here, we utilize UBS for band grouping. To effectively suppress the interference of background information, LR and sparse representations are applied to the band group. The benefits of data decomposition for a few bands are twofold: 1) it can improve the efficiency of decomposition and 2) the effect of spectral variability increases when the number of bands increases, and constraining it with a smaller rank leads to an increase in the remaining information after extracting the LR component. This results in increased interference from noise on the interesting anomaly targets and makes it more difficult to distinguish $\mathbf{S}$ and $\mathbf{N}$, ultimately reducing the accuracy of extracting $\mathbf{S}$. Decomposition for a few bands can reduce the interference of nonanomaly components. The effective separation of the sparse component $\mathbf{S}$ from the background component $\mathbf{L}$ is crucial for reducing false alarm rates in AD on $\mathbf{S}$ [55]

$$\min_{\mathbf{L_B},\mathbf{S_B}} \|\mathbf{L_B}\|_* + \beta\|\mathbf{S_B}\|_{2,1}$$
$$\text{s.t. } \mathbf{X_B} = \mathbf{L_B} + \mathbf{S_B} + \mathbf{N_B} \qquad (8)$$

where $\mathbf{B} = \{b_1, b_2, \ldots, b_i\}$ represents a band group selected via UBS.

In the implementation of the proposed AD method, the key parameter j used for band group selection should be predetermined. According to [39] and [56], j is considered the rank of the sparse matrix $\mathbf{S}$, and its value can be calculated using MX-SVD [57]. The virtual dimension (VD) [58] can be used to estimate p, for representing the number of total base vectors of the LR matrix $\mathbf{L}$ and the sparse matrix $\mathbf{S}$, i.e., the rank p of $\mathbf{L} + \mathbf{S}$ is specified by the VD, and $p = m + j$. This means that the value of m can be obtained by $m = p - j$.

It should be noted that both HFC [59], [60] and NWHFC [58], [59] can be used to estimate the value of VD. Here, the VD values estimated by these two methods are denoted as VD^{HFC} and VD^{NWHFC}.

B. Piecewise Smoothing and LR Priors by CTV Regularization

Anomalies are usually composed of a few pixels, and large areas of information in HS data are generally considered nonanomalous and they are spatially continuous. The method based on LR and sparse preserves the main information of the data by minimizing the kernel norm but does not consider the spatial factors of the image, resulting in imperfect extraction of the nonanomalous LR component. This will lead to two situations: 1) some components that should be anomalous are incorrectly identified as background, reducing the detection rate of AD or 2) some background components are incorrectly recognized as anomalous, leading to an increase in false alarm rates. The TV regularization term can effectively characterize the piecewise smoothness properties that have been successfully applied in many fields [47], [49], [61], [62] and achieved remarkable results. However, more regular terms imply that more parameters need to be tuned. Therefore, in this article, we take advantage of the ability of the CTV model to encode both the LR and local smoothing properties of the gradient map and introduce it into the AD task to promote the separation of the nonanomalous LR components. At the same time, incorporating a mixed norm that satisfies the column-wise sparsity to characterize the sparse component $\mathbf{S}$, to preserve the anomalous targets of interest. Finally, the proposed objective model is formulated as

$$\min_{\mathbf{L_B},\mathbf{S_B}} \sum_{i=1}^{3} \|\nabla_i\mathbf{L_B}\|_* + \beta\|\mathbf{S_B}\|_{2,1}$$
$$\text{s.t. } \mathbf{X_B} = \mathbf{L_B} + \mathbf{S_B} + \mathbf{N_B}, i = 1, 2, 3 \qquad (9)$$

where ∇_1, ∇_2, and ∇_3 represent the first-order difference operators in the horizontal, vertical, and spectral directions, respectively.

In order to make the objective function separable, an auxiliary variable $\mathbf{J}$ is introduced, and with $\mathbf{J} = \nabla \mathbf{L_B}$, the first term $\sum_{i=1}^{3} \|\nabla_i \mathbf{L_B}\|_*$ in the objective function (9) can be rewritten as $\sum_{i=1}^{3} \|\mathbf{J}_i\|_*$. Without loss of generality, $\mathbf{X}$, $\mathbf{L}$, $\mathbf{S}$, and $\mathbf{N}$ are substituted for $\mathbf{X_B}$, $\mathbf{L_B}$, $\mathbf{S_B}$, and $\mathbf{N_B}$ to derive the solution process. Based on this, the question posed can be formulated as

$$\min_{\mathbf{X},\mathbf{S}} \sum_{i=1}^{3} \|\mathbf{J}_i\|_* + \beta \|\mathbf{S}\|_{2,1}$$
$$\text{s.t.} \mathbf{X} = \mathbf{L} + \mathbf{S} + \mathbf{N}$$
$$\mathbf{J}_i = \nabla_i(\mathbf{L}) \quad (10)$$

and using the ALM method to solve (10), it can be rewritten to minimize the following augmented Lagrange function:

$$\min_{\mathbf{L},\mathbf{S}} \ell(\mathbf{L}, \mathbf{S}, \mathbf{J}, \Gamma_i)$$
$$= \min_{\mathbf{L},\mathbf{J},\mathbf{S},\Gamma} \sum_{i=1}^{3} \left(\|\mathbf{J}_i\|_* + \frac{\lambda}{2}\|\nabla_i \mathbf{L} - \mathbf{J}\|_F^2 + <\Gamma_i, \nabla_i \mathbf{L} - \mathbf{J}_i> \right)$$
$$+ <\Gamma, \mathbf{X} - \mathbf{L} - \mathbf{S}> + \frac{\lambda}{2}\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2 + \beta\|\mathbf{S}\|_{2,1} \quad (11)$$

where λ represents the penalty parameter, Γ is the Lagrange multiplier, and the inner product of the two matrices is represented by $< \cdot, \cdot >$. A typical way to solve (11) is to alternately optimize one variable, while the other variables remain fixed. In detail, in the $t + 1$ iteration, the update of each variable is

$$\mathbf{J}_i^{t+1} = \underset{\text{rank}(\mathbf{J}_i) \leq r}{\arg\min} \ell(\mathbf{J}, \mathbf{L}^t, \mathbf{S}^t, \Gamma^t)$$
$$\mathbf{L}^{t+1} = \underset{\mathbf{L}}{\arg\min} \ell(\mathbf{J}^{t+1}, \mathbf{L}, \mathbf{S}^t, \Gamma^t)$$
$$\mathbf{S}^{t+1} = \underset{\mathbf{S}}{\arg\min} \ell(\mathbf{J}^{t+1}, \mathbf{L}^{t+1}, \mathbf{S}, \Gamma^t)$$
$$\Gamma^{t+1} = \Gamma^t + \lambda(\mathbf{X} - \mathbf{L}^{t+1} - \mathbf{S}^{t+1}). \quad (12)$$

By dividing (11) into three major subproblems in (12), based on these subproblems, the following equation can be derived.

1) The update of subproblem $\mathbf{J}$ can be obtained by solving the following equation [63], [64]:

$$\mathbf{J}_i^{t+1} = \underset{\text{rank}(\mathbf{J}_i) \leq r}{\arg\min} \ell(\mathbf{J}, \mathbf{L}^t, \mathbf{S}^t, \Gamma^t)$$
$$= \underset{\text{rank}(\mathbf{J}_i) \leq r}{\arg\min} \|\mathbf{J}_i\|_* + \frac{\lambda}{2}\|\nabla_i \mathbf{L} - \mathbf{J}\|_F^2 + <\Gamma_i, \nabla_i \mathbf{L} - \mathbf{J}_i>$$
$$= \underset{\text{rank}(\mathbf{J}_i) \leq r}{\arg\min} \|\mathbf{J}_i\|_* + \frac{\lambda}{2}\left\|\nabla_i \mathbf{L} - \mathbf{J}_i + \frac{\Gamma_i}{\lambda}\right\|_F^2. \quad (13)$$

The solution to problem (13) can be expressed as

$$\begin{cases} \mathbf{J}_i^{t+1} = \mathbf{U}\mathbf{S}_{1/\lambda}(\Sigma)\mathbf{V}^T \\ \mathbf{U}\Sigma\mathbf{V}^T = \text{svd}(\nabla_i \mathbf{L}^t + \Gamma_i/\lambda, 'econ') \end{cases} \quad (14)$$

2) Then, the subproblem $\mathbf{S}$ is updated as follows:

$$\mathbf{S}^{t+1} = \underset{\mathbf{S}}{\arg\min} \ell(\mathbf{J}^{t+1}, \mathbf{L}^{t+1}, \mathbf{S}, \Gamma^t)$$
$$= \underset{\mathbf{S}}{\arg\min} \beta\|\mathbf{S}\|_{2,1} + <\Gamma, \mathbf{X} - \mathbf{L} - \mathbf{S}>$$
$$+ \frac{\lambda}{2}\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2$$

$$= \underset{\mathbf{S}}{\arg\min} \mu\|\mathbf{S}\|_{2,1} + \frac{\lambda}{2}\left\|\mathbf{X} - \mathbf{L} - \mathbf{S} + \frac{\Gamma_4}{\lambda}\right\|_F^2. \quad (15)$$

Equation (15) can be solved with the help of the following equation [65]:

$$[\mathbf{S}]_{i,:} = \begin{cases} \dfrac{\left\|[\mathbf{E}]_{i,:}\right\|_2 - \beta_2}{\left\|[\mathbf{E}]_{i,:}\right\|_2}[\mathbf{E}]_{i,:}, & \beta_2 < \left\|[\mathbf{E}]_{i,:}\right\|_2 \\ 0, & \text{otherwise} \end{cases} \quad (16)$$

where $\mathbf{E} = \mathbf{X} - \mathbf{L} + \frac{\Gamma_4}{\lambda}$ and $\beta_2 = \mu/\lambda$.

3) The subproblem related to $\mathbf{L}$ can be formulated as follows:

$$\mathbf{L}^{t+1} = \underset{\mathbf{L}}{\arg\min} \ell(\mathbf{J}_i^{t+1}, \mathbf{L}, \mathbf{S}^t, \Gamma^t)$$
$$= \underset{\mathbf{L}}{\arg\min} \|\mathbf{J}_i\|_* + \frac{\lambda}{2}\|\nabla_i \mathbf{L} - \mathbf{J}\|_F^2 + <\Gamma_i, \nabla_i \mathbf{L} - \mathbf{J}_i>$$
$$+ <\Gamma, \mathbf{X} - \mathbf{L} - \mathbf{S}> + \frac{\lambda}{2}\|\mathbf{X} - \mathbf{L} - \mathbf{S}\|_F^2$$
$$= \underset{\mathbf{L}}{\arg\min} \sum_{i=1}^{3} \frac{\lambda}{2}\left\|\nabla_i \mathbf{L} - \mathbf{J}_i^{t+1} + \frac{\Gamma_i^t}{\lambda}\right\|_F^2$$
$$+ \frac{\lambda}{2}\left\|\mathbf{X} - \mathbf{L} - \mathbf{S}^{t+1} + \frac{\Gamma_4^t}{\lambda}\right\|_F^2. \quad (17)$$

Equation (17) can be updated by the following formula:

$$\left(\lambda \mathbf{I} + \lambda \sum_{i=1}^{3} \nabla_i^T \nabla_i\right)\mathbf{L}$$
$$= \lambda(\mathbf{X} - \mathbf{S}^{t+1}) + \Gamma_4^t + \lambda \sum_{i=1}^{3} \nabla_i^T(\mathbf{J}_i^{t+1}) - \nabla_n^T(\Gamma_n^t) \quad (18)$$

where ∇_i^T denotes the transpose operator of $\nabla_i(\cdot)$. $\nabla_i^T \nabla_i$ can be diagonalized by a 3-D fast Fourier transform (FFT) matrix. A closed solution for $\mathbf{L}^{t+1}$ can be derived by Fourier transforming both sides of (18) and using the convolution theorem

$$\begin{cases} \mathbf{H} = \sum_{i=1}^{3} F(\mathbf{D}_n)^* \odot F\left(\text{fold}(\mu \mathbf{J}_i^{t+1} - \Gamma_i^t)\right) \\ \mathbf{T}_x = |F(\mathbf{D}_1)|^2 + |F(\mathbf{D}_2)|^2 + |F(\mathbf{D}_3)|^2 \\ \mathbf{L}^{t+1} = F^{-1}\left(\dfrac{F\left(\text{fold}(\mu \mathbf{X} - \lambda \mathbf{S}^{t+1} + \Gamma_4^t)\right) + \mathbf{H}}{\lambda(\mathbf{1} + \mathbf{T}_x)}\right) \end{cases} \quad (19)$$

where $\odot$ and $|\cdot|^2$ in (19) are element-wise multiplication and square operations, respectively, and $F(\cdot)$ denotes the Fourier transform.

4) Finally, the update of the Lagrange multiplier can be expressed in the following equation:

$$\Gamma_i^{t+1} = \Gamma_i^t + \lambda(\nabla_i \mathbf{L}^{t+1} - \mathbf{J}_i^{t+1})$$
$$\Gamma_4^{t+1} = \Gamma_4^t + \lambda(\mathbf{X} - \mathbf{L}^{t+1} - \mathbf{S}^{t+1})$$
$$\lambda = \lambda\rho \quad (20)$$

where ρ denotes a constant, which is actually greater than 1.

C. Iterative Feedback Mechanism

Although data decomposition based on band group $\mathbf{B}$ can reduce the calculation workload during the decomposition process, UBS is not a BS method for anomaly targets, which will limit the performance of AD to a certain extent. The guidance of prior information is more conducive to improving AD performance. To solve this problem, we introduce the idea of iterative feedback into the proposed efficient decomposition model. Inspired by [6], anomaly detectors typically produce only a single AD map, unable to provide sufficient spatial and spectral information. In order to utilize the information of anomalous targets reflected in the AD map, FGmap is defined. Specifically, the AD map is thresholded by Otsu [66] to obtain the binary map Bmap. In FGmap, the original AD map information is retained according to Bmap to guide the separation of the sparse part, and the background part is set to 0. Thus, the FGmap is defined as follows:

$$\text{FGmap}^{(k+1)} = \begin{cases} \text{FGmap}_{r,c}^{(k)} = 0 & \text{if } \text{Bmap}_{r,c}^{(k)} = 0 \\ \text{FGmap}_{r,c}^{(k)} = \text{ADmap}_{r,c}^{(k)} & \text{if } \text{Bmap}_{r,c}^{(k)} = 1. \end{cases} \tag{21}$$

Finally, the FGmap is used for feedback information to guide the next decomposition process, i.e., $\mathbf{X}_{\mathbf{B}}^{(k)} = \mathbf{X}_{\mathbf{B}}^{(k-1)} \cup \text{FGmap}^{(k)}$, where k denotes the kth iteration. The above process is executed iteratively until the stopping rule [6] in (22) is satisfied

$$\text{TI}^{(k)} = \frac{|\text{Bmap}^{(k)} \cap \text{Bmap}^{(k-1)}|}{|\text{Bmap}^{(k)} \cup \text{Bmap}^{(k-1)}|} > \varphi. \tag{22}$$

The band group has been extended according to the feedback information of FGmap and iteratively updating the AD map. At the same time, it avoids the limitations on AD performance caused by the selected band subset due to the lack of prior knowledge of anomalies. The proposed method can make use of the sufficient information contained in the HS data cube and reduce the computational workload effectively.

D. Algorithms

Algorithm 1 summarizes the processing of the proposed FBGVLR-AD. It can be viewed as two stages: the separation of background and anomalies, and the feedback of the obtained AD map information. In the first stage, the optimization process is terminated when the relative error or the maximum number of iterations is met. Similarly, in the second stage, the iterative process is stopped when the $\text{TI}^{(k)}$ value of two consecutive Bmaps is greater than a given threshold φ or the iteration maximum number of iterations k is reached.

IV. EXPERIMENT RESULTS AND PERFORMANCE ANALYSIS

A. Real HSI AD Datasets

Five benchmark datasets were used to evaluate the detection performance of the method proposed and the baseline methods. In this section, we provide a detailed introduction to the datasets used.

Algorithm 1 FBGVLR for HSI AD

Input: Observed HSI $\mathcal{X} \in \mathbb{R}^{h \times w \times b}$ is unfolded as $\mathbf{X} \in \mathbb{R}^{n \times b}$, $\lambda = 1\sqrt{hw}$, threshold φ, and $k = 0$.

Initialization: $\Omega_{\mathbf{X}}^{(k)} = \text{HSI}$, $\mathbf{L} = \text{randn}(hw, b/j)$, $\mathbf{S} = \mathbf{0}$.

1: **while** $0 \leq k \leq$ maxiter or meet the stop conditions **do**
2:     Select band group: $\mathbf{X}_{\mathbf{B}}^{(k)} = \text{UBS}_{\text{HSI}}$
3:     **while** $0 \leq t \leq$ maxiter or meet the convergence **do**
4:         Updating $\mathbf{J}^t$ via (14)
5:         Updating $\mathbf{S}^t$ via (16)
6:         Updating $\mathbf{L}^t$ via (19)
7:         Updating Γ^t via (20)
8:         check the convergence conditions

$$\left\|\mathbf{X} - \mathbf{L}^{t+1} - \mathbf{S}^{t+1}\right\|_F^2 / \|\mathbf{X}\|_F^2 \leq \tau_1,$$
$$\left\|\nabla_i \mathbf{L}^{t+1} - \mathbf{J}_i^{t+1}\right\|_F^2 / \|\mathbf{X}\|_F^2 \leq \tau_2, i = 1, 2, 3$$

9:         $t = t + 1$;
10:    **end while**
11:     Implement $\delta_{\mathbf{S}_j}^{\text{RAD}}(\mathbf{r}_{\mathbf{S}_j})$ on $\mathbf{B}_{\mathbf{S}}^{(k)}$ to produce $\text{ADmap}^{(k)}$;
12:     using Otsu's to threshold $\text{ADmap}^{(k)}$ to produce $\text{Bmap}^{(k)}$;
13:     Check the conditions by (22)
    $\text{TI}^{(k)} = \frac{|\text{Bmap}^{(k)} \cap \text{Bmap}^{(k-1)}|}{|\text{Bmap}^{(k)} \cup \text{Bmap}^{(k-1)}|} > \varphi$
14:     Create a foreground map $\text{FGmap}^{(k)}$ by

$$\text{FGmap}^{(k+1)} = \begin{cases} \text{FGmap}_{r,c}^{(k)} = 0 & \text{if } \text{Bmap}_{r,c}^{(k)} = 0 \\ \text{FGmap}_{r,c}^{(k)} = \text{ADmap}_{r,c}^{(k)} & \text{if } \text{Bmap}_{r,c}^{(k)} = 1 \end{cases}$$

15:     $k = k + 1$;
16:     Expand the current data cube by $\mathbf{X}_{\mathbf{B}}^{(k)} = \mathbf{X}_{\mathbf{B}}^{(k-1)} \cup \text{FGmap}^{(k)}$
17: **end while**

Output: Anomaly detection map

---

1) HYDICE Panel Scene: The first dataset was collected by the hyperspectral digital imagery collection experiment (HYDICE) sensor in 1995. The scene was captured at an altitude of 10 000 ft with a ground sampling distance of approximately 1.56 m. It has been extensively studied in various literature [67], [68]. The data dimension is $64 \times 64 \times 169$. The scene includes 15 square panels with three different sizes: 3×3 m, 2×2 m, and 1×1 m. Among them, each panel in the first column of rows 2–5 contains two panel pixels highlighted in red, while the remaining 11 panels contain one red panel pixel. Therefore, there are a total of 19 red panel pixels in this scene. Their precise spatial locations are marked in Fig. 2(b), where yellow pixels represent panel pixels blended with the background. The pseudocolored image, ground truth, and spectral features of this scene are shown in Fig. 2(a)–(c).

2) HYDICE Urban Scene: The second dataset used in the experiment was also collected by the HYDICE sensor. The original scene has a size of 307×307 pixels, and for the purpose of AD studies, an 80×100 region is selected. This scene mainly consists of highways, rooftops, vehicles, and factories, with vehicles marked as anomaly targets, totaling 21 anomaly pixels and 188 spectral bands. Fig. 3(a) displays
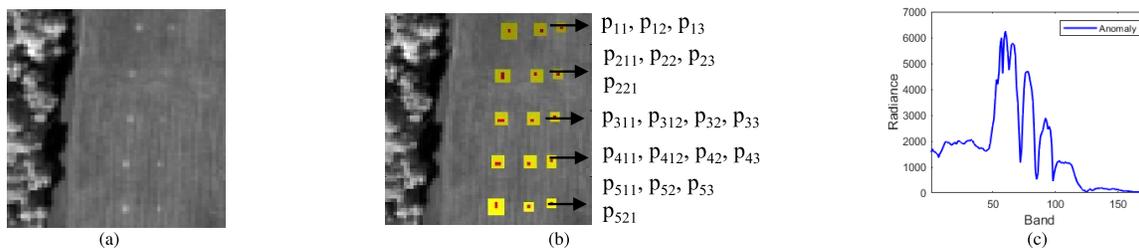
Fig. 2. Pseudocolored image, ground truth, and spectral features of HYDICE15-panel scene. (a) HYDICE15-panel scene (band 59). (b) Ground-truth map of 19R panel pixels. (c) Ground-truth map of 19R panel pixels.
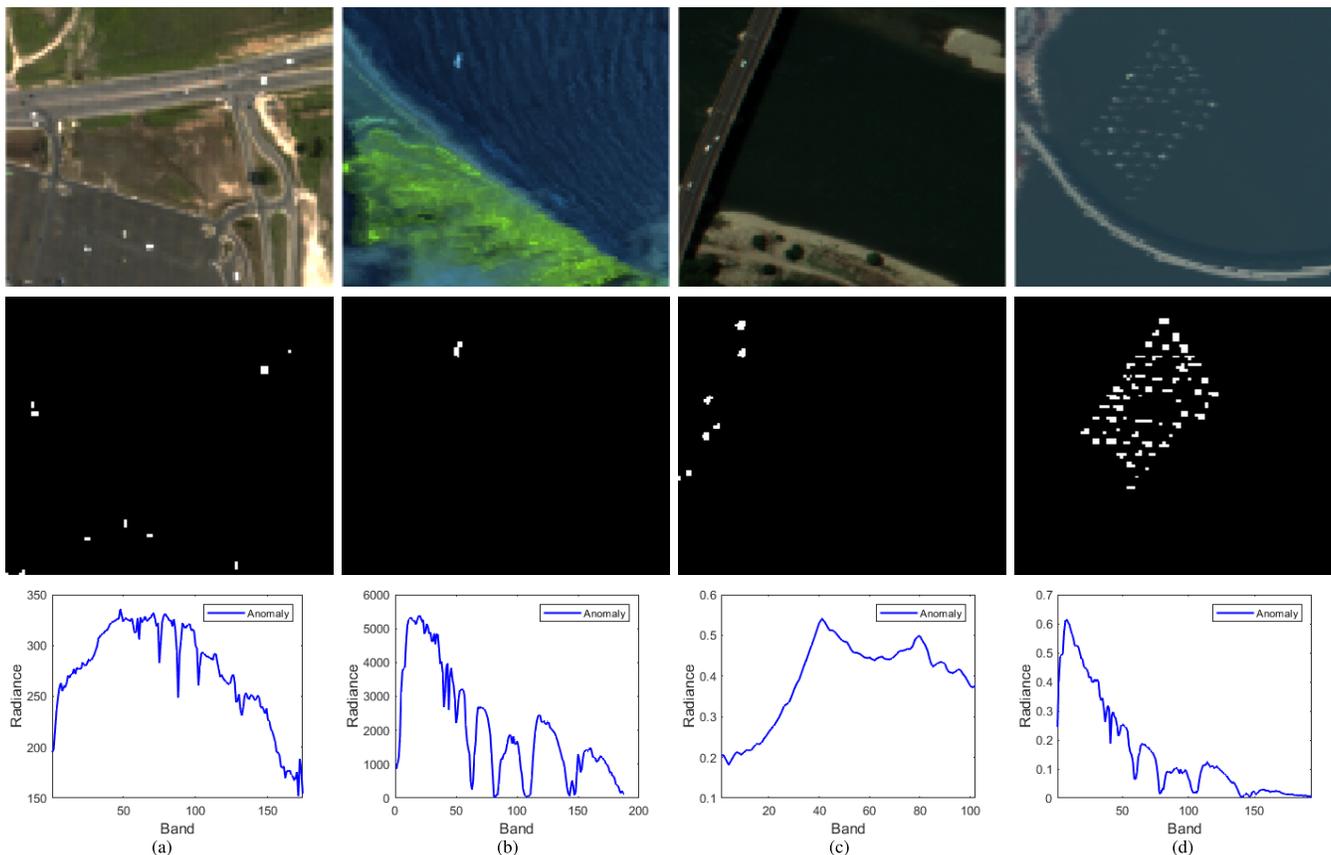

Fig. 3. Pseudocolor images, ground-truth maps, and spectral signature of four experimental datasets. (a) HYDICE Urban (R-42, G-33, and B-19). (b) ABU-beach3 (R-58, G-46, and B-32). (c) Pavia Center (R-56, G-33, and B-22). (d) ABU-beach2 (R-26, G-12, B-8).

the pseudocolored image, ground truth, and spectral signature map corresponding to this dataset.

3) Airport-Beach-Urban (ABU)-Beach3 Scene: The third dataset was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, containing 188 spectral bands and comprising 100×100 pixels. Among these pixels, 11 are considered anomaly targets. This dataset is commonly used in AD studies to validate the performance [69]. The corresponding pseudocolored image, ground truth, and spectral signature are shown in Fig. 3(b).

4) Pavia Center Scene: The fourth dataset was obtained using the Reflective Optics System Imaging Spectrometer (ROSIS) sensor while flying over the city of Pavia [70]. These HS data have 1096×1096 pixels with a spatial resolution of 1.3 m and retain spectral information from 102 bands. For the AD task, a subscene of size 150×150 pixels is used, where anomalies mainly consist of vehicles on bridges and

bridge supports, totaling 68 anomaly pixels. The corresponding pseudocolored image, ground truth, and spectral signature of this scene are depicted in Fig. 3(c).

5) ABU-Beach2 Scene: The fifth dataset, also captured by the AVIRIS sensor, comprises 100×100 pixels, with 202 pixels considered as anomaly targets. It includes 193 spectral bands and is commonly used to validate the performance in AD studies [69]. The corresponding pseudocolored image, ground truth, and spectral signature of this dataset are shown in Fig. 3(d).

In addition, the estimates for j across different datasets by VD and MX-SVD are presented in Table I.

B. Metrics

Given the strengths and weaknesses of different methods, for a more comprehensive assessment of the proposed method's performance, we employ three metrics: 2-D receiver
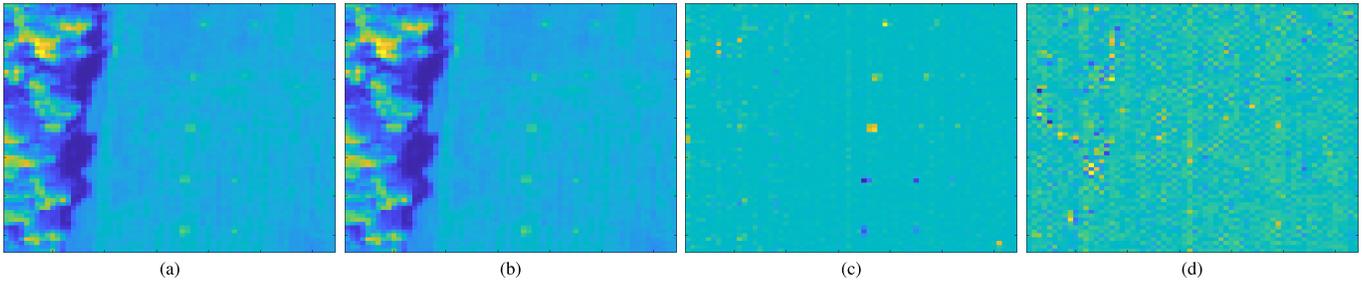
Fig. 4. Visualization results of various components for the full bands. (a) **X.** (b) **L.** (c) **S.** (d) **N.**
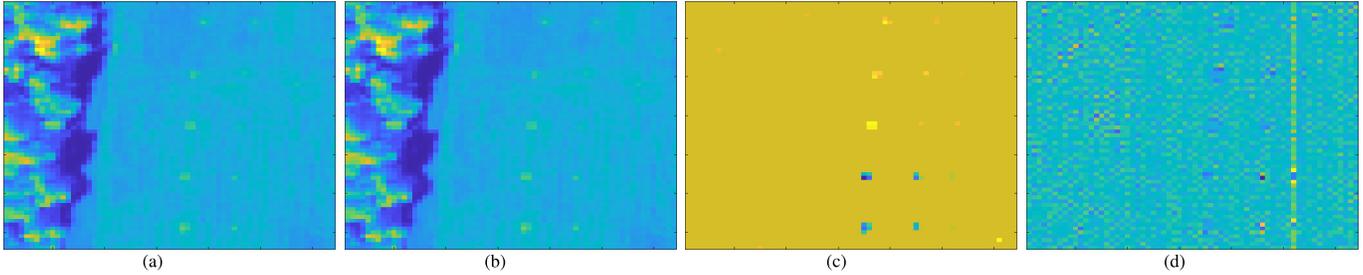

Fig. 5. Visualization results of various components for the band group. (a) $\mathbf{X_B}$. (b) $\mathbf{L_B}$. (c) $\mathbf{S_B}$. (d) $\mathbf{N_B}$.

TABLE I
ESTIMATED VALUES OF j FOR VARIOUS DATASETS

HSI	VD	p	m by $p - j$	j by MX-SVD
HYDICE Panel	HFC, PF$\leq 10^{-4}$	9	5	4
	NWHFC, PF$\leq 10^{-4}$	13	7	6
HYDICE Urban	HFC, PF$\leq 10^{-4}$	9	5	4
	NWHFC, PF$\leq 10^{-4}$	13	7	6
ABU-beach3	HFC, PF$\leq 10^{-4}$	13	9	4
	NWHFC, PF$\leq 10^{-4}$	15	9	6
Pavia Center	HFC, PF$\leq 10^{-4}$	10	3	7
	NWHFC, PF$\leq 10^{-4}$	7	3	4
ABU-beach2	HFC, PF$\leq 10^{-4}$	12	8	4
	NWHFC, PF$\leq 10^{-4}$	12	8	4

operating characteristic (ROC) curves, 3-D ROC [71], and the area under the curve (AUC). These metrics facilitate a quantitative evaluation of the method. In particular, the 3-D ROC curve extends the representation capability of ROC by incorporating a specific threshold τ as the third independent variable to generate (P_D, P_F, τ). Here, we utilize four AUC values, denoted as AUC(P_D, P_F), AUC(P_D, τ), AUC(P_F, τ), and AUC$_{SNPR}$, to conduct a comprehensive evaluation and analysis of the experimental results. The 2-D ROC curve is composed of (P_D, P_F), (P_D, τ), and (P_F, τ). The AUC value reflects the area enclosed by the ROC curve and the coordinate axes. Therefore, larger values of AUC(P_D, P_F), AUC(P_D, τ), and AUC$_{SNPR}$ indicate better detection performance, while a smaller AUC$_{SNPR}$ corresponds to more effective background suppression.

C. Discussion AD for Full Bands and Band Group

To validate the characterization capabilities of various components (**L**, **S**, and **N**) after data decomposition on both the full bands and band group, in this section, we perform AD on each component. Figs. 4 and 5 show the visualization results of various components of the band group with $(p, m, j) = (13, 7, 6)$ and the full bands of HYDICE Panel data, respectively. Table II and Figs. 6 and 7 present the AUC values and AD maps obtained for each component.

The visualization results from Figs. 4 and 5 reveal that decomposition on a band group is more conducive to separating the sparse components from the background. This advantage becomes particularly evident in the comparison of Figs. 4(c) and 5(c). This shows that the decomposition on a band subset is more conducive to anti-various interfering components in the AD task.

Based on the detection results of each component, a horizontal comparison in Table II shows that **S** of band group demonstrates higher AUC (P_D, P_F), AUC (P_D, τ), and AUC SNPR values, as well as lower AUC (P_F, τ). This implies that decomposition on a band group is more conducive to anomaly target detection, particularly evident in the improvement of AUC $(P_D$ and $P_F)$ value and the reduction of AUC $(P_F$ and $\tau)$ value. The AD maps in Figs. 6 and 7 further validate the above analysis. The AD maps on various components of the band subset show a cleaner background suppression than that of the full bands.

These results highlight the advantage of data decomposition on a band group for distinguishing anomalies from the background. This is attributed to the spectral variability of the full-band HS data, which results in increased residual information after extracting the LR component, especially when the rank is small. This increased residual information complicates the differentiation between noise or error **N** and sparse **S** components, affecting the accuracy of **S** extraction and ultimately limiting AD performance. From AD comparison results of **N** and **S** in the band group and the full bands, **S** has a higher AUC $(P_D$ and $P_F)$ and **N** has a smaller AUC $(P_D$ and $P_F)$ in the band group. It shows that decomposition

TABLE II
AD RESULTS OF VARIOUS COMPONENTS FOR BOTH FULL BANDS AND BAND GROUP

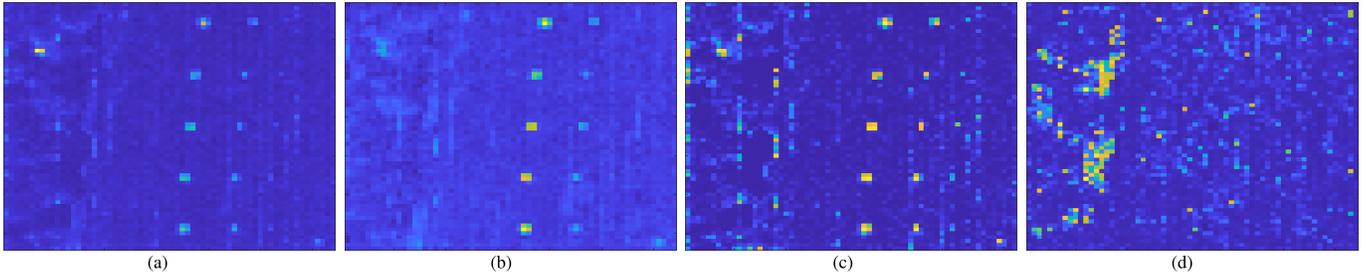|	Full bands				Band group			
	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
X	0.9897	0.345	0.0434	7.9486	0.9798	0.3711	0.0193	19.2017
L	0.9512	0.4926	0.0936	5.2643	0.9948	0.4953	0.0521	9.4996
S	0.9942	**0.7616**	0.0409	18.6401	**0.9977**	0.6905	**0.0087**	**79.1162**
N	0.3039	0.0387	0.0770	0.5022	0.005	0.0050	0.1995	0.0251

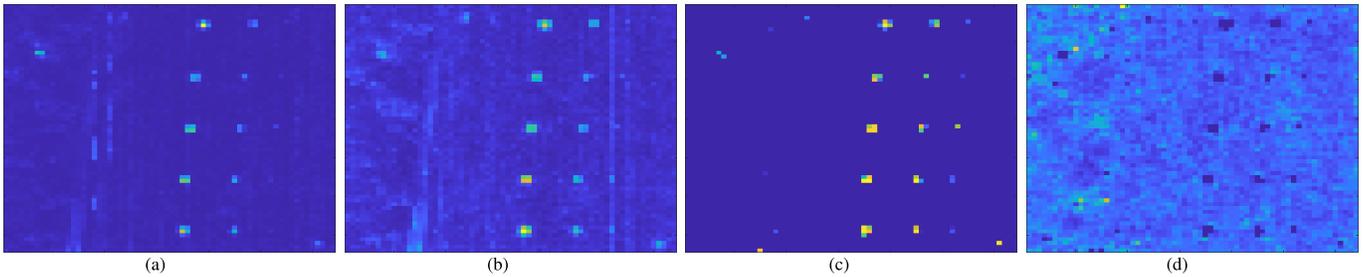Fig. 6. AD results of different components for the full bands. (a) **X.** (b) **L.** (c) **S.** (d) **N.**

Fig. 7. AD results of various components for the band group. (a) $\mathbf{X_B}$. (b) $\mathbf{L_B}$. (c) $\mathbf{S_B}$. (d) $\mathbf{N_B}$.

on band groups is more conducive to distinguishing **N** and **S**. Therefore, decomposition on a band group emerges as a crucial strategy for reducing false alarms.

Furthermore, the vertical comparison in Table II indicates that higher AUC (P_D and P_F), AUC (P_D and τ), and AUC SNPR and lower AUC (P_F and τ) values are realized on sparse space **S** in both full bands and band group than other subspaces, which means that sparse space is more conducive to the capturing of anomaly targets of interest.

Finally, from the third column of the HYDICE Panel data, the third-column anomalies consist of a red panel pixel and smaller yellow panel pixels, representing typical weak anomaly features. The proposed method can more accurately separate the third-column anomaly signature [Figs. 5(c) and (c)]. It indicates that the proposed decomposition strategy based on band group can effectively suppress the background and reduce the impact of spectral variability, that is, facilitating the extraction of weak anomaly targets.

In addition, we further analyze the computational time for the decomposition processing on the full bands and the band group. Table III shows the comparison of the time spent in the data decomposition processing step for the full bands and the band group. It can be observed that decomposition on the band group effectively reduces the computational time,

TABLE III
TIME OF DATA DECOMPOSITION FOR FULL
BANDS AND BAND GROUP (SECOND)

HSI	Full bands	Band group
HYDICE Panel	3.98	1.11
		0.39
HYDICE Urban	26.09	4.31
		2.76
ABU-beach3	13.43	3.36
		2.08
ABU-beach3	5.02	0.80
		1.18
ABU-beach3	8.84	1.94
		1.94

demonstrating the effectiveness of dividing the full bands into smaller scale band group in improving computational efficiency.

D. Comparision Methods and Parameter Setting

The proposed method is compared with 12 baseline methods, including classical RXD [25] and RAD [29], five LR and
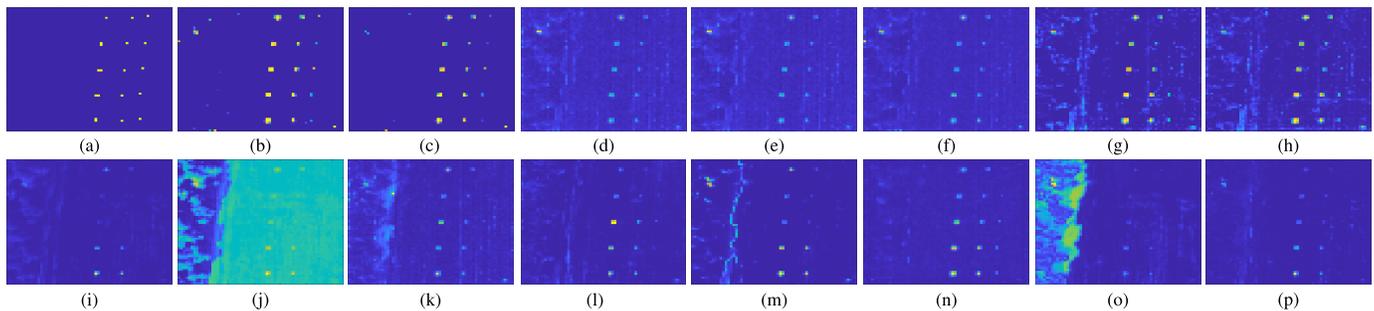
Fig. 8. AD maps of all compared methods for HYDICE Panel data. (a) Ground truth. (b) Proposed-j_1. (c) Proposed-j_2. (d) RXD. (e) RAD. (f) CTV. (g) OSP-GoDec-j_1. (h) OSP-GoDec-j_2. (i) AHMID. (j) PTA. (k) IFEBP. (l) RhyDe. (m) SSIIFD. (n) CRD. (o) RGAE. (p) GAED.

sparse methods [CTV [49], OSP-GoDec [39], RhyDe [46], IFEBP [38], and AHMID [30]], tensor representation-based PTA [37], isolation forest-based SSIIFD [72], collaborative representation-based CRD [73], and two AD methods based on autoencoder networks (GAED [21] and RGAE [16]).

Among these, RXD is the baseline for AD. The RAD detector operates in the original data space and is an AD method based on the correlation matrix R. As a variant of RAD, OSP-GoDec provides the results for background suppression and anomaly target detection in different subspaces. Here, in order to compare the AD performance under the same conditions, both anomaly target detection and background suppression of OSP-GoDec are performed in sparse space, that is, the OSP-GoDec detector is $\delta_S^{R-AD}(\mathbf{r_S})$. RhyDe introduces a denoising and AD framework. A comparison with RhyDe shows the effectiveness of the proposed method compared to AD methods that also consider denoising. CTV is a method for HSI noise removal based on LR and sparse models, and comparative experiments are performed here by performing AD on the sparse matrix S decomposed by CTV. By comparing with CTV, the ability to preserve and extract anomaly targets can be verified. In addition, the proposed method compares with other state-of-the-art AD techniques on different datasets. These methods are based on various strategies to detecting anomalous targets and have achieved satisfactory results. For instance, the PTA method is based on tensor representation. Its considerations include an LR prior for the background and a spatially sparse prior for anomalies. Comparing it with the PTA allows for an assessment of the proposed method's balanced capabilities in background suppression and anomaly target detection. In CRD, collaborative representation is applied to AD, and the ability of the proposed method to express anomalies can be seen by comparison with CRD. IFEBP combines feature extraction with background purification, leveraging the capabilities of fractional Fourier transform in feature extraction and the advantages of LRaSR in background suppression for performing AD tasks. Comparison with it can reflect the anomalies characterization and background suppression capabilities of the proposed method. ADMID has devised a hierarchically mutual noninterfering discriminative learning strategy for modeling three primary components of the data. The effectiveness of the proposed method in distinguishing between background and anomalies can be demonstrated through a comparison with ADMID. The SSIFD method is founded on the assumption that anomalous

pixels are more readily isolated than background pixels. It addresses the performance deficiencies of isolation forests in detecting local anomalies and anomalies in high-dimensional data. The proposed method's representation and stability regarding anomalous targets across different datasets can be assessed through a comparison with SSIFD. In the RGAE framework, a robust AE with an $L_{2,1}$ norm is proposed. Simultaneously, the method integrates graph regularization based on superpixel segmentation into the AE to preserve the geometric structure and local spatial consistency of HSI simultaneously. The comparison with RGAE can highlight the anomaly and background separation ability of the proposed method. To enhance the characterization of background features, GAED incorporates a guided module into the network, thereby diminishing the feature representation of anomalous targets through feedback information. The guided images are computed based on the local spatial structure of HSI. A comparison with GAED can help clarify the advantages of the feedback-based decomposition strategy of the proposed method.

It is noteworthy that, to better showcase the performance of these methods across different datasets and ensure fairness in comparative experiments, parameters for all baseline methods are adjusted based on the parameter ranges and default settings provided in their papers.

Furthermore, since the division of the band group in this article is based on the rank j of the sparse component to guide the selection. As shown in Table I, two sets of j values can be calculated for each data based on the value of p and the MX-SVD method. Therefore, we denote the experimental results of all methods involving j values as j_1 and j_2. In other words, the proposed method is represented as Proposed-j_1 and Proposed-j_2, while OSP-Godec is denoted as OSP-Godec-j_1 and OSP-Godec-j_2.

E. Real-Data Experiment Result and Analysis

The detection results of the 11 detectors on five real datasets are illustrated in Figs. 8–12. For a quantitative study of detection performance, the ROC curves of different methods are presented in Figs. 13–17. Four quantitative evaluation metrics are also provided in Tables IV–VIII. The bold and underlined indicate the best- and second-best metrics, respectively.

1) HYDICE Panel: In the detection results shown in Fig. 8, it can be observed that the proposed method is capable of
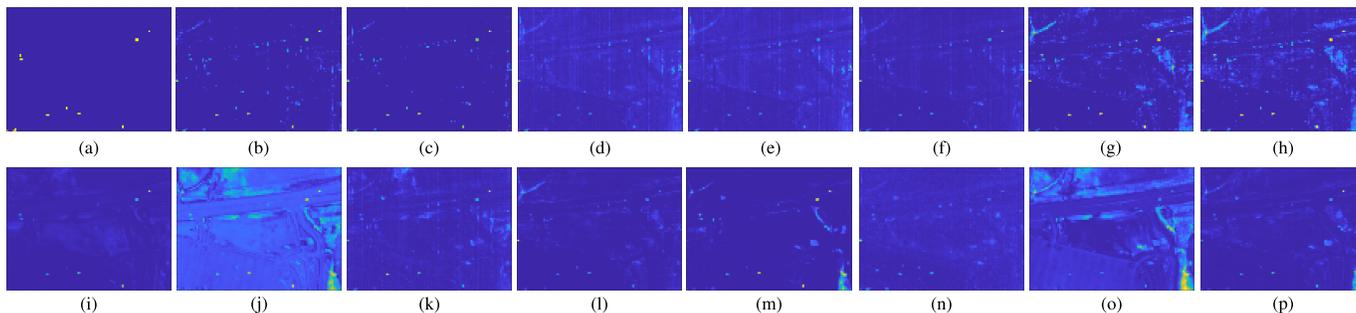
Fig. 9. AD maps of all compared methods for HYDICE Urban data. (a) Ground truth. (b) Proposed-j_1. (c) Proposed-j_2. (d) RXD. (e) RAD. (f) CTV. (g) OSP-GoDec-j_1. (h) OSP-GoDec-j_2. (i) AHMID. (j) PTA. (k) IFEBP. (l) RhyDe. (m) SSIIFD. (n) CRD. (o) RGAE. (p) GAED.

Fig. 10. AD maps of all compared methods for ABU-beach3 data. (a) Ground truth. (b) Proposed-j_1. (c) Proposed-j_2. (d) RXD. (e) RAD. (f) CTV. (g) OSP-GoDec-j_1. (h) OSP-GoDec-j_2. (i) AHMID. (j) PTA. (k) IFEBP. (l) RhyDe. (m) SSIIFD. (n) CRD. (o) RGAE. (p) GAED.

Fig. 11. AD maps of all compared methods for Pavia Center data. (a) Ground truth. (b) Proposed-j_1. (c) Proposed-j_2. (d) RXD. (e) RAD. (f) CTV. (g) OSP-GoDec-j_1. (h) OSP-GoDec-j_2. (i) AHMID. (j) PTA. (k) IFEBP. (l) RhyDe. (m) SSIIFD. (n) CRD. (o) RGAE. (p) GAED.

Fig. 12. AD maps of all compared methods for ABU-beach2 data. (a) Ground truth. (b) Proposed-j_1. (c) Proposed-j_2. (d) RXD. (e) RAD. (f) CTV. (g) OSP-GoDec-j_1. (h) OSP-GoDec-j_2. (i) AHMID. (j) PTA. (k) IFEBP. (l) RhyDe. (m) SSIIFD. (n) CRD. (o) RGAE. (p) GAED.
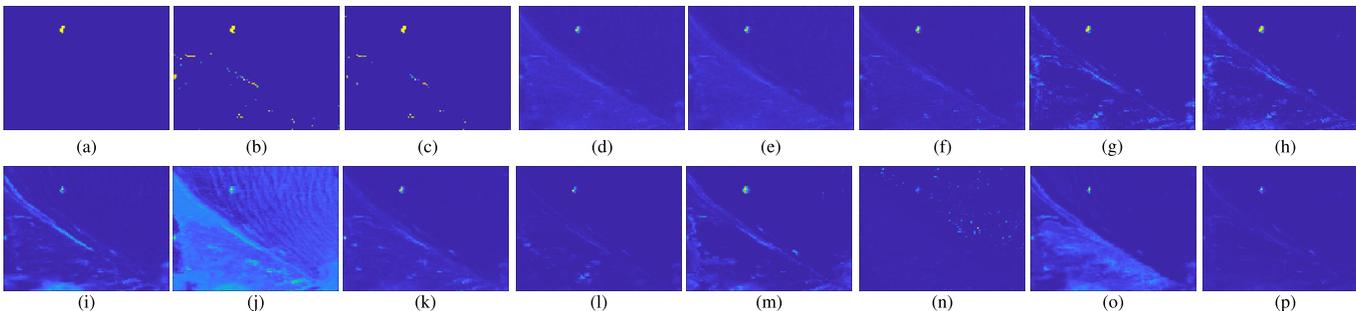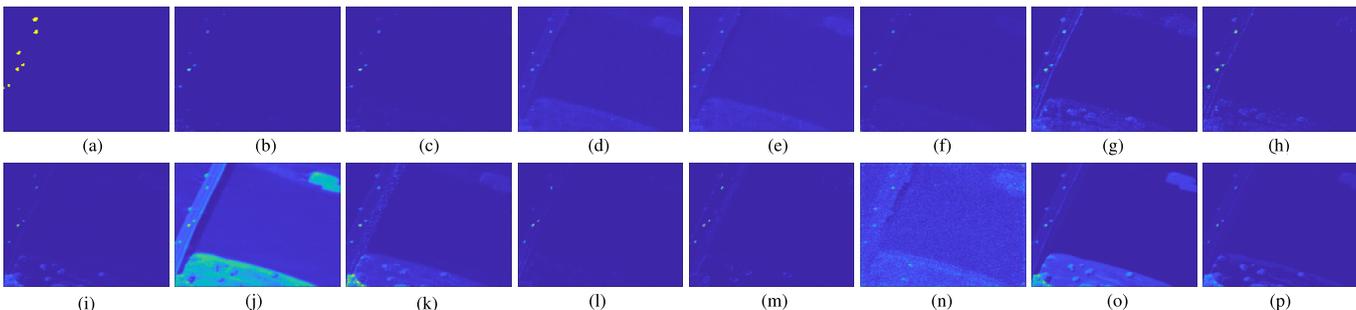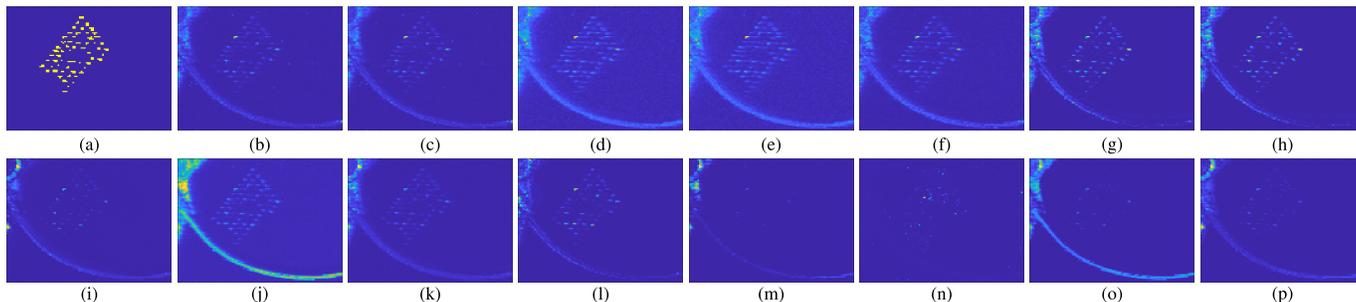
detecting all panel pixels of different sizes, and the background is also satisfactorily suppressed. The third column of the detection results for AHMID, PTA, SSIIFD, RGAE, and GAED exhibits missing anomaly pixels. Although the IFEBP method is capable of detecting anomaly pixels in the third column, there is still a small portion where anomaly pixels are not identified. The background of RhyDe is relatively cleaner compared to other methods. However, the anomaly pixels in the third column are also suppressed. This is because the third column represents a typical weak anomaly target, contributing less to the image signal. The subspace operation of RhyDe retains the main components in HS data, which
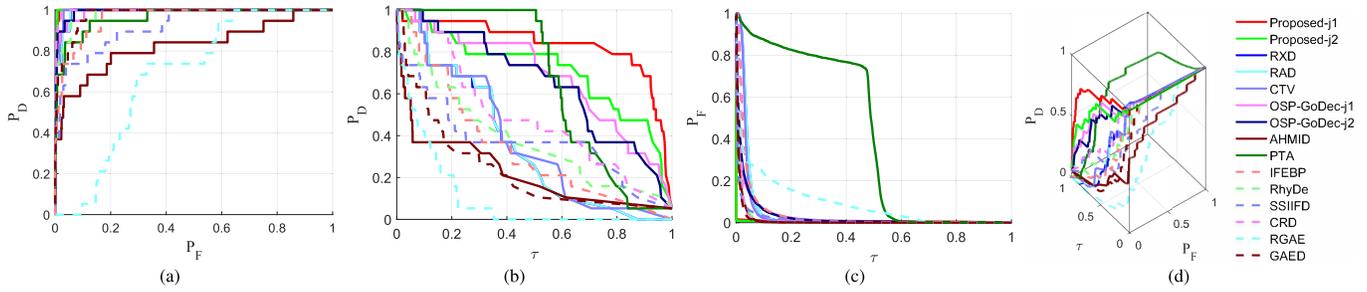
Fig. 13. ROC curves of different methods for HYDICE Panel. (a) 2-D ROC curves ($P_{D,}$). (b) 2-D ROC curves ($P_{D,}$). (c) 2-D ROC curves ($P_{F,}$). (d) 3-D ROC curves ($P_{D,}$).
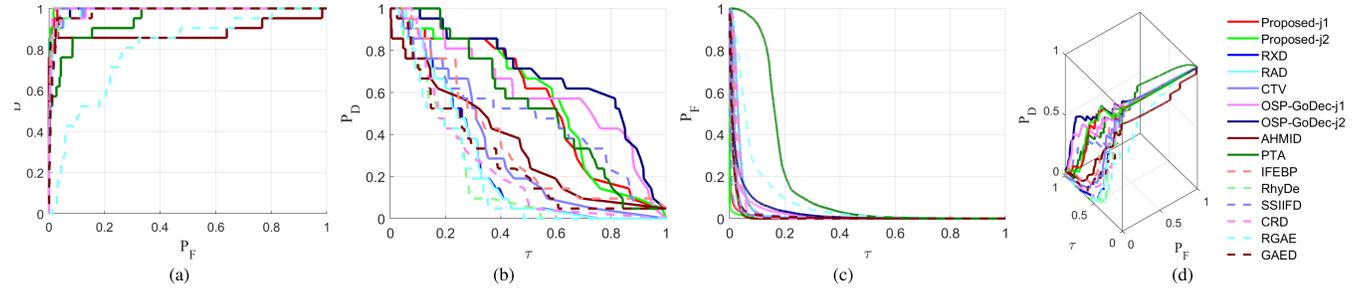

Fig. 14. ROC curves of different methods for HYDICE Urban. (a) 2-D ROC curves ($P_{D,}$). (b) 2-D ROC curves ($P_{D,}$). (c) 2-D ROC curves ($P_{F,}$). (d) 3-D ROC curves ($P_{D,}$).
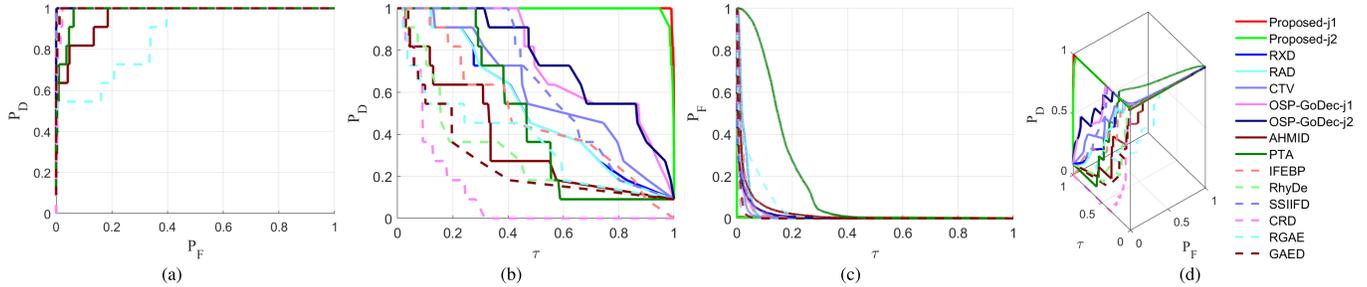

Fig. 15. ROC curves of different methods for ABU-beach3. (a) 2-D ROC curves ($P_{D,}$). (b) 2-D ROC curves ($P_{D,}$). (c) 2-D ROC curves ($P_{F,}$). (d) 3-D ROC curves ($P_{D,}$).

is disadvantageous for handling weak anomaly targets. RXD, RAD, CTV, OSP-GoDec, and CRD, compared to several other compare algorithms, exhibit good detection of anomaly targets in the third column. However, their suppression effectiveness on the background is inadequate, leading to some anomaly targets being submerged within the background.

2) HYDICE Urban: According to Fig. 9, under various values of (p, m, j), the proposed method detects almost all anomaly targets. Due to the relatively small size of anomaly targets, RhyDe does not leverage its advantages for such data. OSP-GoDec, PTA, and RGAE did not effectively suppress the background, especially PTA. CRD, CTV, IFEBP, and GAED had some anomaly targets submerged in background pixels. SSIIFD performed well on this dataset, but there were still a few background pixels mistakenly identified as anomaly targets. On the other hand, RXD and RAD exhibited poor performance on this dataset. The AHMID method demonstrated outstanding background suppression, although there were some missing anomaly targets.

3) ABU-Beach3: From the AD map shown in Fig. 10, it is evident that RhyDe and AHIMD can only detect a small portion of anomalies. IFEBP, RXD, CTV, and RAD exhibit

similar situations. CRD almost fails to detect any anomaly targets. PTA shows poor background suppression in the detection results, which adversely affects its overall performance. Similarly, the RGAE exhibits a weak background suppression capability, leading to a negative impact on detection performance. In contrast, GAED has a cleaner background, effectively highlighting anomalous targets. Compared to other algorithms, OSP-GoDec and SSIIFD have an advantage in the detection results, although there are minor losses in anomaly pixels. In contrast, the proposed method can clearly detect anomalies, and the size and shape of the anomalies closely resemble the ground truth. This suggests that the consideration of spatial factors and feedback strategies on band groups in the proposed method is effective in preserving the characteristics of anomalous targets.

4) Pavia Center: Fig. 11 displays the experimental results of different methods on the Pavia Center dataset. As shown in Fig. 11, PTA captures the majority of anomalies, but it also exhibits poor background suppression capabilities. The AD map of CRD contains excessive interference information. In the results of IFEBP, a small portion of background pixels is falsely identified as anomalies. Compared to PTA, IFEBP,

TABLE IV
AUC VALUES OF DIFFERENT METHODS FOR HYDICE PANEL SCENE

Method	p, m, j	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
Proposed	9,5,4	0.9975	**0.8332**	0.0118	70.7145
	13,7,6	**0.9977**	0.7042	**0.0089**	**79.1248**
RXD	-	0.9898	0.3445	0.0429	8.0308
RAD	-	0.9897	0.3450	0.0434	7.9486
CTV	-	0.9927	0.3639	0.0381	9.5514
OSP-GoDec	9,5,4	0.9894	0.6732	0.0282	23.8370
	13,7,6	0.9921	0.6458	0.0299	21.6120
AHMID	-	0.8225	0.2011	0.0120	16.7099
PTA	-	0.9601	0.6542	0.4155	1.5746
IFEBP	-	0.9635	0.3282	0.0388	8.4682
RhyDe	-	0.9841	0.3684	0.0139	26.5739
SSIIFD	-	0.8820	0.3600	0.0175	20.5726
CRD	-	0.9968	0.4732	0.0208	22.7360
RGAE	-	0.6331	0.0976	0.0902	1.0826
GAED	-	0.9742	0.2074	0.0100	20.7099

TABLE V
AUC VALUES OF DIFFERENT METHODS FOR HYDICE URBAN SCENE

Method	p, m, j	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
Proposed	9,5,4	0.9970	0.5490	0.0094	58.5090
	13,7,6	**0.9983**	0.5583	**0.0076**	**73.1856**
RXD	-	0.9864	0.2507	0.0360	6.9631
RAD	-	0.9864	0.2483	0.0359	6.9237
CTV	-	0.9905	0.3093	0.0207	14.9212
OSP-GoDec	9,5,4	0.9889	0.6064	0.0245	24.7038
	13,7,6	0.9921	**0.6810**	0.0328	20.7531
AHMID	-	0.8997	0.3281	0.0120	27.3230
PTA	-	0.9497	0.5405	0.1789	3.0205
IFEBP	-	0.9918	0.3652	0.0290	12.5810
RhyDe	-	0.9889	0.1931	0.0122	15.7977
SSIIFD	-	0.9912	0.4757	0.0124	38.2220
CRD	-	0.9958	0.2360	0.0330	7.1446
RGAE	-	0.8018	0.1945	0.0878	2.2162
GAED	-	0.9835	0.2814	0.0172	16.3280

RAD, and RXD, OSP-GoDec and AHMID demonstrate better background suppression capabilities but still show slight residue. The detection performance improvement of RGAE and GAED is compromised due to partial background interference, thereby affecting the accuracy of the detection results. In comparison, the detection results of CTV, RhyDe, and the proposed method exhibit more prominent anomalies and a cleaner background. This is particularly evident at the edges of bridges. The primary reason for this is the numerous bands in HS data, leading to strong "interference" among them. The edge positions typically involve a mixture of various materials, resulting in more complex spectral characteristics in the edge pixels and greater variability. These variations significantly impact the performance of AD tasks. By decomposing band groups, the proposed method can mitigate the mutual "interference" caused by the full-band HS data. This approach is more conducive to distinguishing anomalies from the background.

5) ABU-Beach2: Fig. 12 displays the detection results of different methods on the ABU-beach2 dataset. From the AD maps in Fig. 12, SSIIFD hardly detects any anomalous targets and incorrectly labels some background pixels as anomalies. RXD, RAD, and PTA's detection results highlight background pixels more than anomalous ones. CTV exhibits better background suppression than RXD, RAD, and PTA but falls short of other comparative methods. CRD and RhyDe detectors have cleaner backgrounds, but they also suppress most of the anomalous pixels. Most of the anomaly pixels in the RGAE and GAED detectors are suppressed, and some backgrounds are mistakenly detected as anomalies. IFEBP and AHMID show a small number of anomalies submerged within the background. In comparison to other methods, OSP-GoDec and the proposed method demonstrate a better balance between background suppression and preserving anomalous targets.

The combination of qualitative and quantitative assessments allows for a more comprehensive evaluation of the performance of different detectors. Figs. 13–17 depict the ROC curves of different methods on five datasets, and the corresponding AUC values are listed in Tables IV–VIII.

6) HYDICE Panel: For the HYDICE panel dataset, combining the quantitative metrics in Fig. 13 and Table IV, the proposed method demonstrates favorable overall performance in both $(p, m, j) = (9, 5, 4)$ and $(p, m, j) = (13, 7, 6)$. All four AUC values achieve optimal or second-best levels. It indicates that the proposed method achieves a well-balanced performance in capturing anomalous targets and suppressing background. The main reason for this is that the detection results on different band groups exhibit consistency and complementarity. The method in this article strengthens anomalous features in the model by iteratively returning existing detection knowledge. In addition, the residual information after extracting the LR part increases due to the spectral variability of full-band data, especially when the rank is low. This makes it more challenging to distinguish between the **N** and **S** components. However, decomposing band groups can effectively alleviate this issue.

7) HYDICE Urban: From the previous data description, the number of anomalous target pixels on this dataset is relatively rare and presents an overall sparse characteristic. As indicated in Fig. 14 and Table V, the proposed method achieves the best and second-best performance on AUC(P_D, P_F), AUC(P_F, τ), and AUC$_{SNPR}$ indexes in the detection of such weak anomaly targets. In addition, the AUC(P_D, τ) value is surpassed only by OSP-GoDec, achieving the second-best result. It highlights the capability of the proposed method to detect weak targets.

8) ABU-Beach3: For the ABU-beach3 dataset, the ROC curve in Fig. 15 and the AUC values in Table VI indicate that the proposed method achieves the highest detection rate. Specifically, in Fig. 15(a), (P_D, P_F) of the proposed method tends to lean toward the top-left corner, and correspondingly, (P_D, τ) leans toward the top-right corner in Fig. 15(b).
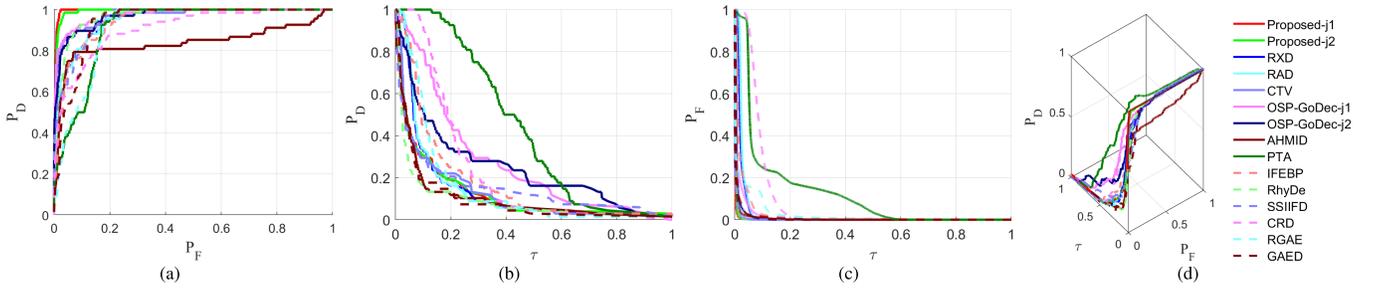
Fig. 16. ROC curves of different methods for Pavia Center. (a) 2-D ROC curves (P_D). (b) 2-D ROC curves (P_D). (c) 2-D ROC curves (P_F). (d) 3-D ROC curves (P_D).
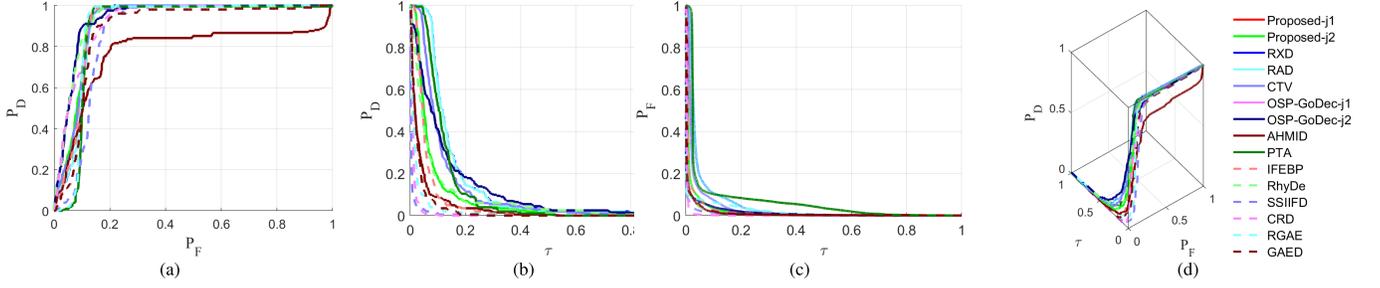

Fig. 17. ROC curves of different methods for ABU-beach2. (a) 2-D ROC curves (P_D). (b) 2-D ROC curves (P_D). (c) 2-D ROC curves (P_F). (d) 3-D ROC curves (P_D).

TABLE VI
AUC VALUES OF DIFFERENT METHODS FOR ABU-BEACH3 SCENE

Method	p, m, j	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
Proposed	13,9,4	0.99995	**0.9955**	0.0087	114.7631
	15,9,6	**0.99996**	0.9891	**0.0072**	**137.5648**
RXD	-	0.99982	0.4918	0.0259	18.9562
RAD	-	0.99983	0.4891	0.0257	19.0360
CTV	-	0.99976	0.5773	0.0152	38.0360
OSP-GoDec	13,9,4	0.99984	0.7300	0.0152	48.0643
	15,9,6	0.99975	0.7400	0.0164	45.0807
AHMID	-	0.96273	0.3400	0.0265	12.8307
PTA	-	0.98493	0.4800	0.1607	2.9865
IFEBP	-	0.99879	0.4686	0.0197	23.7754
RhyDe	-	0.99722	0.2918	0.0076	38.6514
SSIIFD	-	0.99985	0.6082	0.0162	37.6164
CRD	-	0.99321	0.1177	0.0076	15.4214
RGAE	-	0.86600	0.3282	0.0405	8.1001
GAED	-	0.99694	0.2355	0.0087	27.1074

TABLE VII
AUC VALUES OF DIFFERENT METHODS FOR PAVIA CENTER SCENE

Method	p, m, j	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
Proposed	10,3,7	**0.9881**	0.1186	**0.0055**	21.7500
	7,3,4	0.9858	0.1201	0.0057	21.1323
RXD	-	0.9513	0.1289	0.0232	5.5540
RAD	-	0.9467	0.1279	0.0274	4.6715
CTV	-	0.9595	0.1221	0.0075	16.3782
OSP-GoDec	10,3,7	0.9658	**0.2654**	0.0140	18.9211
	7,3,4	0.9315	0.2427	0.0088	**27.6259**
AHMID	-	0.8710	0.0964	0.0077	12.5812
PTA	-	0.9122	0.4298	0.1224	3.5116
IFEBP	-	0.9422	0.1733	0.0237	7.3135
RhyDe	-	0.9205	0.0827	0.0061	13.5342
SSIIFD	-	0.8283	0.1424	0.0063	22.7300
CRD	-	0.9074	0.2238	0.0921	2.4307
RGAE	-	0.9118	0.1490	0.0263	5.6610
GAED	-	0.9204	0.0892	0.0105	8.5309

It demonstrates that the proposed method possesses a superior ability to preserve and detect anomalous targets. In addition, the smaller (P_F, τ) values obtained by the proposed method at $(p, m, j) = (15, 9, 6)$ reflect its enhanced background suppression capability.

9) Pavia Center: ROC curves and corresponding AUC values for different detectors on the Pavia Center dataset are presented in Fig. 16 and Table VII, respectively. The results indicate that the proposed method significantly outperforms other comparative methods in metric AUC(P_D, P_F). Although

it does not exhibit superiority in AUC(P_D, τ), it achieves the minimum value for AUC(P_F, τ) and secures the third position for AUC$_{SNPR}$, thus showcasing good overall performance. Its observation is further supported by the detection results shown in Fig. 7.

10) ABU-Beach2: The ROC curve for the AD results on this dataset is illustrated in Fig. 17. Through the analysis of the AUC values presented in Table VIII, the proposed method demonstrates excellent overall performance. With a superior AUC(P_D, P_F) value compared to other comparative methods,

TABLE VIII
AUC VALUES OF DIFFERENT METHODS FOR ABU-BEACH2 SCENE

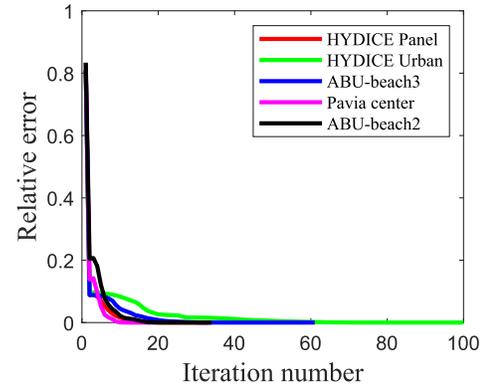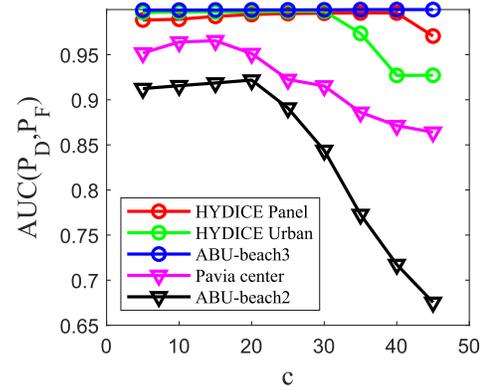| Method	p, m, j	AUC of (P_D, P_F)	AUC of (P_D, τ)	AUC of (P_F, τ)	AUC of SNPR
Proposed	12,8,4	**0.9200**	0.0814	0.0173	4.7089
	12,8,4	**0.9200**	0.0814	0.0173	4.7089
RXD	-	0.9100	**0.1520**	0.0488	3.1137
RAD	-	0.9102	<u>0.1515</u>	0.0484	3.1302
CTV	-	<u>0.9183</u>	0.1202	0.0308	3.8973
OSP-GoDec	12,8,4	0.9060	0.1273	0.0174	**7.3339**
	12,8,4	0.9060	0.1273	0.0174	**7.3339**
AHMID	-	0.7446	0.0386	0.0126	3.0473
PTA	-	0.8935	0.1209	0.0667	1.8137
IFEBP	-	0.9112	0.0582	0.0176	3.3193
RhyDe	-	0.8971	0.0805	0.0119	<u>6.7736</u>
SSIIFD	-	0.5105	0.0086	<u>0.0108</u>	0.7895
CRD	-	0.6788	0.0111	**0.0062**	1.8077
RGAE	-	0.6820	0.0210	0.0227	0.9283
GAED	-	0.7665	0.0278	0.0134	2.0696

Fig. 18. Relative errors versus iterations on five datasets.

Fig. 19. AUC(P_D, P_F) value of five datasets under different parameters c.

it stands at an optimal level, indicating that the proposed method achieves a well-balanced compromise between target preservation and background anomaly.

Based on the qualitative and quantitative assessment results mentioned above, the proposed method demonstrates competitive performance compared to all the contrastive algorithms, especially in the evaluation of the AUC(P_D, P_F) metric. The proposed method achieved the highest AUC(P_D, P_F) values on all five datasets and the lowest AUC(P_F, τ) values on four of them. It also achieved the best AUC$_{SNPR}$ and AUC(P_D, τ) scores on three datasets and demonstrated overall good performance. These results indicate that the data decomposition on the band group and iterative feedback strategy of the proposed method is effective in HS data AD. They not only effectively reduce computational time and minimize background interference but also retain sufficient anomaly information to maintain the detection capability of weak anomaly features.

F. Convergence, Complexity, and Parameter Analysis

1) Convergence Analysis: The numerical convergence is studied on five datasets, including HyDICE Panel, HYDICE Urban, ABU-beach3, Pavia Center, and ABU-beach2. Fig. 18 illustrates the relationship between the relative error in Algorithm 1 and the number of iterations. It can be observed that, as the number of iterations increases, the relative error consistently decreases across the five datasets until the convergence condition is met. This observation suggests that the proposed FBGVLRS-AD exhibits the expected convergence behavior.

2) Complexity Analysis: According to the procedure of Algorithm 1, the computational burden of the proposed FBGVLRS-AD primarily consists of the following steps. For **J**, **S**, **L**, and $\mathbf{\Gamma}$, the primary computational complexity for each iteration includes FFT, SVD, and soft threshold operations.

Specifically, assuming that the size of the data is $M \times N$, the following conditions hold: 1) solving the subproblem (14) requires three SVD operations, each with a computational complexity of $O(MN^2)$; 2) the computational complexity for subproblem (16) is $O(MN)$; and 3) the **L** subproblem requires one FFT, with a computational complexity of $O(MN \log(M))$. Therefore, the total computational complexity of FBGVLR-AD is $O(MN \log(M) + 3MN^2 + MN)$.

3) Parameter Analysis: In this section, we will explore the key parameters of the proposed method: the regularization parameter β in (9), which plays a crucial role in balancing the LR and sparsity terms closely tied to detection accuracy. To explore the influence of the regularization parameter on detection performance, Fig. 19 illustrates the variation in the AUC(P_D, P_F) values of the proposed method across five datasets when $\beta = c/\text{sqrt}(h \times w)$, where the value of c varies within the range {5, 10, 15, 20, 25, 30, 35, 40, 45}. It is evident that, with c fixed at 40, the proposed method consistently attains favorable values of AUC(P_D, P_F) values on the HYDICE Panel and ABU-beach3. When c is fixed at 15 and 20, the proposed method demonstrates optimal AUC(P_D, P_F) values on Pavia Center and ABU-beach2, respectively. The examination of the detection performance of the proposed method across various datasets indicates that, in cases where the pixels occupied by anomalous targets are relatively scarce in the data, a higher value of "c" is necessary for the proposed method to prevent the suppression of more anomalous pixels. When there are more pixels associated with

anomalous targets in the data, a smaller "c" value is more advantageous for achieving better detection results.

V. CONCLUSION

In this article, we propose an AD method based on feedback band group and variation LR sparse model. Decomposition on band group can reduce the interference of nonanomaly components and the computation time of decomposition process. In order to promote the spatial piecewise smoothing of HS data and resist the interference of noise, CTV is used to constrain the background components and promote the separation of the background components and sparse components. At the same time, in order to preserve the anomaly information from being weakened, the $L_{2,1}$ norm is used to constrain the sparse components. To avoid the limitation of AD performance caused by the lack of prior knowledge of BS, we design an iterative feedback strategy to promote the separation of sparse components from the background and retain sufficient anomaly information. Finally, an ALM-based optimization solver is designed for FBGVLRS-AD. Experimental results on multiple real data verify the effectiveness of the proposed method, especially for weak anomaly targets.

Furthermore, in future work, we tend to consider further extending the proposed model to other subtasks, such as classification and object detection.

REFERENCES

[1] Z. Lv, P. Zhang, W. Sun, J. A. Benediktsson, J. Li, and W. Wang, "Novel adaptive region spectral–spatial features for land cover classification with high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5609412.

[2] D. Liao, C. Shi, and L. Wang, "A spectral–spatial fusion transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5515216.

[3] S. Feng, R. Feng, J. Liu, C. Zhao, F. Xiong, and L. Zhang, "An attention-based multiscale spectral–spatial network for hyperspectral target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[4] L. Gao, J. Li, K. Zheng, and X. Jia, "Enhanced autoencoders with attention-embedded degradation learning for unsupervised hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509417.

[5] Z. Li, Y. Zhang, and J. Zhang, "Hyperspectral anomaly detection for spectral anomaly targets via spatial and spectral constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511515.

[6] C.-I. Chang, C.-Y. Lin, P.-C. Chung, and P. F. Hu, "Iterative spectral–spatial hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504330.

[7] Z. Li, Y. Zhang, and J. Zhang, "Tensor approximation with low-rank representation and kurtosis correlation constraint for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533713.

[8] B. Tu, Z. Wang, H. Ouyang, X. Yang, J. Li, and A. Plaza, "Hyperspectral anomaly detection using the spectral–spatial graph," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542814.

[9] Z. Li and Y. Zhang, "Hyperspectral anomaly detection via image super-resolution processing and spatial correlation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2307–2320, Mar. 2021.

[10] 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.

[11] 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.

[12] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.

[13] Q. Zhang, Y. Dong, Q. Yuan, M. Song, and H. Yu, "Combined deep priors with low-rank tensor factorization for hyperspectral image restoration," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[14] Y. Xiao, Q. Yuan, K. Jiang, J. He, C.-W. Lin, and L. Zhang, "TTST: A top-k token selective transformer for remote sensing image super-resolution," *IEEE Trans. Image Process.*, vol. 33, pp. 738–752, 2024.

[15] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7406–7417, Oct. 2020.

[16] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511314.

[17] N. Huyan, X. Zhang, D. Quan, J. Chanussot, and L. Jiao, "AUD-Net: A unified deep detector for multiple hyperspectral image anomaly detection via relation and few-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 27, 2022, doi: 10.1109/TNNLS.2022.3213023.

[18] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, Jul. 2020.

[19] K. Jiang, W. Xie, Y. Li, J. Lei, G. He, and Q. Du, "Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 5224–5236, Jul. 2020.

[20] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, "Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022.

[21] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, "Hyperspectral anomaly detection with guided autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538818.

[22] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.

[23] S. Lin, M. Zhang, X. Cheng, L. Shi, P. Gamba, and H. Wang, "Dynamic low-rank and sparse priors constrained deep autoencoders for hyperspectral anomaly detection," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–18, 2024.

[24] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513412.

[25] 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.

[26] H. Kwon and N. M. Nasrabadi, "Hyperspectral anomaly detection using kernel RX-algorithm," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 5, Oct. 2004, pp. 3331–3334.

[27] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.

[28] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, vol. 3, Mar. 2004, p. 1824.

[29] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.

[30] T. Guo, L. He, F. Luo, X. Gong, Y. Li, and L. Zhang, "Anomaly detection of hyperspectral image with hierarchical antinoise mutual-incoherence-induced low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510213.

[31] Q. Xiao, L. Zhao, S. Chen, and X. Li, "Enhanced tensor low-rank representation learning for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[32] J. Peng et al., "Low-rank and sparse representation for hyperspectral image processing: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 10–43, Mar. 2022.

[33] Y. Fu, Y. Zheng, H. Huang, I. Sato, and Y. Sato, "Hyperspectral image super-resolution with a mosaic RGB image," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5539–5552, Nov. 2018.

[34] M. Zhao, W. Li, L. Li, P. Ma, Z. Cai, and R. Tao, "Three-order tensor creation and tucker decomposition for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5000216.

[35] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6916–6930, Nov. 2022.

[36] Y. Chen, W. He, N. Yokoya, and T. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[37] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.

[38] Y. Ma, G. Fan, Q. Jin, J. Huang, X. Mei, and J. Ma, "Hyperspectral anomaly detection via integration of feature extraction and background purification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 8, pp. 1436–1440, Aug. 2021.

[39] C.-I. Chang, H. Cao, S. Chen, X. Shang, C. Yu, and M. Song, "Orthogonal subspace projection-based go-decomposition approach to finding low-rank and sparsity matrices for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2403–2429, Mar. 2021.

[40] S. Chen, C.-I. Chang, and X. Li, "Component decomposition analysis for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5516222.

[41] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[42] L. Li, M. Song, Q. Zhang, and Y. Dong, "Hyperspectral denoising via global variation and local structure low-rank model," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5525513.

[43] F. Xiong, J. Zhou, J. Zhou, J. Lu, and Y. Qian, "Multitask sparse representation model-inspired network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5518515.

[44] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, "Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 6, 2023, doi: 10.1109/TNNLS.2023.3278866.

[45] X. Cheng, Y. Xu, J. Zhang, and D. Zeng, "Hyperspectral anomaly detection via low-rank decomposition and morphological filtering," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[46] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500117.

[47] 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.

[48] J. Wang, P. Huang, K. Zhang, and Q. Wang, "Hyperspectral anomaly detection via S1/2 regularized low rank representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[49] J. Peng, Y. Wang, H. Zhang, J. Wang, and D. Meng, "Exact decomposition of joint low rankness and local smoothness plus sparse matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5766–5781, May 2023.

[50] X. Shang, C. Cui, and X. Sun, "Spectral–spatial hypergraph-regularized self-representation for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[51] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.

[52] X. Shang, M. Song, Y. Wang, and H. Yu, "Residual-driven band selection for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[53] W. Xie, Y. Li, J. Lei, J. Yang, C.-I. Chang, and Z. Li, "Hyperspectral band selection for spectral–spatial anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3426–3436, May 2020.

[54] X. Shen, H. Liu, J. Nie, and X. Zhou, "Matrix factorization with framelet and saliency priors for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504413.

[55] C.-I. Chang, Y.-M. Kuo, and P. Fuming Hu, "Unsupervised rate distortion function-based band subset selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5517218.

[56] 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.

[57] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.

[58] C.-I. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, vol. 1. New York, NY, USA: Springer, 2003.

[59] J. C. Harsanyi, W. Farrand, and C.-I. Chang, "Detection of subpixel spectral signatures in hyperspectral image sequences," in *Proc. Amer. Soc. Photogram. Remote Sens. (ASPRS)*, vol. 1, 1994, pp. 236–247.

[60] C.-I. Chang, "A review of virtual dimensionality for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, Apr. 2018.

[61] J. Xue, Y. Zhao, Y. Bu, J. C. Chan, and S. G. Kong, "When Laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13887–13901, Dec. 2022.

[62] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.

[63] J. Peng, H. Wang, X. Cao, X. Liu, X. Rui, and D. Meng, "Fast noise removal in hyperspectral images via representative coefficient total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5546017.

[64] H. Zeng, J. Xue, H. Q. Luong, and W. Philips, "Multimodal core tensor factorization and its applications to low-rank tensor completion," *IEEE Trans. Multimedia*, vol. 25, pp. 7010–7024, 2023.

[65] M. A. Saunders, "Solution of sparse rectangular systems using LSQR and Craig," *BIT Numer. Math.*, vol. 35, no. 4, pp. 588–604, Dec. 1995.

[66] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man, Cybern.*, vol. SMC-9, no. 1, pp. 62–66, Jan. 1979.

[67] C.-I. Chang and J. Chen, "Hyperspectral anomaly detection by data sphering and sparsity density peaks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5526321.

[68] Chein-I. Chang, "Hyperspectral anomaly detection: A dual theory of hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511720.

[69] Z. Huang, X. Kang, S. Li, and Q. Hao, "Game theory-based hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2965–2976, Apr. 2020.

[70] 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.

[71] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.

[72] X. Song, S. Aryal, K. M. Ting, Z. Liu, and B. He, "Spectral–spatial anomaly detection of hyperspectral data based on improved isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5516016.

[73] 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.

Lan Li received the B.S. degree in software engineering from Chaohu University, Chaohu, China, in 2018, and the M.S. degree in computer technology from Dalian Maritime University, Dalian, China, in 2021, where she is currently pursuing the Ph.D. degree with the School of Information Science and Technology.
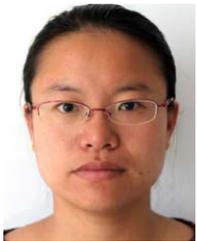
Her research interest is remote sensing information processing.

Qiang Zhang (Member, IEEE) received the B.E. degree in surveying and mapping engineering and the M.E. and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2017, 2019, and 2022, respectively.

He is currently an Associate Professor with the Center of Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. He has published more than ten journal articles on IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Earth System Science Data* (ESSD), and ISPRS *Journal of Photogrammetry and Remote Sensing* (ISPRS P&RS). His research interests include remote sensing information processing, computer vision, and machine learning. More details could be found at https://qzhang95.github.io.

Meiping Song (Member, IEEE) received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006.

She has been a Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China, since 2020. Her research interests include remote sensing and hyperspectral image processing.

Chein-I Chang (Life Fellow, IEEE) received the B.S. degree in mathematics from Soochow University, Taipei, Taiwan, in 1973, the M.S. degree in mathematics from the Institute of Mathematics, National Tsing Hua University, Hsinchu, Taiwan, in 1975, the M.A. degree in mathematics from The State University of New York at Stony Brook, Stony Brook, NY, USA, in 1977, the M.S. and M.S.E.E. degrees from the University of Illinois at Urbana–Champaign, Urbana, IL, USA, in 1982, and the Ph.D. degree in electrical engineering from the University of Maryland at College Park, College Park, MD, USA, in 1987.

He has been with the University of Maryland at Baltimore County (UMBC), Baltimore, MD, USA, since 1987, where he is currently a Professor with the Department of Computer Science and Electrical Engineering. He is also a Yushan Scholar Chair Professor with National Cheng Kung University, Tainan, Taiwan. He has been holding a Chang Jiang Scholar Chair Professorship and the Director of the Center for Hyperspectral Imaging in Remote Sensing (CHIRS), Dalian Maritime University, Dalian, China, since 2016. He has authored four books: *Hyperspectral Imaging: Techniques for Spectral Detection and Classification* (Kluwer Academic Publishers, 2003), *Hyperspectral Data Processing: Algorithm Design and Analysis* (John Wiley & Sons, 2013), *Real Time Progressive Hyperspectral Image Processing: Endmember Finding and Anomaly Detection* (Springer, 2016), and *Recursive Hyperspectral Sample and Band Processing: Algorithm Architecture and Implementation* (Springer, 2017). He is working on his fifth book *Hyperspectral Target and Anomaly Detection: Statistical Signal and Image Processing Perspectives* (Wiley, 2024). In addition, he also edited two books, *Recent Advances in Hyperspectral Signal and Image Processing* (2006) and *Hyperspectral Data Exploitation: Theory and Applications* (John Wiley & Sons, 2007), and coedited with A. Plaza a book *High Performance Computing in Remote Sensing* (CRC Press, 2007). His research interests include multispectral/hyperspectral image processing, automatic target recognition, and medical imaging.

Dr. Chang is a fellow of the Society for Photo-Optical Instrumentation Engineers (SPIE). He received the National Research Council Senior Research Associateship Award from 2002 to 2003 sponsored by the U.S. Army Soldier and Biological Chemical Command, Edgewood Chemical and Biological Center, Aberdeen Proving Ground, MD, USA. He is serving as an Associate Editor for *Remote Sensing* and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He has been a guest editor of several special issues.