



**HAL**  
open science

# Hyperspectral Anomaly Detection via Sparsity of Core Tensor under Gradient Domain

Wenting Shang, Jiangjun Peng, Zebin Wu, Yang Xu, Mohamad Jouni, Mauro Dalla Mura, Zhihui Wei

► **To cite this version:**

Wenting Shang, Jiangjun Peng, Zebin Wu, Yang Xu, Mohamad Jouni, et al.. Hyperspectral Anomaly Detection via Sparsity of Core Tensor under Gradient Domain. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61, pp.1-16. 10.1109/TGRS.2023.3297627 . hal-04170061

**HAL Id: hal-04170061**

**<https://hal.science/hal-04170061>**

Submitted on 25 Jul 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Hyperspectral Anomaly Detection via Sparsity of Core Tensor under Gradient Domain

Wenting Shang, Jiangjun Peng, Zebin Wu, *Senior Member, IEEE*, Yang Xu, *Member, IEEE*, Mohamad Jouni, *Member, IEEE*, Mauro Dalla Mura, *Senior Member, IEEE*, and Zihui Wei

**Abstract**—Hyperspectral anomaly detection (AD) task is a typical binary classification problem, and utilizing background prior knowledge is a key technique to solving such problems. The two most commonly used priors for hyperspectral images are low-rank and local smooth properties. Most traditional matrix-based methods use two regularizations to model these two types of priors and integrate them into one model, which makes these two regularizations unable to maximize their effectiveness. In addition, the matrix method also destroys the structure of the hyperspectral images (HSI). To address these issues, this study identified a unique sparsity property in the gradient tensor of HSI. Specifically, the core tensor resulting from the Tucker decomposition of the gradient tensor was observed to exhibit sparsity. This sparsity property, referred to as GCS (the sparsity on the core tensor of the gradient map), effectively captures the structural information of HSI and improves detection performance. The GCS regularization offers the following advantages: 1) GCS regularization uses one term to simultaneously capture both low-rankness and local smoothness, the size of the core tensor represents the low-rank prior to the background, and the  $\ell_1$  norm describes the sparsity of gradient map, i.e., the local smoothness of the original data; 2) GCS is a constrained regularization, allowing for the full utilization of information from different dimensions of the HSI when updating the core tensor, i.e., utilizing the spatial and spectral information carried by three-factor matrices of the Tucker decomposition. Finally, extensive experiments validate the superiority of our proposed methods.

**Index Terms**—Hyperspectral anomaly detection, sparsity measure, Tucker decomposition, GCS regularization.

## I. INTRODUCTION

**B**ENEFITING from the rapid development of hyperspectral remote sensing technology, hyperspectral images (HSIs) are able to provide hundreds of narrow and contiguous spectral

bands. The abundant spectral and spatial information makes the precise recognition of material substance targets a reality, and hyperspectral anomaly detection (AD), an unsupervised target detection, is one of the most popular directions [1]. It has received scant attention in research due to its real-world application, such as food safety [2], biomedical [3], mineral exploration [4], and environmental monitoring [5]. Hyperspectral AD aims to detect objects of interest, separating anomalies from the surrounding natural background. The criterion for labeling a pixel as an anomaly or a background pixel is that the spectral signature of the anomaly is significantly different from the surrounding background. Nevertheless, AD tasks suffer from unknown prior knowledge of both the target and background spectral signatures, which brings about a challenging task [6], [7].

In recent years, there has been a growing literature on hyperspectral AD. The methods can be roughly divided into two categories: data-based method [8]–[27] and model-based method [5]–[7], [12], [28]–[39], [39].

Data-based methods apply deep learning techniques in part or in whole. Some data-based methods can directly learn a mapping from observation data to background or outliers using a neural network from the paired dataset [8], [9]. These methods do not require a priori information but have problems such as dependence on labeled data and model mismatch. In order to not rely on paired data, some data-based methods use neural networks, such as autoencoders (AE) [19]–[24] and generative adversarial networks (GAN) [16]–[18], [25]–[27], [40]–[42], to effectively reconstruct the background component in the original HSI data in an unsupervised manner while the anomaly part could not be reconstructed successfully. Then the anomalies appear as reconstruction errors or some postprocessing is used to obtain the detection results. In order to further characterize the prior information, some constraints are introduced, such as sparsity constraints on encoded latent variables or residuals of decoded reconstructed data of AE [43] and Gaussian distribution constraints on the latent feature discriminated by the GAN [17]. Besides, neural networks may also be used to solve a subproblem of AD or as a preprocessing step to obtain better results. For example, some methods try to use the extracted features by neural networks to design the subsequent detector [12]–[15] or use neural networks to obtain intermediate abundance or a denoise prior [16]–[18]. Although data-based methods have achieved good results, they still have challenges regarding generalization ability and prior characterization of HSI data. The model-driven or non-deep learning methods focus more on characterizing priors on HSI

This work was supported in part by the National Natural Science Foundation of China under Grant 61772274, Grant 62071233, and grant 61701238, in part by the Jiangsu Provincial Natural Science Foundation of China under grant BK20211570, grant BK20180018, and grant BK20170858, in part by the Fundamental Research Funds for the Central Universities under grant 30917015104, grant 30919011103, grant 30919011402, and grant 30921011209, and in part by the China Postdoctoral Science Foundation under grant 2017M611814. (*Corresponding author: Zebin Wu.*)

W. Shang, Y. Xu, Z. Wu and Z. Wei are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: shangwt1011@gmail.com; xuyangth90@njust.edu.cn; wuzb@njust.edu.cn; gswei@njust.edu.cn).

Jiangjun Peng is with School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, Shaan'xi, China. E-mail: andrew.pengjj@gmail.com.

M. Jouni and M. Dalla Mura are with Univ. Grenoble Alpes, CNRS, Grenoble INP\*, GIPSA-Lab, 38000 Grenoble, France (\*Institute of Engineering Univ. Grenoble Alpes). M. Dalla Mura is also with the Institut Universitaire de France (IUF). (e-mail: mohamad.jouni@gipsa-lab.fr; mauro.dalla-mura@gipsa-lab.fr)

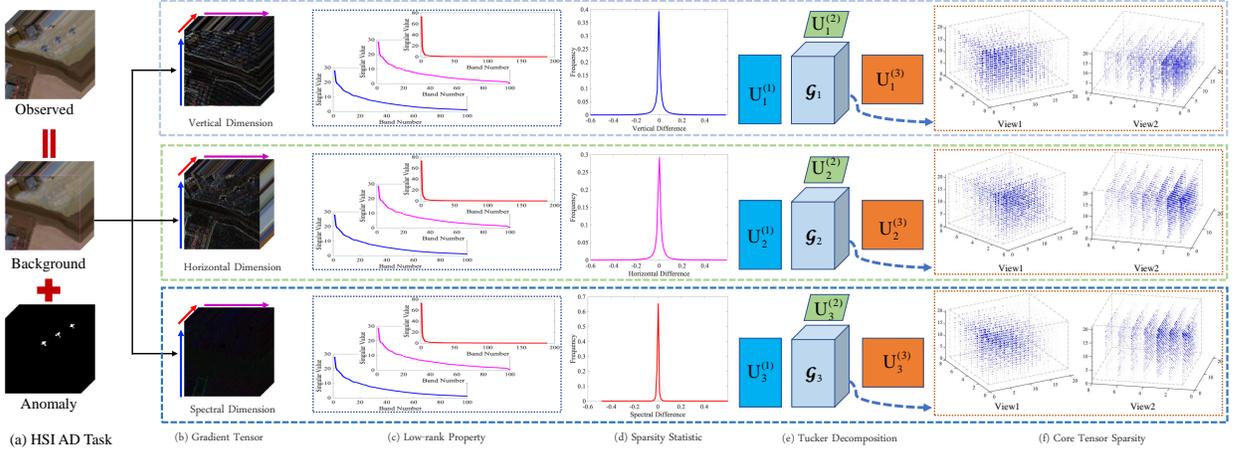


Fig. 1: Illustration of the GCS regularization term. (a) The HSI AD task is to model the background to separate the anomaly and background; (b) The Gradient tensors of background  $\mathcal{B}$  in the vertical, horizontal, and spectral directions, represented as  $\nabla_n \mathcal{B}$ ,  $n = 1, 2, 3$ , respectively; (c) The singular values curves of the three gradient tensors along the three unfolding directions, which indicate a low-rank property of the three gradient tensors; (d) Frequency histograms of elements' values in three gradient tensors, respectively; (e) Tucker decomposition of the gradient tensors in (b); (f) Core tensors  $\mathcal{G}_n \in \mathbb{R}^{20 \times 20 \times 7}$  of gradient tensors  $\nabla_n \mathcal{B}$ ,  $n = 1, 2, 3$ . Where the deeper color of the elements imply a larger value of core tensors, noting that the elements' values of three core tensor decrease from the upper left corner to the lower right corner of tensors, and we named this view as view 1. Besides, we Observe the distribution of elements' values in three core tensors from another view, which we named view 2.

data finely via proper regularizations.

The model-driven or non-deep learning methods focus more on characterizing priors on HSI data finely via proper regularizations. Among all priors, the spectral global correlation property and spatial local smoothness are the two most commonly used properties. For the characterization of spectral global correlation, there are three main categories of methods, namely, Gaussian distribution-based models, sparse representation-based models, and low-rank decomposition-based models. Among the models based on Gaussian distribution, the typical ones are Reed Xiaoli (RX), global RX (GRX), local RX (LRX), weighted-RX, and kernel-RX algorithms [28]–[30]. These methods assume that the background obeys Gaussian distribution, and use Mahalanobis distance to measure whether the test pixel is an abnormal point. However, the assumption of a specific distribution for the background is not valid in real-world scenarios and hinders the improvement of AD accuracy. Representation-based methods, including Sparse Representation (SR)-based [12] and Collaborative Representation (CR)-based [31], [32], which represent the background as a sparse combination of redundant dictionary elements. Models based on low-rank decomposition, including robust principle component analysis (RPCA) [33], Canonical Polyadic (CP) decomposition [34], and Tucker decomposition [5], [35], these methods can decompose the background HSI into two low-rank matrices/tensors, thus can characterize the global correlation of the spectrum. However, only low-rank priors can not bring the best performance for AD tasks. Besides, the low rank of the spectrum, local smoothness is another frequently used prior. The local-smooth prior is often encoded by total variation (TV) regularization [44] on the spatial and spectral domain of

the HSI and often combined into a low-rank matrix/tensor decomposition framework, such as [6], [7], [36]–[39], to improve the performance of models. Whereas, the performance of all the above matrix-based and tensor-based methods is highly affected by the trade-off parameter imposed between the low-rank and local smooth regularizers. It is fairly difficult to build a general rule for finely tuning the balancing parameter in real scenarios [45]. Moreover, the [46] and the [39] respectively utilize the Tucker decomposition and the CP decomposition on the original data and then unfold each gradient map in the tensor manner to the matrix, which destroys the topological structure of the pixel space.

To alleviate the above issues, this paper proposes a new prior characterization paradigm for fusing low-rank and local smoothness properties by exploring the sparsity of the core tensor of the background tensor's gradient map. Specifically, for an original background tensor, we can obtain the gradient map of the background tensor through the difference operator, so that the local smoothness of the original tensor is transferred to the sparsity of its gradient map. Since the difference operator is a linear operator and does not change the low-rank nature of tensors, the gradient maps also have low-rank properties, as seen in Figure 1. Therefore, the difference operation transfers the low rank of the original data and the sparsity of its differential map (that is, the local smoothness of the original data) to the low rank and sparsity of the gradient map, which makes it possible for us to characterize both types of priors simultaneously with one regularizer on the gradient map. In fact, a lot of work has been done based on gradient map modeling before, such as [47]–[49], but these works are all based on the matrix, which destroys the data structure to a certain extent. In order not to

break the structure of the background tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , here we use Tucker decomposition to explicitly encode the low-rank property of the gradient map, i.e.,  $\mathcal{X} = C \times_1 U_1 \times_2 U_2 \times_3 U_3$ , where  $C \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  are core tensor,  $U_i \in \mathbb{R}^{n_i \times r_i}$  are orthogonal factor matrices. In Tucker decomposition, the size of the core tensor determines the low rank of the original tensor in each dimension, and the lower the value in a dimension, the stronger the low rank of the tensor in that dimension. Further, since the gradient map tensor is sparse, and three-factor matrices are orthogonal, it can be deduced that only the sparse core tensor can induce the sparsity of the gradient map. In fact, the core tensor is indeed sparse, as shown in Figure 1, the number of non-zero elements in the core tensor is extremely small. In addition, since the value of the core tensor can be regarded as the representation coefficient under three sets of orthogonal bases  $\{U_i\}_{i=1}^3$ , adding sparsity to the core tensor will make better use of the structural information (i.e.,  $\{U_i\}_{i=1}^3$ ) of the data than adding sparsity to the original tensor, similar analysis can be found in [50]. Based on the above analysis, we can achieve the goal of simultaneously encoding the low rank and local smoothness of the background tensor by controlling the sparsity of the core tensor of the gradient map. For the convenience of description, we refer to this new prior as gradient map core tensor sparsity (GCS). Furthermore, we propose the  $\ell_1$ -norm to encode the GCS prior for the background tensor, and the anomaly tensor is characterized by the  $\ell_1$ -norm to separate the anomaly from the background more accurately.

The main contributions of this paper can be summarized as follows:

- 1) In terms of prior mining. We observe a new prior called GCS prior. Unlike the low-rank or local smoothness priors that have been discovered before, this prior can characterize the low-rank and local smoothness of tensor data at the same time, which allows us to use a simple regularization to efficiently describe low-rank and local smoothness, such as  $\ell_1$ -norm. Besides, with the help of the physical meaning of the core tensor, the GCS prior can also deliver the representation sparsity under Tucker decomposition.
- 2) In terms of algorithm design. Since the GCS prior is obtained through Tucker decomposition under the gradient map, it is cumbersome to solve. Here we design a fast ADMM algorithm to solve  $\ell_1$ -norm on gradient core tensor based on HOSVD and Fast Fourier Transform. Compared with other tensor decomposition methods, the runtime of our algorithm is also comparable.
- 3) In terms of experiment performance. Experimental results on five real-world datasets, with 3D ROC metrics and illustrations, demonstrate the superiority of the proposed method.

The remainder of this paper is organized as follows. In Section II, we introduce some notations and preliminaries of the tensor. Section III describes the motivation and our proposed GCS regularization for hyperspectral AD. Section IV is the proposed method and the optimization algorithm. Evaluation of experimental results and discussions constitute Section V. Finally, we summarize our work in Section VI.

## II. NOTATIONS AND PRELIMINARIES

In this section, we introduce some mathematical notations and definitions of the tensors referred to in the paper to describe the proposed method clearly. The paper uses bold lowercase symbols for vectors, e.g.,  $x$ , and capital letters for matrices, e.g.,  $X$ . Scalar is written as  $x$ . We use bold Euler script letters to denote the third-order, e.g.,  $\mathcal{X}$ . Given a  $N$  mode tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , here,  $I_n, \forall n = \{1, \dots, N\}$  is the number of  $n$ -th mode.

**Definition 1. (The mode- $n$  unfolding and folding of a tensor)** The "unfold" operation along mode- $n$  on a  $N$ -mode tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is defined as  $\text{unfold}_n(\mathcal{X}) = \mathcal{X}_{(n)} \in \mathbb{R}^{I_n \times (I_1 \dots I_{n-1} I_{n+1} \dots I_N)}$ . Its inverse operation is the mode- $n$  folding, denoted as  $\mathcal{X} = \text{fold}_n(\mathcal{X}_{(n)})$ .

**Definition 2. (Tucker decomposition)** A Tucker decomposition is a form of higher-order principal component analysis in which it decomposes a tensor into a core tensor multiplied by a matrix along each mode. Here, we give an example by a third-order tensor  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ , then

$$\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_p \circ b_q \circ c_r = \llbracket \mathcal{G}; A, B, C \rrbracket, \quad (1)$$

here,  $A \in \mathbb{R}^{I \times P}$ ,  $B \in \mathbb{R}^{J \times Q}$ , and  $C \in \mathbb{R}^{K \times R}$  are the factor matrices, which can be thought of as the principal components in each mode. The core tensor  $\mathcal{G} \in \mathbb{R}^{P \times Q \times R}$ , and its entries show the level of interaction between the different components.

**Definition 3. (SSTV)** The spatial and spectral total variation regularization term of tensor  $\mathcal{X}$  is defined as:

$$\|\mathcal{X}\|_{\text{SSTV}} = \|\mathbf{D}_1 \mathcal{X}\| + \|\mathbf{D}_2 \mathcal{X}\| + \|\mathbf{D}_3 \mathcal{X}\|, \quad (2)$$

$\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3$  are the first-order difference operators with respect to the horizontal, vertical, and spectral directions of the background. Here, the sparsity measurement of  $\|\mathcal{X}\|_{\text{SSTV}}$  is by summing the  $\ell_1$ -norm of the gradient maps along three modes. To simplify Eq. (2), we rewritten it as  $\sum_{n=1}^3 \|\nabla_n \mathcal{X}\|_1$ ,  $\nabla_n, (n = 1, 2, 3)$ , is a linear operator on  $\mathcal{X}$  along the  $n$ -th mode,  $\|\cdot\|_1$  is the tensor  $\ell_1$ -norm.

**Definition 4. (The tensor  $\ell_1$ -norm)** The  $\ell_1$ -norm of the tensor is similar to the matrix, for the given third-order tensor  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ , its  $\ell_1$ -norm is  $\|\mathcal{X}\|_1 = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K |x_{ijk}|$ .

**Definition 5. (The tensor Frobenius norm)** Given a third-order tensor  $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ , its Frobenius norm is defined as  $\|\mathcal{X}\|_F = \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$ .

## III. SPARSITY MEASURE ON THE GRADIENT MAP TENSOR UNDER TUCKER DECOMPOSITION REGULARIZATION

In this section, we first develop the problem statement to present the motivation for the background model of hyperspectral AD. The GCS regularization and the proposed method are illustrated afterward.

An HSI can be naturally treated as a third-order tensor, which avoids spatial and spectral information loss in a tensor manner. Given an observed HSI tensor, it can be decomposed

into a background tensor and an anomaly tensor. Then, the model can be expressed as follows:

$$\mathcal{Y} = \mathcal{B} + \mathcal{E}, \quad (3)$$

where  $\mathcal{Y} \in \mathbb{R}^{H \times W \times D}$  is the observed HSI,  $\mathcal{B} \in \mathbb{R}^{H \times W \times D}$ , and  $\mathcal{E} \in \mathbb{R}^{H \times W \times D}$  are the background, and anomaly respectively.  $H$ ,  $W$ , and  $D$  represent the height, width, and number of bands of each tensor. The purpose of AD is to yield superior performance by recovering an accurate background image that allows the anomaly to be more accurately separated from the background.

### A. Motivation

Since the background scene is simple and usually consists of a few materials, it indicates that the spectral bands of the pixels belonging to the background are highly correlated and therefore lie in a low-rank subspace. Besides, similar substances distribute in the same region, which implies a local spatial and spectral smoothness of the background. Here, the SSTV regularization is usually utilized to encode the spatial-spectral smoothness. The SSTV regularization on  $\mathcal{B}$  can be written as:

$$\|\mathcal{B}\|_{\text{SSTV}} = \sum_{n=1}^3 \|\nabla_n \mathcal{B}\|_1, \quad (4)$$

$\nabla_n$ , ( $n = 1, 2, 3$ ) are the first-order difference operators with respect to the horizontal, vertical, and spectral directions of the background. Here,  $\|\mathcal{B}\|_{\text{SSTV}}$  is by summing the  $\ell_1$ -norm of the gradient maps along three directions. To simplify Eq. (4), we rewritten it as  $\sum_{n=1}^3 \|\nabla_n \mathcal{B}\|_1$ ,  $\nabla_n$  is a linear operator on  $\mathcal{B}$  along the  $n$ -th mode,  $\|\cdot\|_1$  is the tensor  $\ell_1$ -norm.

In Fig. 1 (b), the gradient tensors obtained by the TV regularizer smooths the background tensor along the vertical, horizontal, and spectral dimensions, respectively. After that, we unfold the three gradient tensors along three different modes, as can be observed in Fig. 1 (c), the singular values curve reflects the spectral dimension has a strong low-rank property, and the other two spatial dimensions also have a low-rank property. Therefore, the phenomenon inspires us to utilize Tucker decomposition to encode the gradient tensors' low-rank property for the background on the three different dimensions, which can be written as:

$$\mathcal{B} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3, \mathbf{U}_n^T \mathbf{U}_n = \mathbf{I}, \quad (5)$$

$$\text{rank}(\mathcal{G}) = [r_1, r_2, r_3], n = 1, 2, 3,$$

here,  $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \times r_3}$  is the core tensor of  $\mathcal{B}$ ,  $\{\mathbf{U}_n\}_{n=1}^3$  are the factor matrices, with the rank of  $\{r_n\}_{n=1}^3$ . Moreover, we define all the factor matrices  $\mathbf{U}_n$  as orthogonal matrices, where  $\mathbf{U}_n^T \mathbf{U}_n = \mathbf{I}$ , ( $n = 1, 2, 3$ ).

It should be noted that the key to achieving accurate AD results depends on the efficiency of the low-rank background tensor model. Combining the low-rank Tucker decomposition with the SSTV term to encode the low-rank and the local spatial-spectral smooth characteristics of the background has been widely developed in HSI processing. Whereas these two properties, as illustrated above, are characterized separately, they cannot exploit some insightful structures in integrity to maximize their effectiveness. There have been some attempts

for HSI recovery [45], target detection [51], denoise [47], and others [52] to model the two priors into one regularizer, which yields excellent performance. Up to now, on the one hand, encoding the low rankness and smoothness properties by a fusion regularizer is blindly in the hyperspectral AD; on the other hand, the existing works [45], [47] are matrix based methods that destroy the inner structure of HSI. To sum up, it facilitates us to mine the beneficial priors underlying the gradient domain of the background.

### B. GCS Regularization

In this section, we focus on illustrating the sparsity of the gradient map's core tensor obtained by tucker decomposition to verify the reason for the proposed GCS regularization term. In addition, Fig. 1 also helps to demonstrate the sparsity of the gradient tensors through visualization of the distribution of the element's value and the frequency histograms.

In Eq. (4), since the difference operator is a linear operator and can be represented as a nearly full rank matrix  $\mathbf{D}$ , where  $\text{rank}(\nabla_i(\mathcal{B})) = \text{rank}(\mathcal{B} \times_i \mathbf{D})$ ,  $\text{rank}(\mathbf{D}\mathcal{B}_{(i)}) \leq \min\{D, \mathcal{B}_{(i)}\}$ , then we have  $\text{rank}(\nabla_n \mathcal{B}) = \text{rank}(\mathcal{B})$ , ( $n = 1, 2, 3$ ). Thus, the gradient tensor  $\nabla_n \mathcal{B}$  obtained by imposing the difference operation on the original background tensor  $\mathcal{B}$  also inherits the low-rank property of the background. Therefore, as shown in Fig. 1, there always has a strong correlation between gradient maps, we thus employ the low-rank Tucker decomposition to decompose the gradient tensors of the background tensor with low-rank properties encoded. We model the gradient tensor independently as follows:

$$\nabla_n \mathcal{B} = \mathcal{G}_n \times_1 \mathbf{U}_n^{(1)} \times_2 \mathbf{U}_n^{(2)} \times_3 \mathbf{U}_n^{(3)}, n = 1, 2, 3 \quad (6)$$

$$\text{s.t. } \mathbf{U}_n^{(i)T} \mathbf{U}_n^{(i)} = \mathbf{I}, \text{rank}(\mathcal{G}_n) = [r^{(1)}, r^{(2)}, r^{(3)}], i = 1, 2, 3,$$

where  $\{\mathbf{U}_n^{(i)}\}_{i=1}^3$ , ( $n = 1, 2, 3$ ) are the factor matrices,  $\mathbf{U}_1^{(i)}$ ,  $\mathbf{U}_2^{(i)}$ ,  $\mathbf{U}_3^{(i)}$ , ( $i = 1, 2, 3$ ) with the size of  $H \times r^{(1)}$ ,  $W \times r^{(2)}$ ,  $D \times r^{(3)}$ , respectively, and we define all the factor matrices as orthogonal matrices.  $\mathcal{G}_n$  is the core tensor of  $\nabla_n \mathcal{B}$ , ( $n = 1, 2, 3$ ), when its rank is  $[r^{(1)}, r^{(2)}, r^{(3)}]$ , its size would be  $r^{(1)} \times r^{(2)} \times r^{(3)}$ . Thus, we can control the low rank of the three-order tensor by presetting the rank of the core tensor.

Above we analyzed the low rank of the gradient map, and then we analyze the sparsity of the gradient map, that is, the local smoothness of the original data. As shown in Fig. 1, since the gradient maps are obtained by performing difference operations on the original data, only the edge of the object is preserved, so the gradient map shows strong local smoothness. There are two methods to describe the sparseness of the gradient map. The first one directly adds an  $\ell_1$ -norm to the gradient map as a whole, that is, the SSTV regularization of the original tensor, but this regularity is a kind of space-indiscriminate way to describe the sparsity, this is because the SSTV regularization fails to make better use of the information of each spectrum, namely, the information of orthogonal matrices  $\{\mathbf{U}_n^{(i)}\}_{i=1}^3$ , ( $n = 1, 2, 3$ ). The second encoding

way is to describe the sparsity of the core tensor. Since the gradient map has a Tucker decomposition representation, the factor matrix is orthogonal and does not have sparsity, so the sparsity naturally falls on the core tensor. Only the core tensor is sparse, and under the representation of Tucker decomposition, the original gradient map is sparse. In fact, the core tensor of the gradient map is indeed extremely sparse, as shown in Fig. 1(e). Therefore, by using the sparsity of the core tensor, the information of the tensor in different dimensions (i.e.,  $\{U_n^{(i)}\}_{i=1}^3$ , ( $n = 1, 2, 3$ )) are used. Secondly, in work [50], the physical meaning of the core tensor is also analyzed, that is, the core tensor coefficient can be regarded as the weight coefficient under the CP decomposition. If we limit the sparsity of the core tensor, then we can also limit the low-rank property under CP decomposition to some extent. Noting that in Fig. 1 (f) the elements' values of three core tensors decrease from the upper left corner to the lower right corner of tensors which further verifies the core tensors' sparsity. Due to the sparsity of the gradient's core maps reflecting the sparsity of the gradient, thereby, the  $\ell_1$  norm is utilized to describe the sparsity of the gradient map (i.e., the local smoothness of the background). The size of the core tensor in Tucker decomposition determines the low rank of the original tensor in each dimension, and the lower the value in a dimension, the stronger the low rank of the tensor in that dimension. Therefore, the sparsity of the core tensor is inversely deducing that the gradient maps are sparse.

In this paper, we construct a novel tensor regularizer by the sparsity measure, namely  $\ell_1$ -norm, on the core tensor of the gradient map under tucker decomposition achieves the purpose of encoding low-rank and local smooth properties at the same time through a regularization. In particular, the size of the core tensor represents the low-rank prior of the background, and the  $\ell_1$  norm describes the sparsity of the gradient map (i.e., the local smoothness of the original data). This new form of tensor-based prior characterization is named gradient map core tensor sparsity (GCS), we define it as following Eq. (7). To simplify notations, the collection of unknowns is union defined  $\Theta = \{\mathcal{B}, \mathcal{G}_n, U_n^{(1)}, U_n^{(2)}, U_n^{(3)}, \mathcal{E}\}$ ,  $n = 1, 2, 3$ , and the background model is finally written as follows:

$$\min_{\Theta} \sum_{n=1}^3 \|\mathcal{G}_n\|_1$$

$$\text{s.t. } \mathcal{Y} = \mathcal{B} + \mathcal{E}, \nabla_n \mathcal{B} = \mathcal{G}_n \times_1 U_n^{(1)} \times_2 U_n^{(2)} \times_3 U_n^{(3)}, U_n^{(i)T} U_n^{(i)} = \mathbf{I},$$

$$\text{rank}(\mathcal{G}_n) = [r^{(1)}, r^{(2)}, r^{(3)}], n = 1, 2, 3, \quad (7)$$

where  $\|\cdot\|_1$  is the tensor  $\ell_1$ -norm.

#### IV. HYPERSPECTRAL AD UNDERLYING GCS REGULARIZATION

##### A. Proposed Method

Since anomalies make up a small proportion of the whole HSI, which is usually assumed with a sparse property, thus, we utilize the  $\ell_1$ -norm to characterize the anomaly tensor's sparsity, aiming to separate the anomaly from the background more accurately. Therefore, based on the model (8), we proposed a

gradient tensors' low-rank Tucker decomposition modeling for hyperspectral AD as follows:

$$\min_{\Theta} \sum_{n=1}^3 \|\mathcal{G}_n\|_1 + \lambda \|\mathcal{E}\|_1$$

$$\text{s.t. } \mathcal{Y} = \mathcal{B} + \mathcal{E}, \nabla_n \mathcal{B} = \mathcal{G}_n \times_1 U_n^{(1)} \times_2 U_n^{(2)} \times_3 U_n^{(3)}, U_n^{(i)T} U_n^{(i)} = \mathbf{I},$$

$$\text{rank}(\mathcal{G}_n) = [r^{(1)}, r^{(2)}, r^{(3)}], n = 1, 2, 3, \quad (8)$$

here,  $\lambda > 0$  is the trade-off parameter that controls the sparsity of the anomaly part.

##### B. Optimization Algorithm

We optimize the model (8) by the alternating direction method of multipliers (ADMM) [53], the augmented Lagrangian function of the model (8) is

$$L(\Theta, \Gamma, \mathcal{M}_n, \mu) = \sum_{n=1}^3 \|\mathcal{G}_n\|_1 + \lambda \|\mathcal{E}\|_1 + \langle \Gamma, \mathcal{Y} - \mathcal{B} - \mathcal{E} \rangle$$

$$+ \frac{\mu}{2} \|\mathcal{Y} - \mathcal{B} - \mathcal{E}\|_F^2 + \sum_{n=1}^3 \langle \mathcal{M}_n, \nabla_n \mathcal{B} - \mathcal{G}_n \times_1 U_n^{(1)} \times_2 U_n^{(2)} \times_3 U_n^{(3)} \rangle$$

$$+ \frac{\mu}{2} \sum_{n=1}^3 \|\nabla_n \mathcal{B} - \mathcal{G}_n \times_1 U_n^{(1)} \times_2 U_n^{(2)} \times_3 U_n^{(3)}\|_F^2, \quad (9)$$

which is under the constraints of  $\nabla_n \mathcal{B} = \mathcal{G}_n \times_1 U_n^{(1)} \times_2 U_n^{(2)} \times_3 U_n^{(3)}, U_n^{(i)T} U_n^{(i)} = \mathbf{I}$ ,  $\text{rank}(\mathcal{G}_n) = [r^{(1)}, r^{(2)}, r^{(3)}]$ , ( $n = 1, 2, 3, i = 1, 2, 3$ ), where  $\mathcal{M}_n$ ,  $n = 1, 2, 3$ ,  $\Gamma$  are the Lagrangian multipliers, and  $\mu$  is a positive scalar in the ADMM algorithm. Here, we can optimize the augmented Lagrangian function (9) by updating one variable while fixing the others. Specifically, the initial value of  $k$  is set to 1. When the update time is  $k + 1$ , the variables refer to the model (8) iterate as follows:

1) *update*  $\mathcal{E}^{k+1}$ : The optimization problem for  $\mathcal{E}^{k+1}$  is

$$\mathcal{E}^{k+1} = \arg \min_{\mathcal{E}} \lambda \|\mathcal{E}\|_1 + \frac{\mu^k}{2} \|\mathcal{Y} - \mathcal{B}^k - \mathcal{E} + \frac{\Gamma^k}{\mu^k}\|_F^2, \quad (10)$$

$\mathcal{E}^{k+1}$  can be efficiently computed by

$$\mathcal{E}^{k+1} = \text{softthre}(\mathcal{Y} - \mathcal{B}^k + \frac{\Gamma^k}{\mu^k}, \frac{\lambda}{\mu^k}), \quad (11)$$

where  $\text{softthre}(\cdot, \cdot)$  is the soft-thresholding function.

2) *Update*  $\{U_n^{(i)k+1}\}_{i=1}^3$ , ( $n = 1, 2, 3$ ): Here, we update them by the classic HOOI algorithm [54].

3) *Update*  $\mathcal{G}_n^{k+1}$ , ( $n = 1, 2, 3$ ): The optimization problem for  $\mathcal{G}_n^{k+1}$  can be written as follows:

$$\mathcal{G}_n^{k+1} = \arg \min_{\mathcal{G}_n} \sum_{n=1}^3 \|\mathcal{G}_n\|_1 +$$

$$\frac{\mu^k}{2} \sum_{n=1}^3 \|\nabla_n \mathcal{B}^k - \mathcal{G}_n \times_1 U_n^{(1)k+1} \times_2 U_n^{(2)k+1} \times_3 U_n^{(3)k+1} + \frac{\mathcal{M}_n^k}{\mu^k}\|_F^2, \quad (12)$$

here, we solve Eq. (12) by the soft-thresholding function [55],

$$\mathcal{G}_n^{k+1} = \text{softthre}(\text{Temp}, \frac{1}{\mu^k}), \quad (13)$$

where  $\text{Temp} = (\nabla_n \mathcal{B}^k + \frac{\mathcal{M}_n^k}{\mu^k}) \times_1 \mathcal{U}_n^{(1)T^k} \times_2 \mathcal{U}_n^{(2)T^k} \times_3 \mathcal{U}_n^{(3)T^k}$ .

4) *Update*  $\mathcal{B}^{k+1}$ : The optimization problem for  $\mathcal{B}^{k+1}$  is:

$$\begin{aligned} \mathcal{B}^{k+1} = \arg \min_{\mathcal{B}} & \frac{\mu^k}{2} \|\mathcal{Y} - \mathcal{B} - \mathcal{E}^{k+1} + \frac{\Gamma^k}{\mu^k}\|_F^2 + \\ & \frac{\mu^k}{2} \sum_{n=1}^3 \|\nabla_n \mathcal{B} - \mathcal{G}_n^{k+1} \times_1 \mathcal{U}_n^{(1)k+1} \times_2 \mathcal{U}_n^{(2)k+1} \times_3 \mathcal{U}_n^{(3)k+1} + \frac{\mathcal{M}_n^k}{\mu^k}\|_F^2, \end{aligned} \quad (14)$$

the above optimization problem can be transformed into the following linear system:

$$\begin{aligned} \left( \sum_{n=1}^3 \nabla_n^T \nabla_n + \mathbf{I} \right) \mathcal{B}^{k+1} &= \mathcal{Y} - \mathcal{E}^{k+1} + \frac{\Gamma^k}{\mu^k} + \\ \sum_{n=1}^3 \nabla_n^T \left( \mathcal{G}_n^{k+1} \times_1 \mathcal{U}_n^{(1)k+1} \times_2 \mathcal{U}_n^{(2)k+1} \times_3 \mathcal{U}_n^{(3)k+1} - \frac{\mathcal{M}_n^k}{\mu^k} \right), & n = 1, 2, 3, \end{aligned} \quad (15)$$

where  $\nabla_n^T$  indicates the ‘transposition’ operator of  $\nabla_n$ . Considering to the block-circulant of matrix corresponding to the operator  $\nabla_n^T \nabla_n$ , ( $n = 1, 2, 3$ ), we diagonalized it by the FFT matrix [56]. Here, referring [57],  $\mathcal{B}^{k+1}$  can be efficiently computed by

$$\begin{cases} \mathbf{H}_{\mathcal{B}} = \sum_{n=1}^3 \nabla_n^T \left( \mathcal{G}_n^{k+1} \times_1 \mathcal{U}_n^{(1)k+1} \times_2 \mathcal{U}_n^{(2)k+1} \times_3 \mathcal{U}_n^{(3)k+1} - \frac{\mathcal{M}_n^k}{\mu^k} \right) \\ + \mathcal{Y} - \mathcal{E}^{k+1} + \frac{\Gamma^k}{\mu^k}, n = 1, 2, 3, \\ \mathbf{T}_{\mathcal{B}} = |\text{fftn}(\mathbf{D}_1)|^2 + |\text{fftn}(\mathbf{D}_2)|^2 + |\text{fftn}(\mathbf{D}_3)|^2, \\ \mathcal{B}^{k+1} = \text{ifftn} \left( \frac{\text{fftn}(\mathbf{H}_{\mathcal{B}})}{\mathbf{T}_{\mathcal{B}}} \right), \end{cases} \quad (16)$$

where  $\text{fftn}$  and  $\text{ifftn}$  are the fast Fourier transform and its inverse transform, respectively. Here,  $|\cdot|^2$  is the elements-wise square, and the division is also performed element-wisely.

5) *Update multipliers*  $\Gamma^{k+1}$ ,  $\mathcal{M}_1^{k+1}$ ,  $\mathcal{M}_2^{k+1}$ ,  $\mathcal{M}_3^{k+1}$ , and  $\mu$ :

$$\begin{cases} \Gamma^{k+1} = \Gamma^k + \mu^k \left( \mathcal{Y} - \mathcal{B}^{k+1} - \mathcal{E}^{k+1} \right), \\ \mathcal{M}_1^{k+1} = \mathcal{M}_1^k + \mu^k \left( \nabla_1 \mathcal{B}^{k+1} - \mathcal{G}_1^{k+1} \times_1 \mathcal{U}_1^{(1)k+1} \times_2 \mathcal{U}_1^{(2)k+1} \times_3 \mathcal{U}_1^{(3)k+1} \right), \\ \mathcal{M}_2^{k+1} = \mathcal{M}_2^k + \mu^k \left( \nabla_2 \mathcal{B}^{k+1} - \mathcal{G}_2^{k+1} \times_1 \mathcal{U}_2^{(1)k+1} \times_2 \mathcal{U}_2^{(2)k+1} \times_3 \mathcal{U}_2^{(3)k+1} \right), \\ \mathcal{M}_3^{k+1} = \mathcal{M}_3^k + \mu^k \left( \nabla_3 \mathcal{B}^{k+1} - \mathcal{G}_3^{k+1} \times_1 \mathcal{U}_3^{(1)k+1} \times_2 \mathcal{U}_3^{(2)k+1} \times_3 \mathcal{U}_3^{(3)k+1} \right), \\ \mu^{k+1} = \min(1.5\mu^k, 1e5). \end{cases} \quad (17)$$

### C. Computational Complexity Analysis

As shown in Algorithm 1, the computational cost of per-iteration underlying the GCS mainly lies in the subproblems of  $\mathcal{B}$  and  $\nabla_n \mathcal{B}$ ,  $n = 1, 2, 3$ . Updating  $\nabla_n \mathcal{B}$  seek for the HOOI algorithm to estimate the  $\mathcal{U}_n^{(i)}$ ,  $n = 1, 2, 3$ ,  $i = 1, 2, 3$ . The time complexity of factor matrices (Update  $\mathcal{U}_n^{(i)}$ ):  $O(3T_{\max}(H^2 r^{(2)} r^{(3)} + (r^{(2)} r^{(3)})^3 + W^2 r^{(1)} r^{(3)} + (r^{(1)} r^{(3)})^3 + D^2 r^{(1)} r^{(2)} + (r^{(1)} r^{(2)})^3))$ . (Update  $\nabla_n \mathcal{B}$ ):  $O(r^{(1)} r^{(2)} r^{(3)} H + r^{(3)} H r^{(2)} W + (H W r^{(3)})^3 D)$ . The update for  $\mathcal{B}$  depends on the computational cost of FFT.

### Algorithm 1 GCS algorithm

**Input:** HSI tensor  $\mathcal{Y}$ , regularized parameters  $\lambda$ .

**Initialization:**  $\mathcal{E}$ ,  $\mathcal{B}$ ,  $\mathcal{G}_n$  are set to zero tensor,  $\mathcal{U}_n^{(1)}$ ,  $\mathcal{U}_n^{(2)}$ ,  $\mathcal{U}_n^{(3)}$ ,  $\Gamma$  are set to matrices,  $\mu = 10^{-2}$ ,  $\mu_{\max} = 10^5$ ,  $T_{\max} = 50$ ,  $k = 1$ , and the residual error  $\xi = 10^{-6}$ .

**Repeat:**

1. Update  $\mathcal{E}$  with Eq. (11).
2. Update  $\mathcal{U}_n^{(1)}$ ,  $\mathcal{U}_n^{(2)}$ ,  $\mathcal{U}_n^{(3)}$  by the HOOI algorithm, and update  $\mathcal{G}_n$ ,  $n = 1, 2, 3$  with Eq. (12).
3. Update  $\mathcal{B}$  with Eq. (14).
4. Update  $\Gamma$ ,  $\mathcal{M}_1$ ,  $\mathcal{M}_2$ ,  $\mathcal{M}_3$ , and  $\mu$  with Eq. (16), and Eq. (17).
5.  $k \leftarrow k + 1$ .

**Until** either  $T_{\max} = 50$  or  $\|\mathcal{B}^{(t+1)} - \mathcal{B}^{(t)}\|_2 / \|\mathcal{B}^{(t+1)}\|_2 < \xi$  are satisfied.

**Output:** anomaly map  $\mathbf{T}$ .

(Update  $\mathcal{B}$ ):  $O(HWD \log(HWD))$ . So the time complexity of Algorithm 1 is  $O((H^2 r^{(2)} r^{(3)} + (r^{(2)} r^{(3)})^3 + W^2 r^{(1)} r^{(3)} + (r^{(1)} r^{(3)})^3 + D^2 r^{(1)} r^{(2)} + (r^{(1)} r^{(2)})^3) + (r^{(1)} r^{(2)} r^{(3)} H + r^{(3)} H r^{(2)} W + (H W r^{(3)})^3 D) + HWD \log(HWD))$ .

## V. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the proposed GCS method was carried out on five real HSI datasets for AD, and the gray values of the five datasets are normalized to the scale of (0–1) before performing the experiments. Where the detailed description is listed in the following subsections, all the experimental algorithms are performed in MATLAB 2016b on a computer with a 64-bit quad-core Intel Xeon 2.40 GHz CPU and 32.0 GB of RAM in Windows 7. The deep learning experimental algorithm is performed in Python 3.7, PyTorch 1.10, and a 6GB of NVIDIA GeForce RTX 2060 GPU.

### A. Experimental Datasets

1) *AVIRIS Airplane Datasets*: The AVIRIS airplane dataset was collected by AVIRIS in San Diego. There are 189 bands retained, while the water absorption regions, low-SNR, and bad bands are removed. As shown in Fig. 2 (a1), the subimage is named the AVIRIS-1 dataset, and it is located in the top-left corner of the AVIRIS image with a spatial size of  $100 \times 100$ . The contained anomaly is the three airplanes, and the ground truth is shown in Fig. 2 (a2). AVIRIS-2 dataset is located in the center of San Diego, as shown in Fig. 2 (b1), and the groundtruth, as shown in Fig. 2 (b2), with the size of  $100 \times 100 \times 186$ .

2) *HYDICE Dataset*: The real data was collected by the hyperspectral digital imagery collection experiment (HYDICE) sensor, and the original image has a size of  $307 \times 307 \times 210$ . In this experiment, there preserved 175 bands when removed the low-SNR and water vapor absorption bands. An  $80 \times 100$  subspace is cropped from the top right of the whole image, and the cars and roofs in the image scene are considered anomalies. The false color image and the corresponding ground truth map are shown in Fig. 2 (c1) and (c2).

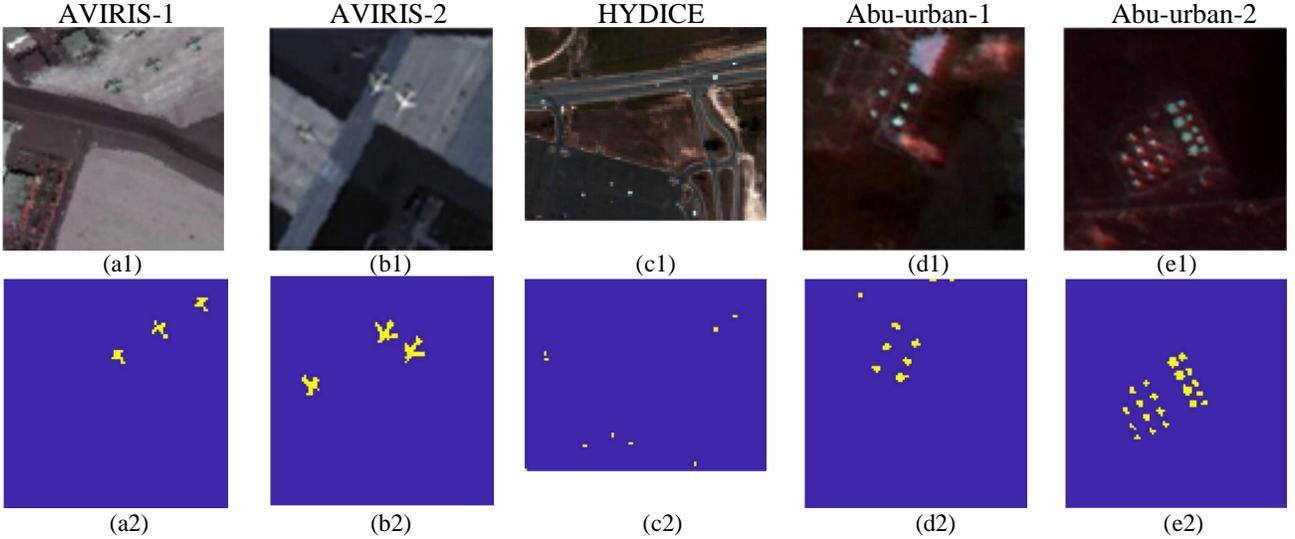


Fig. 2: The False color image and Ground-truth map of the anomalies on the AVIRIS-1, AVIRIS-2, HYDICE, Urban-1, and Urban-2 datasets.

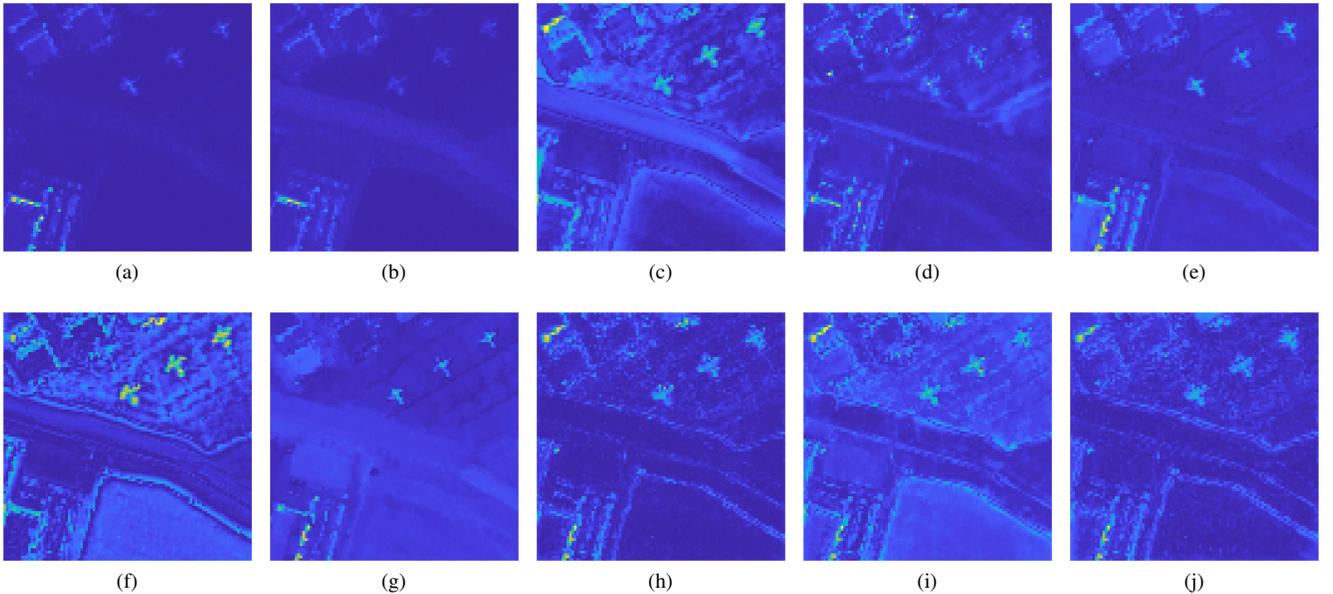


Fig. 3: 2D plots of the detection results obtained by all competing methods on the AVIRIS-1 dataset. (a) RX. (b) RPCA. (c) LRTV. (d) GTVLRR. (e) E-3DTV. (f) PTA. (g) GVAE. (h) TRPCA. (i) LRTDTV. (j) GCS.

3) *Urban (ABU) Datasets*: The Urban dataset was collected with AVIRIS sensors and contained five images of five different scenes. In this paper, we select the Urban-1 image and Urban-2 image, captured at different locations on the Texas Coast, to perform the experiment. The spatial size of the Urban-1 dataset is  $100 \times 100$ , the number of spectral bands is 204, and its false-color image and the ground truth are presented in Fig. 2 (d1) and (d2). For the Urban-2 dataset with a size of  $100 \times 100 \times 207$ , Fig. 2 (e1) and (e2) are the corresponding false color image and the ground truth.

### B. Evaluation Metrics and Parameter Settings

Nine SOTA hyperspectral AD methods are selected as the comparison algorithms to assess the performance of the proposed GCS method. In order to demonstrate GCS is efficient in performing the Tucker decomposition on the gradient maps of the tensor background, we apply the LRTDTV method [38] for AD that is ever used for HSI restoration. The RX [58] is the benchmark method based on the statistical model of AD. RPCA [59], LRTV [60], GTVLRR [6], and E-3DTV [47] assume the background has a low-rank property while the anomaly is sparse. Based on the assumption, GTVLRR employs a TV regularization term with the background spatial smoothness

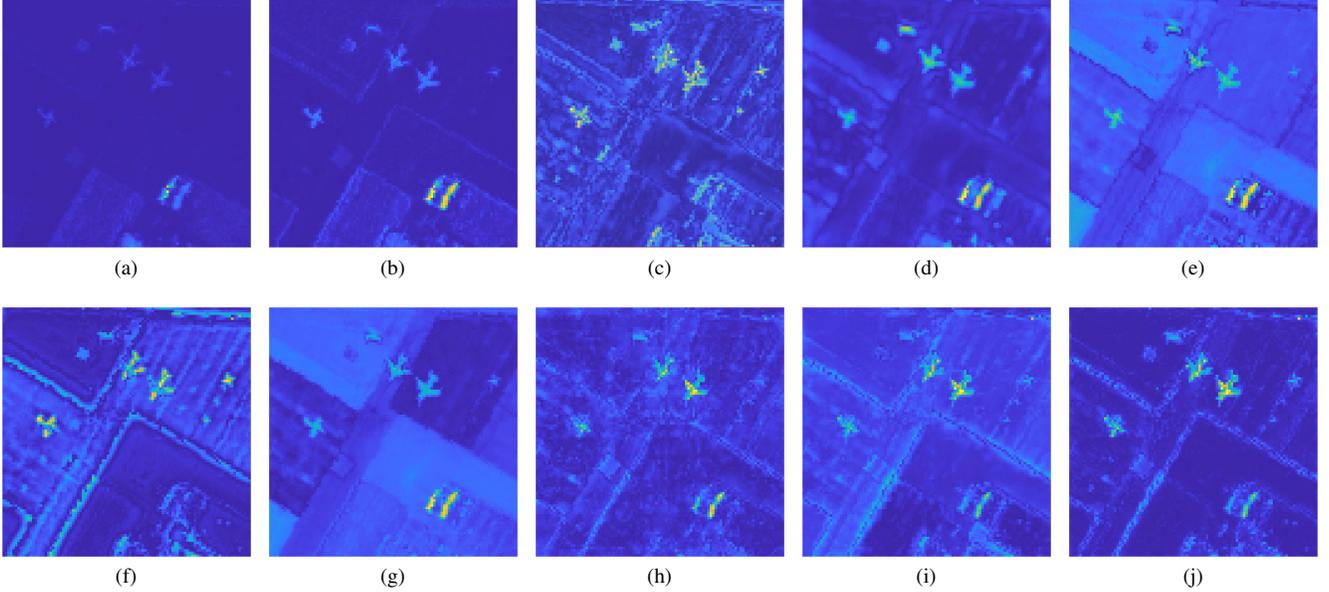


Fig. 4: 2D plots of the detection results obtained by all competing methods on the AVIRIS-2 dataset. (a) RX. (b) RPCA. (c) LRTV. (d) GTVLRR. (e) E-3DTV. (f) PTA. (g) GVAE. (h) TRPCA. (i) LRTDTV. (j) GCS.

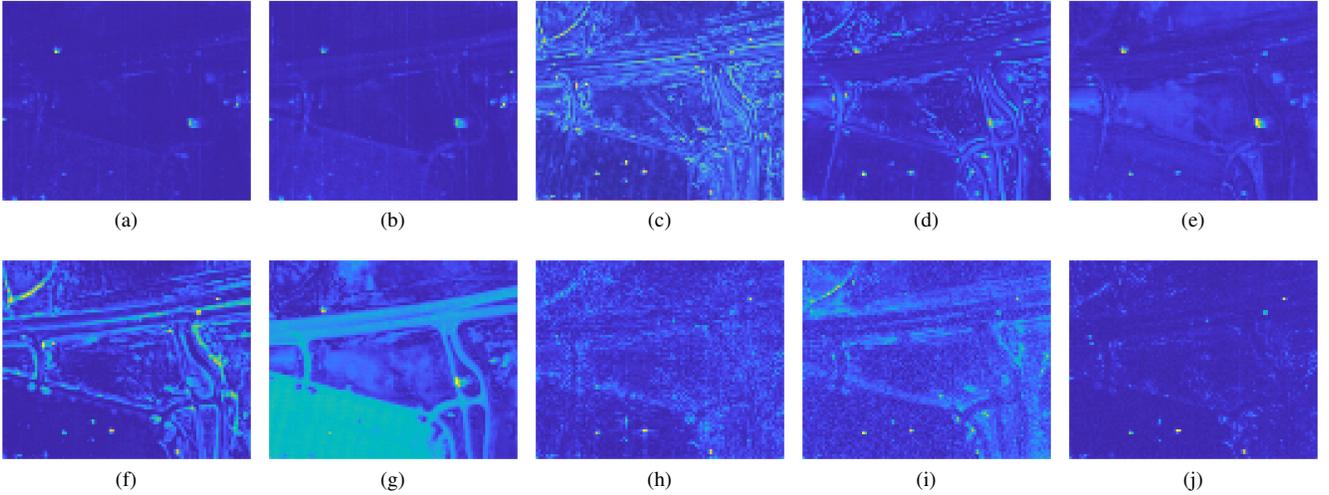


Fig. 5: 2D plots of the detection results obtained by all competing methods on the HYDICE dataset. (a) RX. (b) RPCA. (c) LRTV. (d) GTVLRR. (e) E-3DTV. (f) PTA. (g) GVAE. (h) TRPCA. (i) LRTDTV. (j) GCS.

encoded. The above methods are Matrix-based operations. In addition, PTA [61], TRPCA [62], the LRTDTV and our proposed GCS are the tensor-based methods underlying the low-rank background tensor and sparse anomaly tensor assumption. The PTA is the SOTA tensor-based AD method adapted for comparison with the tensor-based methods. Significantly, the matrix-based method E-3DTV and the tensor-based method GCS both focus on exploiting the property of the gradient maps underlying the background. Moreover, GVAE [9], a SOTA deep learning method, is adopted as the comparison method.

To effectively evaluate the performance of the hyperspectral AD detectors as above referred, the 3-D Receiver Operating Characteristic (3D-ROC) curve  $(P_D, P_F, \tau)$  [63], and it further generates three 2D-ROC curves of  $(P_D, P_F)$ ,  $(P_D, \tau)$ ,  $(P_F, \tau)$  with

their corresponding AUC values that are denoted as  $AUC_{(P_D, P_F)}$ ,  $AUC_{(P_D, \tau)}$ ,  $AUC_{(P_F, \tau)}$  are introduced. In addition, we employed a new AUC measure overall detection probability (OD), which is denoted as

$$AUC_{(OD)} = AUC_{(P_D, P_F)} + AUC_{(P_D, \tau)} - AUC_{(P_F, \tau)}. \quad (18)$$

The following are the description for the different metrics.

- $(P_D, P_F)$  assesses the overall detection performance
- $(P_D, \tau)$  evaluates the target preservation
- $(P_F, \tau)$  evaluates the background suppression
- (OD) for the overall performance evaluation

As such, the performance tends to be favorable as  $AUC_{(P_D, P_F)} \rightarrow 1$ ,  $AUC_{(P_D, \tau)} \rightarrow 1$ ,  $AUC_{(P_F, \tau)} \rightarrow 0$ , and  $AUC_{(OD)} \rightarrow 2$ . Besides, it is desired that the curves of  $(P_D, P_F)$ ,

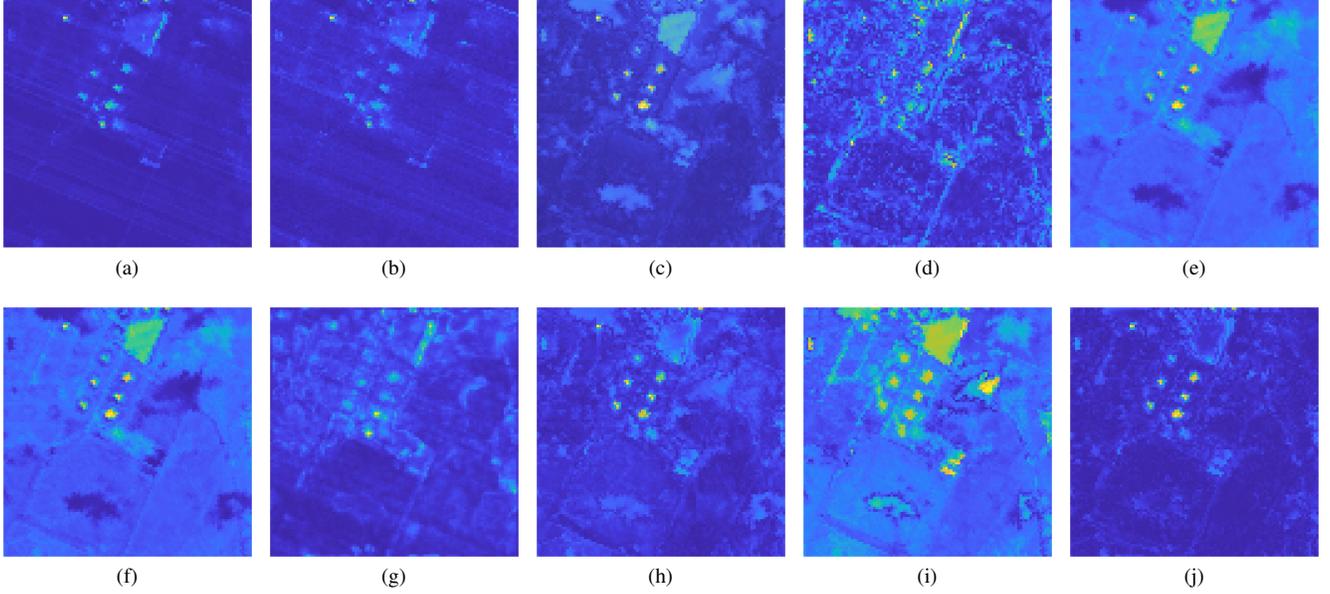


Fig. 6: 2D plots of the detection results obtained by all competing methods on the Urban-1 dataset. (a) RX. (b) RPCA. (c) LRTV. (d) GTVLRR. (e) E-3DTV. (f) PTA. (g) GVAE. (h) TRPCA. (i) LRTDTV. (j) GCS.

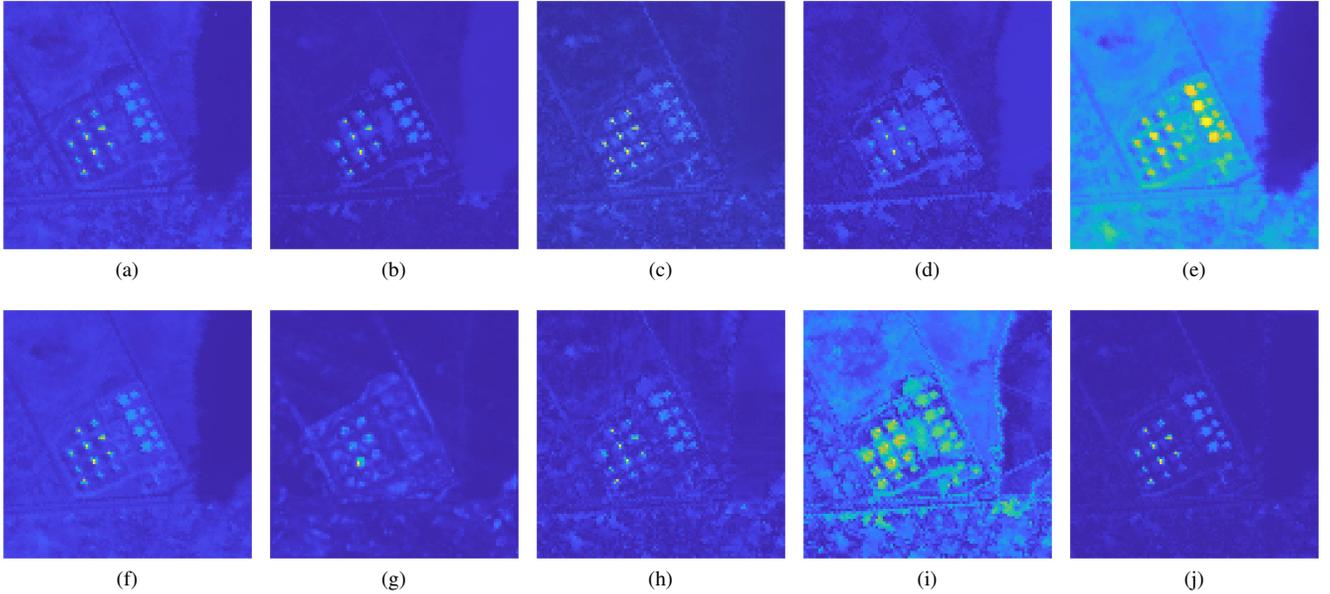


Fig. 7: 2D plots of the detection results obtained by all competing methods on the Urban-2 dataset. (a) RX. (b) RPCA. (c) LRTV. (d) GTVLRR. (e) E-3DTV. (f) PTA. (g) GVAE. (h) TRPCA. (i) LRTDTV. (j) GCS.

$(P_D, \tau)$ , and  $(P_F, \tau)$  are close to the upper left, upper right, and lower left corners of the coordinate axis, respectively.

In the proposed GCS,  $\lambda$  is a parameter to control the sparsity of the anomaly and varies in  $[1, 0.001]$ , which needs to be carefully tuned. The rank constraint of the gradient map tensors  $\text{rank}(\mathcal{G}_n) = [r^{(1)}, r^{(2)}, r^{(3)}]$ ,  $n = 1, 2, 3$ , that is, the  $r^{(3)}$  of the  $\mathcal{G}_n$  on the spectral dimension are vary in  $[5, 10]$  due to the high spectral bands correlation. The rank  $r^{(1)}, r^{(2)}$ , of the spatial dimension, highly depends on the structural complexity in the background scene of the experimental dataset.

### C. Detection Performance

We investigate the performance of RX, RPCA, LRTV, GTVLRR, E-3DTV, PTA, GVAE, TRPCA, LRTDTV, and GCS, ten methods from the quantitative and qualitative views. Specifically, Fig. 3-Fig. 7 shows the 2-D plots of the detection maps from the referring methods on the five datasets. Fig. 8 includes five subfigures, showing the ROC curves of the five datasets underlying the referring methods from (a) to (d). Each subfigure includes a  $(P_D, P_F, \tau)$  curve and its generative three 2D-ROC curves of  $(P_D, P_F)$ ,  $(P_D, \tau)$ ,  $(P_F, \tau)$ . Fig. 9 shows

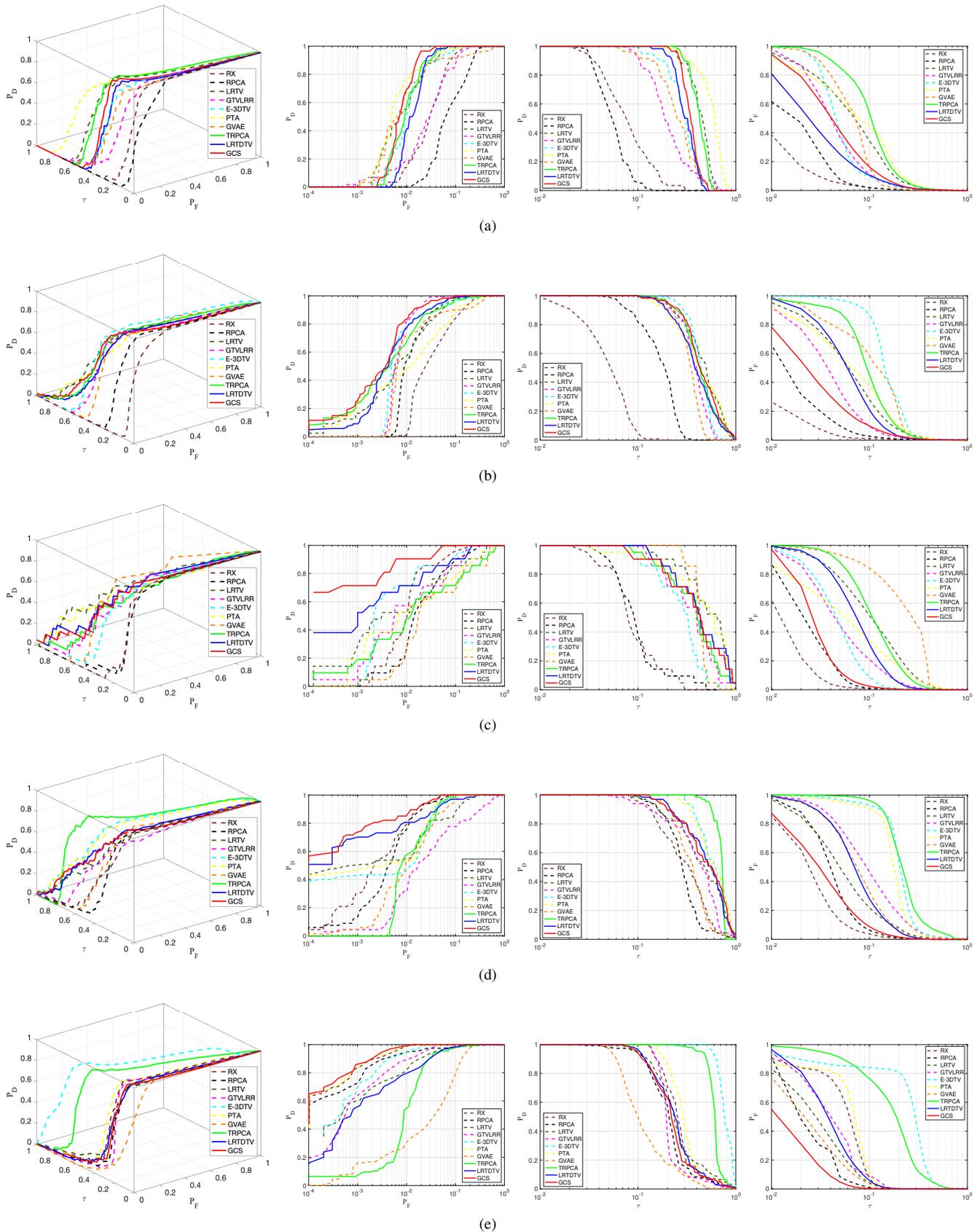


Fig. 8: The ROC curves of different comparison methods on the five datasets. (a) AVIRIS-1. (b) AVIRIS-2. (c) HYDICE. (d) Abu-urban-1. (e) Abu-urban-2. (Left to right) 3D-ROC curve, 2D-ROC curve of  $(P_D, P_F)$ , 2D-ROC curve of  $(P_D, \tau)$ , and 2D-ROC curve of  $(P_F, \tau)$ .

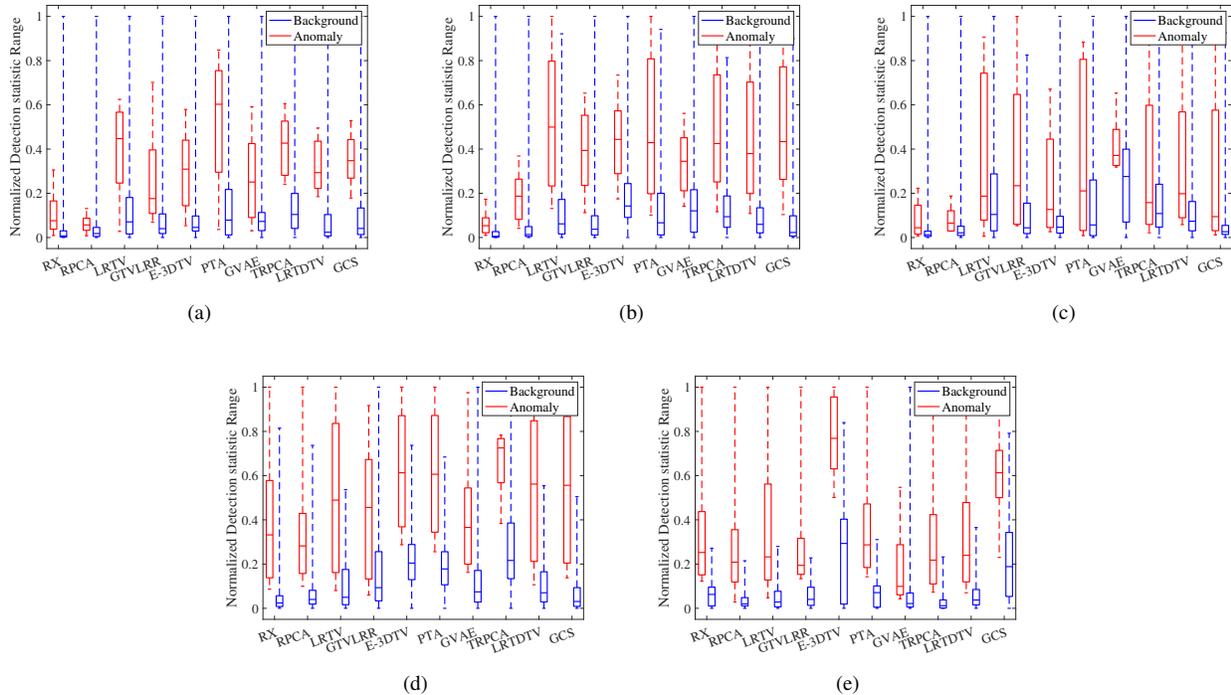


Fig. 9: Box and whisker plots of the different methods under comparison for the five real datasets: (a) AVIRIS-1. (b) AVIRIS-2. (c) HYDICE. (d) Abu-Urban-1. (e) Abu-Urban-2.

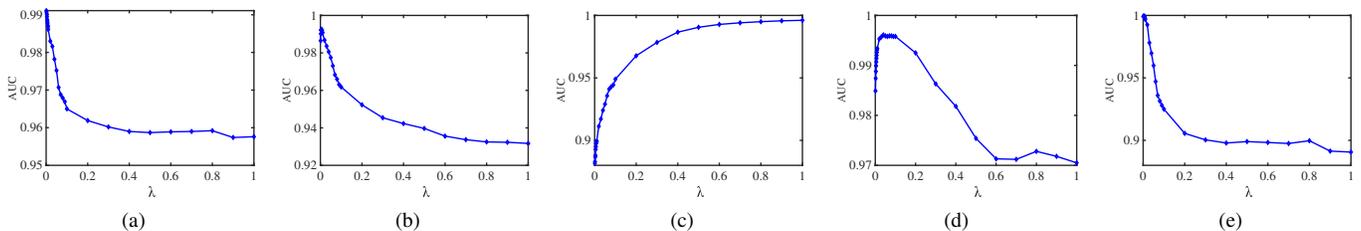


Fig. 10: The AUC values vary with the different  $\lambda$  on the five datasets, respectively. (a) AVIRIS-1. (b) AVIRIS-2. (c) HYDICE. (d) Abu-urban-1. (e) Abu-urban-2.

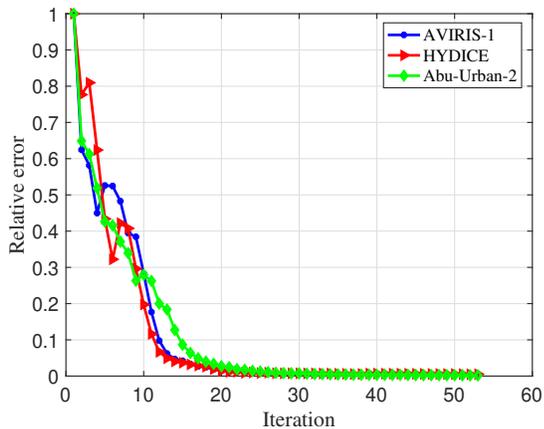


Fig. 11: Convergence error curves of proposed method on the AVIRIS-1, HYDICE, and Abu-urban-2 datasets.

the box and whisker plots of the different methods under comparison for the five real datasets aiming to demonstrate their efficiency in separating the anomalies from the background. Beyond the AUC values metrics mentioned above, TABLE I also lists the  $AUC_{(OD)}$  value to account for the performance of the ten methods comprehensively. Moreover, TABLE II compares the computation times(s) of the ten algorithms on the five datasets. It needs to figure out the running time of GVAE is the Training time(s) + Test time(s).

Here, we give the AVIRIS-2 as an example to demonstrate the superiority of the proposed GCS method.

1) *AVIRIS-2*: When we set the Tucker decomposition rank of the gradient tensors as [35,35,6], the proposed GCS achieved the best performance.

As Fig. 2 shows, the anomalies are the three airplanes. The 2D plot detection maps obtained by ten methods can be observed in Fig. 4. Referring to the color bar in Fig. 4 (i), the proposed GCS aims to detect the three airplanes accurately.

TABLE I: The AUC values obtained by different AD algorithms on the five real datasets.

Datasets	AUC Values	Matrix-based operations						DL	Tensor-based operations		
		RX	RPCA	LRTV	GTVLRR	E-3DTV	PTA	GVAE	TRPCA	LRTDTV	GCS
AVIRIS-1	$AUC_{(P_D, P_F)} \uparrow$	0.9638	0.8935	0.9484	0.9583	0.9834	0.9876	0.9509	0.9872	0.9843	<b>0.9909</b>
	$AUC_{(P_D, \tau)} \uparrow$	0.0926	0.0585	0.4027	0.2328	0.3086	<b>0.5772</b>	0.3423	0.4249	0.3152	0.3512
	$AUC_{(P_F, \tau)} \downarrow$	<b>0.0172</b>	0.0255	0.1024	0.0574	0.0608	0.1080	0.1278	0.1187	0.0472	0.0625
	$AUC_{(OD)} \uparrow$	1.0392	0.9265	1.2486	1.1337	1.2312	<b>1.4568</b>	1.1654	1.2934	1.2523	1.2796
AVIRIS-2	$AUC_{(P_D, P_F)} \uparrow$	0.9488	0.9824	0.9818	0.9913	0.9830	0.9831	0.9596	0.9590	0.9890	<b>0.9928</b>
	$AUC_{(P_D, \tau)} \uparrow$	0.0560	0.1797	0.5034	0.3908	0.4420	0.4756	0.3423	0.4622	0.4319	<b>0.4818</b>
	$AUC_{(P_F, \tau)} \downarrow$	<b>0.0127</b>	0.0265	0.0855	0.0530	0.1614	0.0952	0.1278	0.1100	0.0736	0.0436
	$AUC_{(OD)} \uparrow$	0.9921	1.1356	1.3997	1.3291	1.2636	1.3400	1.1734	1.3353	1.3473	<b>1.4310</b>
HYDICE	$AUC_{(P_D, P_F)} \uparrow$	0.9768	0.9591	0.9421	0.9659	0.9773	0.9221	0.8981	0.9029	0.9725	<b>0.9957</b>
	$AUC_{(P_D, \tau)} \uparrow$	0.1053	0.1235	0.5720	0.4038	0.3049	<b>0.5396</b>	0.4306	0.4327	0.4933	0.4675
	$AUC_{(P_F, \tau)} \downarrow$	<b>0.0175</b>	0.0293	0.1380	0.0704	0.0555	0.1027	0.2567	0.1311	0.0887	0.0335
	$AUC_{(OD)} \uparrow$	1.0646	1.0533	1.3761	1.2993	1.2267	1.3590	1.0720	1.2045	1.3771	<b>1.4297</b>
Abu-urban-1	$AUC_{(P_D, P_F)} \uparrow$	0.9934	0.9916	0.9705	0.9093	0.9794	0.9852	0.9778	0.9823	0.9870	<b>0.9961</b>
	$AUC_{(P_D, \tau)} \uparrow$	0.3442	0.3117	0.5098	0.4403	0.6078	0.6130	0.3729	<b>0.6956</b>	0.5432	0.5494
	$AUC_{(P_F, \tau)} \downarrow$	<b>0.0329</b>	0.0496	0.0766	0.1261	0.2374	0.1869	0.0925	0.2450	0.0868	0.0451
	$AUC_{(OD)} \uparrow$	1.3047	1.2537	1.4038	1.2235	1.3498	1.4113	1.2582	1.4329	1.4434	<b>1.5004</b>
Abu-urban-2	$AUC_{(P_D, P_F)} \uparrow$	0.9946	0.9960	0.9914	0.9967	0.9382	0.9992	0.9828	0.9456	0.9904	<b>0.9994</b>
	$AUC_{(P_D, \tau)} \uparrow$	0.2880	0.2347	0.2481	0.2295	<b>0.8871</b>	0.3213	0.1400	0.6048	0.2668	0.2545
	$AUC_{(P_F, \tau)} \downarrow$	0.0611	0.0283	0.0239	0.0487	0.5681	0.0672	0.1976	0.0325	0.0459	<b>0.0176</b>
	$AUC_{(OD)} \uparrow$	1.2215	1.2024	1.2224	1.1775	1.2572	1.2533	0.9252	<b>1.5179</b>	1.2113	1.2362

TABLE II: Comparison of computation times(s) on the five real datasets.

Dataset	Matrix-based methods					Deep learning	Tensor-based methods			
	RX	RPCA	LRTV	GTVLRR	E-3DTV	GVAE	TRPCA	LRTDTV	PTA	GCS
AVIRIS-1	0.490	6.786	34.916	333.311	28.332	$0.40 \times 10000 + 0.569$	153.577	37.720	20.017	133.745
AVIRIS-2	0.395	6.328	30.449	363.714	25.799	$0.36 \times 10000 + 0.575$	153.443	27.182	19.708	147.349
HYDICE	0.741	6.668	22.410	279.832	20.550	$0.33 \times 10000 + 0.385$	113.790	21.967	15.810	111.197
Abu-urban-1	0.448	8.002	44.171	364.148	29.377	$0.68 \times 10000 + 0.425$	181.422	33.803	22.041	126.255
Abu-urban-2	0.719	8.938	43.772	362.679	31.215	$0.68 \times 10000 + 0.518$	189.801	31.619	22.147	129.852

As shown in Fig. 8(b), the ROC curve ( $P_D, P_F$ ) located in the upper left corner, and the corresponding  $AUC_{(P_D, P_F)}$  in the TABLE I further prove its superiority AD performance. Under the evaluation of the target preservation and the background suppression performance, the  $AUC_{(OD)}$  and the 3D ROC curve ( $P_D, P_F, \tau$ ) of the GCS comprehensively demonstrate an excellent overall performance when compared with the other nine comparison methods. The performance also proved that exploiting the property of gradient tensors, such as low rankness and sparseness, contributes to an efficient AD method. The AUC values in the TABLE I show that the tensor-based methods, including GCS, LRTDTV, TRPCA, and PTA, are competitive, thanks to the tensor that preserve the spatial and spectral structure information. Whereas in TABLE II, it can be observed that the computation time of TRPCA is higher than the GCS, PTA costs a low time due to its matrix-based operation. As can be observed in Fig. 4(g) and (i), the background suppression of the background is better than the

GCS, while the anomaly detection and the noise removal are both inferior to the GCS, the AUC values in the TABLE I also demonstrate it. In TABLE II, the running time of the LRTDTV is shorter than the GCS, whereas, Fig. 4(i), Fig. 4(j), and the TABLE I the detection result of the GCS is better than the LRTDTV since GCS prior describes the sparsity of the core tensor of the gradient map under Tucker decomposition that fuses spatial information to fully deliver the information of tensors in different dimensions. Fig. 4 and the TABLE I from the quality and quantity illustrate the efficiency of the proposed GCS method in anomaly detection, background suppression, and noise removal. The LRTV, GTVLRR, E-3DTV, PTA, and GCS impose the TV regularization on the background; in Fig. 4, it can be observed that the TV term is good at smoothing away the noise and then strength the outlier between the background and the anomaly. As the Fig. 4, Fig. 8, TABLE I, and TABLE II show GVAE, a SOTA deep learning method for AD, presents poor generalization

performance, while GCS based on model modeling can better characterize the background and thus have better AD results. In Fig. 9 (b), the red box represents anomalous targets, while the blue one represents the background pixels. It is concluded that among all the comparison methods, the interval between the red box and the blue box of our proposed GCS is larger than the other nine algorithms, which indicates that the GCS achieves the best separability of anomaly from the background.

2) *AVIRIS-1*: The best detection result was achieved when the tucker decomposition rank of the gradient tensors is [35, 35, 7]. In reference to the color bar, the anomaly airplane can be clearly observed in Fig. 3(i) of the PTA method, which indicates good anomaly preservation. Otherwise, the background also clearly appeared that refer to a poor background suppression performance. In our proposed GCS method, not only the anomaly airplane can be clearly seen in Fig. 3(j), but the background is well suppressed. Besides, as shown in Fig. 8(a) and TABLE I, the ROC curve  $(P_D, P_F)$  of PTA located on the top left corner with the highest  $AUC_{(P_D, P_F)}$  value achieved, and the ROC curve  $(P_D, \tau)$  of GCS is closest to the up right corner and obtained a best  $AUC_{(P_D, \tau)}$  value. The phenomenon demonstrates that underlying the low rank and sparse assumption, imposing the SSTV regularization term on the background efficiently smoothes away the sparse anomaly, yielding a superior performance of GCS. As can be observed in Fig. 4(h) and (j), the background suppression of the background is better than the GCS, while the anomaly detection and the noise removal are both inferior to the GCS, the AUC values in the TABLE I also demonstrate it. LRTV also behaves a good performance in terms of the  $AUC_{(OD)}$ , the low computation time in TABLE II due to it being a matrix-based operation. GVAE achieves a general performance, and the test time is short while the training time is high. As shown in Fig. 9(a), the interval between the red and blue boxes of LRTV, E-3DTV, PTA, TRPCA, LRTDTV, and GCS are better than the RX, RPCA, GTVLRR, GVAE, which indicate that former methods separating the anomalies from the background are better than the latter ones. The interval between our proposed GCS's red and blue boxes is larger than the LRTV, E-3DTV, PTA, TRPCA, and LRTDTV, further demonstrating that the GCS performs the best separation performance.

3) *HYDICE*: The best performance of the proposed GCS was achieved when we set the tucker decomposition rank of the gradient tensors as [70, 70, 5]. Considering Fig. 2 the ground truth as a comparison, the 2D plot detection map of the proposed GCS in Fig. 5(j) exhibits an accurate AD result. As shown in Fig. 8, the location of 3D ROC curve and three 2D ROC curves  $(P_D, P_F)$ ,  $(P_D, \tau)$ ,  $(P_F, \tau)$  reflect that the GCS perform an excellent performance on the overall detection, anomaly preservation, and background suppression. Meanwhile, the TABLE I further proves its efficiency by computing corresponding AUC values. Even though the PTA has the highest  $AUC_{(OD)}$  among all the matrix-based operations, the  $(P_D, P_F)$ ,  $(P_F, \tau)$  curves and values in Fig. 5(i) and TABLE I shows it's at too low a level among all the competitive methods. The performance of the GVAE method on the HYDICE dataset is inferior to the other comparison methods in view of quality and quantity. In addition, as shown in Fig. 9(c), the RX, RPCA,

and proposed GCS achieve the best ability to separate anomalies from the background than other algorithms. Whereas, the ROC curves shown in Fig. 8 and the corresponding auc result in TABLE I confirm that the GCS yields the best performance.

4) *Abu-Urban-1*: When we set the tucker decomposition rank of the gradient tensors as [30, 30, 7], the best AD performance is achieved. In Fig. 6(j), the anomalies are clearly detected, and the background is better suppressed. It also can be observed in Fig. 8 that the  $(P_D, P_F)$  located in the upper left corner, yet the  $(P_F, \tau)$  curve close to lower left corner. In the TABLE I, the  $AUC_{(P_D, P_F)}$  and  $AUC_{(P_F, \tau)}$  also prove its efficiency, it worth to note that the  $AUC_{(OD)}$  comprehensively demonstrates the GCS performed an excellent performance. Beyond that, both PTA and TRPCA achieved a high  $AUC_{(OD)}$ , whereas, their overall detection performance and background suppression are inferior to the GCS. The LRTDTV method is competitive with our proposed method comprehensively, yet all of the four AUC measures for AD performance are not as good as the GCS. In addition, in TABLE I, the  $AUC_{(P_D, P_F)}$  of RX and RPCA are very high, nevertheless, the low values for target preservation result in an unsatisfactory AD result. As shown in Fig. 9(d), the interval between the red box and blue box of TRPCA is larger than GCS, whereas, the detection results in TABLE I demonstrate that our proposed method is the best algorithm.

5) *Abu-Urban-2*: When we set the tucker decomposition rank of the gradient tensors as [20, 20, 7], the best AD performance is achieved. As shown in Fig. 7(j) and TABLE I, the  $AUC_{(OD)}$  of TRPCA is the highest, yet the value of the  $(P_D, P_F)$  is inferior to the others. Besides, the computation times of TRPCA are much higher than the average time. In TABLE I, even though the  $AUC_{(OD)}$  values of PTA are not the best, the performance of the PTA in terms of the other three 2D ROC values is good. As observed in the Fig. 7(i), the background is well suppressed, the  $AUC_{(P_D, P_F)}$  is highest and the  $AUC_{(P_D, P_F)}$  shown in TABLE I verify the phenomenon. Whereas, the high background suppression may result in the anomaly being detected as the background, which generates inferior target preservation and an overall performance evaluation. Compared to the LRTDTV with the GCS, as shown in Fig. 7(h) and (j), the background of GCS is very clear in Fig. 8. Besides, the  $(P_F, \tau)$  curve and value in Fig. 8(e) reflects the excellent performance of GCS for background suppressing. As shown in Fig. 9(d), the separability of E-3DTV is comparable with our proposed GCS, whereas, the  $AUC_{(P_D, P_F)}$  of GCS is much better than E-3DTV.

#### D. Discussion

We perform the proposed GCS method and nine comparison methods on the five datasets and summarize the efficiency of the proposed method as follows:

1) *Effectiveness*: In this paper, we observe a new prior named GCS prior. Different from the traditional methods that characterize the low-rank and local smoothness priors separately, the GCS can characterize the low-rank and local smoothness of tensor data at the same time, which allows us to use one regularization term to efficiently describe low-rank and local smoothness. Besides, with the help of the

physical meaning of the core tensor, the GCS prior can also deliver the representation sparsity under Tucker decomposition. As shown in Fig. 3-Fig. 7 and TABLE I, the proposed method underlying five datasets performs the best background suppression performance, which demonstrates the efficiency of background modeling.

2) *Performance*: Nine comparison algorithms are presented to sufficiently demonstrate the proposed method's performance. Unlike the low-rank and local smoothness priors that have been discovered before, such as LRTV, GTVLR, LRTDTV, which characterize these two priors separately; the observed GCS prior can better characterize the low rank and local smoothness of the background with one regularizer. Even though the E-3DTV can encode the two priors under one regularization term, the matrix-based method unavoidably destroys the spatial or spectral information. Our proposed method encodes the GCS prior in a tensor manner, preserving the inner structure. PTA is a tensor-based but matrix-based operation method that is competitive with the proposed method, but its performance cannot be generalized to all datasets. The deep learning method GVAE has a low test time, whereas the training time assumption is high, and the AD performance is general.

3) *Parameter Analysis*: We discuss the robustness of the parameter  $\lambda$  referring to the proposed GCS method on the five datasets. Fig. 10 reflects the sensitivity of the overall detection performance within the scope of  $\lambda$ . It shows that the best performance is achieved on the five datasets when  $\lambda$  setting as 0.001, 0.006, 1, 0.05, and 0.003, respectively. Fig. 10 (a), (b), (d), and (e) with a similar trend that is the curve first reaching the peak and then, downward to stable. Whereas the curve in Fig. 10 (c) keeps going up, and then it tends to be stable.

4) *Convergence Analysis*: Fig. 11 displays the relative errors versus the iteration numbers on the AVIRIS-1, HYDICE, and ABU-Urban-2 three datasets. It can be observed from the curves that when the number of iteration numbers increases from 1 to 20, the corresponding relative errors go down very sharply. After that, the downward trend of the curves becomes stable, and relative error values approach zero.

## VI. CONCLUSION

In this paper, we find a new prior named GCS prior. GCS prior describes the sparsity of core tensor of gradient map under tucker decomposition, which is the kind of representation sparsity that fuses spatial information so that the information of tensors in different dimensions can be more fully delivered. Different from the traditional method that utilizes the low rank and smoothness priors separately, here, our observed prior can fuse the two priors and be encoded with one regularizer, which is general and maximizes their effectiveness. To further separate the background from the anomaly, the anomaly sparsity is encoded by the  $\ell_1$ -norm. The experimental results on the five real hyperspectral image datasets show that the proposed method can effectively suppress the background and highlight the target. The excellent performance of the proposed method also verifies the effectiveness and robustness compared with other methods in the view of quality and quantity.

## REFERENCES

- [1] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. PP, no. 99, pp. 2–28, 2021.
- [2] S. Sawant, Y. Patole, C. Sawant, and D. Kumawat, "Food safety inspection and control using hyperspectral imaging: A review," 2020.
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- [4] Willett, R., Duarte, M., Baraniuk, and Davenport, "Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 116–126, 2014.
- [5] Y. Qu, W. Wang, R. Guo, B. Ayhan, C. Kwan, S. Vance, and H. Qi, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–15, 2018.
- [6] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 391–406, 2019.
- [7] T. Cheng and B. Wang, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–15, 2020.
- [8] Y. Xu, L. Zhang, B. Du, and L. Zhang, "Hyperspectral anomaly detection based on machine learning: An overview," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3351–3364, 2022.
- [9] J. Wei, J. Zhang, Y. Xu, L. Xu, Z. Wu, and Z. Wei, "Hyperspectral anomaly detection based on graph regularized variational autoencoder," *IEEE Geoscience and Remote Sensing Letters*, 2022.
- [10] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–12, 2020.
- [11] L. Zhang, B. Cheng, S. Tan, and Y. Wang, "Fractional fourier transform based joint adaptive subspace detection for hyperspectral anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 19.
- [12] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2523–2533, 2015.
- [13] W. Shang, M. Jouni, Z. Wu, Y. Xu, M. Dalla Mura, and Z. Wei, "Hyperspectral anomaly detection based on regularized background abundance tensor decomposition," *Remote Sensing*, vol. 15, no. 6, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/6/1679>
- [14] N. Ma, Y. Peng, S. Wang, and D. Liu, "Hyperspectral image anomaly targets detection with online deep learning," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2018, pp. 1–6.
- [15] S. Song, H. Zhou, Y. Yang, and J. Song, "Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3637–3649, 2019.
- [16] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "Unsupervised pixel-wise hyperspectral anomaly detection via autoencoding adversarial networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [17] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4666–4679, 2020.
- [18] W. Xie, B. Liu, Y. Li, J. Lei, C.-I. Chang, and G. He, "Spectral adversarial feature learning for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2352–2365, 2019.
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [20] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4589–4604, 2017.
- [21] S. Chang, B. Du, and L. Zhang, "A sparse autoencoder based hyperspectral anomaly detection algorithm using residual of reconstruction error," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5488–5491.

- [22] J. Zhang, Y. Xu, T. Zhan, Z. Wu, and Z. Wei, "Anomaly detection in hyperspectral image using 3d-convolutional variational autoencoder," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 2512–2515.
- [23] D. Zhu, B. Du, and L. Zhang, "Edlad: An encoder-decoder long short-term memory network-based anomaly detector for hyperspectral images," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 4412–4415.
- [24] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7406–7417, 2020.
- [25] W. Xie, B. Liu, Y. Li, J. Lei, and Q. Du, "Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5416–5427, 2020.
- [26] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1527–1537, 2019.
- [27] 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 5224–5236, 2020.
- [28] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, "Analysis and optimizations of global and local versions of the rx algorithm for anomaly detection in hyperspectral data," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 6, no. 2, pp. 801–814, 2013.
- [29] D. Borghys, I. Käsen, V. Achard, and C. Perneel, "Comparative evaluation of hyperspectral anomaly detectors in different types of background," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, vol. 8390. International Society for Optics and Photonics, 2012, p. 83902J.
- [30] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.
- [31] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Transactions on geoscience and remote sensing*, vol. 53, no. 3, pp. 1463–1474, 2014.
- [32] Z. Hou, W. Li, R. Tao, P. Ma, and W. Shi, "Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection," *Science China Information Sciences*, vol. 65, no. 1, pp. 1–12, 2022.
- [33] W. Yao, L. Li, H. Ni, W. Li, and R. Tao, "Hyperspectral anomaly detection based on improved rpca with non-convex regularization," *Remote Sensing*, vol. 14, no. 6, p. 1343, 2022.
- [34] L. T. Thanh, K. Abed-Meraim, N. L. Trung, and A. Hafiane, "Robust tensor tracking with missing data and outliers: Novel adaptive cp decomposition and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4305–4320, 2022.
- [35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [36] X. Song, L. Zou, and L. Wu, "Detection of subpixel targets on hyperspectral remote sensing imagery based on background endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2365–2377, 2020.
- [37] C. Zhao, C. Li, S. Feng, and X. Jia, "Enhanced total variation regularized representation model with endmember background dictionary for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.
- [38] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2018.
- [39] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6916–6930, 2022.
- [40] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5174–5189, 2019.
- [41] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3071–3084, 2019.
- [42] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6017–6028, 2020.
- [43] E. Bati, A. Çalışkan, A. Koz, and A. A. Alatan, "Hyperspectral anomaly detection method based on auto-encoder," in *Image and Signal Processing for Remote Sensing XXI*, vol. 9643. Spie, 2015, pp. 220–226.
- [44] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [45] H. Wang, J. Peng, W. Qin, J. Wang, and D. Meng, "Guaranteed tensor recovery fused low-rankness and smoothness," *CoRR*, vol. abs/2302.02155, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2302.02155>
- [46] 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. 13 887–13 901, 2022. [Online]. Available: <https://doi.org/10.1109/TCYB.2021.3140148>
- [47] J. Peng, Q. Xie, Q. Zhao, Y. Wang, L. Yee, and D. Meng, "Enhanced 3dtv regularization and its applications on hsi denoising and compressed sensing," *IEEE Transactions on Image Processing*, vol. 29, pp. 7889–7903, 2020.
- [48] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5766–5781, 2023.
- [50] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1888–1902, 2018.
- [51] W. Shang, M. Jouni, Z. Wu, Y. Xu, M. D. Mura, and Z. Wei, "Regularized tensor representative coefficient model for hyperspectral target detection," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1, 2023.
- [52] W. Shang, Z. Wu, J. Wei, Y. Xu, L. Qian, Z. Wei, J. Chanussot, and A. L. Bertozzi, "Gas plume detection in hyperspectral video sequence using tensor nuclear norm," in *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Sep. 2018, pp. 1–4.
- [53] G. Zhang, R. Heusdens, and W. B. Kleijn, "On the convergence rate of the bi-alternating direction method of multipliers," in *IEEE International Conference on Acoustics*, 2014.
- [54] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *Siam Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [55] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. P.613–627, 1995.
- [56] M. K. Ng, R. H. Chan, and W. C. Tang, "A fast algorithm for deblurring models with neumann boundary conditions," *SIAM Journal on Scientific Computing*, 1999.
- [57] D. Krishnan and R. Fergus, "Fast image deconvolution using hyper-laplacian priors," in *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, 2009.
- [58] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE transactions on acoustics, speech, and signal processing*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [59] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.
- [60] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2016.
- [61] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.
- [62] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using mahalanobis distance-regularized tensor rpca," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2919–2930, 2018.
- [63] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3d receiver operating characteristic curve analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5131–5153, 2020.

## REFERENCES

- [1] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. PP, no. 99, pp. 2–28, 2021.
- [2] S. Sawant, Y. Patole, C. Sawant, and D. Kumawat, "Food safety inspection and control using hyperspectral imaging: A review," 2020.
- [3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.
- [4] Willett, R., Duarte, M., Baraniuk, and Davenport, "Sparsity and structure in hyperspectral imaging : Sensing, reconstruction, and target detection," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 116–126, 2014.
- [5] Y. Qu, W. Wang, R. Guo, B. Ayhan, C. Kwan, S. Vance, and H. Qi, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–15, 2018.
- [6] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 391–406, 2019.
- [7] T. Cheng and B. Wang, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–15, 2020.
- [8] Y. Xu, L. Zhang, B. Du, and L. Zhang, "Hyperspectral anomaly detection based on machine learning: An overview," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 15, pp. 3351–3364, 2022.
- [9] J. Wei, J. Zhang, Y. Xu, L. Xu, Z. Wu, and Z. Wei, "Hyperspectral anomaly detection based on graph regularized variational autoencoder," *IEEE Geoscience and Remote Sensing Letters*, 2022.
- [10] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. PP, no. 99, pp. 1–12, 2020.
- [11] L. Zhang, B. Cheng, S. Tan, and Y. Wang, "Fractional fourier transform based joint adaptive subspace detection for hyperspectral anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 19.
- [12] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2523–2533, 2015.
- [13] W. Shang, M. Jouni, Z. Wu, Y. Xu, M. Dalla Mura, and Z. Wei, "Hyperspectral anomaly detection based on regularized background abundance tensor decomposition," *Remote Sensing*, vol. 15, no. 6, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/6/1679>
- [14] N. Ma, Y. Peng, S. Wang, and D. Liu, "Hyperspectral image anomaly targets detection with online deep learning," in *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2018, pp. 1–6.
- [15] S. Song, H. Zhou, Y. Yang, and J. Song, "Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, pp. 3637–3649, 2019.
- [16] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "Unsupervised pixel-wise hyperspectral anomaly detection via autoencoding adversarial networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2021.
- [17] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 4666–4679, 2020.
- [18] W. Xie, B. Liu, Y. Li, J. Lei, C.-I. Chang, and G. He, "Spectral adversarial feature learning for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2352–2365, 2019.
- [19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [20] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 10, pp. 4589–4604, 2017.
- [21] S. Chang, B. Du, and L. Zhang, "A sparse autoencoder based hyperspectral anomaly detection algorithm using residual of reconstruction error," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5488–5491.
- [22] J. Zhang, Y. Xu, T. Zhan, Z. Wu, and Z. Wei, "Anomaly detection in hyperspectral image using 3d-convolutional variational autoencoder," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 2512–2515.
- [23] D. Zhu, B. Du, and L. Zhang, "Edlad: An encoder-decoder long short-term memory network-based anomaly detector for hyperspectral images," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*. IEEE, 2021, pp. 4412–4415.
- [24] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, pp. 7406–7417, 2020.
- [25] W. Xie, B. Liu, Y. Li, J. Lei, and Q. Du, "Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5416–5427, 2020.
- [26] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1527–1537, 2019.
- [27] 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 Transactions on Geoscience and Remote Sensing*, vol. 58, no. 7, pp. 5224–5236, 2020.
- [28] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, "Analysis and optimizations of global and local versions of the rx algorithm for anomaly detection in hyperspectral data," *IEEE journal of selected topics in applied earth observations and remote sensing*, vol. 6, no. 2, pp. 801–814, 2013.
- [29] D. Borghys, I. Käsen, V. Achard, and C. Perneel, "Comparative evaluation of hyperspectral anomaly detectors in different types of background," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, vol. 8390. International Society for Optics and Photonics, 2012, p. 83902J.
- [30] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.
- [31] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Transactions on geoscience and remote sensing*, vol. 53, no. 3, pp. 1463–1474, 2014.
- [32] Z. Hou, W. Li, R. Tao, P. Ma, and W. Shi, "Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection," *Science China Information Sciences*, vol. 65, no. 1, pp. 1–12, 2022.
- [33] W. Yao, L. Li, H. Ni, W. Li, and R. Tao, "Hyperspectral anomaly detection based on improved rpca with non-convex regularization," *Remote Sensing*, vol. 14, no. 6, p. 1343, 2022.
- [34] L. T. Thanh, K. Abed-Meraim, N. L. Trung, and A. Hafiane, "Robust tensor tracking with missing data and outliers: Novel adaptive cp decomposition and convergence analysis," *IEEE Transactions on Signal Processing*, vol. 70, pp. 4305–4320, 2022.
- [35] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [36] X. Song, L. Zou, and L. Wu, "Detection of subpixel targets on hyperspectral remote sensing imagery based on background endmember extraction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2365–2377, 2020.
- [37] C. Zhao, C. Li, S. Feng, and X. Jia, "Enhanced total variation regularized representation model with endmember background dictionary for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.
- [38] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 4, pp. 1227–1243, 2018.
- [39] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 11, pp. 6916–6930, 2022.
- [40] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5174–5189, 2019.
- [41] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 5, pp. 3071–3084, 2019.
- [42] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network

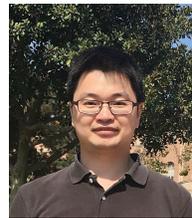
- for hyperspectral anomaly detection,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6017–6028, 2020.
- [43] E. Bati, A. Çalışkan, A. Koz, and A. A. Alatan, “Hyperspectral anomaly detection method based on auto-encoder,” in *Image and Signal Processing for Remote Sensing XXI*, vol. 9643. Spie, 2015, pp. 220–226.
- [44] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [45] H. Wang, J. Peng, W. Qin, J. Wang, and D. Meng, “Guaranteed tensor recovery fused low-rankness and smoothness,” *CoRR*, vol. abs/2302.02155, 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2302.02155>
- [46] 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. 13 887–13 901, 2022. [Online]. Available: <https://doi.org/10.1109/TCYB.2021.3140148>
- [47] J. Peng, Q. Xie, Q. Zhao, Y. Wang, L. Yee, and D. Meng, “Enhanced 3d tv regularization and its applications on hsi denoising and compressed sensing,” *IEEE Transactions on Image Processing*, vol. 29, pp. 7889–7903, 2020.
- [48] 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 Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5766–5781, 2023.
- [50] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, “Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1888–1902, 2018.
- [51] W. Shang, M. Jouni, Z. Wu, Y. Xu, M. D. Mura, and Z. Wei, “Regularized tensor representative coefficient model for hyperspectral target detection,” *IEEE Geoscience and Remote Sensing Letters*, pp. 1–1, 2023.
- [52] W. Shang, Z. Wu, J. Wei, Y. Xu, L. Qian, Z. Wei, J. Chanussot, and A. L. Bertozzi, “Gas plume detection in hyperspectral video sequence using tensor nuclear norm,” in *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Sep. 2018, pp. 1–4.
- [53] G. Zhang, R. Heusdens, and W. B. Kleijn, “On the convergence rate of the bi-alternating direction method of multipliers,” in *IEEE International Conference on Acoustics*, 2014.
- [54] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *Siam Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [55] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. P.613–627, 1995.
- [56] M. K. Ng, R. H. Chan, and W. C. Tang, “A fast algorithm for deblurring models with neumann boundary conditions,” *SIAM Journal on Scientific Computing*, 1999.
- [57] D. Krishnan and R. Fergus, “Fast image deconvolution using hyper-laplacian priors,” in *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada.*, 2009.
- [58] I. S. Reed and X. Yu, “Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution,” *IEEE transactions on acoustics, speech, and signal processing*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [59] E. J. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.
- [60] W. He, H. Zhang, L. Zhang, and H. Shen, “Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 1, pp. 178–188, 2016.
- [61] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, “Prior-based tensor approximation for anomaly detection in hyperspectral imagery,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2020.
- [62] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, “Joint reconstruction and anomaly detection from compressive hyperspectral images using mahalanobis distance-regularized tensor rpca,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 5, pp. 2919–2930, 2018.
- [63] C.-I. Chang, “An effective evaluation tool for hyperspectral target detection: 3d receiver operating characteristic curve analysis,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 6, pp. 5131–5153, 2020.



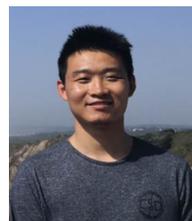
**Wenting Shang** received a BE degree from Zhengzhou University, Henan Province, China, in 2014. She is currently working toward a Ph.D. degree at the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. Her research interests include the areas of hyperspectral image classification, target detection, and anomaly detection.



**Jiangjun Peng** Jiangjun Peng received a BS degree in mathematics from Northwest University, Xi’an, China, in 2015, and an MS degree in mathematics from Xi’an Jiaotong University, Xi’an, China, in 2018. From 2018 to 2020, he worked as a researcher at Tencent in Shenzhen, China. Currently, he is pursuing a Ph.D. degree at the School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an. His current research interests involve low-rank matrix/tensor analysis, theoretical machine learning, and model-driven deep learning.



**Zebin Wu** (Senior Member, IEEE) received B.Sc. and Ph.D. degrees in computer science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2003 and 2007, respectively. He was a Visiting Scholar with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain, from June 2014 to June 2015. He was a Visiting Scholar with the Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, USA, from August 2016 to September 2016 and from July 2017 to August 2017. He was a Visiting Scholar with the GIPSA-Lab, Grenoble INP, the Univ. Grenoble Alpes, Grenoble, France, from August 2018 to September 2018. He is a Professor at the School of Computer Science and Engineering, Nanjing University of Science and Technology. He has authored over 76 publications, including 30 Journal Citation Report (JCR) journal papers (22 in IEEE journals), and over 31 peer-reviewed conference proceedings (25 in IEEE conferences). He has reviewed more than 60 articles for over 15 different journals. His research interests include hyperspectral imaging processing, parallel computing, and remotely sensed big data processing. Dr. Wu served as the Vice Chair for IEEE Geoscience and Remote Sensing Society (GRSS) Nanjing Chapter from 2016–2020. He received the Best Reviewers of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (IEEE JSTARS). He serves as an Associate Editor of IEEE JSTARS and the Chair of IEEE GRSS Nanjing Chapter.



**Yang Xu** (Member, IEEE) received a B.Sc. degree in applied mathematics and a Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2011 and 2016, respectively. He is currently an Associate Professor with the School of Computer Science and Engineering, NUST. His research interests include hyperspectral image classification, hyperspectral detection, image processing, and machine learning.



**Mohamad Jouni** (Member, IEEE) received a B.Eng. degree in Computer and Communications Engineering from the Lebanese University, Beirut, Lebanon, in 2016, and his M.Sc. and Ph.D. degrees in Signal and Image Processing from the University of Grenoble Alpes, Grenoble, France, in 2017 and 2021, respectively. In 2019, he was a visiting researcher for 10 weeks at Tokyo Institute of Technology, Tokyo, Japan. Since 2021, he has been a Postdoctoral Researcher at Grenoble Institute of Technology, Grenoble, France. His research interests include tensor algebra, artificial

intelligence, computational imaging, and applications of multimodal and hyperspectral data analysis.



**Mauro Dalla Mura** (Senior Member, IEEE) received B.Sc. and M.Sc. degrees in telecommunication engineering from the University of Trento, Italy, in 2005 and 2007, respectively, and a joint Ph.D. degree in information and communication technologies (telecommunications area) from the University of Trento, Italy, and in electrical and computer engineering from the University of Iceland, Iceland, in 2011. He was a Research Fellow with Fondazione Bruno Kessler, Trento, Italy, in 2011, and conducted research in the field of computer vision.

Since 2012, he has been an Assistant Professor with Grenoble Institute of Technology (Grenoble INP), France. He is conducting his research at Grenoble Images Speech Signals and Automatics Laboratory (GIPSA-Lab). He has been appointed as a "Specially Appointed Associate Professor" at the School of Computing, Tokyo Institute of Technology, Japan, from 2019-2022. His research interests include remote sensing, image processing, pattern recognition, mathematical morphology, classification, and multivariate data analysis. He was a recipient of the IEEE GRSS Second Prize in the Student Paper Competition of the 2011 IEEE IGARSS 2011 and a corecipient of the Best Paper Award of the International Journal of Image and Data Fusion in 2012-2013 and the Symposium Paper Award for IEEE IGARSS in 2014. Since 2016, he has been the President of the IEEE GRSS French Chapter (he previously served as Secretary from 2013-2016). In 2017, as a member of the IEEE GRSS French Chapter, he was a recipient of the IEEE GRSS Chapter Award and the "Chapter of the year 2017" award from the IEEE French Section. Since 2016, he has been on the Editorial Board of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS).



**Wei Zhihui** was born in Jiangsu, China, in 1963. He received B.Sc. and M.Sc. degrees in applied mathematics and a Ph.D. degree in communication and information systems from Southeast University, Nanjing, China, in 1983, 1986, and 2003, respectively. He is a Professor and a Doctoral Supervisor with the Nanjing University of Science and Technology (NUST), Nanjing. His research interests include inverse problems in image processing, machine learning, intelligent information processing, and scientific computing.