

Hyperspectral Anomaly Detection Fused Unified Nonconvex Tensor Ring Factors Regularization

Wenjin Qin, Hailin Wang, *Student Member, IEEE*, Hao Shu, Feng Zhang, Jianjun Wang, *Member, IEEE*, Xiangyong Cao, *Member, IEEE*, Xi-Le Zhao, and Gemine Vivone, *Senior Member, IEEE*

Abstract—In recent years, tensor decomposition-based approaches for *hyperspectral anomaly detection* (HAD) have gained significant attention in the field of remote sensing. However, existing methods often fail to flexibly and effectively extract both the global correlations and local smoothness of the background components in *hyperspectral images* (HSIs). To mitigate this critical issue, we put forward a novel HAD method named HAD-EUNTRFR, which incorporates an enhanced unified nonconvex tensor ring (TR) factors regularization. In the HAD-EUNTRFR framework, the raw HSIs are first decomposed into background and anomaly components using the idea of tensor robust principal component analysis. The TR decomposition is then employed to capture the spatial-spectral correlations within the background component. Additionally, we introduce a unified and efficient nonconvex regularizer, induced by *tensor singular value decomposition* (T-SVD), to simultaneously encode the low-rankness and sparsity of the 3-D gradient TR factors into a unique concise form. The above characterization scheme enables the interpretable gradient TR factors to inherit the low-rankness and smoothness of the original background. To further enhance anomaly detection, we design a generalized nonconvex regularization term to exploit the group sparsity of the anomaly component. Based upon the above, we ultimately propose a scalable and reliable nonconvex HAD model. To solve the resulting doubly nonconvex model, we develop a highly efficient optimization algorithm based on the *alternating direction method of multipliers* (ADMM) framework. Theoretical results on convergence analysis for the proposed algorithm are derived. Experimental results on several benchmark datasets demonstrate that our proposed method outperforms existing *state-of-the-art* (SOTA) approaches in terms of detection accuracy.

Index Terms—*Hyperspectral anomaly detection, tensor decom-*

This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFA1008502; in part by the National Natural Science Foundation of China's Regional Innovation Development Joint Fund under Grant U24A2001; in part by the Natural Science Foundation of Chongqing, China, under Grant CSTB2023NSCQLZX0044; in part by the National Natural Science Foundation of China under Grant 12301594, Grant 12201505, Grant 12101512; in part by the Chongqing Talent Project, China, under Grant cstc2021ycjh-bgzxm0015; in part by the Fundamental Research Funds for the Central Universities under Grant SWU-KR25013; and in part by the Initiative Projects for Ph.D. in China West Normal University under Grant 22KE030. (Corresponding author: Jianjun Wang.)

Wenjin Qin, Feng Zhang, and Jianjun Wang are with the School of Mathematics and Statistics, Southwest University, Chongqing 400715, China (e-mail: qinwenjin2021@163.com, zfm@swu.edu.cn, wjj@swu.edu.cn).

Hailin Wang and Hao Shu are with the School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wang-hailin97@163.com, haoshu812@gmail.com).

Xiangyong Cao is with School of Computer Science and Technology and Ministry of Education Key Lab For Intelligent Networks and Network Security, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: caoxiangyong@mail.xjtu.edu.cn).

Xi-Le Zhao is with the School of Mathematical Sciences/Research Center for Image and Vision Computing, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: xlzhao122003@163.com).

Gemine Vivone is with the Institute of Methodologies for Environmental Analysis, CNR-IMAA, 85050 Tito Scalo, Italy, and also with the National Biodiversity Future Center (NBFC), 90133 Palermo, Italy (e-mail: gemine.vivone@imaa.cnr.it).

position, prior characterization, gradient map modeling, unified nonconvex factors regularization, ADMM algorithm

I. INTRODUCTION

Hyperspectral images (HSIs) capture a wealth of spatial and spectral information, with numerous spectral bands that provide a continuous and detailed representation of a scene. These rich spectral-spatial information serves as a strong foundation for analyzing and identifying various ground objects. As a result, HSIs have found widespread applications in fields such as fusion [1], [2], restoration [3], [4], denoising [5], [6], super-resolution [7], [8], unmixing [9]–[11], classification [12], [13], and anomaly detection [14]–[16]. Among these, *hyperspectral anomaly detection* (HAD) has gained significant attention due to its importance in both civilian and military applications [17], [18]. The primary goal of HAD is to detect anomalies by distinguishing them from the surrounding natural background. This process involves classifying a pixel as either an anomaly or part of the background based on the criterion that the spectral signature of an anomaly significantly deviates from that of the surrounding background. However, HAD task is inherently challenging because it often lacks prior knowledge of the spectral signatures for both the target anomaly and the background.

A. Related Works

In recent years, HAD literature has seen significant growth, with methods evolving across various advanced technical frameworks. These methods can be grouped into the following four main categories: (i) approaches using statistical theory [19]–[23], (ii) approaches using deep learning [24]–[34], (iii) approaches using matrix decomposition/representation [35]–[51], and (iv) approaches using tensor decomposition/representation [52]–[63].

The typical model of the first category is the *Reed-Xiaoli* (RX) algorithm [19], which adheres to the fundamental premise that background pixels obey the Gaussian distribution, whereas anomalous pixels do not. Additionally, Liu et al. [20] introduced two adaptive anomaly detectors induced by statistical knowledge to fulfill the HAD task in the presence of Gaussian noise. To address the challenge of detecting anomalies in HSIs with complex backgrounds, some strengthened versions of RX methods have been proposed, such as the locally adaptable RX detector [21], the kernel-RX [22], and the weighted-RXD [23]. These improvements have significantly boosted the effectiveness of RX-based approaches in identifying complex anomalies.

The second type of anomaly detectors can be further divided into supervised, unsupervised, and self-supervised categories.

Supervised methods utilize *convolutional neural networks* (CNN) to extract deep features and identify anomalies using labeled data [24]. However, the efficacy of this approach depends heavily on the availability of ground-truth data for training. Consequently, several innovative unsupervised architectures have been proposed for the HAD tasks, including *autoencoder* (AE) networks and *generative adversarial networks* (GANs). For example, Jiang et al. [25] suggested an unsupervised discriminative reconstruction constrained GAN for HAD. Xiang et al. [27] developed a novel guided-AE-based HAD method to reduce the feature representation for the anomaly targets. Li et al. [29] put forward an interpretable deep unfolding network, called LRR-Net, for the HAD task. At last, self-supervised methods [30]–[32] generate pseudolabels by predicting certain tasks without labeled data. Typically, novel blind-spot self-supervised reconstruction networks were devised for the HAD task in PDBSNet [30] and BS³LNet [31]. Generally speaking, this category of HAD methods has significant advantages in extracting features from HSIs, attributable to the strong learning capabilities of neural networks.

The third type of HAD methods include *collaborative representation* (CR)-based approaches [41]–[43], *sparse representation* (SR)-based approaches [44], [45], and *low-rank representation* (LRR)-based approaches [35]–[40], [46]–[51]. Depending on whether the strategy of background dictionary construction is employed, LRR-based approaches can be further subdivided into two branches: *low-rank and SR* (LRaSR) methods [46]–[51] and *low-rank and sparse matrix decomposition* (LRaSMD) methods [35]–[40]. The core idea behind LRaSR methods is modeling the background as the product of a background dictionary and a coefficient matrix, while applying regularization constraints to the coefficient matrix. In contrast, LRaSMD-based methods decompose HSI data directly into a background matrix, an anomaly matrix, and a noise matrix, applying various prior constraints to each of these components to achieve the HAD task. However, these algorithms often involve transforming a 3-D HSI cube into a 2-D matrix, resulting in the loss of crucial structural information embedded in the HSI cube.

Compared to the representation in vector/matrix structure, tensor tends to more faithfully and accurately uncover the intrinsic multidimensional structural information within HSI data [58], [64]–[69]. Motivated by the advantages of tensor representation, the fourth kind of HAD methods has emerged as a captivating research focus. The key to solving this kind of HAD problem is how to excavate the prior structures of HSI data finely, and encode them as certain regularization items for guiding a sound separation of the background and anomaly components. Among all tensor priors, the *global low-rankness* (denoted as “L”) property is particularly important. Nevertheless, different tensor decompositions define various notions of tensor rank. The most mainstream ones are related to the *CANDECOMP/PARAFAC* (CP) decomposition [55], the Tucker decomposition [52], [54], [70], the *Tensor Singular Value Decomposition* (T-SVD) [56], [57], [60]–[63], and the *Tensor Ring* (TR) decomposition [58], [71]. As a result, many researchers have explored different strategies to characterize the L prior in the context of HAD. For instance, Wang et al.

[60] originally introduced the concept of tensor LRR into HAD task, which exploits the L prior of background tensor via the weighted *tensor nuclear norm* (TNN). To improve detection capacity, the weighted tensor Schatten- p ($0 < p < 1$) norm was utilized to estimate the low-rank background in the HAD method named S²ELR [56]. To make full use of temporal continuity and spatial correlation, a novel HAD method based on nonconvex tensor Gamma-norm was proposed in [72]. In addition, Qin et al. [61] replaced the convex TNN with a generalized nonconvex surrogate and developed an effective low-rank tensor representation model for the HAD task.

In addition to investigating the low-rank characteristics, several studies have also focused on exploring the local smoothness (denoted as “S”) property of the background to jointly boost detection performance [52], [54], [57], [58], [62], [63], [70], [71], [73]. For instance, Li et al. [52] designed a new HAD algorithm, called PTA, which leverages the spectral low-rank property and spatial smoothness of the background. Zhao et al. [70] proposed a novel HAD approach based on tensor adaptive reconstruction cascaded with global and local feature fusion. By utilizing TNN and *total variation* (TV) regularization constraints, Sun et al. [62] suggested a novel HAD algorithm, named LARTVAD. Furthermore, TR factorization and TV regularization constraints were introduced to explore the low-rankness and piecewise-smoothness of the background in all dimensions [71]. Nevertheless, the performance of these methods is highly affected by the trade-off parameters imposed between L and S regularizers. To alleviate the above issue, under the high-order T-SVD framework [64], [65], Wang et al. [57] developed a new regularizer, named T-CTV, to simultaneously encode the L+S priors of the background with a unique concise term. Similarly, another method in [54] discovered the sparsity of the core tensor from the Tucker decomposition of the gradient tensor and developed a regularization term that simultaneously captures the L+S priors of the background. Despite these advances, the mentioned HAD methods still have limitations and cannot fully and flexibly extract prior information from complex backgrounds.

B. Research Motivations

1) *Why consider utilizing TR framework?*: The TR factorization [74] approximates a high-dimensional tensor as a multilinear products over a sequence of cyclically contracted low-dimensional cores. In practice, TR factorization can enhance the compression capability and also improve the interpretability of latent factors. The TR rank remains consistently invariant regardless of the cyclic permutation of the factors, thereby flexibly and effectively capturing the underlying intermodal redundancy within the tensor data. Compared with other tensor decompositions, the TR decomposition possesses more flexible, powerful and generalized modeling capacities in a wide range of applications [7], [58], [71], [74]–[80]. Motivated by these merits, in this paper, we consider applying the advanced TR factorization to the HAD problem.

2) *Why consider devising enhanced unified nonconvex TR factors regularization?*: The TR rank minimization and low-rank factorization schemes [7], [58], [71], [74]–[80] have become a hot research topic on the purpose of accurately

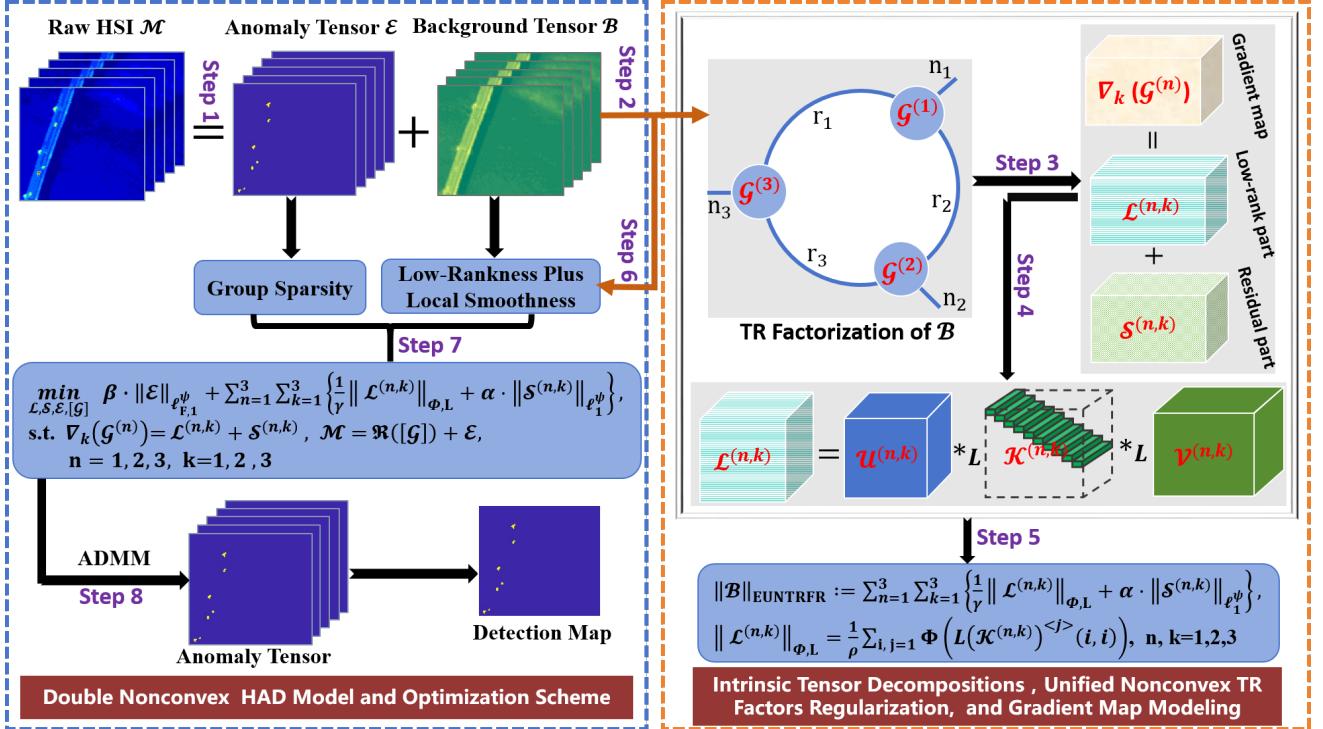


Fig. 1: The detailed flowchart of the proposed HAD-EUNTRFR method. **Module 1 (Right side):** Under a new prior representation paradigm, this module focuses on investigating novel unified nonconvex tensor ring factor regularization strategies; **Module 2 (Left side):** Based on the new regularization methods tailored for the background and anomaly tensors, the second module primarily aims to design effective, scalable and reliable HAD model, as well as optimization algorithm with convergence guarantees.

uncovering the latent space of TR factors. For example, the nuclear norm regularization of the third TR cores with mode-2 unfolding was introduced to further exploit the global spectral low-rank property of high-resolution HSIs [7]. Moreover, to boost the recovery performance, Yuan et al. [75] imposed the the nuclear norm regularization on the unfolding matrices of each TR factor. Recently, Wu et al. [80] empirically and theoretically investigated the physical interpretation of gradient TR factors, and found that the mode-2 unfoldings of gradient TR factors inherit the L+S properties of the original tensor. Inspired by this exploration, they consider the low-rankness and sparsity priors of gradient factors to boost the performance and robustness of TR-based model. However, although acceptable performance can be achieved from the above TR-factors-based modeling methods, they generally unfolded the TR factors to mode- n matrix which may result in loss of optimality in the representation. Therefore, this paper considers investigating a more essential regularization method to reveal the structural information of TR factors, and then applies it to *hyperspectral anomaly detection* task.

Existing HAD methods (e.g., [54], [71]) based on low-rank tensor modeling suffer from some issues, such as insufficient priors representation, loose convex approximation, and the absence of unified regularization form. The method we propose is intended to mitigate the aforementioned problems. Driven by the proven effectiveness and advantages of prior characterization through gradient maps-based modeling strategy [57], [81], [82], we consider imposing an effective regularizer that resembles the T-CTV constraint [57] on the 3-D TR factors.

This regularizer is intended to deeply capture the internal structure of TR factors, serving as a more advantageous option compared to the matrix rank minimization or low-rank matrix factorization techniques employed in [7], [75], [80]. In other words, the key techniques underlying our regularization item will draw on the idea of the TCTV-based HAD method [57], thus enabling it to simultaneously characterize the L+S priors of the background tensor. However, the convex T-CTV regularization scheme still has some room for further improvement. Firstly, the T-CTV does not coordinate the low-rankness and sparsity of gradient maps well, and its robustness requires further enhancement. Secondly, the T-CTV equally treats each singular component of the background tensor in the gradient domain and neglects the physical significance of the different singular values, which may lead to a biased approximation and fail to retain some major information. To resolve the above issues, with the help of a novel nonconvex regularization paradigm, we ultimately consider establishing a generalized and efficient method to encode intrinsic prior structures underlying gradient TR factors. Within our prior characterization scheme, the interpretable gradient TR factors will adeptly inherit the L+S priors of the original background. More importantly, how to enhance the robustness of the proposed regularization term is also taken into account.

C. Proposed HAD Method

In this study, we put forward a novel unified nonconvex HAD method termed HAD-EUNTRFR. The detailed flowchart of the proposed framework is provided in Figure 1. Specifi-

cally, in **Step 1**, the original HSI \mathcal{M} are separated into the background and anomaly components (i.e., \mathcal{B} and \mathcal{E}) using the concept of *tensor robust principal component analysis* (TRPCA). The TR decomposition is then performed on the background tensor \mathcal{B} in **Step 2**, i.e., $\mathcal{B} = \mathfrak{R}([\mathcal{G}])$, $[\mathcal{G}] := \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)}\}$, where $\mathfrak{R}([\mathcal{G}])$ is the TR decomposition of \mathcal{B} , $\{\mathcal{G}^{(k)}\}_{k=1}^3$ are the TR factors. In **Step 3**, a joint low-rank plus sparse decomposition is performed on each gradient TR factor (namely, TR factors in the gradient domain), i.e., $\nabla_k(\mathcal{S}^{(n)}) = \mathcal{L}^{(n,k)} + \mathcal{S}^{(n,k)}$, $n, k = 1, 2, 3$. The low-rank part is further decomposed by T-SVD in **Step 4**, i.e., $\mathcal{L}^{(n,k)} = \mathcal{U}^{(n,k)} *_{\mathfrak{L}} \mathcal{K}^{(n,k)} *_{\mathfrak{L}} (\mathcal{V}^{(n,k)})^T$, where $\mathcal{U}^{(n,k)}, \mathcal{V}^{(n,k)}$ are orthogonal tensors, $\mathcal{K}^{(n,k)}$ is a f-diagonal tensor, $*_{\mathfrak{L}}$ denotes the *tensor-tensor product*. In **Step 5**, we devise an enhanced unified nonconvex TR factors regularization (*please see III-A for more details*). The regularization scheme discussed in this text is designed to be robust against the TR rank selection, while leveraging the prior information of gradient TR factors effectively. This approach enables the interpretable gradient TR factors to skillfully inherit the low-rankness and smoothness of the original background. The procedure outlined in **step 2-5** constitutes a new joint L+S prior characterization paradigm for the background tensor. In **Steps 6-7**, the method combines the proposed EUNTRFR regularizer with another anomaly-sparsity regularizer, leading to the formulation of a new generalized nonconvex HAD model (*please see III-B for more details*). The formulated model is then solved in **Step 8** using Algorithm 3, which is derived from the ADMM framework, as detailed in Section III-D.

D. Main Contributions

The main contributions of this article are as follows:

1) We introduce an innovative unified nonconvex HAD method, termed HAD-EUNTRFR, by integrating several crucial technologies such as TRPCA, tensor decompositions, tensor-correlated TV regularization, and low-rank plus sparse gradient map modeling. This method concisely and effectively captures both global correlations and local smoothness in the background within the spectral-spatial domains, while also addressing the structured sparsity of anomalies.

2) Building on the powerful representation capabilities of TR factorization, we propose a novel generalized nonconvex HAD model. This model incorporates a unified nonconvex regularization term known as UNTRFR, and its enhanced version, EUNTRFR (*please see III-A for more details*). These components efficiently encode both low-rankness and sparsity in the gradient TR factors, providing a concise and effective approach for anomaly detection.

3) In the algorithmic development, we provide new solution paradigm tailored for key subproblems involving a family of generalized nonconvex functions. We then derive an optimization algorithm based on the ADMM framework, enabling the efficient solution of the formulated HAD model. Theoretical results on convergence analysis for the proposed nonconvex algorithm are provided. Experimental results on extensive HSI datasets show that our approach significantly outperforms current state-of-the-art methods, effectively suppressing the background and enhancing the detection of anomalous targets.

II. NOTATIONS AND PRELIMINARIES

Some frequently-employed symbols and preliminaries are summarized in this section.

We use a , \mathbf{a} , \mathbf{A} , and \mathcal{A} to denote scalars, vectors, matrices, and tensors, respectively. For a matrix $\mathbf{A} \in \mathbb{R}^{n_1 \times n_2}$, $\mathbf{I}_n \in \mathbb{R}^{n \times n}$, $\text{Tr}(\mathbf{A})$, $\mathbf{A}^H(\mathbf{A}^T)$, $\langle \mathbf{A}, \mathbf{B} \rangle = \text{Tr}(\mathbf{A}^H \cdot \mathbf{B})$ and $\|\mathbf{A}\|_* = (\sum_i |\sigma_i(\mathbf{A})|)$ denote its identity matrix, trace, conjugate transpose (transpose), inner product and nuclear norm, respectively. For an order- N tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \dots \times n_N}$, $\mathcal{A}_{i_1, \dots, i_N}$ denotes its (i_1, \dots, i_N) -th element, $\mathcal{A}_{(k)} \in \mathbb{R}^{n_k \times \prod_{j \neq k} n_j}$ denotes its *classical mode-k unfolding*, $\mathcal{A}_{<k>} \in \mathbb{R}^{n_k \times (n_{k+1} \dots n_{N-1} \dots n_{k-1})}$ denotes its *reversed mode-k unfolding*, and $\mathcal{A}^{<j>} := \mathcal{A}(:, :, i_3, \dots, i_N), j = \sum_{a=4}^N (i_a - 1)\prod_{b=3}^{a-1} n_b + i_3$ is called as its *face slice*. The *inner product* of two tensors \mathcal{A} and \mathcal{B} with the same size is defined as $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{j=1}^{n_3 n_4 \dots n_N} \langle \mathcal{A}^{<j>}, \mathcal{B}^{<j>} \rangle$. The ℓ_1 -norm, Frobenius norm, infinity norm and $\ell_{F,1}$ -norm of \mathcal{A} are defined as $\|\mathcal{A}\|_1 = (\sum_{i_1 \dots i_N} |\mathcal{A}_{i_1 \dots i_N}|)$, $\|\mathcal{A}\|_F = (\sum_{i_1 \dots i_N} |\mathcal{A}_{i_1 \dots i_N}|^2)^{\frac{1}{2}}$, $\|\mathcal{A}\|_\infty = \max_{i_1, \dots, i_N} |\mathcal{A}_{i_1, \dots, i_N}|$, $\|\mathcal{A}\|_{F,1} = \sum_{i_1, i_2} \|\mathcal{A}_{i_1 i_2, \dots,}\|_F$, respectively. The mode- k product of tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \dots \times n_k \times \dots \times n_N}$ with matrix $\mathbf{M} \in \mathbb{R}^{J \times n_k}$ is $\mathcal{B} = \mathcal{A} \times_k \mathbf{M} \in \mathbb{R}^{n_1 \times \dots \times J \times \dots \times n_N}$, where $\mathcal{B} = \mathcal{A} \times_n \mathbf{M} \iff \mathcal{B}_{(n)} = \mathbf{M} \cdot \mathcal{A}_{(n)}$.

T-SVD framework: Let $\mathfrak{L}(\mathcal{A})$ or $\mathcal{A}_{\mathfrak{L}}$ represent the result of invertible linear transforms \mathfrak{L} on $\mathcal{A} \in \mathbb{R}^{n_1 \times \dots \times n_N}$, i.e.,

$$\mathfrak{L}(\mathcal{A}) = \mathcal{A} \times_3 \mathbf{U}_{n_3} \times_4 \mathbf{U}_{n_4} \dots \times_N \mathbf{U}_{n_N}, \quad (1)$$

where the transform matrices $\mathbf{U}_{n_i} \in \mathbb{C}^{n_i \times n_i}$ of \mathfrak{L} satisfies:

$$\mathbf{U}_{n_i} \cdot \mathbf{U}_{n_i}^H = \mathbf{U}_{n_i}^H \cdot \mathbf{U}_{n_i} = \alpha_i \mathbf{I}_{n_i}, \forall i \in \{3, \dots, N\}, \quad (2)$$

in which $\alpha_i > 0$ is a constant. The inverse operator of $\mathfrak{L}(\mathcal{A})$ is defined as $\mathfrak{L}^{-1}(\mathcal{A}) = \mathcal{A} \times_N \mathbf{U}_{n_N}^{-1} \times_{N-1} \mathbf{U}_{n_{N-1}}^{-1} \dots \times_3 \mathbf{U}_{n_3}^{-1}$, and $\mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{A})) = \mathcal{A}$.

The *tensor-tensor product* of \mathcal{A} and \mathcal{B} under transform \mathfrak{L} is defined as $\mathcal{C} = \mathcal{A} *_{\mathfrak{L}} \mathcal{B} = \mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{X}) \triangle \mathfrak{L}(\mathcal{Y}))$, where \triangle denotes the face-wise product ($\mathcal{P} = \mathcal{M} \triangle \mathcal{N}$ implies $\mathcal{P}^{<j>} = \mathcal{M}^{<j>} \cdot \mathcal{N}^{<j>}$, $j = 1, \dots, n_3 \dots n_N$) [64], [65]. Transform \mathfrak{L} -based T-SVD of \mathcal{A} is denoted as $\mathcal{A} = \mathcal{U} *_{\mathfrak{L}} \mathcal{K} *_{\mathfrak{L}} \mathcal{V}^T$, where \mathcal{U} and \mathcal{V} are orthogonal, \mathcal{K} is f-diagonal, $\mathcal{V}^H(\mathcal{V}^T)$ denotes conjugate transpose (transpose) [64], [65].

TR algebraic framework: For an order- N tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times \dots \times n_N}$ with TR rank $[r_1, \dots, r_N]$, its TR decomposition is represented as a sequence of circularly contracted core tensor $\mathcal{G}^{(k)} \in \mathbb{R}^{r_k \times n_k \times r_{k+1}}$, $k = 1, 2, \dots, N$, with $r_{N+1} = r_1$. Specifically, the element-wise relation of tensor \mathcal{A} and its TR core tensors $\{\mathcal{G}^{(k)}\}_{k=1}^N$ is defined as

$$\mathcal{A}_{i_1 \dots i_N} = \text{Tr} \left(\prod_{k=1}^N \mathcal{G}^{(k)}(:, i_k, :) \right),$$

where $\mathcal{G}^{(k)}(:, i_k, :) \in \mathbb{R}^{r_k \times r_{k+1}}$ is the i_k -th slice matrix of $\mathcal{G}^{(k)}$ along mode-2. In this article, the TR decomposition is abbreviated as $\mathcal{A} = \mathfrak{R}([\mathcal{G}])$, $[\mathcal{G}] := \{\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(N)}\}$.

Given a TR decomposition $\mathcal{A} = \mathfrak{R}([\mathcal{G}])$, its mode- k unfolding matrix can be written as

$$\mathbf{A}_{<k>} = \mathbf{G}_{(2)}^{(k)} \cdot (\mathbf{G}_{<2>}^{(\neq k)})^T, \quad (3)$$

where $\mathcal{G}^{(\neq k)}$ is a subchain tensor obtained by merging all cores except the k -th core $\mathcal{G}^{(k)}$ [74].

III. UNIFIED NONCONVEX HAD FRAMEWORK

A. Unified Nonconvex Regularizers

1) Unified Anomaly-Sparsity Regularizer: In order to effectively recognize the anomalous component existed in HSI data, we hence consider defining the following unified nonconvex regularization penalty, i.e.,

$$\Upsilon(\mathcal{E}) := \psi(h(\mathcal{E})), \quad (4)$$

where $\psi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ stands for a generalized nonconvex function. Here, two types of sparse anomalies are taken into account. If the anomaly tensor \mathcal{E} has structured sparsity on the tubes, i.e., $h(\cdot) = \|\cdot\|_{F,1}$, then we have

$$\|\mathcal{E}\|_{\ell_{F,1}^\psi} := \psi(\|\mathcal{E}\|_{F,1}) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \psi(\|\mathcal{E}_{i_1 i_2}\|_F). \quad (5)$$

If the tensor \mathcal{E} is an entry-wise anomaly tensor, i.e., $h(\cdot) = \|\cdot\|_1$, then we have

$$\|\mathcal{E}\|_{\ell_1^\psi} := \psi(\|\mathcal{E}\|_1) = \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_2} \sum_{i_3=1}^{n_3} \psi(|\mathcal{E}_{i_1 i_2 i_3}|). \quad (6)$$

Assumption III.1. *The generalized nonconvex function $\psi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ satisfies the following assumptions:*

- **(I):** $\psi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is proper, lower semi-continuous and symmetric with respect to y-axis;
- **(II):** $\psi(\cdot)$ is concave and monotonically increasing on $[0, \infty)$ with $\psi(0) = 0$.

Some popular nonconvex penalty functions satisfying the above Assumption III.1 are summarized as follows, where we only consider the parameters of the nonconvex function on $[0, \infty)$ since they are symmetric with respect to y-axis.

- **Lp** [83]: the ℓ_p penalty function is defined by $\psi^{\text{Lp}}(x) = |x|^p$, where $0 < p < 1$.
- **Log** [84]: the logarithmic (Log) function is defined by $\psi^{\text{Log}}(x) = \log(1 + \frac{|x|}{\theta})$, where $\theta > 0$.
- **MCP** [85]: the minimax concave penalty (MCP) function is defined by

$$\psi^{\text{MCP}}(x) = \begin{cases} x - \frac{x^2}{2\eta}, & 0 \leq x \leq \eta, \\ \frac{\eta}{2}, & x > \eta > 0. \end{cases}$$

- **Capped Lp:** the capped ℓ_p penalty function is defined by $\psi^{\text{CapLp}}(x) = \min\{1, \frac{x^p}{v^p}\}$, where $v > 0$, $0 < p < 1$.
- **Capped Log:** the capped logarithm function is defined by $\psi^{\text{CapLog}}(x) = \min\{1, \frac{1}{\psi^{\text{Log}}(v)} \psi^{\text{Log}}(x)\}$, where $v > 0$.
- **Capped MCP:** the capped MCP function is defined by

$$\psi^{\text{CapMCP}}(x) = \min\left\{1, \frac{2\eta}{v(2\eta - v)} \psi^{\text{MCP}}(x)\right\}, 0 < v < \eta.$$

It is worth noting that the above-mentioned capped folded concave functions originate from the reference [86]. Besides, the proximal operator of v with respect to $\psi(\cdot)$ is defined as

$$\text{Prox}_{\psi,\mu}(v) = \arg \min_x \left\{ \mu\psi(x) + \frac{1}{2}(x - v)^2 \right\}, \quad (7)$$

where $\mu > 0$ is a penalty parameter. For the special forms of the nonconvex function $\psi(\cdot)$, their proximal mappings often have analytical expressions, which can be found in the relevant references [83]–[86].

2) Novel Regularizer Encoding the Prior Structures of TR Factors

Based on the characteristics of HSI datasets, we put forward a novel *unified nonconvex TR factors regularization* (UNTRFR). The proposed regularization scheme, along with its enhanced variant, is specifically tailored to encode the crucial prior structures (i.e., **L+S**) of the background tensor $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Before formally introducing this novel regularizer, we first define a *generalized nonconvex tensor correlated total variation* (GNCTV) norm and its enhanced version under the T-SVD framework.

Definition III.1. (Gradient tensor [57]) For $\mathcal{B} \in \mathbb{R}^{n_1 \times \dots \times n_d}$, its gradient tensor along the k -th mode is defined as

$$\nabla_k(\mathcal{B}) := \mathcal{B} \times_k \mathbf{D}_{n_k}, \quad k = 1, 2, \dots, d, \quad (8)$$

where \mathbf{D}_{n_k} is a row circulant matrix of $(-1, 1, 0, \dots, 0)$.

Definition III.2. For any $\mathcal{B} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denote $\Gamma \in \{1, 2, 3\}$ as a priori set consisting of directions along which \mathcal{B} equips **L+S** priors, and $\nabla_k(\mathcal{B})$, $k \in \Gamma$ as its correlated gradient tensors. Then, its GNTCTV norm is defined as

$$\begin{aligned} \|\mathcal{B}\|_{\text{GNCTV}} &:= \frac{1}{\gamma} \sum_{k=1}^3 \|\nabla_k(\mathcal{B})\|_{\Phi, \mathfrak{L}}, \\ &= \frac{1}{\gamma} \sum_{k=1}^3 \left(\frac{1}{\rho} \sum_{i=1}^{\min(n_1, n_2)} \sum_{j=1}^{n_3} \Phi(\sigma_{ij}^{(k)}) \right), \end{aligned} \quad (9)$$

where $\gamma := \#\{\Gamma\}$ equals to the cardinality of Γ , $\rho > 0$ is a constant determined by transform \mathfrak{L} satisfying Formulas (1)-(2), $\mathcal{K}^{(k)}$ is the T-SVD component of $\nabla_k(\mathcal{B}) = \mathcal{U}^{(k)} *_{\mathfrak{L}} \mathcal{K}^{(k)} *_{\mathfrak{L}} (\mathcal{V}^{(k)})^T$, $\sigma_{ij}^{(k)} = \mathfrak{L}(\mathcal{K}^{(k)})^{<j>}_{(i,i)}$, and $\Phi(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is a generalized nonconvex function, which has the same properties as the function $\psi(\cdot)$ defined in (4).

Remark III.1. Drawing inspiration from the principles of correlated total variation (CTV) [81], [82] and tensor CTV (T-CTV) [57] methods, we came up with the aforementioned GNTCTV regularizer (9), which can simultaneously characterize both **L** and **S** priors of HSI's background with a unique term. This single regularizer can promote the two priors concurrently, and overcome the difficulty of tuning the trade-off parameter imposed between these two regularizers. In addition, compared with the convex T-CTV constraint, the GNTCTV regularizer involving a class of nonconvex penalty functions considers the importance of different singular components of each gradient tensor, thus providing better approximation to the background tensor. Note that when the nonconvex function $\Phi(\cdot)$ is set to be an ℓ_1 penalty function, the GNTCTV can be degenerated to the T-CTV regularizer proposed in [57]. When the order of tensor \mathcal{B} is 2, the regularizer (9) is equivalent to

$$\|\mathcal{B}\|_{\text{GNCTV}} := \|\nabla_1(\mathcal{B})\|_{\Phi} = \sum_{i=1}^{\min(n_1, n_2)} \Phi(S(i,i)), \quad (10)$$

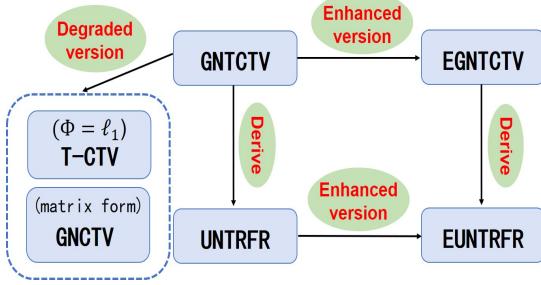


Fig. 2: The relationship of various regularization methods. where S is the SVD component of $\nabla_1(\mathcal{B}) = \mathbf{U}S\mathbf{V}^T$. In this article, the regularizer (10) is called as the generalized nonconvex CTV (GNCTV) of matrix \mathcal{B} , which can be viewed as the nonconvex extensions of CTV norm [82].

To further enhance the robustness of the formulated regularizer (9), we first proposed decomposing the gradient tensor $\nabla_k(\mathcal{B})$, $k = 1, 2, 3$, into a low-rank term \mathcal{L}_k and a sparse term \mathcal{S}_k , i.e., $\nabla_k(\mathcal{B}) = \mathcal{L}_k + \mathcal{S}_k$. Then, the generalized nonconvex regularizers are utilized to encode them. In other words, considering that the background tensor \mathcal{B} may not have exact low-rank structure in the gradient domain, we introduced a residual term on the basis of the regularizer (9), i.e.,

$$\|\mathcal{B}\|_{\text{EGNTCTV}} := \frac{1}{\gamma} \sum_{k=1}^3 \|\mathcal{L}_k\|_{\Phi, \mathfrak{L}} + \alpha \sum_{k=1}^3 \|\mathcal{S}_k\|_{\ell_1^\psi}, \quad (11)$$

where $\alpha > 0$ is a balancing parameter, $\|\cdot\|_{\ell_1^\psi}$ and $\|\cdot\|_{\Phi, \mathfrak{L}}$ are defined according to Formulas (6) and (9), respectively. Note that the regularizer (11) is named as the *enhanced GNTCTV* (EGNTCTV) of \mathcal{B} .

Existing studies [71], [80] have shown that TR factors possess both global low-rankness and sparsity in gradient domain (i.e., the L+S properties of TR factors itself). Differently from these methods, we investigate the low-rankness and sparsity of gradient TR factors under the T-SVD framework, rather than the matrix unfolding scheme that may result in loss of optimal representation. That is to say, treating the 3-D TR factors as the low-tubal-rank tensors and borrowing the idea of the GNTCTV regularizer (9), we propose the following regularization item, called UNTRFR, to measure the L+S properties of TR factors of the background tensor:

$$\|\mathcal{B}\|_{\text{UNTRFR}} := \frac{1}{\gamma} \sum_{n=1}^3 \sum_{k=1}^3 \|\nabla_k(\mathcal{G}^{(n)})\|_{\Phi, \mathfrak{L}}, \quad (12)$$

where $\mathfrak{R}([\mathcal{G}])$ ($[\mathcal{G}] := \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \mathcal{G}^{(3)}\}$) is the TR decomposition of \mathcal{B} , $\{\mathcal{G}^{(k)}\}_{k=1}^3$ are the TR factors.

Similarly to the definition of the EGNTCTV regularizer (11), we put forward an enhanced version of UNTRFR regularizer (12), known as EUNTRFR. Specifically, we factorize the gradient TR factor $\nabla_k(\mathcal{G}^{(n)})$, $n, k = 1, 2, 3$ into a low-rank component $\mathcal{L}^{(n,k)}$ and a sparse component $\mathcal{S}^{(n,k)}$, i.e., $\nabla_k(\mathcal{G}^{(n)}) = \mathcal{L}^{(n,k)} + \mathcal{S}^{(n,k)}$. Then, the EUNTRFR regularizer is defined as follows

$$\|\mathcal{B}\|_{\text{EUNTRFR}} := \frac{1}{\gamma} \sum_{n=1}^3 \sum_{k=1}^3 \|\mathcal{L}^{(n,k)}\|_{\Phi, \mathfrak{L}} + \alpha \sum_{n=1}^3 \sum_{k=1}^3 \|\mathcal{S}^{(n,k)}\|_{\ell_1^\psi}. \quad (13)$$

Note that the above regularizer (13) can be viewed as a robust extension of the regularization item (12).

Figure 2 clarifies the relationships among the various nonconvex regularization methods (including GNCTV, GNTCTV, EGNCTCTV, UNTRFR, EUNTRFR) designed above.

B. Generalized Nonconvex Model

In this subsection, a novel HAD model using unified nonconvex factors regularization is proposed on the basis of Subsection III-A, i.e.,

$$\min_{\mathcal{E}, \mathcal{B}} \|\mathcal{B}\|_{\text{EUNTRFR}} + \beta \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi}, \text{ s.t. } \mathcal{M} = \mathcal{B} + \mathcal{E}, \quad (14)$$

where \mathcal{M} , \mathcal{B} and \mathcal{E} respectively denote observed HSI, the background component and the abnormal component, the definitions of $\|\cdot\|_{\ell_{F,1}^\psi}$ and $\|\cdot\|_{\text{EUNTRFR}}$ are consistent with Formulas (5) and (13), and $\alpha, \beta > 0$ are the trade-off parameters.

The proposed HAD method differs from the previous researches fused L and S properties (e.g., [52], [54], [56]–[59], [61], [63], [71], [87], [88]). For instance, in TRDFT-VAD method [71], TR factorization and TV regularization constraints were introduced to explore the L+S natures of the background tensor. However, this convex TRDFTVAD method did not take the prior characteristics of the TR factors into consideration, and fail to possess a more unified, efficient and concise form. The proposed UNTRFR and EUNTRFR regularization strategies alleviate these limitations. Different from the nonconvex low-rank regularization methods that previously acted on the original domain [56], [61], [63], [87], [88], the method we propose is modeled from the perspective of gradient mapping, enabling it to simultaneously mine the L+S priors in a more concise yet effective form.

In previous work [58], a circular TR unfolding strategy was used alongside nonconvex constraints on the gradient maps of the background tensor to develop a generalized and effective regularizer. This approach effectively extracts both the L and S priors in a concise and unified manner. Whereas, this paper proposes novel unified nonconvex TR factors regularization methods, whose key techniques (i.e., GNTCTV, EGNCTCTV) differ from the convex TCTV-based HAD approach [57]. By applying the new representation paradigm to the 3-D gradient TR factors derived from the background tensor, the proposed method can encode both L and S properties simultaneously. More importantly, to achieve an improved and more comprehensive solution for background and anomaly modeling, a family of generalized nonconvex functions and their corresponding modified versions (i.e., capped folded concave functions) are included in our HAD model (14). Therefore, previous related methods [57], [58] and the proposed one are entirely different regularization schemes for extracting L+S prior features.

C. Generalized solution paradigm for Key Subproblems

This subsection mainly presents the solution method of two kinds of key subproblems involved in our model (14), i.e.,

$$\arg \min_{\mathcal{E}} \lambda \cdot \psi(h(\mathcal{E})) + \frac{1}{2} \|\mathcal{E} - \mathcal{A}\|_F^2, \quad (15)$$

$$\arg \min_{\mathcal{L}} \tau \cdot \|\mathcal{L}\|_{\Phi, \mathfrak{L}} + \frac{1}{2} \|\mathcal{L} - \mathcal{A}\|_F^2. \quad (16)$$

Algorithm 1: Generalized thresholding operator.

```

1 Input:  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ ,  $\lambda > 0$ ,  $\psi(\cdot)$ .
2 if  $h(\cdot) = \|\cdot\|_1$  then
3    $\hat{\mathcal{E}}_{i_1 i_2 i_3} = \text{Prox}_{\psi, \lambda}(\mathcal{A}_{i_1 i_2 i_3})$ ;
4 end
5 else if  $h(\cdot) = \|\cdot\|_{F,1}$  then
6   for  $i = 1, 2, \dots, n_1$  do
7     for  $j = 1, 2, \dots, n_2$  do
8        $\hat{\mathcal{E}}_{ij} := \frac{\mathcal{A}_{ij}}{\|\mathcal{A}_{ij}\|_F} \cdot \text{Prox}_{\psi, \lambda}(\|\mathcal{A}_{ij}\|_F)$ ;
9     end
10 end
Output:  $\hat{\mathcal{E}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ .

```

Algorithm 2: GNTSVT operation, $\mathcal{D}_{\Phi, \tau}(\mathcal{A}, \mathfrak{L})$.

```

1 Input:  $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ , transform:  $\mathfrak{L}$ ,  $\Phi(\cdot)$ ,  $\tau > 0$ .
2 Compute the results of  $\mathfrak{L}$  on  $\mathcal{A}$ , i.e.,  $\mathfrak{L}(\mathcal{A})$ ;
3 for  $v = 1, 2, \dots, n_3$  do
4    $[\mathfrak{L}(\mathcal{U})^{<v>}, \mathfrak{L}(\mathcal{S})^{<v>}, \mathfrak{L}(\mathcal{V})^{<v>}] = \text{svd}(\mathfrak{L}(\mathcal{A})^{<v>})$  ;
5    $\hat{\mathcal{S}} = \text{Prox}_{\Phi, \tau}[\text{diag}(\mathfrak{L}(\mathcal{S})^{<v>})]$ ;
6    $\mathfrak{L}(\mathcal{C})^{<v>} = \mathfrak{L}(\mathcal{U})^{<v>} \cdot \text{diag}(\hat{\mathcal{S}}) \cdot (\mathfrak{L}(\mathcal{V})^{<v>})^T$ ;
end
Output:  $\mathcal{D}_{\Phi, \tau}(\mathcal{A}, \mathfrak{L}) \leftarrow \mathfrak{L}^{-1}(\mathfrak{L}(\mathcal{C}))$ .

```

1) **Solve problem (15) via Generalized Nonconvex Thresholding/Shrinkage Operator:** The optimal solution to the subproblem (15) can be computed by the generalized nonconvex thresholding/shrinkage operator in an element-wise or tube-wise manner. Please see Algorithm 1 for more details.

2) **Solve problem (16) via Generalized Nonconvex Tensor Singular Value Thresholding (GNTSVT) Operator:** The close-form solution to the problem (16) can be obtained via the following key Theorem III.1. Note that the computation process of GNTSVT operation can be found in Algorithm 2.

Theorem III.1. (GNTSVT) [68], [89] Let the T-SVD decomposition of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ be $\mathcal{A} = \mathcal{U} *_{\mathfrak{L}} \mathcal{S} *_{\mathfrak{L}} \mathcal{V}^T$, $m = \min(n_1, n_2)$. For any $\tau \geq 0$, then a global optimal solution to the optimization problem (16) is given by

$$\mathcal{D}_{\Phi, \tau}(\mathcal{A}, \mathfrak{L}) = \mathcal{U} *_{\mathfrak{L}} \tilde{\mathcal{Y}} *_{\mathfrak{L}} \mathcal{V}^T,$$

where $\tilde{\mathcal{Y}}$ is an f -diagonal tensor, $\mathfrak{L}(\tilde{\mathcal{Y}})^{<t>}(i, i) = \text{Prox}_{\Phi, \tau}(\mathfrak{L}(\mathcal{S})^{<t>}(i, i))$, $t \in \{1, \dots, n_3\}$, $i \in \{1, \dots, m\}$, and $\text{Prox}_{\Phi, \tau}(\cdot)$ denotes the proximity operator of nonconvex penalty function $\Phi(\cdot)$, which has the same properties as $\psi(\cdot)$.

D. Optimization Algorithm

In this subsection, the well-known ADMM framework [90] is adopted to solve the proposed model (14). The nonconvex model (14) can be equivalently reformulated as follows:

$$\begin{aligned}
& \min_{\mathfrak{G}^{(n)}, \mathcal{L}^{(n,k)}, \mathcal{S}^{(n,k)}, \mathcal{E}} \sum_{n=1}^3 \sum_{k=1}^3 \left\{ \frac{1}{\gamma} \|\mathcal{L}^{(n,k)}\|_{\Phi, \mathfrak{L}} + \alpha \cdot \|\mathcal{S}^{(n,k)}\|_{\ell_1^\psi} \right\} \\
& + \beta \cdot \|\mathcal{E}\|_{\ell_{F,1}}, \\
& \text{s.t. } \mathfrak{M} = \mathcal{B} + \mathcal{E} = \mathfrak{R}([\mathfrak{G}]) + \mathcal{E}, \\
& \nabla_k(\mathfrak{G}^{(n)}) = \mathcal{L}^{(n,k)} + \mathcal{S}^{(n,k)}, \quad n, k \in \{1, 2, 3\}.
\end{aligned} \tag{17}$$

The augmented Lagrangian function of (17) is

$$\begin{aligned}
& \mathcal{F}(\mathcal{L}^{(n,k)}, \mathcal{S}^{(n,k)}, \mathcal{Q}^{(n,k)}, [\mathfrak{G}], \mathcal{E}, \mathcal{Y}) = \sum_{n=1}^3 \sum_{k=1}^3 \left\{ \frac{1}{\gamma} \|\mathcal{L}^{(n,k)}\|_{\Phi, \mathfrak{L}} \right. \\
& + \alpha \cdot \|\mathcal{S}^{(n,k)}\|_{\ell_1^\psi} + \langle \mathcal{Q}^{(n,k)}, \nabla_k(\mathfrak{G}^{(n)}) - \mathcal{L}^{(n,k)} - \mathcal{S}^{(n,k)} \rangle \\
& + \frac{\mu}{2} \|\nabla_k(\mathfrak{G}^{(n)}) - \mathcal{L}^{(n,k)} - \mathcal{S}^{(n,k)}\|_F^2 \Big\} + \beta \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi} \\
& + \langle \mathcal{Y}, \mathfrak{M} - \mathfrak{R}([\mathfrak{G}]) - \mathcal{E} \rangle + \frac{\mu}{2} \|\mathfrak{M} - \mathfrak{R}([\mathfrak{G}]) - \mathcal{E}\|_F^2,
\end{aligned} \tag{18}$$

where μ is the regularization parameter, and $\mathcal{Q}^{(n,k)}$, \mathcal{Y} are Lagrange multipliers. It can be further expressed as

$$\begin{aligned}
& \mathcal{F}(\mathcal{L}^{(n,k)}, \mathcal{S}^{(n,k)}, \mathcal{Q}^{(n,k)}, [\mathfrak{G}], \mathcal{E}, \mathcal{Y}) = \sum_{n=1}^3 \sum_{k=1}^3 \left\{ \frac{1}{\gamma} \|\mathcal{L}^{(n,k)}\|_{\Phi, \mathfrak{L}} + \right. \\
& \alpha \cdot \|\mathcal{S}^{(n,k)}\|_{\ell_1^\psi} + \frac{\mu}{2} \|\nabla_k(\mathfrak{G}^{(n)}) - \mathcal{L}^{(n,k)} - \mathcal{S}^{(n,k)} + \frac{\mathcal{Q}^{(n,k)}}{\mu}\|_F^2 \Big\} \\
& + \beta \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi} + \frac{\mu}{2} \|\mathfrak{M} - \mathfrak{R}([\mathfrak{G}]) - \mathcal{E} + \mathcal{Y}/\mu\|_F^2 + C,
\end{aligned} \tag{19}$$

where C is only the multipliers dependent squared items. Below, we show how to solve the subproblems for each involved variable.

Module 1: Update $(\mathfrak{G}^{(n)})^{\{\nu+1\}}$, $n = 1, 2, 3$: Through the matrix representation of TR decomposition (please see Formula (3) for details) and let $\mathbf{R}_1 = (\mathbf{G}_{<2>}^{(\neq n)})^T$. Then, the subproblem with respect to $\mathbf{G}_{(2)}^{(n)}$ can be constructed as

$$\begin{aligned}
& (\mathbf{G}_{(2)}^{(n)})^{\{\nu+1\}} = \arg \min_{\mathbf{G}_{(2)}^{(n)}} \mu^{\{\nu\}} / 2 \|\mathbf{M}_{<n>} - \mathbf{G}_{(2)}^{(n)} \mathbf{R}_1^{\{\nu\}} \\
& - \mathbf{E}_{<n>}^{\{\nu\}} + \mathbf{Y}_{<n>}^{\{\nu\}} / \mu^{\{\nu\}}\|_F^2 + \mu^{\{\nu\}} / 2 \|\mathbf{D}_n \mathbf{G}_{(2)}^{(n)} - (\mathbf{L}_{(2)}^{(n,2)})^{\{\nu\}} \\
& - (\mathbf{S}_{(2)}^{(n,2)})^{\{\nu\}} + (\mathbf{Q}_{(2)}^{(n,2)})^{\{\nu\}} / \mu^{\{\nu\}}\|_F^2.
\end{aligned} \tag{20}$$

Taking the derivative in (20) with respect to $\mathbf{G}_{(2)}^{(n)}$, it obtains the following linear system:

$$\begin{aligned}
& \mathbf{D}_n^T \mathbf{D}_n \mathbf{G}_{(2)}^{(n)} + \mathbf{G}_{(2)}^{(n)} \mathbf{R}_1^{\{\nu\}} (\mathbf{R}_1^{\{\nu\}})^T = (\mathbf{M}_{<n>} - \mathbf{E}_{<n>}^{\{\nu\}} \\
& + \mathbf{Y}_{<n>}^{\{\nu\}} / \mu^{\{\nu\}}) (\mathbf{R}_1^{\{\nu\}})^T + \mathbf{D}_n^T ((\mathbf{L}_{(2)}^{(n,2)})^{\{\nu\}} + (\mathbf{S}_{(2)}^{(n,2)})^{\{\nu\}} \\
& - (\mathbf{Q}_{(2)}^{(n,2)})^{\{\nu\}} / \mu^{\{\nu\}}).
\end{aligned}$$

The above formulation is a classical Sylvester matrix equation, which can be efficiently solved in the light of the reference [91]. Eventually, the updated TR factor tensor $\mathfrak{G}^{(n)}$ is achieved by the ‘fold’ operation, i.e.,

$$(\mathfrak{G}^{(n)})^{\{\nu+1\}} = \text{fold}_2((\mathbf{G}_{(2)}^{(n)})^{\{\nu+1\}}), \tag{21}$$

where $\text{fold}_k(\cdot)$ denotes the inverse operator of mode- k unfolding, which satisfies $\mathcal{A} = \text{fold}_k(\mathbf{A}_{(k)})$, $\mathbf{A}_{(k)} = \text{unfold}_k(\mathcal{A})$.

Module 2: Update $(\mathcal{S}^{(n,k)})^{\{\nu+1\}}$, $n, k = 1, 2, 3$: Similarly, extracting all items containing $\mathcal{S}^{(n,k)}$ from (19), we have

$$\begin{aligned}
& (\mathcal{S}^{(n,k)})^{\{\nu+1\}} = \arg \min_{\mathcal{S}^{(n,k)}} \alpha \cdot \|\mathcal{S}^{(n,k)}\|_{\ell_1^\psi} + \frac{\mu^{\{\nu\}}}{2} \|\mathcal{S}^{(n,k)} \\
& + \nabla_k(\mathfrak{G}^{(n)})^{\{\nu+1\}} - (\mathcal{L}^{(n,k)})^{\{\nu\}} + \frac{(\mathcal{Q}^{(n,k)})^{\{\nu\}}}{\mu^{\{\nu\}}}\|_F^2.
\end{aligned} \tag{22}$$

Module 3: Update $(\mathcal{L}^{(n,k)})^{\{\nu+1\}}$, $n, k = 1, 2, 3$: For each $n = 1, 2, 3$, $k = 1, 2, 3$, extracting all items containing $\mathcal{L}^{(n,k)}$ from (19), we have

$$\begin{aligned} (\mathcal{L}^{(n,k)})^{\{\nu+1\}} &= \arg \min_{\mathcal{L}^{(n,k)}} \frac{1}{\gamma} \|\mathcal{L}^{(n,k)}\|_{\Phi, \mathfrak{L}} + \frac{\mu^{\{\nu\}}}{2} \| - \mathcal{L}^{(n,k)} \\ &+ \nabla_k(\mathcal{G}^{(n)})^{\{\nu+1\}} - (\mathcal{S}^{(n,k)})^{\{\nu+1\}} + \frac{(\mathcal{Q}^{(n,k)})^{\{\nu\}}}{\mu^{\{\nu\}}} \|_F^2. \end{aligned} \quad (23)$$

This key subproblem is analogous to the form of the minimization problem (16), and its close-form solution can be obtained by utilizing the GNTSVT operation (see III-C for details).

Module 4: Update $\mathcal{E}^{\{\nu+1\}}$: The optimization subproblem with respect to \mathcal{E} can be written as

$$\begin{aligned} \mathcal{E}^{\{\nu+1\}} &= \arg \min_{\mathcal{E}} \beta \|\mathcal{E}\|_{\ell_{F,1}^\psi} + \mu^{\{\nu\}} / 2 \|\mathcal{M} - \mathfrak{R}([\mathcal{G}]) \\ &- \mathcal{E} + \mathcal{Y}^{\{\nu\}} / \mu^{\{\nu\}}\|_F^2. \end{aligned} \quad (24)$$

The subproblems (22), (24) are equivalent to the form of the minimization problem (15), and their close-form solutions can be obtained by utilizing the generalized nonconvex shrinkage operator. Please see III-C for details.

Module 5: Update $(\mathcal{Q}^{(n,k)})^{\{\nu+1\}}$, $\mathcal{Y}^{\{\nu+1\}}$, and $\mu^{\{\nu+1\}}$ (Lagrange multipliers and penalty parameter) Based on the rule of the ADMM framework, the lagrange multipliers are updated by the following equations:

$$\mathcal{Y}^{\{\nu+1\}} = \mathcal{Y}^{\{\nu\}} + \mu^{\{\nu\}} (\mathcal{M} - \mathfrak{R}([\mathcal{G}]) - \mathcal{E}^{\{\nu+1\}}), \quad (25)$$

$$(\mathcal{Q}^{(n,k)})^{\{\nu+1\}} = (\mathcal{Q}^{(n,k)})^{\{\nu\}} + \mu^{\{\nu\}} \cdot \mathcal{N}^{\{\nu\}}, \quad (26)$$

$$\mu^{\{\nu+1\}} = \min \left(\mu^{\max}, \vartheta \mu^{\{\nu\}} \right), \quad (27)$$

where $\mathcal{N}^{\{\nu\}} = \nabla_k(\mathcal{G}^{(n)})^{\{\nu+1\}} - (\mathcal{L}^{(n,k)})^{\{\nu+1\}} - (\mathcal{S}^{(n,k)})^{\{\nu+1\}}$, ϑ stands for the growth rate.

Module 6: Update the detection map $\mathbf{T}^{\{\nu+1\}}$: Finally, the anomaly detection map can be obtained by

$$\mathbf{T}^{\{\nu+1\}}(i_1, i_2) = \sqrt{\sum_{i_3=1}^{n_3} |\mathcal{E}^{\{\nu+1\}}(i_1, i_2, i_3)|^2}. \quad (28)$$

The entire ADMM optimization framework is summarized in Algorithm 3.

E. Complexity Analysis

Given an input HSI data with the size of $n_1 \times n_2 \times n_3$, we analyze the per-iteration complexity of Algorithm 3, which mainly needs to update $\{\mathcal{G}^{(n)}\}_{n=1}^3$, $\{\mathcal{L}^{(n,k)}\}_{n,k=1}^3$, $\{\mathcal{S}^{(n,k)}\}_{n,k=1}^3$, \mathcal{E} , $\{\mathcal{Q}^{(n,k)}\}_{n,k=1}^3$, and \mathcal{Y} , respectively. For simplicity, we set $N = n_1 = n_2 = n_3$, $R = r_1 = r_2 = r_3$, and $R < N$. Specifically, **1**) the time complexity of calculating the first term that mainly involves SVD, FFT, and matrix multiplication is $\mathcal{O}(N^3 R^2 + R^6)$; **2**) The update of the second term, which mainly involves the GNTSVT operation and matrix multiplication, equals to $\mathcal{O}(N^2 R^2)$; **3**) the time complexity of computing the third term is $\mathcal{O}(R^2 N^2)$; **4**) Upadating \mathcal{E} requires to perform the generalized thresholding operation and the matrix multiplication with a complexity of

Algorithm 3: Solve the HAD model (14) via ADMM.

Input: HSI data: \mathcal{M} , $\Phi(\cdot)$, $\psi(\cdot)$, α , β , \mathfrak{L} , TR rank: $[r_1, r_2, r_3]$.

1 Initialize: For $n, k = 1, 2, 3$, $(\mathcal{L}^{(n,k)})^{\{0\}} = (\mathcal{S}^{(n,k)})^{\{0\}} = (\mathcal{Q}^{(n,k)})^{\{0\}} = \mathcal{E}^{\{0\}} = \mathcal{Y}^{\{0\}} = \mathbf{0}$, random sample $\mathcal{G}^{(n)}$ by distribution $N \sim (0, 1)$, ϑ , $\mu^{\{0\}}$, μ^{\max} , ϖ , $\nu_{\max} = 500$;

2 while not converged **do**

 3 Update $(\mathcal{G}^{(n)})^{\{\nu+1\}}$ by computing (21);

 4 Update $(\mathcal{S}^{(n,k)})^{\{\nu+1\}}$ by computing (22);

 5 Update $(\mathcal{L}^{(n,k)})^{\{\nu+1\}}$ by computing (23);

 6 Update $\mathcal{E}^{\{\nu+1\}}$ by computing (24);

 7 Update $(\mathcal{Q}^{(n,k)})^{\{\nu+1\}}$, $\mathcal{Y}^{\{\nu+1\}}$, $\mu^{\{\nu+1\}}$ by computing (25)-(27);

 8 Update $\mathcal{B}^{\{\nu+1\}}$ by computing $\mathcal{B}^{\{\nu+1\}} = \mathfrak{R}([\mathcal{G}])$;

 9 Obtain the detection map $\mathbf{T}^{\{\nu+1\}}$ by (28);

 10 Check the convergence conditions
 Error = $\|\mathcal{B}^{\{\nu+1\}} - \mathcal{B}^{\{\nu\}}\|_F / \|\mathcal{B}^{\{\nu\}}\|_F \leq \varpi$;

11 end

Output: background \mathcal{L} , anomalies \mathcal{E} , detection map \mathbf{T} .

5) the time complexity of computing the fifth term is $\mathcal{O}(R^2 N^2)$; **6**) \mathcal{Y} can be updated by a low consumed algebraic computation.

F. Convergence Guarantee

In this subsection, we provide the convergence analysis of Algorithm 3. The detailed proofs of relevant theories and lemmas can be found in our Github: <https://github.com/Qinwenjinwu/QWSWU-Convergence-Theory-Proof>.

Lemma III.1. *The sequences $\{\mathcal{Y}^{\{\nu\}}\}$ and $\{\mathcal{Q}^{(n,k)}\}^{\{\nu\}}$ ($n, k = 1, 2, 3$) generated by Algorithm 3 are bounded.*

Lemma III.2. *Suppose that the sequences $\{\mathcal{Y}^{\{\nu\}}\}$ and $\{\mathcal{Q}^{(n,k)}\}^{\{\nu\}}$ ($n, k = 1, 2, 3$) generated by Algorithm 3 are bounded, then the sequences $\{(\mathcal{L}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{S}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{G}^{(n)})^{\{\nu\}}\}$, and $\mathcal{E}^{\{\nu\}}$ are bounded.*

Theorem III.2. *Suppose that the sequences $\{\mathcal{Y}^{\{\nu\}}\}$ and $\{\mathcal{Q}^{(n,k)}\}^{\{\nu\}}$ ($n, k = 1, 2, 3$) generated by Algorithm 3 are bounded. Then, for any $n, k \in \{1, 2, 3\}$, the sequences $\{(\mathcal{L}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{S}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{G}^{(n)})^{\{\nu\}}\}$, and $\mathcal{E}^{\{\nu\}}$ satisfy:*

- 1) $\lim_{\nu \rightarrow \infty} \|\mathcal{M} - \mathfrak{R}([\mathcal{G}]) - \mathcal{E}^{\{\nu+1\}}\|_F = \|\mathcal{N}^{\{\nu\}}\|_F = 0$;
- 2) $\lim_{\nu \rightarrow \infty} \|(\mathcal{L}^{(n,k)})^{\{\nu+1\}} - (\mathcal{L}^{(n,k)})^{\{\nu\}}\|_F = 0$;
- 3) $\lim_{\nu \rightarrow \infty} \|\mathcal{E}^{\{\nu+1\}} - \mathcal{E}^{\{\nu\}}\|_F = 0$;
- 4) $\lim_{\nu \rightarrow \infty} \|(\mathcal{S}^{(n,k)})^{\{\nu+1\}} - (\mathcal{S}^{(n,k)})^{\{\nu\}}\|_F = 0$;
- 5) $\lim_{\nu \rightarrow \infty} \|(\mathcal{G}^{(n)})^{\{\nu+1\}} - (\mathcal{G}^{(n)})^{\{\nu\}}\|_F = 0$.

Theorem III.3. *For any $n, k \in \{1, 2, 3\}$, let $\{\mathcal{Y}^{\{\nu\}}\}$, $\{\mathcal{Q}^{(n,k)}\}^{\{\nu\}}$, $\{(\mathcal{L}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{S}^{(n,k)})^{\{\nu\}}\}$, $\{(\mathcal{G}^{(n)})^{\{\nu\}}\}$, and $\mathcal{E}^{\{\nu\}}$ be the sequences generated by Algorithm 3. Suppose that the sequences $\{\mathcal{Y}^{\{\nu\}}\}$ and $\{\mathcal{Q}^{(n,k)}\}^{\{\nu\}}$ are bounded. Then, any accumulation point of the sequence $\{\{\mathcal{Y}^{\{\nu\}}\}\}$,*

$\{\Omega^{(n,k)}\}^{\{\nu\}}, \{(\mathcal{L}^{(n,k)})\}^{\{\nu\}}, \{(\mathcal{S}^{(n,k)})\}^{\{\nu\}}, \{(\mathcal{G}^{(n)})\}^{\{\nu\}}, \mathcal{E}^{\{\nu\}}$ is a Karush-Kuhn-Tucker (KKT) point of (19).

IV. EXPERIMENTAL RESULTS

In this section, we perform extensive experiments on several HSI datasets to substantiate the superiority and effectiveness of the proposed HAD approach. Some of the experimental results are presented in the supplementary materials. All the experiments are run on the following platforms: 1) Windows 11 and Matlab (R2023b) with 12th Gen Intel(R) Core(TM) i7-12700 2.1GHz CPU and 20GB memory; 2) Windows 11 and Python 3.13//Matlab (R2020b) with NVIDIA GeForce RTX 4090//Intel(R) Core(TM) i9-14900KF 4.56GHz CPU.

A. Experimental Datasets

The first Salinas ($126 \times 150 \times 204$) is a synthetic dataset offered by Verdoja and Grangetto [92], which is generated based on the Salinas real HSI dataset. The second Pavia dataset ($105 \times 120 \times 102$) is captured by the *Reflection Optical System Imaging Spectroradiometer* (ROSIS) above the center of Pavia, Italy [43]. The third Hyperion dataset ($100 \times 100 \times 145$) is taken by the Hyperion sensor carried by EO-1 in the Okavango Delta in 2001 [59]. The fourth HYDICE-Urban dataset ($80 \times 100 \times 175$) is taken by the *Hyperspectral Digital Imagery Collection Experiment* (HYDICE) sensor [41]. The fifth San Diego dataset ($100 \times 100 \times 189$) is collected by the *Airborne Visible/Infrared Imaging Spectrometer* (AVIRIS) sensor over the State of California, USA [46]. The sixth to the eleventh HSI datasets (i.e., Airport-4: $100 \times 100 \times 191$, Urban-3: $100 \times 100 \times 191$, Urban-4: $100 \times 100 \times 205$, Urban-5: $100 \times 100 \times 205$, Beach-3: $100 \times 100 \times 188$, Beach-4: $150 \times 150 \times 102$) are from the Airport-Beach-Urban (ABU [93])¹ dataset. The remaining ones are several larger-scale datasets: Qingpu-I dataset ($740 \times 400 \times 250$), Qingpu-II dataset ($400 \times 600 \times 250$), Avon dataset ($400 \times 400 \times 360$), Cri dataset ($400 \times 400 \times 46$), beach dataset ($452 \times 3072 \times 108$), Sequoia National Park (SNP) dataset ($1116 \times 2499 \times 13$), synthetic dataset ($600 \times 2400 \times 90$). Detailed descriptions of these large-scale datasets can be found in references [94], [95].

The pseudocolor map and the corresponding ground-truth map of the afore-mentioned HSIs are shown in Figure 3.

B. Comparative Experiments

1) **Comparison Methods:** To showcase the effectiveness of the proposed HAD algorithm, some baselines are introduced as compared detectors in our experiments, which contain one classical statistics-based method: RX [19], one collaborative representation based method: CRD [41], three deep learning-based methods: GAED [27], PDBSNet [30], GT-HAD [34], three matrix-based methods: LSMAD [36], LRASR [46] and GTVLRR [47], eight tensor-based methods: PTA [52], PCA-TLRSR [60], T-CTV [57], MERAETC [59], GCS [54], TRDFTVAD [71], GNLTR [61], and GNBRL [63].

¹https://ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

2) **Experimental Settings:** In our experiments, each raw HSI data is conducted with band-by-band normalization operation. The order of the tensor ring is consistent with the dimensions of the raw hyperspectral data. For the proposed HAD method, the nonconvex combinations $\Phi(\cdot)+\psi(\cdot)$ are set to be the same ($\Phi=\psi=\text{Capped-Log}$), with the exception of Hyperion ($\Phi=\psi=\text{MCP}$), San-Diego ($\Phi=\psi=\text{Capped-Lp}$), and Urban-3 ($\Phi=\psi=\text{MCP}$). For the Pavia, Airport-4, and Urban-5 datasets, the TR rank (r_1, r_2, r_3) is set to be $(6, 6, 6)$. For the other remaining datasets, we assign $(r_1, r_2, r_3) = (6, 16, 6)$. Besides, the trade-off parameters α and β are both searched from $\{10^{-6}, 5 \cdot 10^{-6}, 10^{-5}, 5 \cdot 10^{-5}, 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 0.1, 0.5, 1, 5, 10\}$, set $\mu^{\{0\}} = 10^{-3}$, $\mu^{\max} = 10^{10}$, $\vartheta = 1.1$, $\varpi = 10^{-5}$, $\nu_{\max} = 500$, $\mathfrak{L} = \text{FFT}$. For large-scale datasets, the TR rank (r_1, r_2, r_3) are chosen to be $\{(10, 200, 10), (10, 300, 10), (10, 500, 10), (15, 300, 15)\}$. To maintain the fairness of our comparative experiments, optimal parameters of competing methods are given according to the relevant references. Please refer the supplementary materials for more details.

3) **Evaluation Metrics:** In our experiments, four commonly used evaluation metrics are adopted, namely, anomaly detection map, 3-D receiver operating characteristic (ROC) curve [96], area under the ROC curve (AUC), and box-whisker map [97]. 3-D ROC describes the relationship between detection probability P_D , false alarm probability P_F , and detection threshold τ . Moreover, three 2-D ROC curves can be decomposed from the 3-D ROC curve, namely, (P_D, P_F) , (P_D, τ) , and (P_F, τ) . Correspondingly, we can get $\text{AUC}_{(P_D, P_F)}$, $\text{AUC}_{(P_D, \tau)}$, and $\text{AUC}_{(P_F, \tau)}$. By combining these three evaluation indicators with each other, the following comprehensive evaluation indicators can be obtained and defined as follows:

$$\begin{cases} \text{AUC}_{(\text{ODP})} = \text{AUC}_{(P_D, P_F)} + \text{AUC}_{(P_D, \tau)} - \text{AUC}_{(P_F, \tau)}, \\ \text{AUC}_{(\text{SNPR})} = \text{AUC}_{(P_D, \tau)} / \text{AUC}_{(P_F, \tau)}, \\ \text{AUC}_{(\text{TDBS})} = \text{AUC}_{(P_D, \tau)} - \text{AUC}_{(P_F, \tau)}, \end{cases}$$

where $\text{AUC}_{(\text{SNPR})}$ evaluates the signal-to-noise ratio of the detector, $\text{AUC}_{(\text{TDBS})}$ evaluates the comprehensive target detection and background suppression capabilities of the detector, and $\text{AUC}_{(\text{ODP})}$ evaluates the overall detection probability of the detector. In theory, an excellent detector should have very small $\text{AUC}_{(P_F, \tau)}$ and very large other AUC values. The box-whisker map is primarily utilized for assessing the discriminability between anomalies and background. The greater the gap between the anomaly box and the background box, the enhanced separation and detection effect of the detector can be achieved.

4) **Experimental Results and Analysis:** The AUC values obtained by various HAD detectors on extensive HSI datasets are reported in Tables I, II. Regarding the overall detection performance index $\text{AUC}_{(\text{ODP})}$, we observed that the proposed algorithm achieves the optimal $\text{AUC}_{(\text{ODP})}$ on the Salinas, Hyperion, HYDICE, Airport-4, Urban-3, Urban-4, Qingpu-I, and AVON datasets but achieves suboptimal $\text{AUC}_{(\text{ODP})}$ on the Pavia, San-Diego, Urban-5, Beach-4, Synthetic, and large-scale Beach datasets. Regarding the metric $\text{AUC}_{(\text{TDBS})}$, the

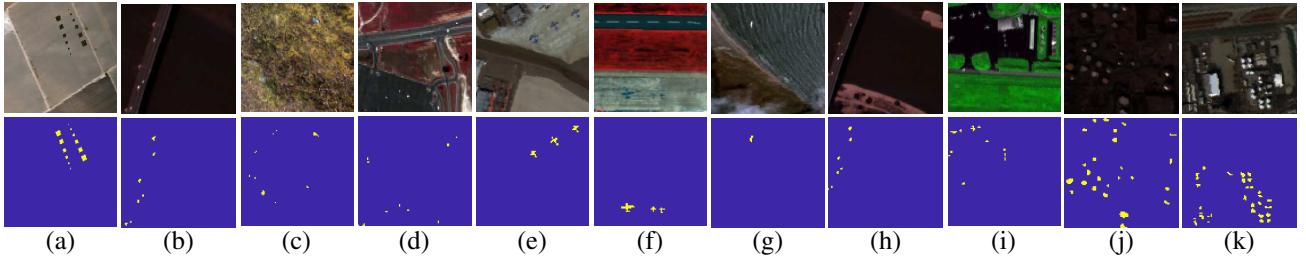


Fig. 3: Pseudocolor images and ground-truth maps of eleven HSI datasets. (a) Salinas. (b) Pavia. (c) Hyperion. (d) HYDICE. (e) San-Diego. (f) Airport-4. (g) Beach-3. (h) Beach-4. (i) Urban-3. (j) Urban-4. (k) Urban-5.

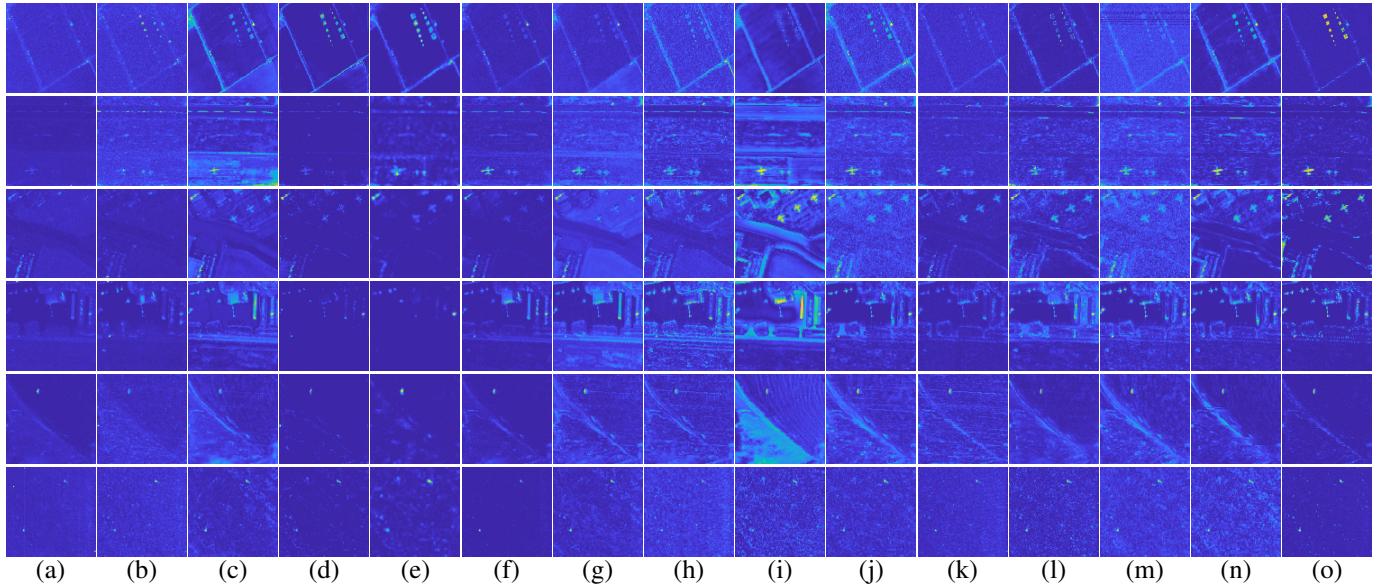


Fig. 4: Anomaly detection map of various HAD methods on five HSI datasets: Salinas, Airport-4, San-Diego, Urban-3, Beach-3, Hyperion (from top to bottom). (a) RX. (b) CRD. (c) GAED. (d) PDBSNet. (e) GT-HAD. (f) LSMAD. (g) LRASR. (h) GNTLRR. (i) PTA. (j) PCA-TLRSR. (k) T-CTV. (l) TRDFTVAD. (m) GNLTR. (n) GNBRL. (o) Proposed.

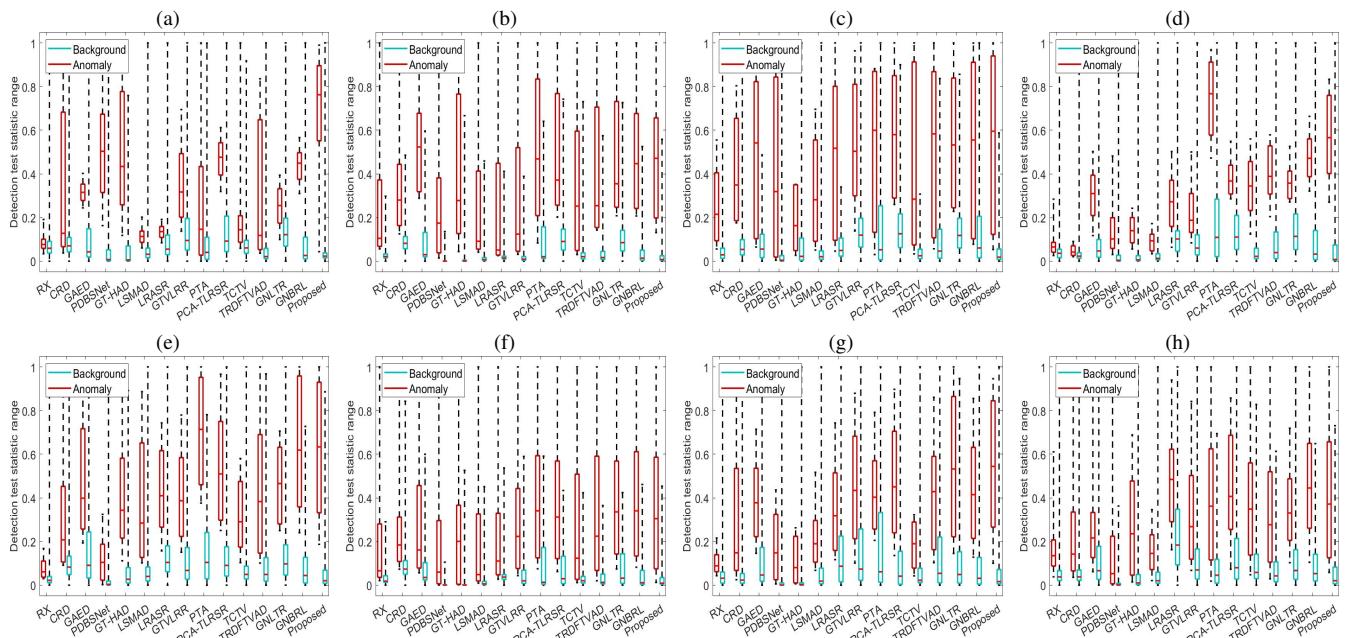


Fig. 5: Separability maps of various HAD methods for different HSI datasets. (a) Salinas. (b) Pavia. (c) HYDICE. (d) San-Diego. (e) Airport-4. (f) Beach-4. (g) Urban-3. (h) Urban-5.

TABLE I: The AUC values and average running time output by various HAD methods on twelve HSI datasets. The best and the second-best results are highlighted in boldface and blue, respectively.

HSI-Name	AUC-Metrics	RX	CRD	GAED	PDBSNet	GT-HAD	LSMAD	LRASR	GTV-LRR	PTA	PCA-TLRSR	TCTV	TRD-FTVAD	GNLTR	GNBRL	Proposed
Salinas	AUC _(P_D, P_F)	0.7437	0.8343	0.9660	0.9957	0.9936	0.9746	0.9227	0.9618	0.8230	0.9916	0.9438	0.9482	0.9369	0.9957	0.9998
	AUC _(P_D, τ)	0.0805	0.2631	0.3161	0.5028	0.4675	0.1134	0.1337	0.3030	0.1943	0.4715	0.1503	0.2437	0.2554	0.4435	0.7654
	AUC _(P_F, τ)	0.0638	0.0754	0.0727	0.0248	0.0252	0.0383	0.0668	0.1136	0.0575	0.1141	0.0653	0.0344	0.1295	0.0489	0.0292
	AUC _(ODP)	0.7604	1.0220	1.2094	1.4737	1.4359	1.0496	0.9897	1.1785	0.9597	1.3489	1.0288	1.1575	1.0629	1.3902	1.7360
	AUC _(SNPR)	1.2623	3.4909	4.3467	20.2521	18.5308	2.9582	2.0028	2.9077	3.3777	4.1321	2.3012	7.0921	1.9727	9.0624	26.1976
	AUC _(TDBS)	0.0167	0.1878	0.2434	0.4781	0.4422	0.0750	0.0670	0.2167	0.1368	0.3574	0.0850	0.2094	0.1259	0.3946	0.7362
Pavia	AUC _(P_D, P_F)	0.9981	0.9963	0.9993	0.9995	0.9986	0.9978	0.9600	0.9907	0.9779	0.9969	0.9867	0.9961	0.9970	0.9985	0.9976
	AUC _(P_D, τ)	0.1884	0.3122	0.5313	0.2260	0.3252	0.1859	0.1449	0.1931	0.4872	0.4375	0.3019	0.3568	0.4242	0.4614	0.4901
	AUC _(P_F, τ)	0.0233	0.0844	0.0472	0.0004	0.0026	0.0091	0.0163	0.0123	0.0475	0.0963	0.0242	0.0247	0.0914	0.0248	0.0161
	AUC _(ODP)	1.1632	1.2241	1.4834	1.2251	1.3212	1.1747	1.0886	1.1715	1.4176	1.3381	1.2644	1.3282	1.3298	1.4351	1.4716
	AUC _(SNPR)	8.0852	3.6981	11.2480	490.1876	123.7455	20.4287	8.8862	15.6551	10.2628	4.5422	12.4909	14.4401	4.6406	18.6028	30.4869
	AUC _(TDBS)	0.1651	0.2278	0.4841	0.2255	0.3226	0.1768	0.1286	0.1808	0.4398	0.3412	0.2778	0.3320	0.3328	0.4366	0.4740
Hyperion	AUC _(P_D, P_F)	0.9828	0.9555	0.9952	0.9879	0.9860	0.9979	0.9927	0.9602	0.9505	0.9975	0.9239	0.9900	0.9975	0.9888	0.9972
	AUC _(P_D, τ)	0.2183	0.3280	0.3849	0.2899	0.3947	0.1976	0.3928	0.3287	0.3566	0.4242	0.3099	0.4116	0.4566	0.4262	0.4065
	AUC _(P_F, τ)	0.0434	0.0621	0.0448	0.0134	0.0259	0.0132	0.0500	0.0896	0.0708	0.0614	0.0574	0.0531	0.0821	0.0904	0.0238
	AUC _(ODP)	1.1577	1.2214	1.3353	1.2644	1.3547	1.1823	1.3355	1.1993	1.2363	1.3603	1.1764	1.3484	1.3721	1.3245	1.3799
	AUC _(SNPR)	5.0321	5.2810	8.5907	21.5669	15.2203	14.9777	7.8601	3.6697	5.0364	6.9071	5.3977	7.7457	5.5649	4.7127	17.0651
	AUC _(TDBS)	0.1749	0.2659	0.3401	0.2765	0.3687	0.1844	0.3429	0.2391	0.2858	0.3628	0.2525	0.3585	0.3746	0.3358	0.3827
HYDICE	AUC _(P_D, P_F)	0.9855	0.9956	0.9600	0.9578	0.9089	0.9925	0.9580	0.9808	0.9236	0.9938	0.9777	0.9564	0.9931	0.9403	0.9921
	AUC _(P_D, τ)	0.2339	0.4091	0.5020	0.3819	0.1870	0.3032	0.4741	0.5156	0.5375	0.5857	0.3976	0.5347	0.5606	0.5490	0.6026
	AUC _(P_F, τ)	0.0351	0.0618	0.0644	0.0126	0.0447	0.0259	0.0577	0.1269	0.0983	0.1361	0.0309	0.0673	0.1253	0.0922	0.0289
	AUC _(ODP)	1.1843	1.3428	1.3977	1.3271	1.0513	1.2698	1.3744	1.3696	1.3628	1.4435	1.3444	1.4238	1.4283	1.3970	1.5658
	AUC _(SNPR)	6.6667	6.6153	7.7972	30.1095	4.1815	11.6882	8.2154	4.0633	5.4662	4.3047	12.8552	7.9402	4.4720	5.9523	20.8786
	AUC _(TDBS)	0.1988	0.3472	0.4376	0.3692	0.1423	0.2772	0.4164	0.3887	0.4392	0.4497	0.3666	0.4674	0.4352	0.4568	0.5737
San-Diego	AUC _(P_D, P_F)	0.8884	0.8577	0.9896	0.9815	0.9867	0.9788	0.9670	0.9529	0.9959	0.9862	0.9963	0.9907	0.9828	0.9953	0.9916
	AUC _(P_D, τ)	0.0685	0.0447	0.3081	0.1179	0.1395	0.0914	0.2717	0.2113	0.7454	0.3759	0.3417	0.3989	0.3597	0.4755	0.5642
	AUC _(P_F, τ)	0.0381	0.0263	0.0589	0.0125	0.0140	0.0193	0.0974	0.0742	0.1341	0.1257	0.0325	0.0612	0.1283	0.0593	0.0345
	AUC _(ODP)	0.9188	0.8761	1.2388	1.0869	1.1121	1.0509	1.1412	1.0900	1.6072	1.2363	1.3054	1.3284	1.2142	1.4116	1.5213
	AUC _(SNPR)	1.7994	1.7001	5.2289	9.4199	9.9285	4.7259	2.7879	2.8491	5.5591	2.9901	10.5072	6.5192	2.8036	8.0227	16.3419
	AUC _(TDBS)	0.0304	0.0184	0.2492	0.1053	0.1254	0.0720	0.1742	0.1371	0.6113	0.2502	0.3092	0.3377	0.2314	0.4162	0.5297
Airport-4	AUC _(P_D, P_F)	0.9525	0.9318	0.9660	0.9644	0.9951	0.9888	0.9889	0.9838	0.9972	0.9931	0.9928	0.9787	0.9925	0.9964	0.9982
	AUC _(P_D, τ)	0.0727	0.2449	0.4579	0.1118	0.3765	0.3545	0.4260	0.4059	0.7045	0.5411	0.3049	0.3950	0.4640	0.6266	0.6314
	AUC _(P_F, τ)	0.0247	0.0899	0.1197	0.0091	0.0391	0.0495	0.1161	0.0885	0.1244	0.1048	0.0569	0.0658	0.1105	0.0621	0.0331
	AUC _(ODP)	1.0005	1.0868	1.3043	1.0672	1.3325	1.2938	1.2988	1.3012	1.5773	1.4294	1.2408	1.3079	1.3460	1.5609	1.5965
	AUC _(SNPR)	2.9409	2.7240	3.8270	12.3651	9.6271	7.1568	3.6697	4.5878	5.6614	5.1652	5.3619	6.0057	4.1997	10.0830	19.0670
	AUC _(TDBS)	0.0480	0.1550	0.3383	0.1027	0.3374	0.3050	0.3099	0.3175	0.5801	0.4363	0.2480	0.3293	0.3535	0.5645	0.5983
Beach-3	AUC _(P_D, P_F)	0.9997	0.9888	0.9863	0.9919	0.9996	0.9992	0.9989	0.9944	0.9109	0.9984	0.9985	0.9910	0.9985	0.9928	0.9991
	AUC _(P_D, τ)	0.4937	0.3590	0.4357	0.3706	0.6138	0.5011	0.5811	0.4945	0.5397	0.6670	0.6158	0.4962	0.6597	0.5297	0.5271
	AUC _(P_F, τ)	0.0259	0.0559	0.0733	0.0077	0.0138	0.0237	0.0702	0.0739	0.1656	0.0947	0.0695	0.0540	0.0804	0.0535	0.0197
	AUC _(ODP)	1.4675	1.2919	1.3486	1.3548	1.5996	1.4766	1.5098	1.4150	1.2850	1.5707	1.5449	1.4332	1.5778	1.4691	1.5065
	AUC _(SNPR)	19.0477	6.4267	5.9408	47.9560	44.2937	21.1441	8.2831	6.6931	3.2596	7.0459	8.8660	9.1903	8.2065	9.9065	26.8211
	AUC _(TDBS)	0.4678	0.3032	0.3624	0.3628	0.4774	0.5110	0.4206	0.3741	0.5723	0.5464	0.4422	0.5793	0.4762	0.5074	
Beach-4	AUC _(P_D, P_F)	0.9538	0.9571	0.9362	0.9870	0.9906	0.9702	0.9341	0.9757	0.9515	0.9643	0.9291	0.9799	0.9696	0.9912	0.9863
	AUC _(P_D, τ)	0.1284	0.2121	0.2235	0.1270	0.2169	0.1133	0.1728	0.2542	0.3732	0.3328	0.2059	0.2791	0.3511	0.3706	0.3290
	AUC _(P_F, τ)	0.0233	0.0787	0.0540	0.0011	0.0013	0.0087	0.0380	0.0311	0.0586	0.0540	0.0236	0.0226	0.0580	0.0260	0.0180
	AUC _(ODP)	1.0589	1.0905	1.1058	1.1131	1.2063	1.0749	1.0688	1.1988	1.2660	1.2431	1.1131	1.2364	1.2626	1.3358	1.2973
	AUC _(SNPR)	5.5066	2.6949	4.1404	125.7441	164.4770	13.0048	4.5427	8.1761	6.3651	6.1572	8.7213	12.3738	6.0506	14.2684	18.3170
	AUC _(TDBS)	0.1051	0.1334	0.1695	0.1261	0.2156	0.1046	0.1348	0.2231	0.3146	0.2787	0.1822	0.2565	0.2930	0.3446	0.3110
Urban-3	AUC _(P_D, P_F)	0.9512	0.9611	0.9785	0.9846	0.9689	0.9656	0.9210	0.9423	0.9227	0.9849	0.9650	0.9447	0.9829	0.9862	0.9916
	AUC _(P_D, τ)	0.0961	0.2581	0.3887	0.1620	0.0995	0.2070	0.3373	0.4393	0.4118	0.4665	0.1856	0.3988	0.5311	0.4255	0.5342
	AUC _(P_F, τ)	0.0350	0.0312	0.0784	0.0049	0.0061	0.0369	0.1071	0.1123	0.1228	0.0644	0.0369	0.0892	0.0696	0.0547	0.0363
	AUC _(ODP)	1.0123	1.1881	1.2889	1.1417	1.0624	1.1358	1.1512	1.2693	1.2116	1.3871	1.1137	1.2542	1.4444	1.3570	1.4895
	AUC _(SNPR)	2.7448	8.2811	4.9608	32.5299	16.5531	5.6121	3.1500	3.9137	3.3542	7.2451	5.				

TABLE II: The AUC values and average running time output by various HAD methods on large-scale HSI datasets. The best and the second-best results are highlighted in boldface and blue, respectively.

HSI-Name	AUC-Metrics	RX	CRD	GAED	PDBSNet	LRASR	GTV-LRR	PTA	PCA-TLRSR	TCTV	GCS	MERA-ETV	TRD-FTVAD	GNLTR	GNBRL	Proposed
Qingpu-I	$AUC_{(P_D, P_F)}$	0.9958	0.9856	0.9968	0.9410	0.9576	0.9993	0.8750	0.9969	0.9898	0.9968	0.9003	0.9976	0.9961	0.9948	0.9997
	$AUC_{(P_D, \tau)}$	0.3375	0.2300	0.4680	0.0286	0.4053	0.5736	0.1626	0.5941	0.5562	0.6185	0.4530	0.4995	0.5813	0.5745	0.5943
	$AUC_{(P_F, \tau)}$	0.0402	0.0315	0.0802	0.0047	0.1407	0.0556	0.0676	0.1315	0.1553	0.1472	0.1259	0.0665	0.1376	0.1327	0.0363
	$AUC_{(ODP)}$	1.2932	1.1842	1.3846	0.9648	1.2222	1.5173	0.9700	1.4596	1.3908	1.4681	1.2275	1.4306	1.4398	1.4365	1.5576
	$AUC_{(SNPR)}$	8.4002	7.3067	5.8323	6.0404	2.8811	10.3144	2.4039	4.5193	3.5819	4.2016	3.5993	7.5084	4.2259	4.3287	16.3678
	$AUC_{(TDBS)}$	0.2974	0.1985	0.3878	0.0238	0.2646	0.5180	0.0950	0.4627	0.4009	0.4713	0.3272	0.4330	0.4437	0.4417	0.5580
Qingpu-II	$AUC_{(P_D, P_F)}$	0.9990	0.9997	1.0000	0.9982	1.0000	0.9999	0.9951	0.9999	0.9999	0.9994	0.9988	0.9977	0.9996	0.9856	1.0000
	$AUC_{(P_D, \tau)}$	0.1175	0.1708	0.5861	0.2988	0.6506	0.5501	0.2728	0.3606	0.3518	0.3490	0.2673	0.1646	0.3820	0.3648	0.3743
	$AUC_{(P_F, \tau)}$	0.0158	0.0182	0.0365	0.0048	0.0746	0.0360	0.0272	0.0586	0.0551	0.0658	0.0551	0.0278	0.0707	0.1800	0.0229
	$AUC_{(ODP)}$	1.1008	1.1524	1.5496	1.2923	1.5759	1.5141	1.2407	1.3019	1.2966	1.2826	1.2110	1.1344	1.3110	1.1704	1.3514
	$AUC_{(SNPR)}$	7.4418	9.4060	16.0620	62.0467	8.7183	15.2814	10.0268	6.1560	6.3875	5.3052	4.8498	5.9125	5.4069	2.0265	16.3476
	$AUC_{(TDBS)}$	0.1017	0.1527	0.5496	0.2940	0.5759	0.5141	0.2456	0.3020	0.2967	0.2832	0.2122	0.1367	0.3114	0.1848	0.3514
AVON	$AUC_{(P_D, P_F)}$	0.8358	0.8047	0.9432	0.8233	0.8884	0.9332	0.8948	0.8994	0.9179	0.8792	0.9028	0.9797	0.9108	0.9128	0.9916
	$AUC_{(P_D, \tau)}$	0.0274	0.0428	0.1037	0.0272	0.1231	0.1896	0.1062	0.1823	0.1813	0.1293	0.1814	0.1521	0.1966	0.1970	0.1715
	$AUC_{(P_F, \tau)}$	0.0033	0.0015	0.0209	0.0034	0.0268	0.0309	0.0276	0.0403	0.0410	0.0265	0.0373	0.0131	0.0521	0.0517	0.0092
	$AUC_{(ODP)}$	0.8599	0.8460	1.0260	0.8471	0.9847	1.0919	0.9734	1.0414	1.0583	0.9821	1.0469	1.1187	1.0553	1.0580	1.1539
	$AUC_{(SNPR)}$	8.3872	28.0867	4.9604	8.0295	4.5876	6.1452	3.8444	4.5237	4.4251	4.8847	4.8608	11.6118	3.7733	3.8065	18.6560
	$AUC_{(TDBS)}$	0.0241	0.0413	0.0828	0.0238	0.0963	0.1588	0.0786	0.1420	0.1404	0.1028	0.1441	0.1390	0.1445	0.1452	0.1623
CRI	$AUC_{(P_D, P_F)}$	0.9781	0.9803	0.9843	0.9876	0.9101	0.9333	0.9750	0.8527	0.9173	0.9517	0.7122	0.8887	0.8254	0.9760	0.9761
	$AUC_{(P_D, \tau)}$	0.0888	0.1115	0.5339	0.3168	0.4329	0.4264	0.7723	0.2607	0.3767	0.4828	0.2201	0.3778	0.3805	0.5042	0.5046
	$AUC_{(P_F, \tau)}$	0.0236	0.0037	0.0916	0.0172	0.1286	0.1785	0.2456	0.0831	0.0993	0.2052	0.1160	0.2120	0.2293	0.1675	0.1717
	$AUC_{(ODP)}$	1.0432	1.0881	1.4266	1.2872	1.2144	1.1813	1.5018	1.0303	1.1947	1.2293	0.8163	0.9766	1.3128	1.3089	
	$AUC_{(SNPR)}$	3.7571	30.0465	5.8285	18.4298	3.3663	2.3892	3.1452	3.1383	3.7925	2.3526	1.8974	1.7824	1.6595	3.0107	2.9379
	$AUC_{(TDBS)}$	0.0651	0.1078	0.4423	0.2996	0.3043	0.2479	0.5268	0.1776	0.2774	0.2776	0.1041	0.1658	0.1512	0.3367	0.3328
Beach	$AUC_{(P_D, P_F)}$	0.8106	0.7462	0.8522	0.9236	0.7452	0.8110	0.9002	0.8048	0.8712	0.9015	0.8321	0.9194	0.9487	0.9715	0.9709
	$AUC_{(P_D, \tau)}$	0.0157	0.0024	0.1993	0.0143	0.1751	0.0716	0.6674	0.0669	0.1160	0.1744	0.0677	0.1624	0.1442	0.2376	0.2182
	$AUC_{(P_F, \tau)}$	0.0032	0.0007	0.0644	0.0006	0.0635	0.0298	0.6171	0.0286	0.0252	0.0578	0.0264	0.0277	0.0318	0.0358	0.0368
	$AUC_{(ODP)}$	0.8231	0.7479	0.9871	0.9373	0.8568	0.8528	0.9506	0.8430	0.9620	1.0182	0.8734	1.0541	1.0611	1.1733	1.1523
	$AUC_{(SNPR)}$	4.8724	3.3326	3.0933	24.8677	2.7562	2.4034	1.0816	2.3390	4.5990	3.0191	2.5603	5.8697	4.5378	6.6310	5.9277
	$AUC_{(TDBS)}$	0.0125	0.0017	0.1349	0.0137	0.1115	0.0418	0.0504	0.0383	0.0908	0.1167	0.0413	0.1347	0.1124	0.2018	0.1814
SNP	$AUC_{(P_D, P_F)}$	0.9969	0.9076	0.9732	0.9966	0.9944	0.9928	0.9879	0.9654	0.9904	0.9936	0.9643	0.9780	0.9922	0.9842	0.9948
	$AUC_{(P_D, \tau)}$	0.0197	0.0102	0.4965	0.4385	0.5239	0.3512	0.4159	0.2480	0.3096	0.3793	0.1605	0.2943	0.3491	0.4263	0.3405
	$AUC_{(P_F, \tau)}$	0.0002	0.0001	0.1296	0.0032	0.0491	0.0486	0.0745	0.0459	0.0248	0.0452	0.0171	0.0550	0.0544	0.0741	0.0226
	$AUC_{(ODP)}$	1.0163	0.9176	1.3402	1.4319	1.4692	1.2953	1.3293	1.1675	1.2752	1.3277	1.1078	1.2173	1.2869	1.3364	1.3126
	$AUC_{(SNPR)}$	88.5452	72.7690	3.8319	137.4273	10.6779	7.2218	5.5829	5.4009	12.4751	8.3901	9.4019	5.3515	6.4163	5.7532	15.0538
	$AUC_{(TDBS)}$	0.0194	0.0100	0.3670	0.4353	0.4748	0.3026	0.3414	0.2021	0.2847	0.3341	0.1434	0.2393	0.2947	0.3522	0.3179
Synthetic	$AUC_{(P_D, P_F)}$	0.7118	0.4997	0.8429	0.9506	0.7032	0.7074	0.6116	0.9078	0.9160	0.6826	0.7917	0.7814	0.7993	0.9192	0.9210
	$AUC_{(P_D, \tau)}$	0.0362	0.0001	0.3668	0.1463	0.2515	0.1936	0.1219	0.2739	0.2812	0.1697	0.1680	0.1095	0.2165	0.2823	0.2813
	$AUC_{(P_F, \tau)}$	0.0147	0.0001	0.1855	0.0097	0.2078	0.1382	0.0846	0.1225	0.1236	0.1146	0.0881	0.0511	0.1256	0.1236	0.1222
	$AUC_{(ODP)}$	0.7332	0.4997	1.0242	1.0871	0.7469	0.7628	0.6489	1.0592	1.0737	0.7377	0.8716	0.8397	0.8902	1.0778	1.0801
	$AUC_{(SNPR)}$	2.4568	0.9921	1.9776	15.0512	1.2104	1.4012	1.4410	2.2355	2.2752	1.4810	1.9073	2.1402	1.7242	2.2835	2.3022
	$AUC_{(TDBS)}$	0.0214	0.0000	0.1813	0.1365	0.0437	0.0554	0.0373	0.1514	0.1576	0.0551	0.0799	0.0583	0.0909	0.1587	0.1591
Average Time (second)		3.13	3349.89	2962.81	834.38	3342.15	5710.69	483.88	2210.92	747.78	529.65	850.76	709.06	3072.29	1091.67	851.96

proposed algorithm outperforms all competitors on the Salinas, Hyperion, HYDICE, Airport-4, Urban-3, Urban-4, Qingpu-I, and AVON datasets, while yielding results that are on a par with those of other algorithms on the remaining datasets. This indicates that our detector possesses excellent comprehensive capabilities in both target detection and background suppression. When it comes to the evaluation indicator $AUC_{(SNPR)}$, the proposed method reaches the optimal value on the Salinas, Airport-4, San-Diego, and Qingpu-I datasets. Although the PDBSNet and GT-HAD methods achieve notably high $AUC_{(SNPR)}$ values on the remaining datasets, their performance in terms of other evaluation metrics (e.g., $AUC_{(ODP)}$, $AUC_{(TDBS)}$ (P_D, τ)) is suboptimal. These quantitative results effectively prove the superiority of the proposed algorithm compared to other competing algorithms.

Figures 4 and 6 display the two-dimensional anomaly detection maps of various HAD methods on extensive HSI datasets. Compared with other detectors, the proposed method strikes a superior balance between accurate anomaly recognition and effective background removal. The detection maps from the Salinas and San-Diego datasets are utilized as illustrative examples to support our conclusion. For the Salinas's detection maps, we find that the PTA detector fails to detect

the abnormal target. GTVLRR and PCA-TLRSR misidentify more backgrounds as abnormal targets, resulting in poor visual effect. The LRASR and LSMAD detectors have better background suppression, but abnormal targets can hardly be detected. The GAED, PDBSNet, GT-HAD, TRDFTVAD, and GNBRL detectors can distinguish the approximate distribution of abnormal targets, yet they suffer from a certain degree of background contamination. For the San-Diego's detection maps, we find that RX and CRD completely are unable to detect three airplanes. The LSMAD, LRASR and GTVLRR can detect three planes but are not in good shape. The PTA, TRDFTVAD and GNBRL can brightly highlight anomalous targets, but it does not suppress the background well. In stark contrast, GAED, PDBSNet, GT-HAD, and T-CTV perform the inverse. Overall, our method effectively suppressed the background while detecting clear anomalous targets.

The box-whisker maps of different detectors on eight datasets are displayed in Figure 5. This figure shows that the proposed method has the widest gap between the anomaly box and the background box in most cases. Besides, the abnormal box of our HAD method is not always at the highest position for the different HSI datasets, but the position of the background box is almost always at the lowest position.

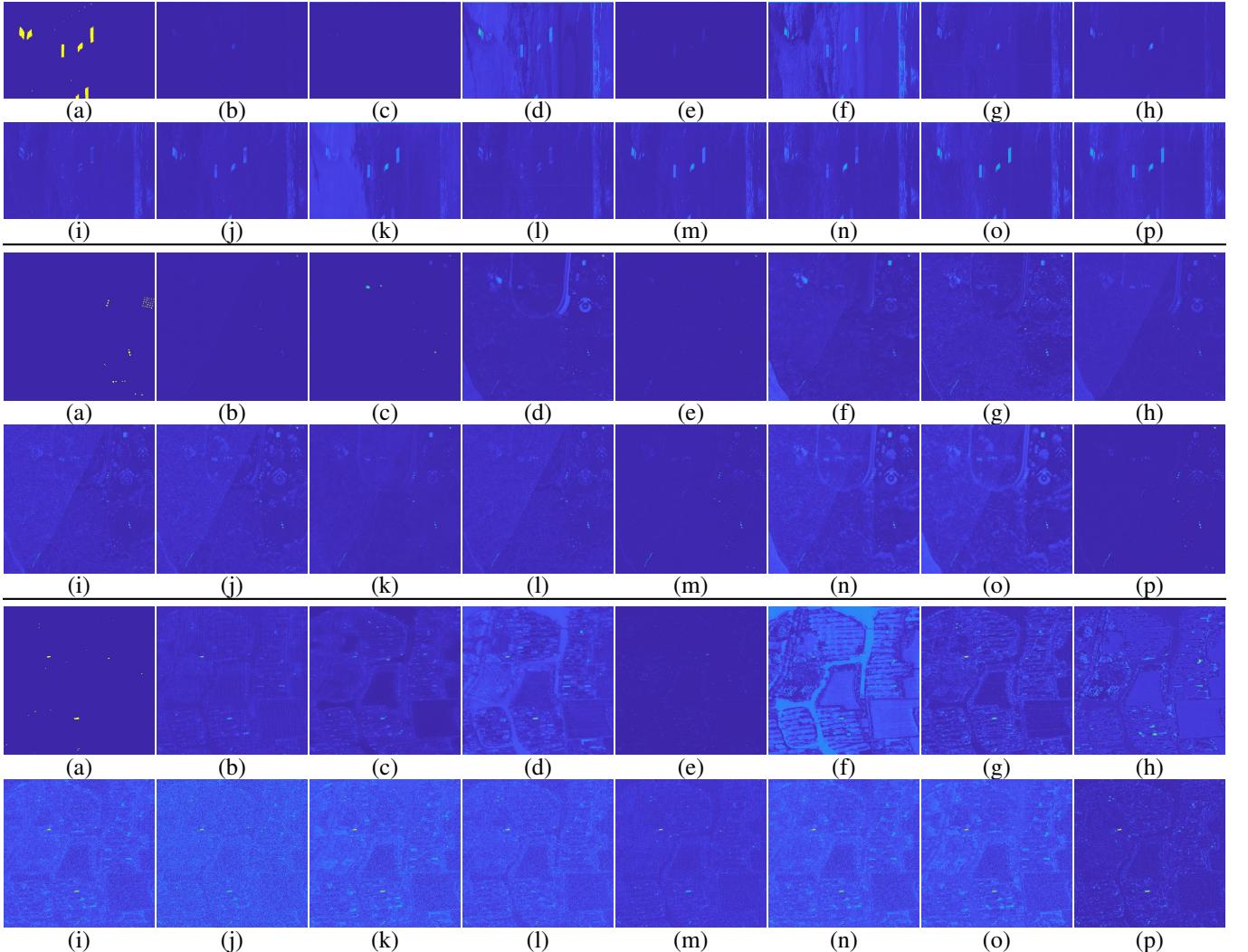


Fig. 6: Anomaly detection map of various HAD methods on large-scale HSI datasets: Beach, AVON, Qingpu-I (from top to bottom). (a) Ground-Truth. (b) RX. (c) CRD. (d) GAED. (e) PDBSNet. (f) LRASR. (g) GTVLRR. (h) PTA. (i) PCA-TLRSR. (j) T-CTV. (k) GCS. (l) MERAETV (m) TRDFTVAD. (n) GNLTR. (o) GNBRL. (p) Proposed.

This imply that the proposed method has a good inhibitory effect on the background, which is consistent with the above visual results. The 3-D ROC curves and 2-D ROC curves of each algorithm on the eight datasets are shown in Figure 7. As shown in Figure 7 (b), the proposed algorithm achieves a higher detection rate than other algorithms under different false alarm rates in most cases. As can be observed from Figure 7 (c), the proposed algorithm beats most of its competitors in the ROC curves on all HSI datasets, exhibiting better target detection performance, especially on the Salinas and Urban-3 datasets. As can be seen from Figure 7 (d), the ROC curves of the proposed algorithm are almost all in the lower left corner on all datasets, which indicates that our detector has a very low false alarm rate under different threshold τ 's. Overall, our HAD algorithm ranks third after PDBSNet and GT-HAD in terms of ROC curves of (P_F, τ) .

C. Parameter Analysis

The proposed HAD algorithm involves five parameters that need to be tuned, namely, the TR rank (r_1, r_2, r_3) , the nonconvex functions $\Phi(\cdot), \psi(\cdot)$, and two tradeoff parameters α and β .

In our experiments, we tune a certain parameter by fixing the others. For brevity, we set the nonconvex functions $\Phi(\cdot)$ and $\psi(\cdot)$ to be the same, $\Phi, \psi \in \{\text{L1}, \text{Lp}, \text{Log}, \text{MCP}, \text{Capped-L1}, \text{Capped-Lp}, \text{Capped-Log}, \text{Capped-MCP}\}$, $\mathfrak{L} = \text{FFT}$. The value ranges for α and β are selected from $\{10^{-6}, 5 \cdot 10^{-6}, 10^{-5}, 5 \cdot 10^{-5}, 10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 0.1, 0.5, 1, 5, 10\}$.

1) Trade-off Parameters α and β : This experiment focuses on investigating the impact of the different trade-off parameters α and β upon anomaly detection performance in the context of given TR rank (r_1, r_2, r_3) and nonconvex combinations $\Phi(\cdot)+\psi(\cdot)$. In this part, The TR rank is empirically fixed as $(6, 6, 6)$. The relevant experimental results are displayed in Figure 8. Due to space constraints, we only present the results achieved by the Capped-Log function. We find that for Salinas, Airport-4, Hyperion, Pavia, and Beach-3, the detection performance demonstrates strong robustness to the parameters α and β , with $AUC_{(ODP)}$ values experiencing only slight fluctuations as α and β vary. The robustness of the detection performance in San-Diego varies moderately with

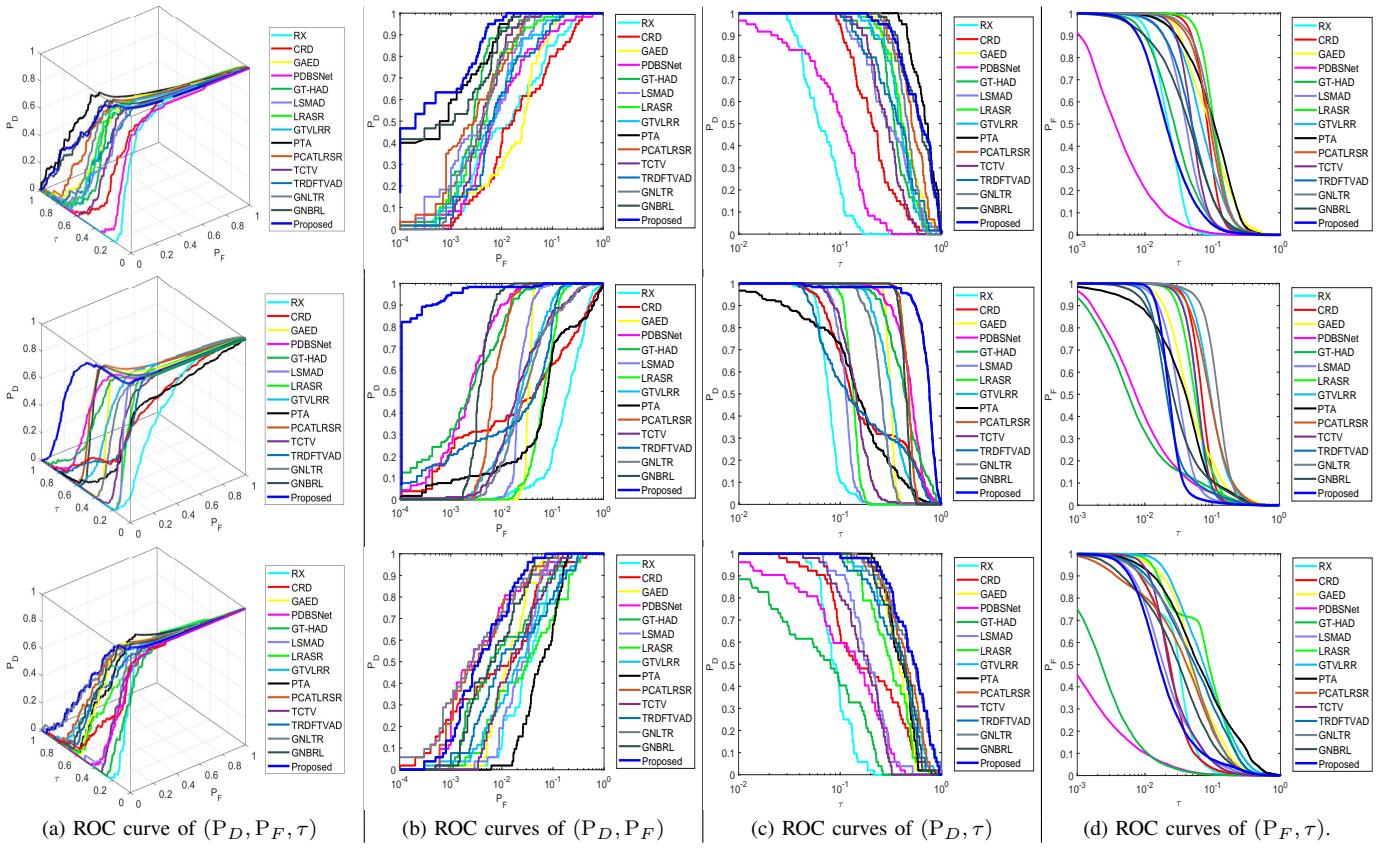


Fig. 7: 3-D and 2-D ROC curves' performance comparison of different HAD approaches on HSI datasets: Airport-4, Salinas and Urban-3 (from top to bottom).

the parameters α and β and the $AUC_{(ODP)}$ value generally shows a decreasing trend with the increase of β . The changing trend of $AUC_{(ODP)}$ with the rise of α is not apparent. For HYDICE and Urban-3, as the parameter β increases, the $AUC_{(ODP)}$ value of the proposed HAD algorithm begins to stabilize. Overall, the detection performance is insensitive to the trade-off parameters α and β .

2) **Nonconvex Functions $\Phi(\cdot)$ and $\psi(\cdot)$:** In this experiment, the performance of our proposed HAD method under different nonconvex functions is verified. Based on the parameters α and β learned in the previous experiment, Table III shows the AUC values obtained by various nonconvex combinations $\Phi(\cdot)+\psi(\cdot)$ for different HSI datasets. We pursue a general surrogate for approximating low rank and sparsity, which can be flexibly chosen according to different scenarios. We can observe that the convex function L1 performs worse than all nonconvex functions in terms of AUC values. We can also observe that the non-Capped-type functions generally perform worse than the corresponding Capped-type functions in terms of AUC values. Among the Capped-type functions, Capped-Log achieve relatively good AUC values in many scenarios.

3) **TR Rank** $[r_1, r_2, r_3]$: The TR rank contains three parameters $r = [r_1, r_2, r_3]$, and the estimation of TR rank is still an open problem. In this experiment, we find that the TR rank is usually needed into a pattern that sets the two sides of the rank smaller and the middle larger. To simplify the complexity of the parameters analysis, we choose $r_1 = r_3$ to maintain the consistency of the spatial TR core tensor. The rank components r_1 and r_3 undergo adjustments

within the range of 2 to 20, while r_2 is varied within the interval of 2 to 42. Figure 9 illustrates the trend of the $AUC_{(ODP)}$ value as it varies with TR rank (r_1, r_2, r_3) . From this figure, we find that under suitable nonconvex functions and trade-off parameters, a relatively low TR rank can yield satisfactory detection performance. On the whole, the anomaly detection performance of the proposed algorithm is robust to the combination of r_1 , r_2 and r_3 .

D. More Discussions

1) **Part 1:** In this part, we conduct related ablation experiments on several benchmark HSI datasets. The purpose of these experiments is to investigate the impact of adopting various nonconvex regularization schemes to encode the prior structures of background tensor on the performance of anomaly detection. Equivalently, to verify the effectiveness of the designed EUNTRFR regularization term, we compare the proposed model (14) with the following four models:

Model 1 (GNLCTV): The regularizer (9) is directly imposed on the background tensor, i.e.,

$$\min_{\mathcal{B}, \mathcal{E}} \|\mathcal{B}\|_{GNLCTV} + \lambda_1 \cdot \|\mathcal{E}\|_{\ell_{F,1}^{\psi}}, \text{s.t. } \mathcal{M} = \mathcal{B} + \mathcal{E}. \quad (29)$$

Model 2 (EGNTCTV): The regularizer (11) is directly imposed on the background tensor, i.e.,

$$\min_{\mathcal{B}, \mathcal{E}} \|\mathcal{B}\|_{EGNTCTV} + \lambda_2 \cdot \|\mathcal{E}\|_{\ell_{F,1}^{\psi}}, \text{s.t. } \mathcal{M} = \mathcal{B} + \mathcal{E}. \quad (30)$$

Model 3 (GNCTV): The regularizer (10) is imposed on the

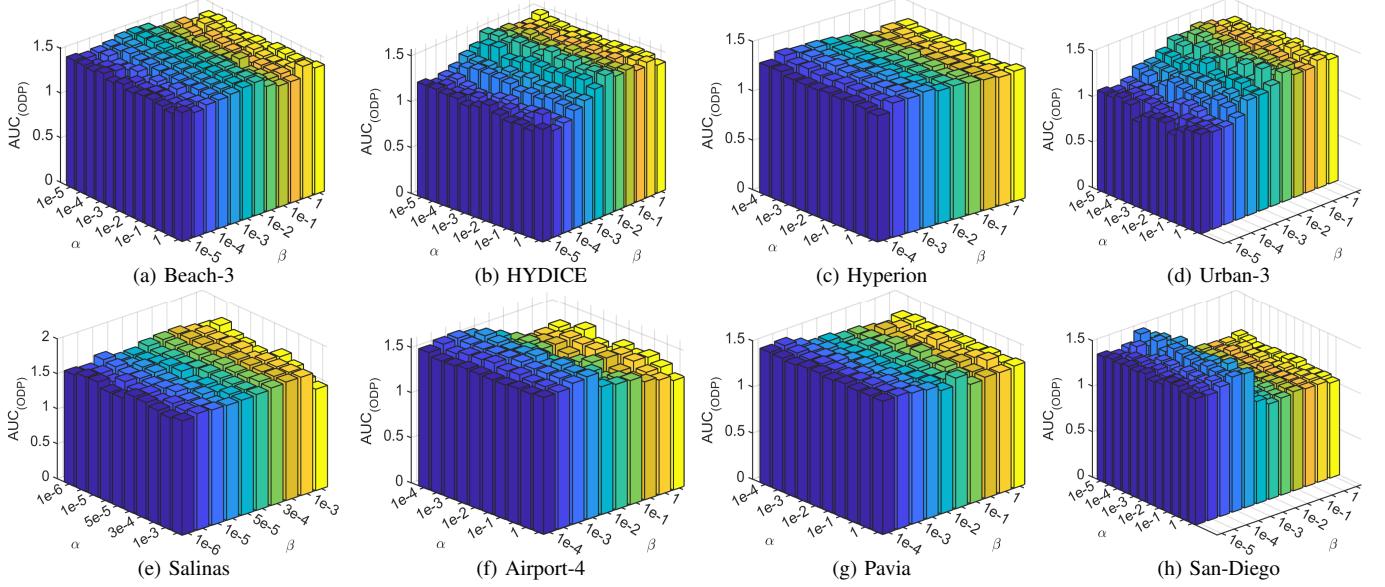


Fig. 8: Sensitivity analysis of the trade-off parameters α and β . For each HSI data, the TR rank is fixed as $(r_1, r_2, r_3) = (6, 6, 6)$, and the nonconvex combination $\Phi+\psi$ is set to be the same, i.e., $\Phi(\cdot) = \psi(\cdot) = \text{Capped-Log}$.

TABLE III: The AUC values obtained by various nonconvex combinations $\Phi(\cdot)+\psi(\cdot)$ for different HSI datasets.

Nonconvex function	AUC-Metrics	Salinas	Pavia	Hyperion	HYDICE	San-Diego	Airport-4	Beach-3	Beach-4	Urban-3	Urban-4	Urban-5
L1	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9986	0.9975	0.9971	0.9949	0.9895	0.9956	0.9980	0.9809	0.9809	0.9954	0.9638
	$\text{AUC}_{(\text{ODP})}$	1.6073	1.3734	1.3081	1.4140	1.4711	1.5619	1.4455	1.2506	1.2828	1.0893	1.2838
	$\text{AUC}_{(\text{SNPR})}$	17.0345	24.6140	13.1800	17.8686	11.1167	10.4967	15.3768	16.2700	7.3307	20.0167	5.6482
	$\text{AUC}_{(\text{TDBS})}$	0.6087	0.3759	0.3110	0.4191	0.4816	0.5663	0.4475	0.2697	0.3019	0.0939	0.3200
L-p	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9986	0.9983	0.9972	0.9968	0.9910	0.9950	0.9973	0.9752	0.9838	0.9953	0.9793
	$\text{AUC}_{(\text{ODP})}$	1.6204	1.4010	1.3092	1.4426	1.4911	1.4771	1.4558	1.2779	1.3470	1.0982	1.2627
	$\text{AUC}_{(\text{SNPR})}$	16.2106	27.4386	14.2062	16.2505	10.2608	12.3109	21.7941	14.2359	8.9213	22.5599	7.5783
	$\text{AUC}_{(\text{TDBS})}$	0.6218	0.4027	0.3120	0.4458	0.5001	0.4821	0.4585	0.3027	0.3632	0.1029	0.2834
MCP	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9991	0.9979	0.9972	0.9947	0.9887	0.9971	0.9988	0.9881	0.9916	0.9948	0.9820
	$\text{AUC}_{(\text{ODP})}$	1.7069	1.4503	1.3799	1.4718	1.5072	1.5474	1.5043	1.2705	1.4895	1.1168	1.2850
	$\text{AUC}_{(\text{SNPR})}$	15.6435	30.0371	17.0651	23.1564	10.6468	14.5926	29.4970	19.4595	14.7299	26.3305	8.1361
	$\text{AUC}_{(\text{TDBS})}$	0.7078	0.4524	0.3827	0.4771	0.5185	0.5503	0.5055	0.2824	0.4979	0.1220	0.3030
Log	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9996	0.9983	0.9970	0.9943	0.9916	0.9977	0.9991	0.9800	0.9890	0.9935	0.9842
	$\text{AUC}_{(\text{ODP})}$	1.6988	1.4248	1.3402	1.4743	1.5059	1.5572	1.4942	1.2905	1.4353	1.0859	1.3078
	$\text{AUC}_{(\text{SNPR})}$	22.3044	29.1209	16.1383	23.1957	8.0262	18.2023	27.7427	16.4094	13.5974	26.1800	9.1143
	$\text{AUC}_{(\text{TDBS})}$	0.6992	0.4265	0.3432	0.4800	0.5143	0.5595	0.4951	0.3105	0.4463	0.0924	0.3236
Capped-L1	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9997	0.9981	0.9933	0.9942	0.9899	0.9941	0.9985	0.9857	0.9888	0.9950	0.9826
	$\text{AUC}_{(\text{ODP})}$	1.7019	1.4645	1.3350	1.4618	1.4862	1.5695	1.4844	1.2657	1.4293	1.1105	1.2833
	$\text{AUC}_{(\text{SNPR})}$	30.1980	29.7396	14.5027	19.5036	10.1049	13.2079	25.3616	16.9770	12.5657	27.6946	7.8313
	$\text{AUC}_{(\text{TDBS})}$	0.7022	0.4664	0.3417	0.4676	0.4963	0.5754	0.4859	0.2800	0.4405	0.1155	0.3007
Capped-Lp	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9997	0.9981	0.9977	0.9921	0.9916	0.9982	0.9990	0.9883	0.9903	0.9962	0.9811
	$\text{AUC}_{(\text{ODP})}$	1.6694	1.4442	1.3661	1.5088	1.5213	1.5711	1.4900	1.2877	1.4087	1.1189	1.3418
	$\text{AUC}_{(\text{SNPR})}$	26.6731	30.0460	14.7704	20.2671	16.3419	19.3292	27.4201	20.7341	13.3398	28.2700	10.2623
	$\text{AUC}_{(\text{TDBS})}$	0.6697	0.4461	0.3684	0.5167	0.5297	0.5729	0.4910	0.2994	0.4184	0.1227	0.3607
Capped-MCP	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9986	0.9986	0.9921	0.9908	0.9900	0.9986	0.9993	0.9851	0.9915	0.9956	0.9822
	$\text{AUC}_{(\text{ODP})}$	1.6512	1.4464	1.3509	1.5333	1.4486	1.5872	1.4912	1.2880	1.4025	1.1185	1.3505
	$\text{AUC}_{(\text{SNPR})}$	24.5450	31.4068	14.8410	19.2955	16.2175	19.9611	26.0997	18.6652	13.5420	23.4183	9.7221
	$\text{AUC}_{(\text{TDBS})}$	0.6526	0.4478	0.3588	0.5425	0.4586	0.5886	0.4919	0.3029	0.4110	0.1229	0.3683
Capped-Log	$\text{AUC}_{(\text{P}_D, \text{P}_F)}$	0.9998	0.9976	0.9941	0.9921	0.9877	0.9982	0.9991	0.9863	0.9901	0.9958	0.9847
	$\text{AUC}_{(\text{ODP})}$	1.7360	1.4716	1.3691	1.5658	1.4781	1.5965	1.5065	1.2973	1.4375	1.1258	1.3254
	$\text{AUC}_{(\text{SNPR})}$	26.1976	30.4869	14.1624	20.8786	12.6244	19.0670	26.8211	18.3170	13.2782	24.6349	10.3161
	$\text{AUC}_{(\text{TDBS})}$	0.7362	0.4740	0.3750	0.5737	0.4904	0.5983	0.5074	0.3110	0.4474	0.1300	0.3407

mode-2 unfolding of each TR factor, i.e.,

$$\begin{aligned} & \min_{[\mathcal{G}], \mathcal{E}} \sum_{n=1}^3 \|\mathcal{G}_{(2)}^{(n)}\|_{\text{GNCTV}} + \lambda_3 \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi}, \\ & \text{s.t. } \mathcal{M} = \mathcal{R}([\mathcal{G}]) + \mathcal{E}. \end{aligned} \quad (31)$$

Model 4 (UNTRFR): The regularizer (12) is imposed on the background tensor. This is equivalent to imposing the GNTCTV regularizer directly on gradient TR factors, i.e.,

$$\begin{aligned} & \min_{[\mathcal{G}], \mathcal{E}} \sum_{n=1}^3 \sum_{k=1}^3 \frac{1}{\gamma} \|\mathcal{T}^{(n,k)}\|_{\Phi, \mathcal{E}} + \lambda_4 \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi}, \\ & \text{s.t. } \mathcal{M} = \mathcal{R}([\mathcal{G}]) + \mathcal{E}, \nabla_k(\mathcal{G}^{(n)}) = \mathcal{T}^{(n,k)}, n, k \in [3], \end{aligned} \quad (32)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ are the regularization parameters.

Proposed Model (EUNTRFR): The regularizer (13) is imposed on the background tensor. This is equivalent to imposing the EGNTCTV regularizer directly on gradient TR factors. Please see the model (14) for more details.

It is worth mentioning that the models (29) and (30) use the GNTCTV and EGNTCTV regularizers induced by the T-SVD framework to encode the L+S priors of HSI's background, respectively. The model (31) utilizes the GNCTV to describe the L and S priors of TR factors of HSI's background. Unlike the model (31) that uses the mode-2 unfolding scheme, the model (32) treats all TR factors as the low-tubal-rank tensors, and then employ the GNCTV regularizer to characterize

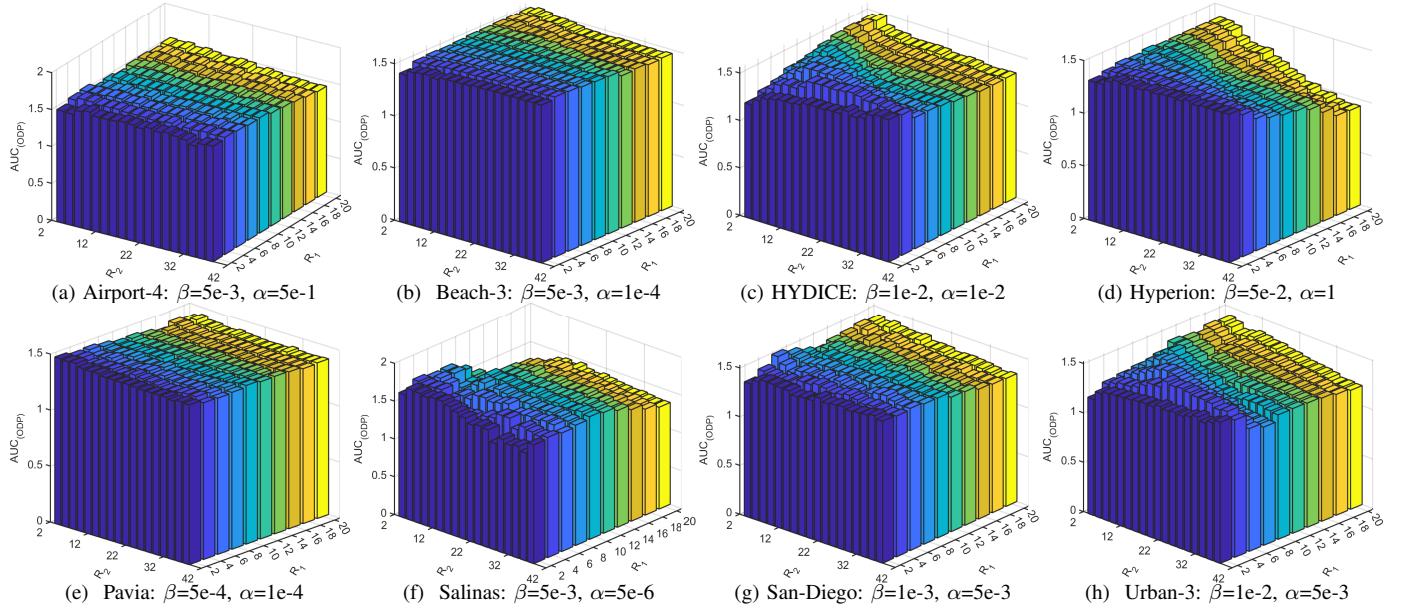


Fig. 9: Sensitivity analysis of the TR rank (r_1, r_2, r_3). All nonconvex combinations are set to Capped-Log, with the exception of Hyperion (MCP), San-Diego (Capped-Lp) and Urban-3 (MCP).

their prior structures in the gradient domain. Based on the model (32), the proposed model (14) additionally introduces a residual term to enhance its robustness. All of the above models are optimized by using ADMM framework.

In our experiments, the trade-off parameter $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ in (29)-(32) are set to be $\lambda = \kappa / \sqrt{\min(n_1, n_2) \cdot n_3}$, κ is selected from $\{0.1, 0.3, 0.5, 0.8, 1, 1.2, 1.5, 1.8, 2, 2.2\}$. The relevant experimental results are shown in Table IV. Among all the T-SVD-based models, we discovered that Model 1 and Model 2 overcome the TCTV model in AUC values, with Model 2 exceeding Model 1. This demonstrates that the GNTCTV regularizer and its enhanced version are more capable of effectively extracting the prior information of the background compared to convex TCTV regularizer. Besides, the EG-NTCTV regularizer outperforms the GNTCTV regularizer. Across all TR-based models, we observed that our model and Model 4 are superior to Model 3. This implies that, compared to the matrix unfolding strategy employed in Model 3, the UNTRFR and EUNTRFR schemes are able to better extract structural information from TR factors, thereby enhancing detection performance.

2) *Part 2:* To verify the validity of the novel prior representation paradigm introduced in our proposed method, we conduct a comparative analysis by juxtaposing our model against the following alternative models:

Model 1 (Pure L prior): Nonconvex regularization scheme is utilized to encode the low-rank property of TR factors, while disregarding their local smoothness characteristics, i.e.,

$$\begin{aligned} & \min_{[\mathcal{G}], \mathcal{E}} \sum_{n=1}^3 \|\mathcal{G}^{(n)}\|_{\Phi, \mathcal{L}} + \lambda_5 \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi}, \\ & \text{s.t. } \mathcal{M} = \mathcal{B} + \mathcal{E} = \Re([\mathcal{G}]) + \mathcal{E}, \end{aligned} \quad (33)$$

where λ_5 is the regularization parameter, $\|\cdot\|_{\ell_{F,1}^\psi}$ and $\|\cdot\|_{\Phi, \mathcal{L}}$ are defined according to Formulas (5) and (9), respectively.

Model 2 (Pure S prior): Nonconvex regularization scheme is utilized to encode the smoothness property of TR factors, while disregarding their low-rankness characteristics, i.e.,

$$\begin{aligned} & \min_{[\mathcal{G}], \mathcal{E}} \sum_{n=1}^3 \|\mathcal{G}^{(n)}\|_{\mathcal{D}_n} + \lambda_6 \cdot \|\mathcal{E}\|_{\ell_{F,1}^\psi}, \\ & \text{s.t. } \mathcal{M} = \mathcal{B} + \mathcal{E} = \Re([\mathcal{G}]) + \mathcal{E}, \end{aligned} \quad (34)$$

where λ_6 is the regularization parameter, $\|\cdot\|_{\ell_{F,1}^\psi}$ is defined according to Formula (6), and \mathcal{D}_n is the first-order difference matrix.

The experimental results are shown in Figure 10, which presents various AUC values obtained from three prior representation paradigms under different nonconvex regularization strategies. Wherein, the AUC values refer to the average AUC values obtained on the 11 HSI datasets. We find that compared with the pure low-rankness prior and the pure smoothness prior, the proposed joint L+S priors can well improve the detection performance.

3) *Part 3:* To verify the effectiveness of the TR decomposition, we replace the low-TR-rank factorization scheme in our HAD model with other tensor decompositions to achieve anomaly detection, such as TT decomposition, Tucker decomposition [54], T-SVD decomposition [57], and *multiscale entanglement renormalization ansatz* (MERA) [59]. In addition, we also compare our HAD algorithm with the recently proposed HAD method [58], which is based on the nonconvex low-TR-rank gradient tensor approximation. This experiment keeps on using the experimental datasets from the previous part. The experimental results are depicted in Figure 11. It can be easily noticed from Figure 11 that, the proposed algorithm achieves optimal AUC values in most cases. This shows that the TR decomposition and the nonconvex regularization strategies can fully exploit the spectral-spatial correlation of the background, thereby improving detection performance.

TABLE IV: The AUC values obtained by the proposed HAD model and its degraded versions for different HSI datasets.

HAD-Model	AUC-Metrics	Salinas	Pavia	Hyperion	HYDICE	San-Diego	Airport-4	Beach-3	Beach-4	Urban-3	Urban-4	Urban-5
Model 1	$AUC_{(P_D, P_F)}$	0.9703	0.9986	0.9935	0.9709	0.9959	0.9897	0.9956	0.9823	0.9730	0.9863	0.9670
	$AUC_{(ODP)}$	1.2238	1.4730	1.3674	1.3891	1.3180	1.4215	1.4466	1.2976	1.1945	1.0389	1.1726
	$AUC_{(SNPR)}$	1.8815	27.2931	13.2716	16.6455	19.0602	3.6785	11.2594	19.2033	8.1959	25.2131	8.1807
	$AUC_{(TDBS)}$	0.2535	0.4744	0.3739	0.4182	0.3221	0.4318	0.4510	0.3153	0.2215	0.0526	0.2056
Model 2	$AUC_{(P_D, P_F)}$	0.9959	0.9985	0.9940	0.9725	0.9933	0.9966	0.9976	0.9800	0.9742	0.9871	0.9766
	$AUC_{(ODP)}$	1.3840	1.4733	1.3704	1.3982	1.3357	1.5237	1.5292	1.2822	1.1681	1.0293	1.2596
	$AUC_{(SNPR)}$	2.9909	27.1495	12.0886	16.7718	17.3890	6.4321	22.9846	17.9461	8.7253	25.7181	4.4842
	$AUC_{(TDBS)}$	0.3882	0.4748	0.3764	0.4257	0.3425	0.5272	0.5316	0.3022	0.1939	0.0421	0.2841
Model 3	$AUC_{(P_D, P_F)}$	0.9955	0.9969	0.9830	0.9514	0.9937	0.9957	0.9962	0.9731	0.9734	0.9940	0.9789
	$AUC_{(ODP)}$	1.5132	1.3938	1.3328	1.4394	1.3696	1.5837	1.4617	1.2483	1.3891	1.0889	1.2797
	$AUC_{(SNPR)}$	18.9065	21.1303	7.3124	6.3971	8.3112	11.4182	11.5632	12.0305	6.0296	8.7531	6.1895
	$AUC_{(TDBS)}$	0.5138	0.3969	0.3498	0.4880	0.3759	0.5880	0.4655	0.2752	0.4157	0.0949	0.3008
Model 4	$AUC_{(P_D, P_F)}$	0.9997	0.9983	0.9943	0.9902	0.9890	0.9978	0.9986	0.9824	0.9819	0.9824	0.9847
	$AUC_{(ODP)}$	1.5915	1.3854	1.3652	1.4799	1.4310	1.5976	1.4955	1.2894	1.4370	1.0556	1.3674
	$AUC_{(SNPR)}$	23.9790	27.7836	13.7848	17.5180	12.5630	17.9849	21.2692	16.4999	11.4464	19.2204	10.4884
	$AUC_{(TDBS)}$	0.5918	0.3871	0.3709	0.4897	0.4421	0.5998	0.4969	0.3070	0.4551	0.0732	0.3826
Proposed	$AUC_{(P_D, P_F)}$	0.9988	0.9975	0.9967	0.9902	0.9927	0.9983	0.9990	0.9846	0.9908	0.9963	0.9821
	$AUC_{(ODP)}$	1.7059	1.4487	1.3705	1.5269	1.5443	1.5961	1.4975	1.2877	1.4974	1.1168	1.3310
	$AUC_{(SNPR)}$	26.3255	28.6802	15.6485	19.6848	16.3717	19.0930	25.8585	18.8186	14.9567	31.0597	10.0293
	$AUC_{(TDBS)}$	0.7071	0.4512	0.3738	0.5367	0.5515	0.5978	0.4985	0.3031	0.5066	0.1205	0.3489

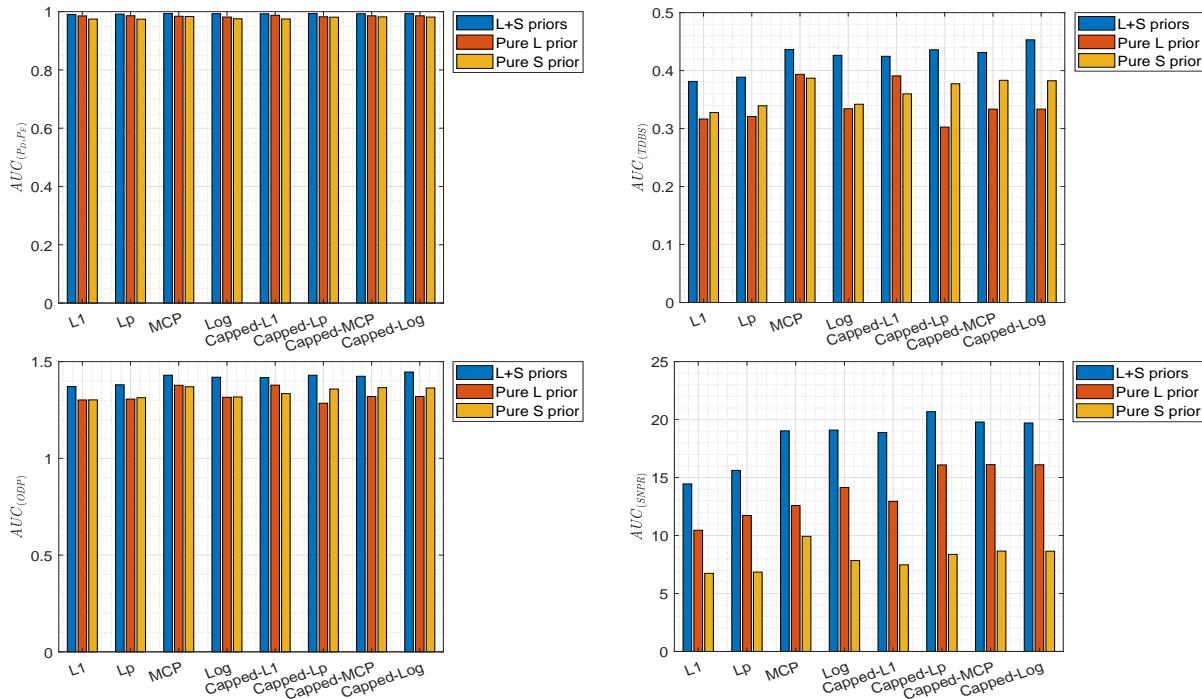


Fig. 10: The influence of different prior representation methods upon anomaly detection performance of the proposed method under various nonconvex regularization schemes.

TABLE V: The AUC values obtained by the proposed HAD model under different nonconvex combinations $\Phi(\cdot)+\psi(\cdot)$ for various HSI datasets with subpixel anomalies.

HSIs-Name	AUC-Metrics	L1		Lp		MCP		Log		Capped-L1		Capped-Lp		Capped-MCP		Capped-Log	
		FFT	DCT	FFT	DCT	FFT	DCT	FFT	DCT	FFT	DCT	FFT	DCT	FFT	DCT	FFT	DCT
Hyperion	$AUC_{(P_D, P_F)}$	0.9953	0.9944	0.9967	0.9956	0.9906	0.9926	0.9951	0.9950	0.9948	0.9934	0.9959	0.9924	0.9962	0.9935	0.9959	0.9935
	$AUC_{(ODP)}$	1.3101	1.3354	1.3027	1.3706	1.3371	1.4157	1.3017	1.3611	1.3221	1.3954	1.3161	1.3653	1.3153	1.4149	1.3099	1.4149
	$AUC_{(SNPR)}$	8.2186	6.4845	8.1385	6.7609	10.0104	9.3219	10.1093	8.6432	8.1611	7.6908	8.4699	9.0822	9.1034	7.8352	9.0871	7.8275
	$AUC_{(TDBS)}$	0.3147	0.3410	0.3061	0.3750	0.3464	0.4231	0.3067	0.3661	0.3274	0.4020	0.3202	0.3729	0.3191	0.4214	0.3141	0.4214
HYDICE-I	$AUC_{(P_D, P_F)}$	0.9964	0.9971	0.9964	0.9972	0.9941	0.9945	0.9955	0.9951	0.9918	0.9926	0.9951	0.9930	0.9953	0.9961	0.9953	0.9935
	$AUC_{(ODP)}$	1.4463	1.5152	1.3923	1.4314	1.4838	1.4800	1.4521	1.4682	1.4353	1.4983	1.4657	1.4972	1.4668	1.4679	1.4662	1.4736
	$AUC_{(SNPR)}$	15.4641	16.1052	19.9096	20.9871	25.0011	23.0837	23.0879	24.3862	19.6751	20.0651	21.1034	21.4722	21.7587	20.5353	21.7211	21.2262
	$AUC_{(TDBS)}$	0.4499	0.5181	0.3959	0.4342	0.4897	0.4855	0.4566	0.4731	0.4435	0.5057	0.4708	0.5042	0.4715	0.4720	0.4709	0.4801
HYDICE-II	$AUC_{(P_D, P_F)}$	0.9914	0.9954	0.9941	0.9910	0.9855	0.9844	0.9947	0.9963	0.9951	0.9918	0.9927	0.9943	0.9901	0.9856	0.9903	0.9914
	$AUC_{(ODP)}$	1.5017	1.5932	1.5591	1.5830	1.5395	1.5179	1.4384	1.5133	1.5165	1.5797	1.3904	1.4317	1.4709	1.4989	1.4731	1.4546
	$AUC_{(SNPR)}$	5.9149	6.9202	7.1408	7.0256	9.5867	12.1078	11.1171	10.6459	10.6961	9.8834	10.6946	10.3506	9.8432	8.2111	9.9273	11.5243
	$AUC_{(TDBS)}$	0.5103	0.5978	0.5651	0.5920	0.5541	0.5335	0.4437	0.5170	0.5214	0.55878	0.3977	0.4374	0.4809	0.5134	0.4828	0.4632
AVIRIS-II	$AUC_{(P_D, P_F)}$	0.9911	0.9963	0.9961	0.9960	0.9942	0.9971	0.9964	0.9981	0.9927	0.9939	0.9942	0.9972	0.9975	0.9955	0.9974	0.9941
	$AUC_{(ODP)}$	1.4763	1.5000	1.5067	1.5047	1.5282	1.5236	1.5287	1.5287	1.5287	1.5348	1.5241	1.5271	1.5259	1.5279	1.5261	1.5265
	$AUC_{(SNPR)}$	34.1055	31.8844	34.1388	34.1388	39.3751	36.4750	34.8498	37.0606	21.5638	32.6637	33.7411	34.7466	33.4456	32.9886	33.4628	33.3617
	$AUC_{(TDBS)}$	0.4852	0.5037	0.5087	0.5087	0.5341	0.5265	0.5268	0.5306	0.5311	0.5410	0.5299	0.5299	0.5285	0.5324	0.5285	0.5325
AVIRIS-I	$AUC_{(P_D, P_F)}$	0.9732	0.9800	0.9705	0.9735	0.9859	0.9881	0.9888	0.9876	0.9721	0.9829	0.9728	0.9846	0.9695	0.9883	0.9687	0.9885
	$AUC_{(ODP)}$	1.4352	1.4456	1.4072	1.4172	1.4584	1.4528	1.4728	1.4702	1.4132	1.4306	1.3659	1.4668	1.3464	1.4575	1.3455	1.4570
	$AUC_{(SNPR)}$	11.3653	11.6347	12.2208	11.5077	22.0709	15.9134	18.5783	22.7939	14.8688	12.5085	14.5268	21.3680	21.2485	16.6486	21.2086	16.4736
	$AUC_{(TDBS)}$	0.4619	0.4656	0.4366	0.4437	0.4725	0.4647	0.4841	0.4825	0.4412	0.4477	0.3932	0.4822	0.3769	0.4693	0.3769	0.4685

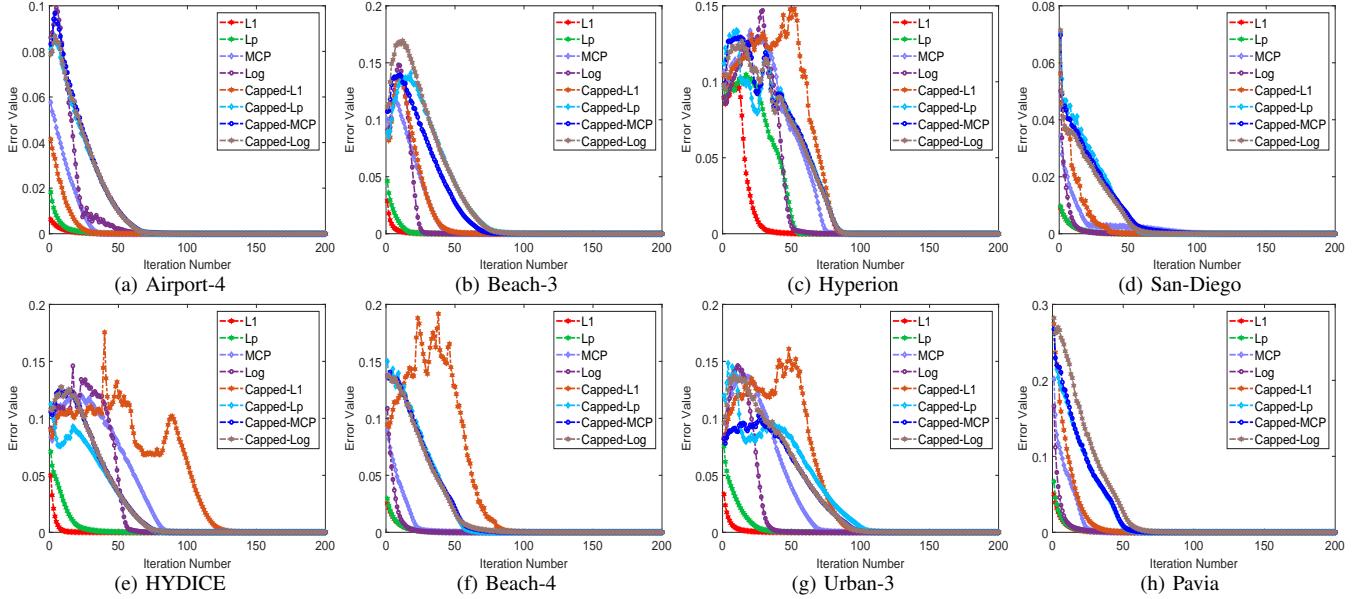


Fig. 12: The convergence behavior of the proposed HAD algorithm under different nonconvex functions.

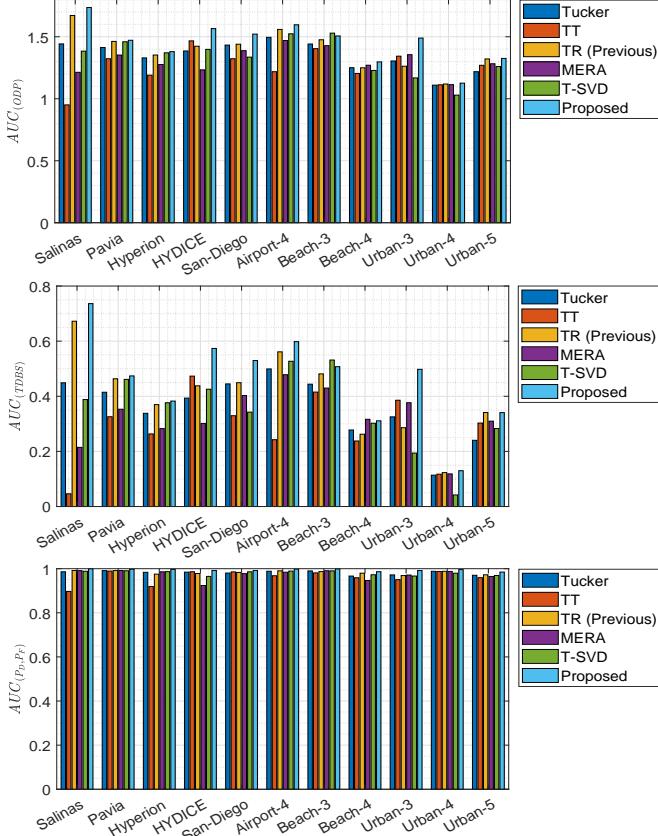


Fig. 11: The influence of various tensor decomposition schemes upon anomaly detection performance of the proposed method under different HSIs datasets.

4) *Part 4:* We conducted experiments on HSI datasets with subpixel anomalies to further verify the effectiveness and superiority of the proposed HAD algorithm. The hyperspectral data utilized in this part are consistent with those used in the literature [31], i.e., HYDICE-I ($80 \times 100 \times 162$), HYDICE-II ($60 \times 80 \times 162$), AVIRIS-I ($100 \times 100 \times 189$), AVIRIS-II ($120 \times 120 \times 204$), and Hyperion ($150 \times 150 \times 155$). The

parameter settings for this experiment are the same as those in section IV-C. The experimental results can be found in Table V, from which we can see that our proposed nonconvex HAD framework exhibits good detection performance in terms of various output AUC values. In general, the *Discrete Cosine Transform* (DCT) is slightly better than the *Fast Fourier Transform* (FFT) in most cases.

E. Convergence Analysis

Under different nonconvex combinations: $\Phi(\cdot)+\psi(\cdot)$, the convergence error curves of the proposed HAD algorithm on all eight tested HSI datasets are presented in Figure 12. For brevity, the nonconvex functions $\Phi(\cdot)$ and $\psi(\cdot)$ are set to be the same. It can be seen that compared to the non-Capped-type functions, the Capped-type functions exhibit a slower rate of convergence. Although there exist some fluctuations during the first 20~50 iterations, they finally converge to zero after at most 100 iterations. Thereby, the proposed HAD algorithm has good convergence and stability.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we have integrated several effective technologies, including joint tensor decompositions, gradient maps modeling, and novel nonconvex sparsity-inducing strategy, to collaboratively devise an innovative HAD method called HAD-EUNTRFR. In our formulated HAD model, the powerful TR decomposition is utilized to fully mine the essential structural information of background component in the spectral and spatial modes. Drawing upon the interpretable gradient TR factor, we further introduce a unified nonconvex regularizer induced by the T-SVD framework. This regularizer effectively captures the inherent low-rankness and the transformed sparsity simultaneously, thereby substantially enhancing the model's performance and robustness. Meanwhile, another generalized nonconvex constraint is incorporated into our model to promote the structured sparsity of anomalous targets. Algorithmically, we deduce a detailed procedure with an ADMM

structure for solving the proposed nonconvex model. A series of experiments considering on both synthetic and real-world HSIs have verified the superiority and effectiveness of our HAD method. In future research, we plan to first develop a novel data-driven regularization strategy for profound characterization, along with a sketching framework aimed at dimensionality reduction. Furthermore, we intend to integrate these key tools into tensor-format deep neural networks to explore more accurate and effective methods for the HAD task.

REFERENCES

- [1] G. Vivone, L.-J. Deng, S. Deng, D. Hong, M. Jiang, C. Li, W. Li, H. Shen, X. Wu, J.-L. Xiao, J. Yao, M. Zhang, J. Chanussot, S. Garcia, and A. Plaza, “Deep learning in remote sensing image fusion: Methods, protocols, data, and future perspectives,” *IEEE Geosci. Remote Sens. Mag.*, vol. 13, no. 1, pp. 269–310, 2025.
- [2] G. Vivone, “Multispectral and hyperspectral image fusion in remote sensing: A survey,” *Inform. Fusion*, vol. 89, pp. 405–417, 2023.
- [3] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, “Non-local meets global: An iterative paradigm for hyperspectral image restoration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, 2020.
- [4] L. Pang, X. Rui, L. Cui, H. Wang, D. Meng, and X. Cao, “Hir-diff: Unsupervised hyperspectral image restoration via improved diffusion models,” in *Proc. IEEE Conf. Comput. Vision. Pattern. Recognit. (CVPR)*, 2024, pp. 3005–3014.
- [5] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, “Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven,” *IEEE Trans. Neural Netw. Learn. Syst.*, 2023.
- [6] H. Wang, J. Peng, X. Cao, J. Wang, Q. Zhao, and D. Meng, “Hyperspectral image denoising via nonlocal spectral sparse subspace representation,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5189–5203, 2023.
- [7] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, “Hyperspectral super-resolution via coupled tensor ring factorization,” *Pattern Recognit.*, vol. 122, p. 108280, 2022.
- [8] H. Xu, Y. Quan, M. Qin, Y. Wang, C. Fang, Y. Li, and J. Zheng, “Nonlinear learnable triple-domain transform tensor nuclear norm for hyperspectral image super-resolution,” *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5514117.
- [9] B. Yang and B. Wang, “Band-wise nonlinear unmixing for hyperspectral imagery using an extended multilinear mixing model,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6747–6762, 2018.
- [10] J. Gu, B. Yang, and B. Wang, “Nonlinear unmixing for hyperspectral images via kernel-transformed bilinear mixing models,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520313.
- [11] M. Li, B. Yang, and B. Wang, “EMLM-Net: An extended multilinear mixing model-inspired dual-stream network for unsupervised nonlinear hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5509116.
- [12] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, “Hyperspectral image classification with markov random fields and a convolutional neural network,” *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, 2018.
- [13] X. Cao, J. Yao, Z. Xu, and D. Meng, “Hyperspectral image classification with convolutional neural network and active learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4604–4616, 2020.
- [14] Y. Xu, L. Zhang, B. Du, and L. Zhang, “Hyperspectral anomaly detection based on machine learning: An overview,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3351–3364, 2022.
- [15] H. Su, Z. Wu, H. Zhang, and Q. Du, “Hyperspectral anomaly detection: A survey,” *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, 2021.
- [16] X. Hu, C. Xie, Z. Fan, Q. Duan, D. Zhang, L. Jiang, X. Wei, D. Hong, G. Li, X. Zeng *et al.*, “Hyperspectral anomaly detection using deep learning: A review,” *Remote Sens.*, vol. 14, no. 9, p. 1973, 2022.
- [17] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, “Anomaly detection from hyperspectral imagery,” *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, 2002.
- [18] N. M. Nasrabadi, “Hyperspectral target detection: An overview of current and future challenges,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, 2013.
- [19] I. S. Reed and X. Yu, “Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, 1990.
- [20] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, “Multipixel anomaly detection with unknown patterns for hyperspectral imagery,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5557–5567, 2021.
- [21] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, “Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, 2013.
- [22] H. Kwon and N. M. Nasrabadi, “Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, 2005.
- [23] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, “Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, 2014.
- [24] W. Li, G. Wu, and Q. Du, “Transferred deep learning for anomaly detection in hyperspectral imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, 2017.
- [25] T. Jiang, Y. Li, W. Xie, and Q. Du, “Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, 2020.
- [26] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, “Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, 2021.
- [27] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, “Hyperspectral anomaly detection with guided autoencoder,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 553818.
- [28] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, “LREN: low-rank embedded network for sample-free hyperspectral anomaly detection,” in *Proc. AAAI Conf. Artif. Intel.*, vol. 35, no. 5, 2021, pp. 4139–4146.
- [29] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, “LRR-Net: an interpretable deep unfolding network for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5513412.
- [30] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, “PDB-SNet: pixel-shuffle down-sampling blind-spot reconstruction network for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, 2023, Art. no. 5511914.
- [31] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, “BS³LNet: a new blind-spot self-supervised learning network for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504218.
- [32] D. Wang, L. Ren, X. Sun, L. Gao, and J. Chanussot, “Nonlocal and local feature-coupled self-supervised network for hyperspectral anomaly detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 6981–6993, 2025.
- [33] X. Cheng, Y. Huo, S. Lin, Y. Dong, S. Zhao, M. Zhang, and H. Wang, “Deep feature aggregation network for hyperspectral anomaly detection,” *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–16, 2024.
- [34] J. Lian, L. Wang, H. Sun, and H. Huang, “GT-HAD: gated transformer for hyperspectral anomaly detection,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 2, pp. 3631–3645, 2025.
- [35] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, “Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery,” *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 083 641–083 641, 2014.
- [36] Y. Zhang, B. Du, L. Zhang, and S. Wang, “A low-rank and sparse matrix decomposition-based mahalanobis distance method for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, 2015.
- [37] L. Li, W. Li, Q. Du, and R. Tao, “Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection,” *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, 2020.
- [38] Y. Zhang, Y. Fan, M. Xu, W. Li, G. Zhang, L. Liu, and D. Yu, “An improved low rank and sparse matrix decomposition-based anomaly target detection algorithm for hyperspectral imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2663–2672, 2020.
- [39] C. Chang, H. Cao, S. Chen, X. Shang, M. Song, and C. Yu, “Orthogonal subspace projection-based GoDec for low rank and sparsity matrix decomposition for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2403–2429, 2021.
- [40] L. Li, Q. Zhang, M. Song, and C.-I. Chang, “Feedback band group and variation low rank sparse model for hyperspectral image anomaly

- detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5508919.
- [41] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, 2014.
- [42] K. Tan, Z. Hou, F. Wu, Q. Du, and Y. Chen, "Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation," *Remote Sens.*, vol. 11, no. 11, p. 1318, 2019.
- [43] Z. Wu, H. Su, X. Tao, L. Han, M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Hyperspectral anomaly detection with relaxed collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533417.
- [44] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, 2015.
- [45] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, 2017.
- [46] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, 2015.
- [47] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, 2019.
- [48] ——, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1472–1486, 2020.
- [49] M. Wang, D. Hong, B. Zhang, L. Ren, J. Yao, and J. Chanussot, "Learning double subspace representation for joint hyperspectral anomaly detection and noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507517.
- [50] T. Guo, L. He, F. Luo, X. Gong, L. Zhang, and X. Gao, "Learnable background endmember with subspace representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5501513.
- [51] L. Ren, L. Gao, M. Wang, X. Sun, and J. Chanussot, "HADGSM: A unified nonconvex framework for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5503415.
- [52] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, 2020.
- [53] S. Feng, D. Wu, R. Feng, and C. Zhao, "Hyperspectral anomaly detection with total variation regularized low rank tensor decomposition and collaborative representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6009105.
- [54] W. Shang, J. Peng, Z. Wu, Y. Xu, M. Jouni, M. Dalla Mura, and Z. Wei, "Hyperspectral anomaly detection via sparsity of core tensor under gradient domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5517816.
- [55] J. Wang, Y. Xia, and Y. Zhang, "Anomaly detection of hyperspectral image via tensor completion," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1099–1103, 2020.
- [56] Y.-P. Zhao, H. Li, Y. Chen, Z. Wang, and X. Li, "Hyperspectral anomaly detection via structured sparsity plus enhanced low-rankness," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5515115.
- [57] H. Wang, J. Peng, W. Qin, J. Wang, and D. Meng, "Guaranteed tensor recovery fused low-rankness and smoothness," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10 990–11 007, 2023.
- [58] W. Qin, H. Wang, F. Zhang, J. Wang, X. Cao, and X.-L. Zhao, "Tensor ring decomposition-based generalized and efficient nonconvex approach for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5539818.
- [59] Q. Xiao, L. Zhao, S. Chen, and X. Li, "Hyperspectral anomaly detection via MERA decomposition and enhanced total variation regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5514919.
- [60] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, 2022.
- [61] H. Qin, Q. Shen, H. Zeng, Y. Chen, and G. Lu, "Generalized nonconvex low-rank tensor representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526612.
- [62] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, "Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8679–8692, 2023.
- [63] Q. Yu and M. Bai, "Generalized nonconvex hyperspectral anomaly detection via background representation learning with dictionary constraint," *SIAM J. Imag. Sci.*, vol. 17, no. 2, pp. 917–950, 2024.
- [64] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo, and T. Huang, "Low-rank high-order tensor completion with applications in visual data," *IEEE Trans. Image Process.*, vol. 31, pp. 2433–2448, 2022.
- [65] W. Qin, H. Wang, F. Zhang, W. Ma, J. Wang, and T. Huang, "Non-convex robust high-order tensor completion using randomized low-rank approximation," *IEEE Trans. Image Process.*, vol. 33, pp. 2835–2850, 2024.
- [66] J. Hou, F. Zhang, H. Qiu, J. Wang, Y. Wang, and D. Meng, "Robust low-tubal-rank tensor recovery from binary measurements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 8, pp. 4355–4373, 2021.
- [67] J. Wang, J. Hou, and Y. C. Eldar, "Tensor robust principal component analysis from multilevel quantized observations," *IEEE Trans. Inf. Theory*, vol. 69, no. 1, pp. 383–406, 2022.
- [68] H. Wang, F. Zhang, J. Wang, T. Huang *et al.*, "Generalized nonconvex approach for low-tubal-rank tensor recovery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3305–3319, 2021.
- [69] X. Liu, J. Hou, J. Peng, H. Wang, D. Meng, and J. Wang, "Tensor compressive sensing fused low-rankness and local-smoothness," in *Proc. AAAI Conf. Artif. Intel.*, vol. 37, no. 7, 2023, pp. 8879–8887.
- [70] X. Zhao, K. Liu, X. Wang, S. Zhao, K. Gao, H. Lin, Y. Zong, and W. Li, "Tensor adaptive reconstruction cascaded with global and local feature fusion for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 607–620, 2025.
- [71] M. Feng, W. Chen, Y. Yang, Q. Shu, H. Li, and Y. Huang, "Hyperspectral anomaly detection based on tensor ring decomposition with factors TV regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5514114.
- [72] X. Zhao, K. Liu, K. Gao, and W. Li, "Hyperspectral time-series target detection based on spectral perception and spatial-temporal tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–12, 2023.
- [73] X. Zhao, J. Huang, Y. Gao, and Q. Wang, "Hyperspectral target detection based on prior spectral perception and local graph fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 13 936–13 948, 2024.
- [74] Q. Zhao, G. Zhou, S. Xie, L. Zhang, and A. Cichocki, "Tensor ring decomposition," *arXiv preprint arXiv:1606.05535*, 2016.
- [75] L. Yuan, C. Li, D. Mandic, J. Cao, and Q. Zhao, "Tensor ring decomposition with rank minimization on latent space: An efficient approach for tensor completion," in *Proc. AAAI Conf. Artif. Intel.*, vol. 33, no. 01, 2019, pp. 9151–9158.
- [76] H. Xu, M. Qin, S. Chen, Y. Zheng, and J. Zheng, "Hyperspectral-multispectral image fusion via tensor ring and subspace decompositions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8823–8837, 2021.
- [77] J. Zhang, L. Zhu, C. Deng, and S. Li, "Hyperspectral and multispectral image fusion via logarithmic low-rank tensor ring decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, 2024.
- [78] Y. Chen, J. Zeng, W. He, X.-L. Zhao, and T.-Z. Huang, "Hyperspectral and multispectral image fusion using factor smoothed tensor ring decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515417.
- [79] H. Zhang, T.-Z. Huang, X.-L. Zhao, W. He, J. K. Choi, and Y.-B. Zheng, "Hyperspectral image denoising: Reconciling sparse and low-tensor-ring-rank priors in the transformed domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502313.
- [80] P.-L. Wu, X.-L. Zhao, M. Ding, Y.-B. Zheng, L.-B. Cui, and T.-Z. Huang, "Tensor ring decomposition-based model with interpretable gradient factors regularization for tensor completion," *Knowl.-Based Syst.*, vol. 259, p. 110094, 2023.
- [81] 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 Trans. Image Process.*, vol. 29, pp. 7889–7903, 2020.
- [82] J. Peng, Y. Wang, H. Zhang, J. Wang, and D. Meng, "Exact decomposition of joint low rankness and local smoothness plus sparse matrices," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 5, pp. 5766–5781, 2022.
- [83] G. Marjanovic and V. Solo, "On ℓ_q optimization and matrix completion," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5714–5724, 2012.
- [84] P. Gong, C. Zhang, Z. Lu, J. Huang, and J. Ye, "A general iterative shrinkage and thresholding algorithm for non-convex regularized optimization problems," in *Proc. Int. Conf. Mach. Learn. (ICML)*. PMLR, 2013, pp. 37–45.

- [85] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, vol. 38, no. 2, pp. 894–942, 2010.
- [86] L. Pan and X. Chen, "Group sparse optimization for images recovery using capped folded concave functions," *SIAM J. Imag. Sci.*, vol. 14, no. 1, pp. 1–25, 2021.
- [87] X. He, J. Wu, Q. Ling, Z. Li, Z. Lin, and S. Zhou, "Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509917.
- [88] S. Liu, C. Zhu, D. Ran, and G. Wen, "Anomaly detection via tensor multi-subspace learning and nonconvex low-rank regularization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 8178–8190, 2023.
- [89] D. Qiu, M. Bai *et al.*, "Nonlocal robust tensor recovery with nonconvex regularization," *Inverse Probl.*, vol. 37, no. 3, p. 035001, 2021.
- [90] S. Boyd, N. Parikh, E. Chu *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
- [91] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, Y. Chen, and W. He, "Double-factor-regularized low-rank tensor factorization for mixed noise removal in hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8450–8464, 2020.
- [92] F. Verdoja and M. Grangetto, "Graph laplacian for image anomaly detection," *Mach. Vis. Appl.*, vol. 31, no. 1, p. 11, 2020.
- [93] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, 2017.
- [94] Z. Wang, K. Tan, X. Wang, and W. Zhang, "Leveraging multi-class background description and token dictionary representation for hyperspectral anomaly detection," *Pattern Recognit.*, p. 111945, 2025.
- [95] S. Garske, B. Evans, C. Artlett, and K. C. Wong, "Erx: A fast real-time anomaly detection algorithm for hyperspectral line scanning," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5503617.
- [96] C.-I. Chang, "Comprehensive analysis of receiver operating characteristic (ROC) curves for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–24, 2022.
- [97] D. F. Williamson, R. A. Parker, and J. S. Kendrick, "The box plot: a simple visual method to interpret data," *Ann. Internal Med.*, vol. 110, no. 11, pp. 916–921, 1989.