



**HAL**  
open science

# Hyperspectral Anomaly Detection Based on Regularized Tensor Decomposition

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

► **To cite this version:**

Wenting Shang, Mohamad Jouni, Zebin Wu, Yang Xu, Mauro Dalla Mura, et al.. Hyperspectral Anomaly Detection Based on Regularized Tensor Decomposition. *Remote Sensing*, 2023, 15 (6), pp.1679. 10.3390/rs15061679. hal-03806912

**HAL Id: hal-03806912**

**<https://hal.science/hal-03806912v1>**

Submitted on 8 Oct 2022

**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 Based on Regularized Tensor Decomposition

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

**Abstract**—The tensor-based anomaly detection (AD) model has attracted increasing interest in the hyperspectral image (HSI) community. Since it is powerful in maintaining spatial and spectral structures, an HSI is essentially a third-order tensor. In this article, we propose a novel AD method based on a low-rank background linear mixing model of the scene background. The obtained abundance maps possess more distinctive features than the raw data, which is beneficial for identifying an anomaly from the background. Specifically, the low-rank tensor background is approximated as the mode-3 product of an abundance tensor and endmember matrix. Due to the spatial sparse and smooth natures of abundance maps, the  $\ell_1$ -norm is introduced to enforce sparseness, and the total variation (TV) regularizer is adopted to encourage spatial smoothness. Moreover, the typical great correlation among abundance vectors implies the low-rank structure of abundance maps. Compared with the rigorous low-rank constraint, a soft low-rank regularization is imposed on the background in order to leverage its spatial homogeneity. Its strictness is controlled by scalar parameters. For the anomaly part, the anomaly spectra account for a small part of the whole scene, and therefore, an  $\ell_{1,1,2}$ -norm is defined to characterize its tube-wise sparsity. Notably, Gaussian noise is integrated into the model to suppress confusion with the anomaly. The experimental results on five real datasets demonstrate the outstanding performance of our proposed method.

**Index Terms**—Anomaly detection, low-rank background linear mixing model, mode-3 product, third-order abundance tensor, TV regularizer,  $\ell_{1,1,2}$ -norm.

## I. INTRODUCTION

**D**IFFERENT from multispectral sensors, hyperspectral sensors enable hundreds of narrow and contiguous bands to be obtained to characterize each pixel in a real scene. They have attracted a great deal of attention from researchers in signal and image processing fields, as hyperspectral images (HSIs) contain valuable spatial and spectral information. Over the past decades, hyperspectral images have been applied in

various areas [1], such as data fusion, unmixing, classification, and target detection [2]–[7].

Anomaly detection (AD) aims to detect interests and to label each pixel as a target or background, such as natural materials, man-made targets, and other interferers. This is one of the most important research fields in HSI processing [8]. Specifically, it is assumed that the spectral signatures of an anomaly are different from those of their surrounding neighbors, which makes it feasible to distinguish anomalies from the background. In essence, AD is an unsupervised classification problem without prior knowledge of the anomaly or background, leading to an incredibly challenging detection task.

To date, a large number of AD methods have been proposed for HSIs, grouped into three main categories: deep learning-based, statistical-based, and geometric modeling-based methods. Deep learning-based AD algorithms usually use deep neural networks to mine deep features of spectra in HSIs. Since no prior information can be exploited in AD, many researchers train the deep model in an unsupervised way for AD tasks [9]–[11]. However, the deep model for off-the-shelf HSI still takes a lot of time to train while exhibiting limited generalization ability. Reed-Xiaoli (RX) [12], a benchmark statistical-based method, assumes that the background follows a Gaussian distribution, and accomplishes AD via the generalized likelihood ratio test (GLRT). In particular, the Mahalanobis distance between a pixel being tested and the surrounding background is calculated to determine whether the pixel is an anomaly. Specifically, many RX-based methods have been developed in [13]–[18]. For example, the global RX (GRX) [13] and local RX (LRX) [14] methods are two typical versions, and they estimate the background statistical variables using the entire image or the surrounding neighbors, respectively. The above methods suffer the limitation of relying on the assumption that the background obeys a single distribution, which is difficult to satisfy in real hyperspectral scenes. Thereafter, the kernel-RX algorithm [15], [16] employs a kernel function and maps the data into a higher-dimensional feature space to characterize non-Gaussian distributions. Geometric modeling-based methods [19]–[28] are another category of AD methods. Representation-based methods [19]–[21] have been successfully applied to AD because they do not need a specific distribution assumption, but they fail to capitalize on the high spectral correlation of HSIs. Afterward, robust principal component analysis (RPCA) assumes that the background is represented by a single subspace aiming to recover low-rank background and separate a sparse anomaly from an

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; xuyangh90@njust.edu.cn; wuzb@njust.edu.cn; gswei@njust.edu.cn).

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)

observed HSI [22], [23]. Nevertheless, the low-rank representation (LRR)-based model assumes that the data lie in multiple subspaces since most pixels are mixed pixels, which yielded the exploration of other methods for HSI AD [24], [25]. An accurate background dictionary construction for LRR is still a challenging task. Thereafter, linear mixing model (LMM) [29]-based methods have attracted considerable attention in AD fields due to their explicit physical descriptions found with background modeling and their ability to avoid advanced background model estimation [26]–[28]. Here, the background mixed pixels can be linearly represented by pure material signatures, which is normally addressed by nonnegative matrix factorization (NMF) [30] and can be approximately written as the product of two nonnegative matrices: an endmember matrix and an abundance matrix. Initially, Qu *et al.* [26] applied spectral unmixing to model original HSI data and regard the obtained abundance maps as the input for LRR. To enhance the HSI spatial structure smoothness, most recently, [28] proposed an enhanced total variation (ETV) model with an endmember background dictionary (EBD) by applying ETV to the row vectors of the representation coefficient matrix. The above-referred methods both achieved promising performance with respect to background modeling using the LMM; whereas, they are matrix-based methods that reshape the 3D HSI into a 2D matrix, and cannot avoid destroying the spatial or the spectral structure of the HSI.

Since an HSI is essentially a cube, the aforementioned matrix-based background modeling fails to explore its intrinsic multidimensional structure [31]. In comparison, the HSI processing based on tensor decomposition can simultaneously preserve the spatial and spectral structure information [32], [33]. Recent developments in tensor-based AD methods have heightened the need for inner structure preservation [34], and exploration of tensor physical characteristics [35]–[37]. The spectral signatures of the background pixels in the homogeneous regions have a high correlation, resulting in the background having a strong spatial linear correlation and therefore admitting a low-rank characteristic. The conventional low-rank tensor decomposition [38] includes the prior-based tensor approximation, Canonical Polyadic (CP) decomposition and Tucker decomposition. Recently, some new tensor decomposition models [39]–[41] are proposed to capture the low-rank structure information in a tensor manner. Li *et al.* [35] proposed a prior-based tensor approximation (PTA) method for hyperspectral AD assuming low rankness of the spectra and piecewise-smoothness in the spatial domain. Note that PTA actually operates on matrices and not tensors, as the hyperspectral cube is unfolded in 2D structures. This does not allow us to preserve the inner structure of the data with these tensor-based approaches. Song *et al.* [36] employs an endmember extraction technology to yield an abundance map for AD. Based on LRR, the dictionary of background is constructed by Tucker decomposition. The endmember extraction methods that are based on tensor have been successfully applied in anomaly detection, whereas they lack the consideration of prior information of the abundance map of backgrounds, such as the spatial smoothness and structure low rankness.

In this paper, motivated by the abundance maps possess-

ing more distinctive features than the raw data, we propose an abundance tensor regularization with low rankness and smoothness based on sparse unmixing (ATLSS) for hyperspectral AD. With proper modeling of the physical conditions, an observed third-order tensor HSI can be decomposed into a *background tensor*, an *anomaly tensor*, and *Gaussian noise*. Motivated by the LMM, the background tensor is approximated as a mode-3 product of an abundance tensor and an endmember matrix. The spectral signatures of the background pixels in the homogeneous regions have a high correlation, which yields a spatial linear correlation and therefore admits a low-rank property. In [42], [43], the authors impose low-rankness on the abundance tensor in order to effectively capture the HSI's low-dimensional structure. In our paper, the abundance tensor is characterized by tensor regularization with low rankness through CP decomposition.

Moreover, each pixel contains limited materials, and neighboring pixels are constituted by similar materials, which indicate sparsity and spatial smoothness properties. The anomaly part accounts for a small amount of the whole scene, and each tube-wise fiber (i.e., the spectral bands of each pixel) contains few nonzero values, which indicates the tube-wise sparsity of the anomaly tensor. In a real observed HSI, the spectra are usually corrupted with noise that is caused by precision limits of the imaging spectrometer and errors in analog to digital conversion, this issue is dealt with by modeling the noise as identically and independently distributed Gaussian random variables [44]–[46]. Here, Gaussian noise is modeled separately with anomalies to suppress the noise confusion with the anomaly.

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

- 1) In view of the LMM, we propose a completely blind tensor-based model where the background is decomposed using the mode-3 product of the abundance tensor and endmember matrix. The abundance tensor of the background is characterized by tensor regularization with imposed low rankness through CP decomposition, smoothness, and sparsity.
- 2) For the anomaly part of HSI AD, the  $\ell_{1,1,2}$ -norm is utilized to characterize the tube-wise sparsity since the anomalies account for a small amount of the scene. Gaussian noise is also incorporated into the model to suppress noise confusion with the anomaly.

The experimental results on five different datasets, with extensive metrics and illustrations, demonstrate that the proposed method significantly outperforms the other competing methods.

The remainder of this article is organized as follows. The third paragraph of section II introduces the problem formulation and proposed method. In Section III, we evaluate the performance of the proposed method and compare it with some traditional and state-of-the-art AD methods using five real hyperspectral datasets. Finally, the conclusion and future works are illustrated in Section IV.

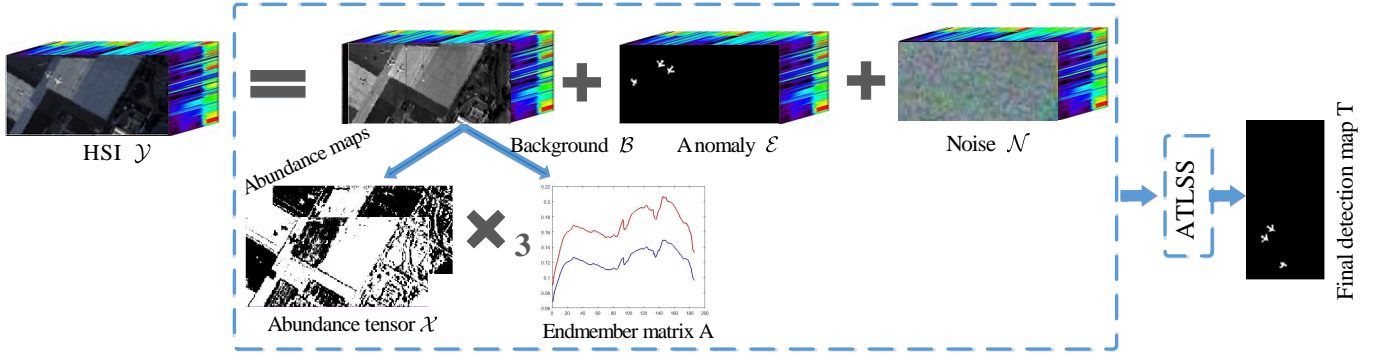


Fig. 1: The flowchart of the proposed ATLSS method, and an AVIRIS-2 HSI is used as an example.

TABLE I: Symbols.

$\mathcal{X}$	Abundance tensor
$\mathcal{E}$	Anomaly tensor
$\mathcal{N}$	Noise tensor
$\mathcal{Y}$	The observed HSI
$\mathbf{A}$	Endmember matrix
$\mathbf{Z}_{(1)}$	The factor matrix of abundance tensor
$\mathbf{z}_1^{(1)}$	The first outproduct vector of factor matrix
$H$	The height of the observed HSI
$W$	The width of the observed HSI
$D$	The number of bands of the observed HSI
$R$	The number of endmember
$K_{\mathcal{X}}$	The rank of abundance tensor

## II. PROBLEM FORMULATION AND PROPOSED METHOD

In this section, we first introduce HSI AD based on the background LMM. The proposed ATLSS algorithm is illustrated afterward. In addition, the overall flowchart of the proposed ATLSS algorithm for the hyperspectral AD method is shown in Fig. 1.

### A. Tensor Notation and Definition

This subsection introduces some mathematical notation and preliminaries of tensors, with our proposed method clearly described. We use lowercase bold symbols for vectors, e.g.,  $\mathbf{x}$ , and capital letters for matrices, e.g.,  $\mathbf{X}$ . The paper denotes a third-order tensor by bold Euler script letters, e.g.,  $\mathcal{X}$ . The scalar is written as  $x$ .

**Definition 1.** *The dimension of a tensor is called the mode, and  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  has  $N$  modes. Slices are two-dimensional sections of a tensor and are defined by fixing all but two indices. For a third-order tensor  $\mathcal{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ ,  $\mathcal{Y}_{\dots:k}$  is the  $k$ -th frontal slice.*

**Definition 2. (The mode- $n$  unfolding and folding of a tensor)** *The "unfold" operation along mode- $n$  on an  $N$ -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 3. (Rank-one tensors)** *An  $N$ -way tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  is a rank-one tensor if it can be written as the outer product of  $N$  vectors.*

**Definition 4. (Mode- $n$  product)** *The mode- $n$  product of tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  and a matrix  $\mathbf{U} \in \mathbb{R}^{J \times I_n}$  is defined as:*

$$(\mathcal{X} \times_n \mathbf{U})_{I_1 \dots I_{n-1} J I_{n+1} \dots I_N} = \sum_{i_n=1}^{I_n} \mathbf{x}_{i_n I_2 \dots I_N} \mathbf{u}_{J i_n}, \quad (1)$$

*In contrast, the  $n$ -mode product can be further computed by matrix multiplication.*

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \iff \mathbf{Y} = \mathbf{U} \mathcal{X}_{(n)}, \quad (2)$$

**Definition 5. (CP decomposition)** *CP decomposition factorizes an  $n$ -order tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  into a sum of component rank-1 tensors as:*

$$\mathcal{X} \approx \llbracket \lambda; \mathbf{B}^{(1)}, \mathbf{B}^{(2)}, \dots, \mathbf{B}^{(N)} \rrbracket \equiv \sum_{r=1}^R \lambda_r \mathbf{b}_r^{(1)} \circ \mathbf{b}_r^{(2)} \circ \dots \circ \mathbf{b}_r^{(N)}, \quad (3)$$

*where  $\lambda \in \mathbb{R}^R$  is the weight vector and  $\mathbf{B}^{(n)} = [\mathbf{b}_1^{(n)} \dots \mathbf{b}_R^{(n)}] \in \mathbb{R}^{I_n \times R} \forall n = \{1, \dots, N\}$  is the  $n$ -th factor matrix.  $R$  is the rank of tensor  $\mathcal{X}$ , and we denote  $\text{rank}(\mathcal{X}) = R$ .*

### B. HSI AD based on Background LMM

Since an HSI cube can be naturally treated as a third-order tensor, we use a tensor-based representation to avoid spatial and spectral information loss. In such an AD application, an HSI tensor can be decomposed into so-called background and anomaly tensors [47]. However, in real-world applications, the scenes are usually corrupted by noise [45]; hence, in this paper, noise is also added into the model to suppress its confusion with the estimated anomaly term, and the model is expressed as follows:

$$\mathcal{Y} = \mathcal{B} + \mathcal{E} + \mathcal{N}, \quad (4)$$

where  $\mathcal{Y} \in \mathbb{R}^{H \times W \times D}$  is the observed HSI and  $\mathcal{B} \in \mathbb{R}^{H \times W \times D}$ ,  $\mathcal{E} \in \mathbb{R}^{H \times W \times D}$ , and  $\mathcal{N} \in \mathbb{R}^{H \times W \times D}$  are the background, anomaly, and noise, respectively.  $H$ ,  $W$ , and  $D$  represent the height, width, and number of bands of each tensor, respectively. The purpose of AD is to reconstruct an accurate background image

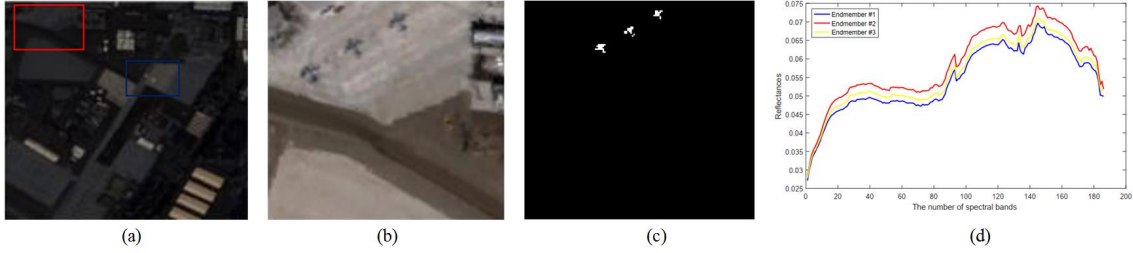


Fig. 2: AVIRIS-1 dataset. (a) False color image of the whole scene. (b) False color image of the detection area corresponding to the area surrounded by the red box in the scene (a). (c) Ground-truth map of the anomalies. (d) Spectra of endmembers #1-#3 of the background, and the estimated number of endmembers is 3.

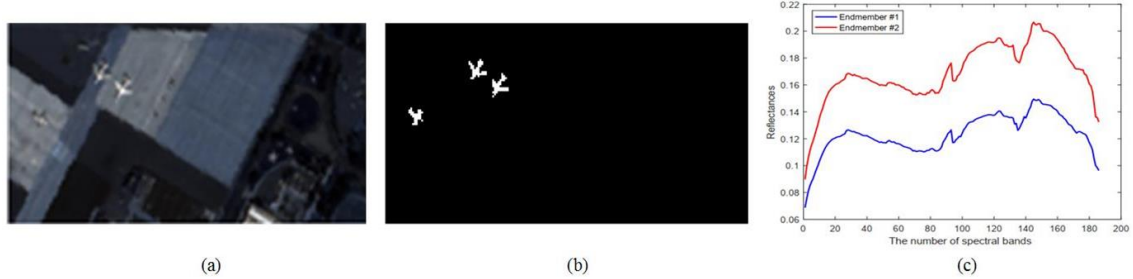


Fig. 3: AVIRIS-2 dataset. (a) False color image of the detection area corresponding to the area surrounded by the blue box in the scene in Fig. 2 (a). (b) Ground-truth map of the anomalies. (c) Spectra of endmembers #1-#2 of the background, and the estimated number of endmembers is 2.

to more accurately separate the anomaly from the background and noise, yielding superior performance.

There are typically several mixed pixels in natural HSI scenes, implying that more than one material participates in constituting each mixing pixel. This refers to an explicit physical interpretation under the assumption of the LMM; that is, the spectrum of each pixel of a low-rank background can be linearly combined with a few numbers of pure spectral endmembers. NMF is an ideal solution for the LMM, as it decomposes the original data into the product of two low-dimensional nonnegative matrices. Alternatively, under tensor notation, the tensor background approximates the mode-3 product of a nonnegative abundance tensor and a nonnegative endmember matrix. Mathematically, eq. (4), inspired by NMF, takes the following form:

$$\begin{aligned} \mathcal{Y} &= \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N} \\ \text{s.t. } \mathcal{X} &\geq 0, \mathbf{A} \geq 0, \end{aligned} \quad (5)$$

where  $\mathcal{X} \in \mathbb{R}^{H \times W \times R}$  is the third-order abundance tensor,  $\mathbf{A} \in \mathbb{R}^{D \times R}$  is the endmember matrix, and  $R$  is the number of endmembers. The third-order abundance tensor is obtained by reshaping each abundance vector into a frontal slice matrix  $\mathcal{X}_{:, :, r} \forall r \in \{1, \dots, R\}$  of dimensions  $H \times W$  and then stacking them along the mode-3 direction. The newly formed abundance tensor implements the underlying inner low-rank structure information to be well characterized.

### C. Proposed method

An HSI usually consists of a few materials, which makes it lie in low-rank subspaces. Moreover, a similar substance is

distributed in the adjacent region, which gives the HSI a locally smooth property. Compared to the background, anomalies are distributed randomly; thus, anomalies are often assumed to be sparse. Therefore, in this paper, we model the problem based on the assumption that the observed HSI is a superposition of a low-rank background, sparse anomalies, and the noise term.

The LMM assumes that each mixed pixel in the background linearly consists of a few endmembers, indicating that many zero entries are contained in the abundance tensor, which can be represented by a sparse property. Remarkably, the  $\ell_0$ -norm can directly minimize nonzero components, but this leads to an NP-hard problem. The  $\ell_1$ -norm, therefore, is introduced to promote the sparsity of the abundance tensor, where the sparsity prior narrows the solution space and achieves accurate abundance tensor estimation. The anomaly pixels occupy a small proportion of the scene, indicating that the anomaly matrix has a column-sparse property and is characterized by the  $\ell_{2,1}$ -norm [48]. Here, we have  $\|\mathcal{E}_{(3)}\|_{2,1} = \|\mathcal{E}\|_{1,1,2}$ ; therefore, due to the physical meaning and the definition of the  $\ell_{1,1,2}$ -norm in [49], it is reasonable to assume that the tensor anomaly has tube-wise sparsity. In addition, an HSI is usually corrupted with the noise, which is assumed to be identically and independently distributed Gaussian random variables [44]; therefore, the noise is modeled as  $\|\mathcal{N}\|_F^2$  to suppress it from being confused with the anomaly. In general, the background model based on the LMM can be rewritten as follows:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t. } & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \quad \mathcal{X} \geq 0, \mathbf{A} \geq 0, \end{aligned} \quad (6)$$

where  $\|\cdot\|_1$  and  $\|\cdot\|_F^2$  denote the  $\ell_1$ -norm and  $\ell_2$ -norm,

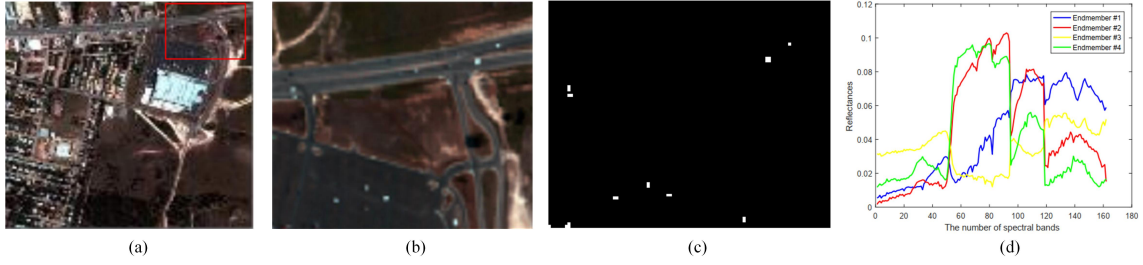


Fig. 4: HYDICE dataset. (a) False color image of the whole scene. (b) False color image of the detection area. (c) Ground-truth map of the anomalies. (d) Spectra of endmembers #1-#4 of the background, and the estimated number of endmembers is 4.

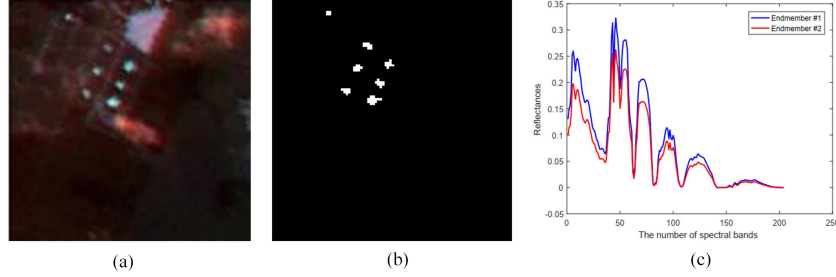


Fig. 5: Urban-1 of the Urban (ABU) dataset. (a) False color image of the detection area. (b) Ground-truth map of the anomalies. (c) Spectra of endmembers #1-#2 of the background, and the estimated number of endmembers is 2.

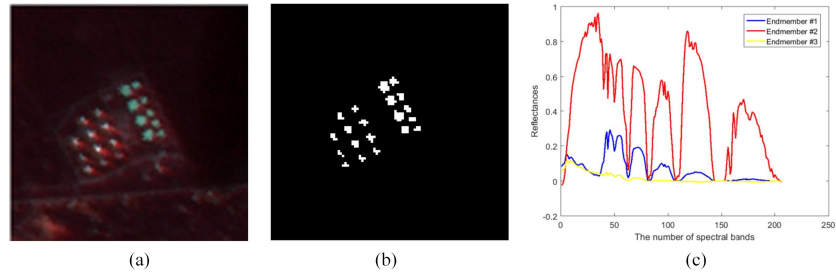


Fig. 6: Urban-2 of the Urban (ABU) dataset. (a) False color image of the detection area. (b) Ground-truth map of the anomalies. (c) Spectra of endmembers #1-#3 of the background, and the estimated number of endmembers is 3.

respectively. Here,  $\|\mathcal{E}\|_{1,1,2} = \sum_{h=1}^H \sum_{w=1}^W \|\mathcal{E}_{h,w,:}\|_F$  is defined as the sum of the Frobenius norm of all tube-wise spectra vectors of  $\mathcal{E}$ .  $\lambda_1$  and  $\beta$  are the trade-off parameters that control the sparsity of the abundance tensor and tube-wise sparsity of the anomaly, respectively.

Beyond the single-pixel sparsity, the spatial correlation between the neighbor pixels also deserves to be exploited. It supposes that there is a high correlation among the spectra of the pixels lying in homogeneous regions. For that, we impose a soft low-rank property on the abundance tensor  $\mathcal{X}$  of the background tensor  $\mathcal{B}$  in order to model the aforementioned high correlation property of homogeneous.

Assuming that  $K_X$  designates the rank of the abundance tensor, the loss function for HSI AD can be written as:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{N}} \quad & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} \quad & \text{rank}(\mathcal{X}) = K_X, \\ & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \mathcal{X} \geq 0, \mathbf{A} \geq 0. \end{aligned} \quad (7)$$

However, the rank of  $\mathcal{X}$  in (7) is a nonconvex problem

in that the optimization is also difficult. Referring to [50], we introduce  $\mathcal{Q}$ , which is assumed to be low-rank and represents a low-rank prior. Then, CP decomposition is employed to measure the low rankness of  $\mathcal{Q}$  via the summation of the  $K_X$  rank-1 components. Consequently, CP decomposition of  $\mathcal{Q}$  can be written as:

$$\mathcal{Q} = \llbracket \mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \mathbf{Z}^{(3)} \rrbracket = \sum_{i=1}^{K_X} \mathbf{z}_i^{(1)} \circ \mathbf{z}_i^{(2)} \circ \mathbf{z}_i^{(3)}, \quad (8)$$

where  $\mathbf{Z}^{(1)} = \llbracket \mathbf{z}_1^{(1)} \dots \mathbf{z}_{K_X}^{(1)} \rrbracket \in \mathbb{R}^{H \times K_X}$ ,  $\mathbf{Z}^{(2)} = \llbracket \mathbf{z}_1^{(2)} \dots \mathbf{z}_{K_X}^{(2)} \rrbracket \in \mathbb{R}^{W \times K_X}$ , and  $\mathbf{Z}^{(3)} = \llbracket \mathbf{z}_1^{(3)} \dots \mathbf{z}_{K_X}^{(3)} \rrbracket \in \mathbb{R}^{D \times K_X}$  are the factor matrices.

Subsequently, we introduce a new regularization term controlled by a low-rank tensor  $\mathcal{Q}$  to enforce a nonstrict constraint on  $K_X$ , as shown in (9). A very large  $K_X$  value will lead to the mixing of anomalies into the background, which undermines the low-rank characterization of the background by the regularization term. Thereby, the parameter  $\lambda_3$  aims to modify the strictness of the low-rank constraint on  $\mathcal{X}$ . Thus, not only is the nonconvex issue addressed, but the small-

scale details that are necessary for the background may also be preserved. The function based on (7) is approximately rewritten as:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{Q}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \frac{\lambda_3}{2} \|\mathcal{X} - \mathcal{Q}\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} & \text{rank}(\mathcal{Q}) = K_{\mathcal{X}}, \\ & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \mathcal{X} \geq 0, \mathbf{A} \geq 0, \end{aligned} \quad (9)$$

where the rank is controlled by  $K_{\mathcal{X}}$ . Then, the model of (11) can be rewritten as:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \\ & \frac{\lambda_3}{2} \|\mathcal{X} - \llbracket \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)} \rrbracket\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \mathcal{X} \geq 0, \mathbf{A} \geq 0. \end{aligned} \quad (10)$$

Notably, the pixels with spatial homogeneity are more likely to contain the same materials, which indicates the fractional abundance of the adjacent pixels that tend to be similar. Here, the spatial-context information of the abundance tensor is characterized by the TV regularizer [51] by encouraging the piecewise smoothness structure with the distinct edges preserved. The TV-based cost function corresponding to (10) can be modeled as:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \lambda_2 \|\mathcal{X}\|_{TV} + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} & \text{rank}(\mathcal{X}) = K_{\mathcal{X}}, \\ & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \mathcal{X} \geq 0, \mathbf{A} \geq 0, \end{aligned} \quad (11)$$

where  $\|\cdot\|_{TV}$  is the TV norm and  $\lambda_2$  is a parameter to adjust the strength of the piecewise smoothness. The spatial TV norm of the abundance tensor  $\mathcal{X}$  is defined as [52]:

$$\|\mathcal{X}\|_{TV} = \|\mathcal{H}\mathcal{X}\|_1 = \|\mathcal{H}_h\mathcal{X}\|_1 + \|\mathcal{H}_v\mathcal{X}\|_1, \quad (12)$$

Let  $\mathcal{X}_{h,w,r} |_{\forall h=\{1,\dots,H\}, w=\{1,\dots,W\}, r=\{1,\dots,R\}}$  indicate the intensity of the voxel  $(h, w, r)$  and  $\mathcal{H}_h$  and  $\mathcal{H}_v$  be the two horizontal and vertical differential operators in the spatial domain. Then, we have:

$$\begin{cases} \mathcal{H}_h\mathcal{X}_{h,w,r} = \mathcal{X}_{h,w+1,r} - \mathcal{X}_{h,w,r} \\ \mathcal{H}_v\mathcal{X}_{h,w,r} = \mathcal{X}_{h+1,w,r} - \mathcal{X}_{h,w,r} \end{cases} \quad (13)$$

Similar to (7), (8), it can be further rewritten as:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{X}\|_1 + \lambda_2 \|\mathcal{X}\|_{TV} + \\ & \frac{\lambda_3}{2} \|\mathcal{X} - \llbracket \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)} \rrbracket\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}, \mathcal{X} \geq 0, \mathbf{A} \geq 0. \end{aligned} \quad (14)$$

#### D. Optimization Procedure

The optimization problem in (14) can be solved by ADMM [53]. We need to introduce three auxiliary variables,

$\mathcal{V}_1, \mathcal{V}_2$ , and  $\mathcal{V}_3$ , and then transform it to the following equivalent problem:

$$\begin{aligned} \min_{\mathcal{X}, \mathbf{A}, \mathcal{E}, \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}, \mathcal{N}} & \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{V}_3\|_1 + \lambda_2 \|\mathcal{V}_2\|_1 + \\ & \frac{\lambda_3}{2} \|\mathcal{X} - \llbracket \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)} \rrbracket\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2} \\ \text{s.t.} & \mathbf{A} \geq 0, \mathcal{X} \geq 0, \mathcal{X} = \mathcal{V}_1, \mathcal{H}\mathcal{V}_1 = \mathcal{V}_2, \mathcal{X} = \mathcal{V}_3, \\ & \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathcal{E} + \mathcal{N}. \end{aligned} \quad (15)$$

The problem in (15) can be solved by ALM [54] to minimize the following augmented Lagrangian function:

$$\begin{aligned} \mathcal{L}(\mathcal{X}, \mathbf{A}, \mathcal{Q}, \mathcal{N}, \mathcal{E}, \mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3, \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4) & \\ = \frac{1}{2} \|\mathcal{N}\|_F^2 + \lambda_1 \|\mathcal{V}_3\|_1 + \lambda_2 \|\mathcal{V}_2\|_1 + \frac{\lambda_3}{2} \|\mathcal{X} - \mathcal{Q}\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2} & \\ + \frac{\mu}{2} (\|\mathcal{X} - \mathcal{V}_1 + \mathcal{D}_1\|_F^2 + \|\mathcal{H}\mathcal{V}_1 - \mathcal{V}_2 + \mathcal{D}_2\|_F^2 + & \\ \|\mathcal{X} - \mathcal{V}_3 + \mathcal{D}_3\|_F^2) + \|\mathcal{Y} - \mathcal{X} \times_3 \mathbf{A} - \mathcal{E} - \mathcal{N} + \mathcal{D}_4\|_F^2 & \\ \text{s.t.} & \mathbf{A} \geq 0, \mathcal{X} \geq 0, \end{aligned} \quad (16)$$

where  $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ , and  $\mathcal{D}_4$  are the Lagrange multipliers and  $\mu$  is the penalty parameter. The above problem can be solved by updating one variable while fixing the others. Specifically, in the  $t+1$ -th iteration, where the problem can be divided into several subproblems, the variables are updated as follows:

1) *Update  $\mathcal{X}$* : The related subproblem is:

$$\begin{aligned} \mathcal{X}^{(t+1)} = \arg \min_{\mathcal{X}} & \frac{\lambda_3}{2} \|\mathcal{X} - \mathcal{Q}^{(t)}\|_F^2 + \frac{\mu}{2} (\|\mathcal{X} - \mathcal{V}_1^{(t)} + \mathcal{D}_1^{(t)}\|_F^2 \\ + \|\mathcal{X} - \mathcal{V}_3^{(t)} + \mathcal{D}_3^{(t)}\|_F^2) + \|\mathcal{Y} - \mathcal{X} \times_3 \mathbf{A}^{(t)} - \mathcal{E}^{(t)} - \mathcal{N}^{(t)} + \mathcal{D}_4^{(t)}\|_F^2 & \\ \text{s.t.} & \mathcal{X} \geq 0, \end{aligned} \quad (17)$$

which can be transformed to the following linear system:

$$\begin{aligned} \mathcal{X}^{(t+1)} = \mathcal{X} \cdot * [ & (\mu \times (\mathcal{Y} - \mathcal{E}^{(t)} - \mathcal{N}^{(t)} + \mathcal{D}_4^{(t)}) \times_3 \mathbf{A}^{(t)T} + \lambda_3 \mathcal{Q}^{(t)} \\ + \mu (\mathcal{V}_1^{(t)} - \mathcal{D}_1^{(t)} + \mathcal{V}_3^{(t)} - \mathcal{D}_3^{(t)}) \cdot (\mu \times \mathcal{X} \times_3 (\mathbf{A}^{(t)T} \mathbf{A}^{(t)}) + & \\ (\lambda_3 + 2\mu)\mathcal{X})]. & \end{aligned} \quad (18)$$

2) *Update  $\mathbf{A}$* : The related subproblem is:

$$\begin{aligned} \mathbf{A}^{(t+1)} = \arg \min_{\mathbf{A}} & \|\mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t)} - \mathcal{E}^{(t)} - \mathcal{N}^{(t)} + \mathcal{D}_4^{(t)}\|_F^2 \\ \text{s.t.} & \mathbf{A} \geq 0. \end{aligned} \quad (19)$$

Similar to (18), we obtain the following solution:

$$\mathbf{A}^{(t+1)} = \mathbf{A} \cdot * (\mathcal{Y} - \mathcal{E}^{(t)} - \mathcal{N}^{(t)} + \mathcal{D}_4^{(t)}) \mathcal{X}^{(t+1)T} \cdot / \mathbf{A} \mathcal{X}^{(t+1)} \mathcal{X}^{(t+1)T}. \quad (20)$$

3) *Update  $\mathcal{B}$* : We combine (18) and (20) and compute the result as follows:

$$\mathcal{B}^{(t+1)} = \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)}. \quad (21)$$

4) *Update  $\mathcal{Q}$* : The related subproblem is:

$$\mathcal{Q}^{(t+1)} = \arg \min_{\mathcal{Q}} \frac{\lambda_3}{2} \|\mathcal{X}^{(t+1)} - \mathcal{Q}\|_F^2. \quad (22)$$

By substituting (11) into (14), the optimization problem becomes:

$$(\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}) = \arg \min_{\mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)}} \frac{\lambda_3}{2} \|\mathcal{X} - \llbracket \mathcal{Z}^{(1)}, \mathcal{Z}^{(2)}, \mathcal{Z}^{(3)} \rrbracket\|_F^2, \quad (23)$$

where CP decomposition is solved by an alternating least squares (ALS) [55] algorithm. Consequently, each factor matrix is calculated through a linear least square approach by fixing the other two matrices.

5) *Update  $\mathcal{N}$* : The related subproblem is:

$$\mathcal{N}^{(t+1)} = \arg \min_{\mathcal{N}} \frac{1}{2} \|\mathcal{N}^{(t)}\|_F^2 + \frac{\mu}{2} \|\mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)} - \mathcal{E}^{(t)} - \mathcal{N}^{(t)} + \mathcal{D}_4^{(t+1)}\|_F^2, \quad (24)$$

for which the solution is:

$$\mathcal{N}^{(t+1)} = \frac{\mu (\mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)} - \mathcal{E}^{(t)} + \mathcal{D}_4^{(t)})}{(\mu + 1)}. \quad (25)$$

6) *Update  $\mathcal{V}_1$* : The related subproblem is:

$$\begin{aligned} \mathcal{V}_1^{(t+1)} = \arg \min_{\mathcal{V}_1} & \|\mathcal{X}^{(t+1)} - \mathcal{V}_1^{(t)} + \mathcal{D}_1^{(t)}\|_F^2 \\ & + \|\mathcal{H}\mathcal{V}_1^{(t)} - \mathcal{V}_2^{(t)} + \mathcal{D}_2^{(t)}\|_F^2. \end{aligned} \quad (26)$$

Now, we solve the subproblem of  $\mathcal{V}_1$ :

$$(\mathcal{H}^T \mathcal{H} + \mathcal{I}) \mathcal{V}_1^{(t+1)} = (\mathcal{X}^{(t+1)} + \mathcal{D}_1^{(t)}) + \mathcal{H}^* (\mathcal{V}_2^{(t)} - \mathcal{D}_2^{(t)}). \quad (27)$$

where  $\mathcal{I}$  is the identity tensor;  $\mathcal{H}$  is a convolution, as defined in (13), that operates in the spatial domain; and  $\mathcal{H}^T$  indicates the adjoint operator of  $\mathcal{H}$ . Therefore,  $\mathcal{V}_1$  can be quickly computed by:

$$\mathcal{V}_1^{(t+1)} = \text{ifft} \left( \frac{\mathcal{X}^{(t+1)} + \mathcal{D}_1^{(t)} + \mathcal{H}^T (\mathcal{V}_2^{(t)} - \mathcal{D}_2^{(t)})}{1 + \text{fft}(\mathcal{H}_h)^T \text{fft}(\mathcal{H}_h) + \text{fft}(\mathcal{H}_v)^T \text{fft}(\mathcal{H}_v)} \right), \quad (28)$$

where *fft* and *ifft* denote the fast Fourier transform [56] and its inverse, respectively.

7) *Update  $\mathcal{V}_2$  and  $\mathcal{V}_3$* : The subproblem for  $\mathcal{V}_2$  is:

$$\mathcal{V}_2^{(t+1)} = \arg \min_{\mathcal{V}_2} \lambda_2 \|\mathcal{V}_2^{(t)}\|_1 + \frac{\mu}{2} \|\mathcal{H}\mathcal{V}_1^{(t+1)} - \mathcal{V}_2^{(t)} + \mathcal{D}_2^{(t)}\|_F^2, \quad (29)$$

which can be solved by a soft-thresholding function. For  $\mathcal{V}_3$ , its update rule is similar to that of  $\mathcal{V}_2$ .

8) *Update  $\mathcal{E}$* : The related subproblem is:

$$\mathcal{E}^{(t+1)} = \arg \min_{\mathcal{E}} \frac{\mu}{2} \|\mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)} - \mathcal{E} - \mathcal{N}^{(t+1)} + \mathcal{D}_4^{(t)}\|_F^2 + \beta \|\mathcal{E}\|_{1,1,2}. \quad (30)$$

Denoting  $\mathcal{F}^{(t)} = \mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)} - \mathcal{N}^{(t+1)} + \mathcal{D}_4^{(t)}$ , the closed-solution of (30) can be achieved as:

$$\mathcal{E}^{(t+1)}(h, w, :) = \max \left( 1 - \frac{\beta}{\mu^{(t)} \|\mathcal{F}^{(t)}(h, w, :)\|_F}, 0 \right) * \mathcal{F}^{(t)}(h, w, :), \quad (31)$$

where  $h \in \{1, \dots, H\}$  and  $w \in \{1, \dots, W\}$ .

9) *Update  $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4$ , and  $\mu$* : The four Lagrange multipliers and the penalty parameter are updated as follows:

$$\mathcal{D}_1^{(t+1)} = \mathcal{D}_1^{(t)} + \mathcal{X}^{(t+1)} - \mathcal{V}_1^{(t+1)}, \quad (32)$$

$$\mathcal{D}_2^{(t+1)} = \mathcal{D}_2^{(t)} + \mathcal{H}\mathcal{V}_1^{(t+1)} - \mathcal{V}_2^{(t+1)}, \quad (33)$$

$$\mathcal{D}_3^{(t+1)} = \mathcal{D}_3^{(t)} + \mathcal{X}^{(t+1)} - \mathcal{V}_3^{(t+1)}, \quad (34)$$

---

### Algorithm 1 ATLSS algorithm

---

**Input:** HSI tensor  $\mathcal{Y}$ , number of background endmembers  $R$ , abundance tensor rank  $K$ , regularized parameters  $\lambda_1, \lambda_2, \lambda_3, \beta, \mu = 10^{-2}, \mu_{\max} = 10^6, T_{\max} = 100$ , and the residual error  $\xi = 10^{-6}$ .

**Output:** anomaly tensor  $\mathcal{E}$  and anomaly map  $\mathbf{T}$ .

1. **Initialization:** endmember matrix  $\mathbf{A}$  and abundance tensor  $\mathcal{X}$ .
  2. **Repeat:**
  3. Update  $\mathcal{X}$  with Equation (17)
  4. Update  $\mathbf{A}$  with Equation (19).
  5. Update  $\mathcal{B}$  with Equation (21).
  6. Update  $\mathcal{Q}$  with Equation (22).
  7. Update  $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \mathbf{Z}^{(3)}$ , with Equation (23).
  8. Update  $\mathcal{N}$  with Equation (24).
  9. Update  $\mathcal{V}_1$  with Equation (26).
  10. Update  $\mathcal{V}_2$  with Equation (29).
  11. Update  $\mathcal{V}_3$  in a way similar to that in step 10.
  12. Update  $\mathcal{E}$  with Equation (30).
  13. Update  $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3$ , and  $\mathcal{D}_4$  with Equation (32), (33), (34), and (35) respectively.
  14. **Until** either  $T_{\max} = 100$  or  $\|\mathcal{B}^{(t+1)} - \mathcal{B}^{(t)}\|_2 / \|\mathcal{B}^{(t+1)}\|_2 < \xi$  are satisfied.
- 

$$\mathcal{D}_4^{(t+1)} = \mathcal{Y} - \mathcal{X}^{(t+1)} \times_3 \mathbf{A}^{(t+1)} - \mathcal{E}^{(t+1)} - \mathcal{N}^{(t+1)} + \mathcal{D}_4^{(t)}, \quad (35)$$

$$\mu^{(t+1)} = \min(\rho\mu^{(t)}, \mu_{\max}). \quad (36)$$

Finally, according to the anomaly  $\mathcal{E}$ , the AD maps  $\mathbf{T}$  can be obtained by  $\mathbf{T}_{h,w} = \|\mathcal{E}_{h,w}\|_F \mid_{\forall h=\{1,\dots,H\}, w=\{1,\dots,W\}}$ . The abovementioned approach can be summarized in Algorithm 1, in which we arrive at an augmented Lagrangian alternating direction method to solve the proposed ATLSS model.

### E. Initialization and Termination Condition

In the proposed ATLSS solver, the input terms are the observed HSI, the basis matrix  $\mathbf{A}$ , the abundance matrix  $\mathcal{X}$ , and the number of endmembers  $R$ . It is worth noting that the initialization of  $\mathbf{A}$ ,  $\mathcal{X}$ , and  $R$  influences the AD results of the proposed method. For that, we initialize the background  $\mathcal{B}$  by the RX algorithm. Then, HySime [57] is employed to estimate the number of background endmembers  $R$ . Afterward, NMF [58] and the related update rules in [53] are utilized to iterate over  $\mathbf{A}$  and  $\mathcal{X}$ , respectively. The CP rank  $K$  is determined by the algorithm referred to in [50] to provide an accurate rank estimation for the proposed method. In general, the iterative process continues until a maximum number of 100 iterations or a residual error is satisfied.

### F. Computational Complexity Analysis

As shown in Algorithm 1, each iteration's computational cost consists of updating all the referred factors. The time complexity of CP decomposition (Update  $\mathcal{Q}$ ):  $O(T_{\max} R^2 (W^2 D^2 + H^2 D^2 + H^2 W^2))$ . Time complexity of FFT (update  $\mathcal{V}_1$ ):  $O(HWR \log(HW))$ . Time complexity of soft threshold operator (update  $\mathcal{V}_2, \mathcal{V}_3$ ) is  $O(HWR)$ . Time complexity of other matrix



TABLE II: Formulations of the ATLSS model and its different degradation models.

Model Name	Formulation
Dm-1	$\min_{\mathcal{X}, \mathbf{A}, \mathbf{E}, \mathbf{N}} \frac{1}{2} \ \mathcal{N}\ _F^2 + \lambda_1 \ \mathcal{X}\ _1 + \beta \ \mathbf{E}\ _{1,1,2}$ $\text{s.t. } \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathbf{E} + \mathbf{N}, \quad \mathcal{X} \geq 0, \mathbf{A} \geq 0$
Dm-2	$\min_{\mathcal{X}, \mathbf{A}, \mathbf{E}, \mathbf{N}} \frac{1}{2} \ \mathcal{N}\ _F^2 + \lambda_1 \ \mathcal{X}\ _1 + \beta \ \mathbf{E}\ _{1,1,2}$ $\text{s.t. } \text{rank}(\mathcal{X}) = K_{\mathcal{X}}, \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathbf{E} + \mathbf{N}, \quad \mathcal{X} \geq 0, \mathbf{A} \geq 0$
ATLSS Model	$\min_{\mathcal{X}, \mathbf{A}, \mathbf{E}, \mathbf{N}} \frac{1}{2} \ \mathcal{N}\ _F^2 + \lambda_1 \ \mathcal{X}\ _1 + \lambda_2 \ \mathcal{X}\ _{TV} + \beta \ \mathbf{E}\ _{1,1,2}$ $\text{s.t. } \text{rank}(\mathcal{X}) = K_{\mathcal{X}}, \mathcal{Y} = \mathcal{X} \times_3 \mathbf{A} + \mathbf{E} + \mathbf{N}, \quad \mathcal{X} \geq 0, \mathbf{A} \geq 0$

multiplication operations (update  $\mathcal{X}$ ,  $\mathbf{A}$ ,  $\mathcal{B}$ ,  $\mathbf{Z}^{(1)}$ ,  $\mathbf{Z}^{(2)}$ ,  $\mathbf{Z}^{(3)}$ ,  $\mathbf{N}$ ,  $\mathcal{E}$ ,  $\mathcal{D}_1$ ,  $\mathcal{D}_2$ ,  $\mathcal{D}_3$ , and  $\mathcal{D}_4$ ):  $O(HWR + DR + HK_{\mathcal{X}} + 4HWD) = O(HW(D + R))$ . So the time complexity of Algorithm 1 is  $O(R^2(W^2D^2 + H^2D^2 + H^2W^2) + HWR \log(HW) + HW(D + 2R))$ .

### III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the proposed ATLSS algorithm was carried out on five real HSI datasets for AD, where the detailed description is listed as follows. 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.

#### A. Experimental Datasets

1) *AVIRIS Airplane Data*: 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 (1 – 6, 33 – 35, 97, 107 – 113, 153 – 166, 221 – 224) are removed. As shown in Fig. 2(b), the subimage is named the AVIRIS-1 dataset, and it is located in the top-left corner of the AVIRIS image with a size of  $150 \times 150 \times 186$ . The contained anomaly is the three air planes, and the ground truth is shown in Fig. 2(c). AVIRIS-2 is located in the center of San Diego, as shown in Fig. 3, and it contains 120 anomaly pixels with a size  $100 \times 200 \times 186$ .

2) *HYDICE Data*: The real data was collected by the hyperspectral digital imagery collection experiment (HYDICE) sensor, and the original image has a size of  $307 \times 307 \times 210$ . After removing the low-SNR and water vapor absorption bands, 162 bands remained. 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 are shown in Fig. 4.

3) *Urban (ABU) Data*: The Urban dataset was collected with AVIRIS sensors and contains 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. 5(a) and (b). For the Urban-2 dataset with a size of  $100 \times 100 \times 207$ , Fig. 6(a) and (b) are the corresponding false-color image and the ground truth.

#### B. Evaluation Metrics and Parameter Settings

The AD performance of the proposed ATLSS method is demonstrated in this section. Table II shows the formulations of ATLSS and its two degradation models, degradation model 1 (Dm-1) and degradation model 2 (Dm-2), which are based on the linear spectral unmixing method. In addition, we choose to compare our models with RX [12], RPCA [22], LRASR [24], GTVLRR [25], GVAE [9], PTA [35], and TRPCA [49]. RXD is a statistical-based method in the HSI AD field, and it is the baseline in almost all reference articles. The methods above are based on matrix modeling. PTA is a tensor-based method, but it is based on matrix operations. GVAE is a deep learning method. TRPCA and the proposed ATLSS algorithm are based on tensor modeling. We classified the RX, RPCA, LRASR, GTVLRR, and PTA methods as matrix-based operations and TRPCA, Dm-1, Dm-2, and ATLSS as tensor-based operations.

To evaluate the detector more effectively, the 3D ROC curve [59] is employed, which introduces the threshold  $\tau$  in addition to the parameters  $P_d$  and  $P_f$  used in the 2D ROC curve [60] to specify a third dimension ( $P_d, P_f, \tau$ ). In addition, the 2D ROC curves ( $P_d, P_f$ ) and ( $P_f, \tau$ ) are used to measure the AD result, where an efficient detector would have a performance with a larger ( $P_d, P_f$ ) ( $\rightarrow 1$ ) value but a smaller ( $P_f, \tau$ ) ( $\rightarrow 0$ ) value, where it is desired that the curves of ( $P_d, P_f$ ) and ( $P_f, \tau$ ) are close to the upper left and lower left corners of the coordinate axis, respectively. In addition, box and whisker plots are used to evaluate the separability between the anomaly and background. The boxes in the box and whisker plot reflect the distribution range of the detection values of the anomaly and background; that is, a larger gap between the anomaly and background boxes indicates better discrimination of the detector.

In the proposed ATLSS method, the number of endmembers  $R$  is first estimated in the initialization phase using the HySime algorithm, whereas the most significant task is to search for the best set of parameters ( $\lambda_1, \lambda_2, \lambda_3, \beta$ ) that need to be carefully identified. Moreover, the  $\lambda_1, \lambda_2, \lambda_3$ , and  $\beta$  range in the set  $\{5 \cdot 10^{-1}, 1 \cdot 10^{-1}, 5 \cdot 10^{-2}, 1 \cdot 10^{-2}, 5 \cdot 10^{-3}, 1 \cdot 10^{-3}, 5 \cdot 10^{-4}, 1 \cdot 10^{-4}\}$ , respectively, while  $\mu$  is selected from  $\{0.01, 0.001\}$ . Here, we carefully searched the five datasets to determine the optimal parameters for all the algorithms. To demonstrate the contribution of the different parameters to AD, we take the example of the AVIRIS-2 airplane data while changing the  $\lambda_1, \lambda_2, \lambda_3$ , and  $\beta$  parameters to illustrate the tuning procedure in detail. We applied ATLSS with different parameter settings

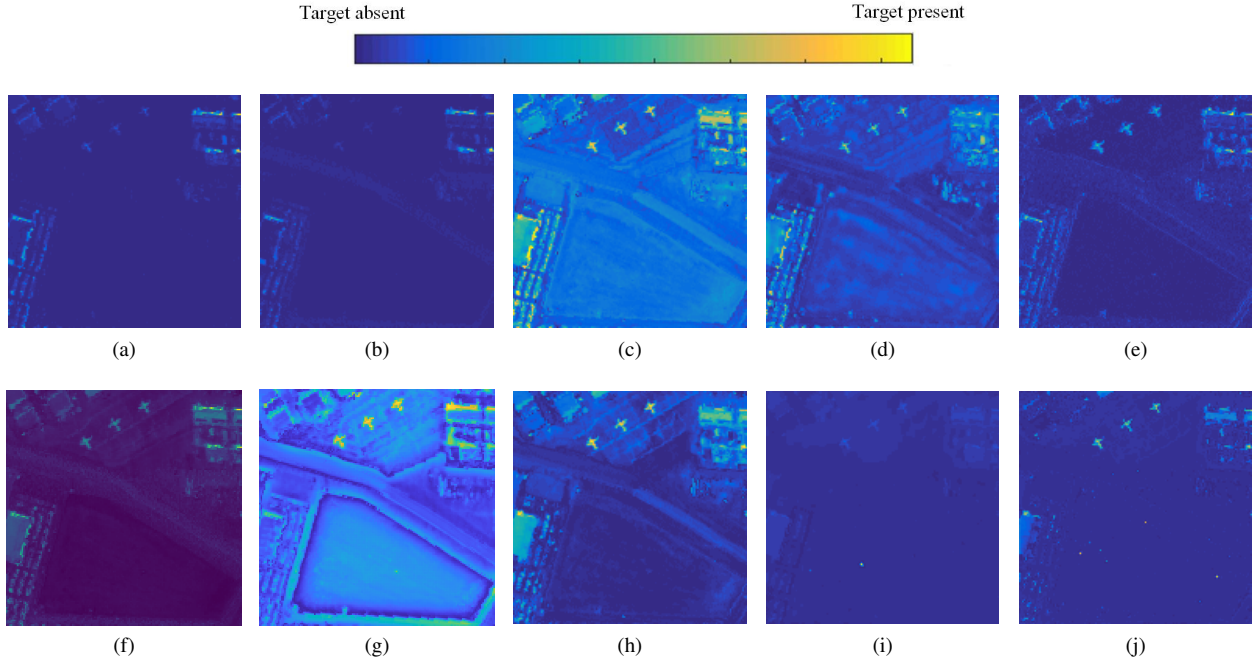


Fig. 7: 2-D plots of the detection results obtained by RX, RPCA, LRASR, TRPCA, GTVLRR, GVAE, PTA, Dm-1, Dm-2, and ATLSS on the AVIRIS-1 dataset. (a) RX. (b) RPCA. (c) LRASR. (d) TRPCA. (e) GTVLRR. (f) GVAE. (g) PTA. (h) Dm-1. (i) Dm-2. (j) ATLSS.

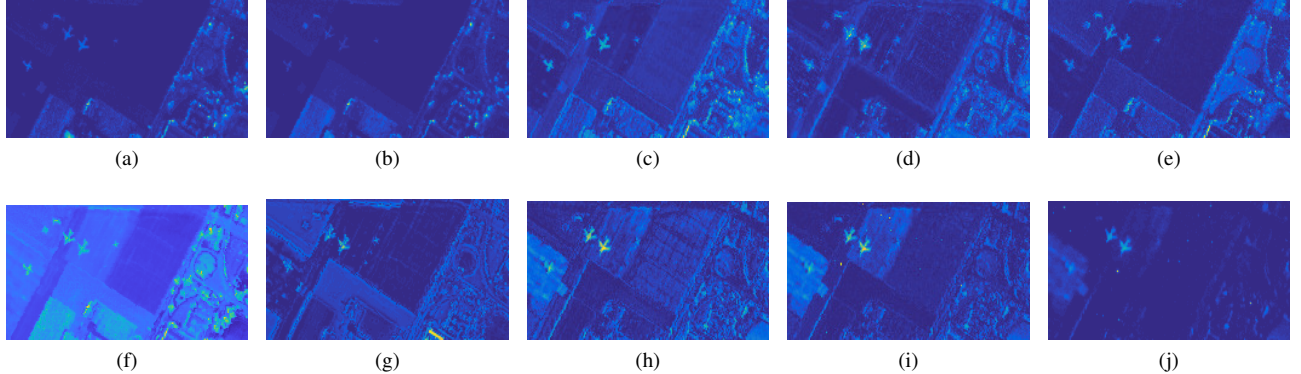


Fig. 8: 2-D plots of the detection results obtained by RX, RPCA, LRASR, TRPCA, GTVLRR, GVAE, PTA, Dm-1, Dm-2, and ATLSS on the AVIRIS-2 dataset. (a) RX. (b) RPCA. (c) LRASR. (d) TRPCA. (e) GTVLRR. (f) GVAE. (g) PTA. (h) Dm-1. (i) Dm-2. (j) ATLSS.

on the AVIRIS airplane datasets (including AVIRIS-1 and AVIRIS-2), HYDICE dataset, and Urban (ABU) datasets (including Urban-1 and Urban-2) in turn to achieve the detection results under the optimal parameter combination.

### C. Detection Performance

We investigate the contribution of the regularization terms, including sparsity regularization, TV regularization, and CP regularization, in the proposed ATLSS method with regard to the accuracy of AD. We refer again to Table II for the ATLSS, Dm-1, and Dm-2 models. Furthermore, we compare the performance of ATLSS, including Dm-1 and Dm-2, with that of RX, RPCA, LRASR, GTVLRR, TRPCA, LSMAD, and PTA. Sequentially, the 2-D plots of the comparison algorithm

detection results on the five datasets are shown in Fig. 7-11. Table III shows the AUC values of  $(P_d, P_f)/(P_f, \tau)$  obtained by different AD algorithms on the five real datasets. Each algorithm is ran ten times on each dataset to avoid randomness, and the average AUC values are used. Fig. 12 and Fig. 13 show the corresponding performance curves and the box and whisker plots for different comparison methods on the five real datasets. We also take the AVIRIS-2 dataset as an example to illustrate the superior performance of the proposed method in detail.

For Dm-1, when the parameter settings are  $\lambda_1 = 0.01$  and  $\beta = 0.1$ , all five datasets can achieve very good results.

1) *AVIRIS-2*: For the AVIRIS-2 dataset, the estimation number of the endmembers for the background is 2, as shown

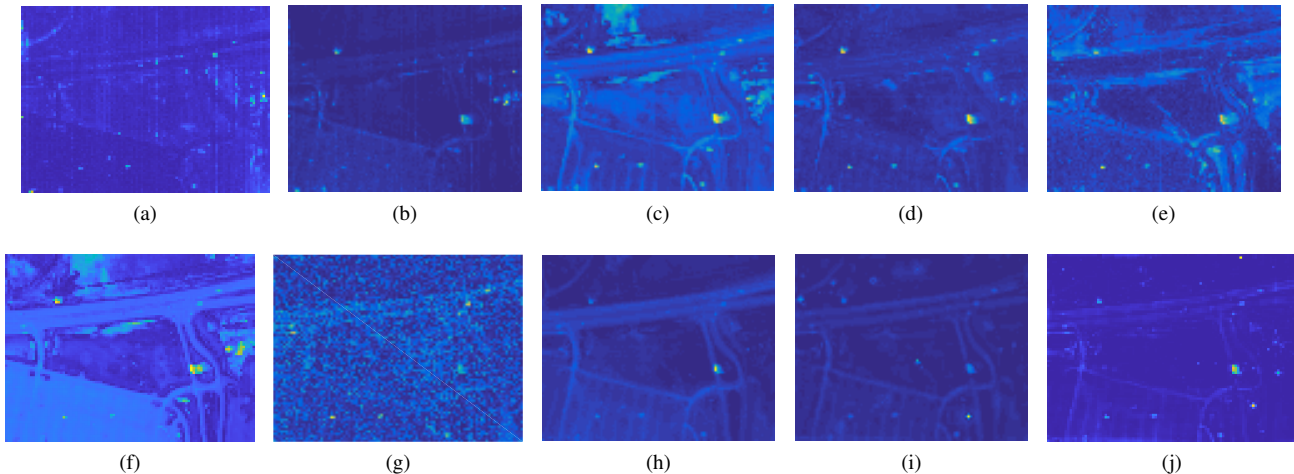


Fig. 9: 2-D plots of the detection results obtained by RX, RPCA, LRASR, TRPCA, GTVLRR, GVAE, PTA, Dm-1, Dm-2, and ATLSS on the HYDICE dataset. (a) RX. (b) RPCA. (c) LRASR. (d) TRPCA. (e) GTVLRR. (f) GVAE. (g) PTA. (h) Dm-1. (i) Dm-2. (j) ATLSS.

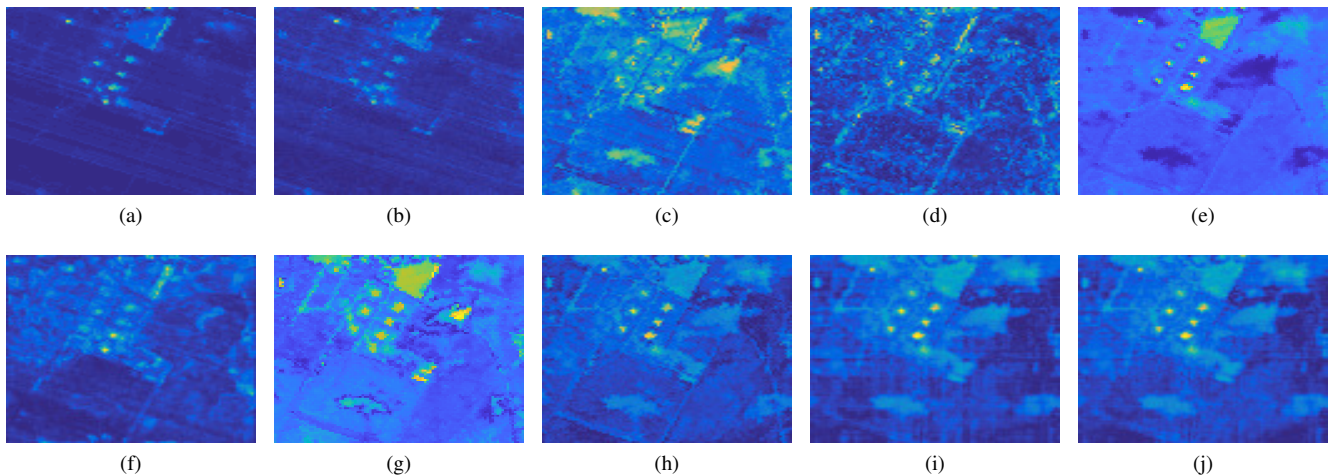


Fig. 10: 2-D plots of the detection results obtained by RX, RPCA, LRASR, TRPCA, GTVLRR, GVAE, PTA, Dm-1, Dm-2, and ATLSS on the ABU-Urban-1 dataset. (a) RX. (b) RPCA. (c) LRASR. (d) TRPCA. (e) GTVLRR. (f) GVAE. (g) PTA. (h) Dm-1. (i) Dm-2. (j) ATLSS.

in Fig. 3(c), and the estimated rank is  $K = 18$ . In the following two paragraphs, we study (1.a) the effects of the regularization terms and (1.b) the comparison with different anomaly detectors.

*(1.a) Effects of the Regularization Terms:* After a large number of parameter traversals, the trade-off parameters referenced in the ATLSS algorithm and Dm-2 both achieve their optimal performance, i.e.,  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.005$ ,  $\lambda_3 = 0.005$ , and  $\beta = 0.005$  for ATLSS and  $\lambda_1 = 0.1$ ,  $\lambda_3 = 0.5$ , and  $\beta = 0.001$  for Dm-2. Fig. 14 shows the detection accuracy of ATLSS on the AVIRIS-2 dataset when one parameter varies within a predefined parameter range and when the other three trade-off parameters are fixed. In Fig. 14(a), it can be observed that when  $\lambda_1$  increases from 0.0001 to 0.5, the curve shows an upward trend, and the highest detection result is obtained at  $\lambda_1 = 0.5$ , which indicates a positive effect on controlling

the sparsity of the abundance tensor. Table III and Fig. 8 provide the certification from the quantitative analysis and visual qualitative analysis. The peak of the curve in Fig. 14(b) is located at  $\lambda_2 = 0.005$  when  $\lambda_2$  is in the interval  $[0.0001, 0.005]$ . The increasing curve implies that TV has a positive effect on suppressing noise, while  $\lambda_2 > 0.005$  imposes an over smoothness constraint on the abundance tensor leading to a dramatic decline in the detection results.  $\lambda_3$  is imposed on  $\mathbf{Q}$  to control its low rankness, and Fig. 14(c) reveals that setting  $\lambda_3 = 0.005$  balances the low-rank regularization with the most important information but captures small-scale details. The curve first steadily increases and then falls as it deviates from the optimal parameter value, implying that parameter values that are too large lead to strict low rankness on  $\mathbf{X}$  and thus result in a significant residual loss of the reconstructed abundance tensor. In Fig. 14(d), we clearly see

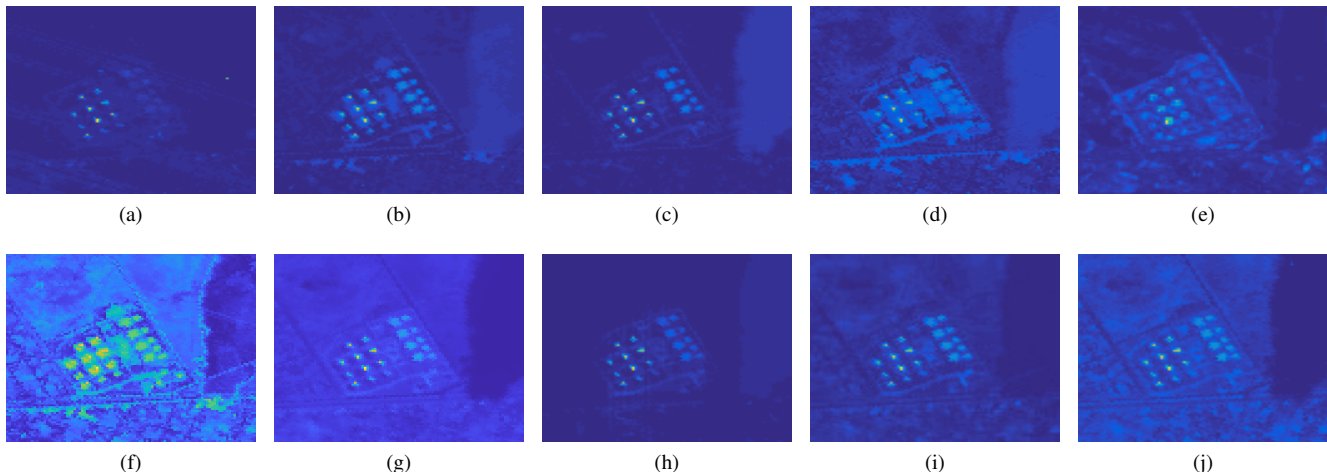


Fig. 11: 2-D plots of the detection results obtained by RX, RPCA, LRASR, TRPCA, GTVLRR, GVAE, PTA, Dm-1, Dm-2, and ATLSS on the ABU-Urban-2 dataset. (a) RX. (b) RPCA. (c) LRASR. (d) TRPCA. (e) GTVLRR. (f) GVAE. (g) PTA. (h) Dm-1. (i) Dm-2. (j) ATLSS.

that the curve is stable when  $\beta \in [0.005, 0.1]$  and exhibits superior performance. Outside this interval, the detection result is sensitive to the change in the parameter and results in the detection performance declining in a straight downward trend. When  $\beta$  is set to 0.005, the best performance is achieved.

(1.b) *Comparison with Different Anomaly Detectors:* Table III demonstrates the AUC values of  $(P_d, P_f)$  and  $(P_f, \tau)$  for the different comparison methods on the five datasets. In Table III, the anomaly detection accuracy  $(P_d, P_f)$  of ATLSS, which obtains the highest score among all the comparison methods when  $\tau = 0.087$ , is illustrated by the ROC curve in the upper left corner, which also shows the efficiency of the model. The  $(P_f, \tau)$  score is lower than the others, and the  $(P_f, \tau)$  ROC curve is also closest to the lower-left corner, as shown in Fig. 12(b). ATLSS performs well on the AVIRIS-2 dataset and achieves optimal AD accuracy and a low false alarm rate. The proposed ATLSS model and the generation DM-1 and DM-2 methods better separate the anomaly and background, which is shown in Fig. 13(b). The 2-D plots of the detection results in Fig. 8(i) show the low-rank structure imposed by the regularization constraint (i.e., CP decomposition on the introduced low-rank prior term), which allows the abundance tensor adequate flexibility to model fine-scale spatial details with most of its spatial distribution preserved. The exploration of the background tensor enables the suppression of the background more efficiently. Moreover, TV regularization, which smooths the estimated abundance map, effectively suppresses Gaussian noise. When compared with the generation DM-1 and DM-2 methods, ATLSS has better background suppression, as shown in Fig. 8, and the best AD performance, as shown in Table III. In Fig. 12(b), except for the curves of  $(P_d, P_f)$  and  $(P_f, \tau)$ , the 3D ROC curve also comprehensively shows the performance of the proposed ATLSS method.

The methods assume that the background and the anomaly have a low-rank and sparse property and performs better than the RX method in terms of the AUC value and

false alarm rate, which can be observed from Table III and Fig. 12(b). The 2-D plots in Fig. 8 also reveal that the background and Gaussian noise are both effectively suppressed; moreover, the anomaly airplanes are clearly detected by ATLSS compared to RX. The GTVLRR, PTA, Dm-2, and ATLSS, utilize TV regularization to smooth away the noise signature while strengthening the outlines of the airplanes. We can observe that GTVLRR performs well among the comparison methods. However, the proposed ATLSS method based on tensor decomposition is completely outstanding compared to the comparison methods in terms of  $(P_d, P_f)/(P_f, \tau)$  and the power of the anomaly and background separation views.

Table IV shows the computational times of the ten referred algorithms on the AVIRIS-2 dataset. We observe that the running time of ATLSS is higher than that of Dm-1 and Dm-2 because the added regular terms increase the running time, but the AD performance improves. The time cost of ATLSS is also lower than that of GTVLRR and TRPCA. Compared to RX, RPCA, LRASR, and PTA, the time cost of ATLSS is much higher because they are based on matrix operations. As shown in Table III, the deep learning method GVAE with the Training time/Test time is  $80 \times 2000 / 25.56$ , which not only has a high time complexity but also the experimental performance is far inferior to our proposed ATLSS.

2) *AVIRIS-1:* For the AVIRIS-1 dataset, the estimation number of the endmember for the background is 3, as shown in Fig. 2(d), and the estimated rank is  $K = 46$ .

When comparing Dm-1, Dm-2, and ATLSS, we can see that the anomaly part includes three planes, which are clearly identified in Fig. 7. In Table III and Fig. 12(a), for ATLSS, we can observe that its value of  $(P_d, P_f)$  is the largest, and the curve is on the upper left corner. The  $(P_f, \tau)$  value of ATLSS is also relatively low. The superior performance of ATLSS demonstrates that it imposes low-rank tensor regularization and TV regularization on the abundance tensor, efficiently suppressing the background and smoothing away the noise. Among all the comparison methods, as shown in Table III,

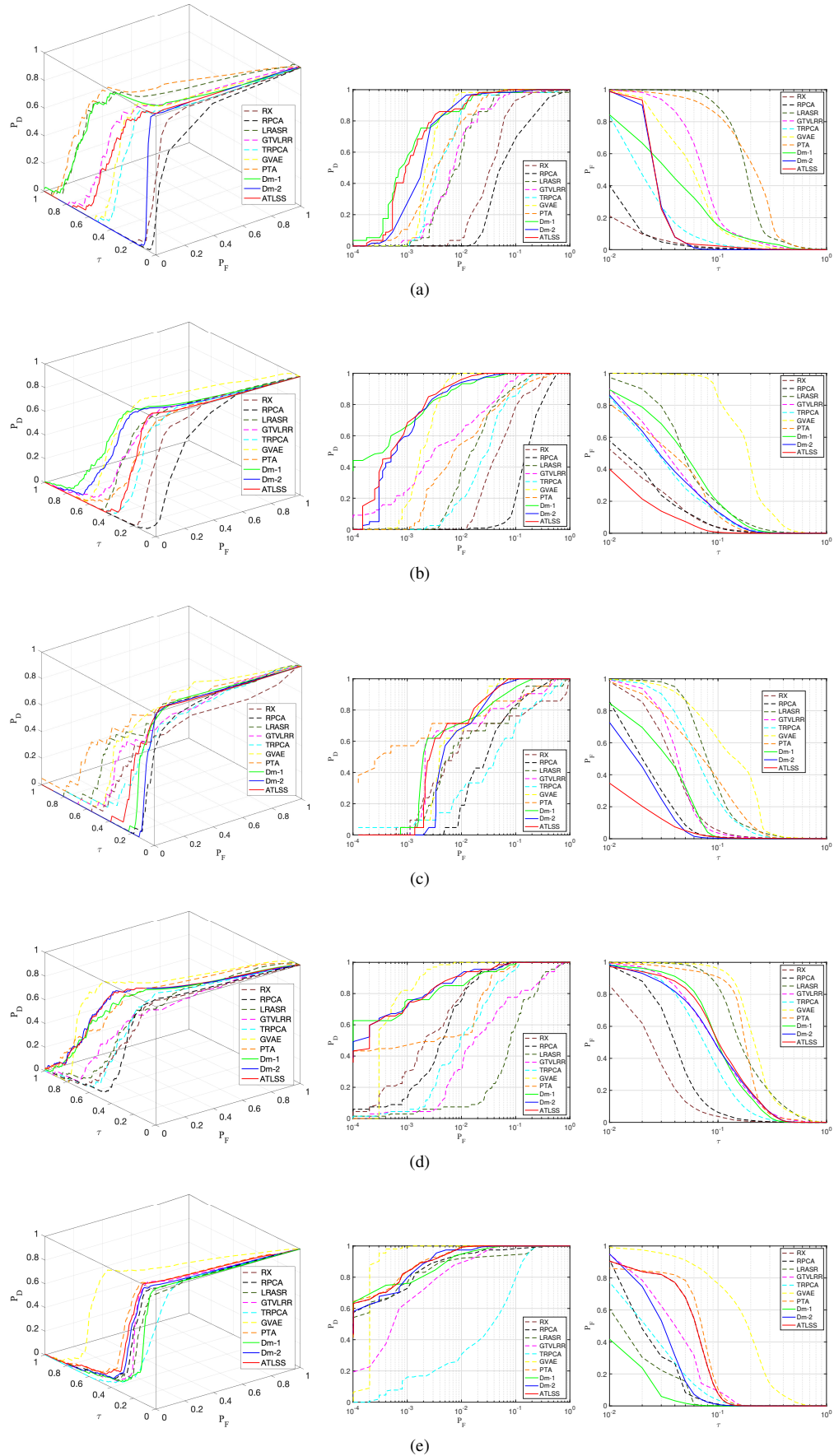


Fig. 12: The ROC curves of different comparison methods on the five datasets. (a) AVIRIS-1. (b) AVIRIS-2. (c) HYDICE. (d) ABU-Urban-1. (e) ABU-Urban-2. (Left to right) 3D-ROC curve and 2D-ROC curve of  $(P_d, P_f)$ , and 2D-ROC curve of  $(P_f, \tau)$ .

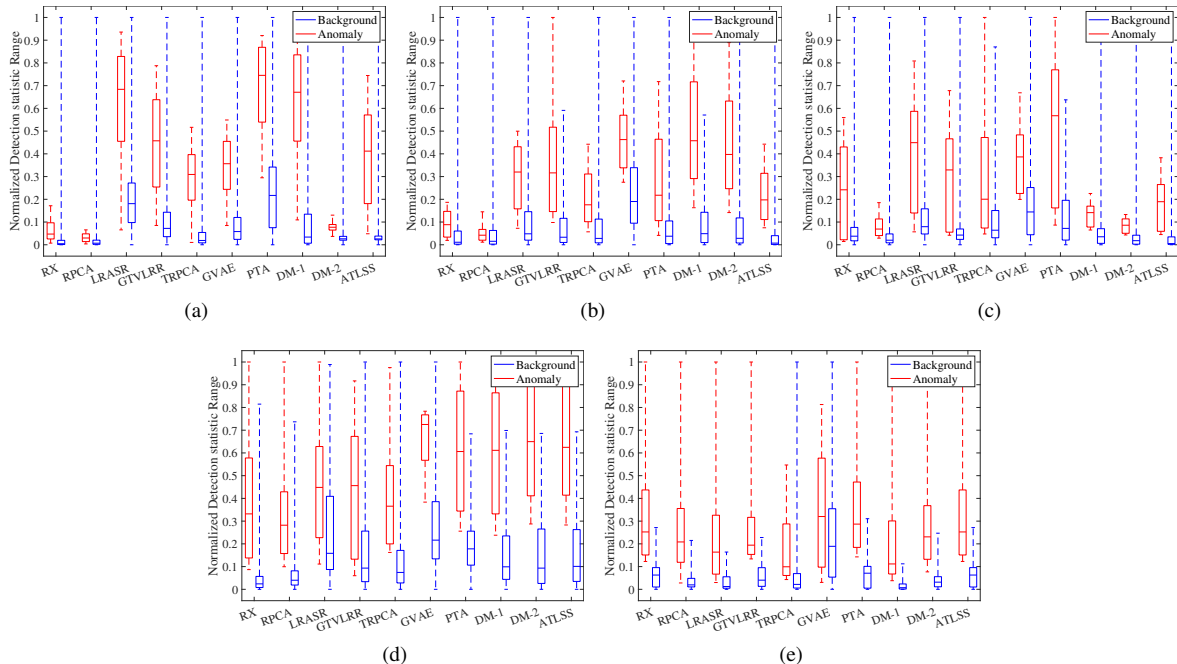


Fig. 13: Box and whisker plots for the different comparison methods on the five real datasets: (a) AVIRIS-1. (b) AVIRIS-2. (c) HYDICE. (d) ABU-Urban-1. (e) ABU-Urban-2.

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

Datasets	AUC values of $(P_d, P_f)/(P_f, \tau)$						
	Matrix-based operations					Deep learning	
	RX	RPCA	LRASR	GTVLRR	PTA	GVAE	TRPCA
AVIRIS-1	0.9551/ <b>0.0118</b>	0.8935/0.0129	0.9716/0.1968	0.9822/0.0893	0.9890/0.2227	0.9860/0.0718	0.9801/0.0304
AVIRIS-2	0.9213/0.0266	0.7969/0.0262	0.9672/0.0711	0.9816/0.0518	0.9609/0.0514	0.9616/0.2009	0.9549/0.0478
HYDICE	0.8511/0.0470	0.9436/0.0277	0.9311/0.0989	0.9393/0.0488	0.9829/0.0935	0.9311/0.1549	0.9600/0.0817
Abu-urban-1	0.9934/ <b>0.0329</b>	0.9916/0.0496	0.8666/0.2096	0.9093/0.1261	0.9852/0.1869	0.9778/0.0925	0.9823/0.2450
Abu-urban-2	0.9946/0.0611	0.9960/0.0283	0.9867/0.0211	0.9967/0.0487	0.9992/0.0672	0.9828/0.1976	0.9456/0.0325

the best detection result was also achieved by ATLSS. We can observe that even though RX has the lowest false alarm rate, the AD accuracy and the separation of the anomaly and background in Fig. 12(a) and Fig. 13(a) demonstrate that its performance is not very good. Notably, our proposed method on the AVIRIS-2 is superior to the tensor-based PTA method in all the measurements. Fig. 12(a) shows that the ROC curve of ATLSS is closer to the top left that those of all the comparison methods, and Table III further proves the remarkable performance of ATLSS.

3) *HYDICE*: For the urban data, the estimated number of endmembers is  $R = 4$ , as shown in Fig. 4(d), where the estimated CP rank is  $K = 67$ .

Compared to Dm-1 and Dm-2, as shown in Table III, the ATLSS method shows a competitive performance, the AD AUC value is the highest, and the false alarm rate is the lowest. For anomaly and background separability, ATLSS is also powerful in separating anomalies and the background, as

Tensor-based operations		
Dm-1	Dm-2	ATLSS
0.9950/0.0649	0.9954/0.0293	<b>0.9967/0.0328</b>
0.9967/0.0655	0.9970/0.0490	<b>0.9982/0.0159</b>
0.9829/0.0388	0.9847/0.0222	<b>0.9893/0.0152</b>
0.9940/0.1214	0.9960/0.1226	<b>0.9962/0.1291</b>
0.9976/ <b>0.0126</b>	0.9989/0.0346	<b>0.9992/0.0611</b>

shown in Fig. 13(c). In Fig. 9(i), the anomaly is accurately detected; furthermore, the background and the noise are both well suppressed, which is a benefit of tensor low-rank regularization and TV smoothness for the background abundance tensor. Compared with PTA, the  $(P_d, P_f)/(P_f, \tau)$  values of ATLSS, as shown in Table III, account for the fact that the

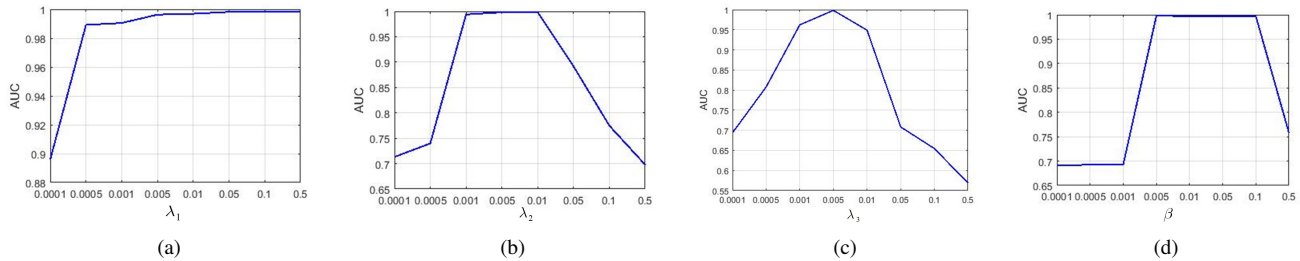


Fig. 14: Detection accuracy of ATLSS on the AVIRIS-2 dataset with different parameter settings. (a)  $\lambda_1$  varies. (b)  $\lambda_2$  varies. (c)  $\lambda_3$  varies. (d)  $\beta$  varies.

TABLE IV: The computational time (seconds) of different AD algorithms on the AVIRIS-2 dataset.

Method	Matrix-based operations					Deep learning	Tensor-based operations			
	RX	RPCA	LRASR	GTVLRR	PTA	GVAE	TRPCA	Dm-1	Dm-2	ATLSS
Time(s)	3.577	15.457	26.399	241.099	241.099	80×2000/25.56	234.465	70.494	149.949	200.732

abundance maps possess more distinctive features than the raw data, which is beneficial for identifying the anomaly from the background and achieving outstanding performance. The qualitative and quantitative results are given in Table III and Fig. 12(c), as well as in Fig. 13(c). These results are obtained due to the background low-rank decomposition, which enables more accurate background reconstruction.

4) *Urban (ABU-1)*: For the ABU-Urban-1 dataset, the number of estimation endmembers is 2, as shown in Fig. 5(c), and the estimated CP rank is  $K = 25$ .

Compared with Dm-1 and Dm-2, the ATLSS background model imposes a CP regularization constraint and TV regularization on the abundance tensor, and an excellent AD performance achieved, as shown in Table III. A higher detection result would lead to a few anomaly pixels being regarded as background pixels, resulting in a lower false alarm rate. Thus, as shown in III and Fig. 12(d), the  $(P_d, P_f)$  values of Dm-1, Dm-2, and ATLSS increase gradually, while the  $(P_f, \tau)$  values in Table III also increased. Furthermore, we can observe that  $(P_f, \tau)$  of RX is the lowest, and  $(P_d, P_f)$  is high among all the comparison methods. However, Fig. 13(d) shows that ATLSS and the generation methods perform a superior separation between the anomaly and background. In general, the evaluation metrics mentioned above validate that the proposed ATLSS method outperforms the other methods in both qualitative and quantitative aspects.

5) *Urban (ABU-2)*: For the ABU-Urban-2 dataset, the number of estimation endmembers is 3, as shown in Fig. 6(c), and the estimated CP rank is  $K = 25$ .

As shown in Table III, the detection result  $(P_d, P_f)$  of ATLSS is higher than those of Dm-1 and Dm-2. In contrast, the  $(P_f, \tau)$  value of ATLSS is slightly higher than those of the other two generative methods. It suffers from a much higher detection result, which causes a few anomaly pixels to be detected as background pixels. However, its efficient separability between the anomaly and background, as shown

in Fig. 7(e), demonstrates that it is still a competitive AD method. For the Urban (ABU-2) dataset, the results of the tensor-based PTA and ATLSS methods are evenly matched, as shown in Table III, Fig. 12, and Fig. 13. However, it is worth noting that the  $(P_f, \tau)$  value of ATLSS is lower than that of PTA, which demonstrates that the abundance maps possess more distinctive features than the original data and enable a more accurate identification of the anomaly and the background; hence, it achieves outstanding performance.

We perform the proposed ATLSS method and extensive comparison experiments on the five datasets and summarize the advantages of the proposed method as follows:

(1) *Effectiveness*: The proposed ATLSS method decomposes the background into an abundance tensor and endmember matrix. The structural characteristics of the abundance tensor are fully explored, i.e., the local spatial continuity and the high abundance vector correlations, which contribute to reconstructing a more accurate abundance tensor for the background. The proposed ATLSS model performs excellently compared to its degradation Dm-1 and Dm-2.

(2) *Performance*: Seven comparison algorithms are presented to sufficiently demonstrate the performance of the proposed method. Compared to RX, ATLSS, based on the low-rank and sparse assumption, has a more accurate AD performance. PTA is a tensor-based but matrix-based operation method that is competitive with the proposed ATLSS method, but its performance cannot be generalized to all datasets. The deep learning method GVAE has a high training time while achieving a general performance. ATLSS, which exploits the abundance tensor's physical meaning and possesses more distinctive features than the raw data, achieves outstanding AD performance.

## IV. CONCLUSION

In this paper, a novel method is proposed for hyperspectral AD. The idea is that the background is decomposed into

the mode-3 product of an abundance tensor and an endmember matrix. Considering that the background pixels contain limited endmembers, we impose the  $\ell_1$ -norm to characterize the sparsity. In addition, low-rank structure regularization is introduced for encoding fine-scale abundance tensor behavior to avoid a strict low-rank constraint. Moreover, a TV regularizer is also incorporated into the model to smooth the noise and anomaly signatures and separate the background more accurately. The sparse anomaly is characterized by the defined  $\ell_{1,1,2}$ -norm that aims to better constrain the tensor tube-wise sparsity. Notably, the noise is assumed to be Gaussian to avoid confusion with the anomaly. The experimental results on the five datasets demonstrated that the proposed method has an excellent AD performance compared to the other methods.

#### REFERENCES

- [1] 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.
- [2] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 2, pp. 29–56, 2017.
- [3] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 2341–2357, 2019.
- [4] M. Jouni, M. Dalla Mura, and P. Comon, "Hyperspectral image classification based on mathematical morphology and tensor decomposition," *Mathematical Morphology-Theory and Applications*, vol. 4, no. 1, pp. 1–30, 2020.
- [5] B. Xi, J. Li, Y. Li, R. Song, Y. Xiao, Y. Shi, and Q. Du, "Multi-direction networks with attentional spectral prior for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.
- [6] J. Xie, N. He, L. Fang, and P. Ghamisi, "Multiscale densely-connected fusion networks for hyperspectral images classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 1, pp. 246–259, 2021.
- [7] N. Nasrabadi, "Hyperspectral target detection : An overview of current and future challenges," *Signal Processing Magazine, IEEE*, vol. 31, pp. 34–44, 01 2014.
- [8] X. Zhang, X. Ma, N. Huyen, J. Gu, X. Tang, and L. Jiao, "Spectral-difference low-rank representation learning for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2021.
- [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. 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.
- [11] L. Zhang and B. Cheng, "Transferred cnn based on tensor for hyperspectral anomaly detection," *IEEE Geoscience and Remote Sensing Letters*, vol. PP, no. 99, pp. 1–5, 2020.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] Y. Gu, Y. Liu, and Y. Zhang, "A selective kpca algorithm based on high-order statistics for anomaly detection in hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 1, pp. 43–47, 2008.
- [17] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-rxd and linear filter-based rxd: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2351–2366, 2014.
- [18] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, "Multipixel anomaly detection with unknown patterns for hyperspectral imagery," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 4, pp. 1990–2000, 2015.
- [25] 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.
- [26] 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. 56, no. 8, pp. 4391–4405, 2018.
- [27] 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.
- [28] 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.
- [29] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE transactions on geoscience and remote sensing*, vol. 39, no. 7, pp. 1392–1409, 2001.
- [30] N. Wang, B. Du, and L. Zhang, "An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 554–569, 2013.
- [31] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "A low-rank tensor regularization strategy for hyperspectral unmixing," in *2018 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2018, pp. 373–377.
- [32] "Hyperspectral image restoration via global l 1-2 spatial-spectral total variation regularized local low-rank tensor recovery," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.
- [33] J. Xue, Y. Zhao, W. Liao, and C. W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–16, 2019.
- [34] Y. Wang, J. Peng, Q. Zhao, D. Meng, Y. Leung, and X. L. Zhao, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," 2017.
- [35] 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.
- [36] S. Song, H. Zhou, L. Gu, Y. Yang, and Y. Yang, "Hyperspectral anomaly detection via tensor-based endmember extraction and low-rank decomposition," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 10, pp. 1772–1776, 2019.



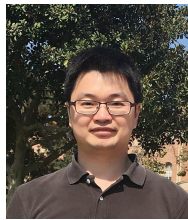
- [37] J. Wang, Y. Xia, and Y. Zhang, "Anomaly detection of hyperspectral image via tensor completion," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 6, pp. 1099–1103, 2020.
- [38] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.
- [39] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Transactions on Image Processing*, pp. 1–1, 2018.
- [40] J. Xue, Y. Zhao, S. Huang, W. Liao, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP, no. 99, pp. 1–15, 2021.
- [41] J. Xue, Y. Zhao, Y. Bu, J. C.-W. Chan, and S. G. Kong, "When laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion," *IEEE Transactions on Cybernetics*, pp. 1–15, 2022.
- [42] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Transactions on Geoscience Remote Sensing*, vol. 55, no. 3, pp. 1776–1792, 2017.
- [43] T. Imbiriba, R. A. Borsoi, and J. Bermudez, "A low-rank tensor regularization strategy for hyperspectral unmixing," 2018, pp. 373–377.
- [44] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1376–1389, 2015.
- [45] 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. 59, no. 2, pp. 1472–1486, 2021.
- [46] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *Journal of Applied Remote Sensing*, vol. 8, no. 1, p. 083641, 2014.
- [47] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4363–4372, 2021.
- [48] F. Nie, H. Huang, X. Cai, and C. Ding, "Efficient and robust feature selection via joint  $\ell_{2,1}$ -norms minimization," *Advances in neural information processing systems*, vol. 23, 2010.
- [49] 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.
- [50] T. Imbiriba, R. A. Borsoi, and J. C. M. Bermudez, "Low-rank tensor modeling for hyperspectral unmixing accounting for spectral variability," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 1833–1842, 2019.
- [51] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical imaging and vision*, vol. 20, no. 1, pp. 89–97, 2004.
- [52] W. He, H. Zhang, and L. Zhang, "Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3909–3921, 2017.
- [53] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.
- [54] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
- [55] D. Mitchell, N. Ye, and H. De Sterck, "Nesterov acceleration of alternating least squares for canonical tensor decomposition: Momentum step size selection and restart mechanisms," *Numerical Linear Algebra with Applications*, vol. 27, no. 4, p. e2297, 2020.
- [56] P. P. Vinchurkar, S. Rathkanthiwar, and S. Kakde, "Hdl implementation of dft architectures using winograd fast fourier transform algorithm," in *2015 Fifth International Conference on Communication Systems and Network Technologies*. IEEE, 2015, pp. 397–401.
- [57] J. M. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [58] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [59] 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.

- [60] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 251–255, 2008.

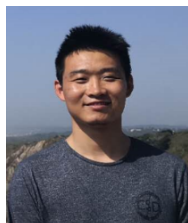


**Wenting Shang** received a BE degree from Zhengzhou University, Henan Province, China, in 2014. She is currently working toward a Ph.D. degree at the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. She was also a Visiting Ph.D. Student with the University of Grenoble Alpes, Grenoble, France, funded by the Chinese Scholarship Council. Her research interests include the areas of hyperspectral image classification and anomaly detection.



**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.



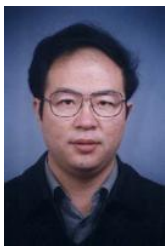
**Mohamad Jouni** (Member, IEEE) received a B.Eng. degree in Computer and Communications Engineering from the Lebanese University, Beirut, Lebanon, in 2016, and his M.Sc. and Ph.D. degrees in Signal and Image Processing from the University of Grenoble Alpes, Grenoble, France, in 2017 and 2021, respectively. 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.