

Generalized Nonconvex Low-rank Tensor Representation for Hyperspectral Anomaly Detection

Hao Qin, Qiangqiang Shen, Haijin Zeng, Yongyong Chen, *Member, IEEE*, Guangming Lu, *Senior Member, IEEE*

Abstract—Low-rank tensor representation (LRTR) methods have attracted great interest for their powerful ability to separate backgrounds and anomalies. However, most of the current LRTR models use the popular and convex surrogate tensor nuclear norm to solve optimization problems, which results in a loose approximation and suboptimal solver for the original problem. Besides, most existing methods solve the nonconvex optimization problems case-by-case, consequently losing one unified solver. To solve the above issues, we propose the Generalized Nonconvex Low-rank Tensor Representation (GNLTR) for hyperspectral anomaly detection (HAD), a unified solver not case-by-case one of existing nonconvex optimization problems. Compared to the tensor nuclear norm, GNLTR contains many popular nonconvex penalty functions as tighter regularizers of the tensor tubal rank to constrain the low rank of the background. Moreover, the $L_{2,1}$ norm has been integrated into the GNLTR model for the sparse anomalies. For the optimization problem, it is handled quickly and efficiently through a well-organized alternating direction method of multipliers (ADMM). The experiments on several real-world hyperspectral data sets demonstrate the superior performance of the GNLTR model in comparison with some state-of-the-art anomaly detection models.

Index Terms—Low-rank tensor representation, hyperspectral anomaly detection, unified anomaly detection nonconvex framework, alternating direction method of multipliers.

I. INTRODUCTION

IMAGING spectroscopy has greatly promoted the capture and processing of hyperspectral images (HSIs). Extensive hyperspectral imaging applications have emerged in recent years, such as military surveillance [1], [2], material identification [3], [4], grassland degradation monitoring [5], [6], invasive plant monitoring [7], [8], and image classification [9], [10]. Anomaly detection (AD) [11] occupies an important

place in these applications since it could separate potentially valuable targets from HSIs. High-value targets considered as anomalies like airplanes and vehicles are usually embedded with a low probability in locally homogeneous backgrounds, which have different shapes, spectral features, and other characteristics from the surrounding pixels. As for the backgrounds like mountains, rivers, and roads, their features are locally similar with a limited variation to each other. Therefore, AD could be regarded as a binary classification problem based on two primary principles: (1) anomalies appear with low probability in the background, while their spectral features differ from the surrounding background's spectral features; (2) the backgrounds are closely connected to each other due to the local similarity and limited variation from surrounding pixels.

Based on the above basic criteria, numerous methods have been formulated during the last decades. The pioneering work is the famous Reed-Xiaoli (RX) algorithm [12] as well as its variants including global RX and local RX [13], dual-window-based RX [14] and Multiple-Window-based RX [15]. Nevertheless, the complexity and variability of hyperspectral data in practical applications usually cause difficulties in satisfying the multivariate normal distribution condition. To overcome this problem, various algorithms such as subspace RX [16], weighted-RXD [17], and linear filter-based RXD [17] have been introduced. The weighted RXD improved background estimation by altering the weights of background samples, anomalous pixels or noisy signals while the linear filter-based RXD filtered wrong anomalies or noise instances by using the probability. The kernel RX (KRX) [18] provided the kernelization procedure, which has the ability for mapping the raw data into high-dimensional space, such that the raw data can be separated linearly. The cluster KRX [19] classified pixels of backgrounds into different clusters with a fast eigendecomposition algorithm to detect anomalies.

In recent research, there is a tendency to focus on the inherent characteristics of HSIs. Prominent characteristics among them are the low rank of backgrounds and the anomalous scarcity. With this in mind, the robust principal component analysis framework (RPCA) [20] decomposed HSIs into a background part and sparse anomalies. The studies in [21], [22] extended RPCA into the low-rank and sparse matrix decomposition (LRaSMD) in which the matrix was transformed into a low-rank matrix, a sparse matrix, and a noise matrix. The LRaSMD-based Mahalanobis distance (LSMAD) detector [23] took full advantage of the low-rankness on the basis of LRaSMD. Niu et al. [24] enhanced the robustness and detec-

This work was supported in part by the National Natural Science Foundation of China under Grant 62106063, by the Guangdong Natural Science Foundation under Grant 2022A1515010819, by the Shenzhen Science and Technology Program under Grant RCBS20210609103708013, and by Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies under Grant 2022B1212010005. (Hao Qin and Qiangqiang Shen contributed equally to this work. Corresponding author: Yongyong Chen).

Hao Qin, Yongyong Chen, and Guangming Lu are with the School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China. Yongyong Chen and Guangming Lu are also with Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies (Email: YongyongChen.cn@gmail.com).

Qiangqiang Shen is with the School of Electronics and Information Engineering, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, Guangdong, China, (Email: 1120810623@hit.edu.cn).

H. Zeng is with Image Processing and Interpretation, imec research group at Ghent University, Belgium (Email: Haijin.Zeng@UGent.be).

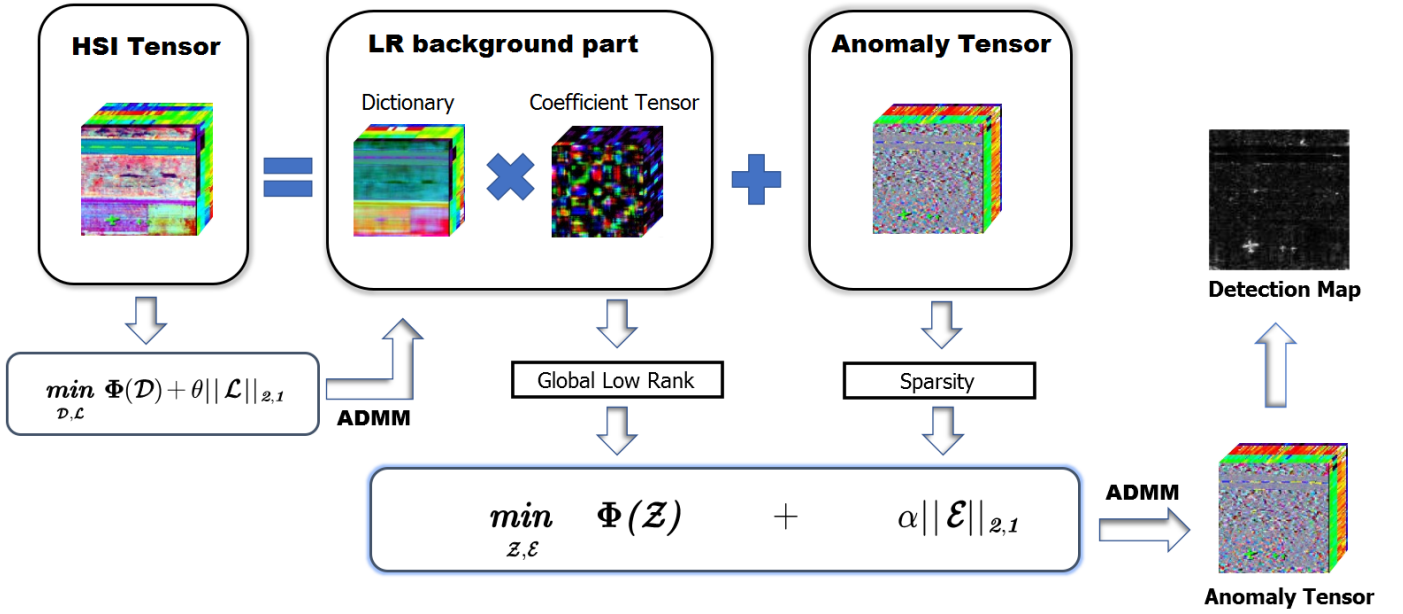


Fig. 1. The flow diagram of our proposed GNLTR model for hyperspectral image anomaly detection.

tion performance of their model with a learned dictionary and low-rank backgrounds. In [25], Xu et al. proposed a sparsity-inducing regularized term into low-rank representation (LRR) with a new strategy to construct a comprehensive dictionary. The LRASMD-based spectral-spatial (LS-SS) method [26] integrated both spatial information and spectral information of the data into the model. Zhu et al. [27] focused on the sparse part and engaged a clustering-based weighting strategy. In [28], the graph regularization and total variation (TV) regularization were embedded into LRR (GTVLRR) for conservation of the spatial relationships and geometrical structure in HSIs. The S^2 ELR [29] method employed Schatten- p norm to approximate the rank and structured sparse norm, which plays an effective role in restricting the low-rank nature of the background and the sparsity of the anomalies.

However, the above-mentioned methods usually transformed the 3-dimensional HSI into a 2-dimensional matrix, losing its intrinsic 3-dimensional characteristics. In contrast, tensor [30], widely implemented in plenty of fields like image fusion [31], [32], image restoration [33], [34], and subspace clustering [35], [36], is capable of preserving the inherent structure of multidimensional data. The first two dimensions of the 3-dimensional HSIs stand for spatial information while the third dimension is for spectral information. This drives researchers to conduct studies into tensor-based AD. Following the above research line, the low-rank tensor decomposition-based AD (LTDD) model [37] obtained a low rank tensor with the Tucker decomposition plus a sparse tensor through an unmixing technology. In [38], tensor principal component analysis (TPCA) was carried out for gaining the principal component part and the residual part of HSIs. While prior-based tensor approximation (PTA) [39] had fully integrated a linear TV norm, a novel truncated nuclear norm (TNNR), and $L_{2,1}$ norm. Moreover, PCA-TLRSR [40] adopted a weighted tensor nuclear norm, a well-designed dictionary, and a princi-

pal component analysis (PCA) method for AD. In addition, many deep learning-based HAD methods also keep the 3-dimensional characteristics of HSI. LREN [41] imported Deep neural networks (DNNs) to construct an unsupervised low-rank embedded network with a specifically designed latent representation. DeCNN-AD [42] proposed a plug-and-play framework by integrating a convolutional neural network (CNN) denoiser and a new dictionary based on clustering. In PDBSNet [43] and BS³Net [44], both of them proposed a novel blind-spot self-supervised reconstruction network which rebuilt the center pixel's spectral signatures through neighbor pixels and considered the pixels that had with high reconstruction errors as anomalous pixels. For PDBSNet, it also used the pixel-shuffle downsampling method to break out the pixels' spatial relevance and enhanced the difference between anomalies and shuffled surrounding pixels.

Simultaneously, most low-rank based models engaged the convex surrogate nuclear norm and its extensions as a rank substitute, but the effect is weaker than that of nonconvex surrogates [45]. Despite the wide exploration of tensor-based anomaly detection, the majority of these tensor-based approaches solve the nonconvex optimization problems case-by-case, lacking a unified framework in anomaly detection. For example, PTA [39] used the truncated nuclear norm while PCA-TLRSR [40] adopted the weighted tensor nuclear norm as the nonconvex surrogates of matrix rank and tensor tubal rank, respectively. For solving their corresponding nonconvex optimization minimization problems, PTA [39] developed the two-step approach while PCA-TLRSR [40] proposed the weighted tensor singular value thresholding. Consequently, a unified framework with nonconvex surrogates for anomaly detection should be put forward.

In this paper, a unified nonconvex framework named Generalized Nonconvex Low-rank Tensor Representation (GNLTR) is proposed for anomaly detection (AD). The flow

diagram of our GNLTR model is illustrated in Fig. 1. GNLTR transforms the HSI data into a dictionary derived from background information, the corresponding coefficient tensor and an anomaly tensor. Our GNLTR provides a unified nonconvex surrogate in place of the familiar convex surrogate nuclear norm to restrict the low rankness of the backgrounds, allowing us to select the adaptive nonconvex penalty functions. For the anomaly tensor, GNLTR utilizes the $L_{2,1}$ norm to measure the dense property and the sparsity in the spectral direction. Finally, the alternating direction method of multipliers (ADMM) offers us an excellent option to solve the nonconvex optimization problem.

Compared with existing HAD models, the main contributions of our paper are as follows:

- 1) We propose the generalized nonconvex low-rank tensor representation framework (GNLTR) for anomaly detection and a unified nonconvex surrogate function to approximate the tensor tubal rank, which fully takes advantage of the low-rank prior knowledge in anomaly detection.
- 2) A nonconvex, novel, and reliable dictionary construction approach is presented. Based on the reality that nonconvex surrogates are more approximate to $rank_t$ compared with convex surrogates, we can better exploit the low-ranking property of the background part, which means a purer background part could be separated from HSIs.
- 3) An iterative algorithm is derived based on the alternating direction method of multipliers and its effectiveness is proven by extensive experiments on real-world HSIs.

The remainder of this paper is organized as follows. Notations and preliminaries are presented in Section II. Section III introduces the unified anomaly detection nonconvex framework, the nonconvex dictionary construction and their details. Section IV shows the experimental results and discussions. Conclusions are drawn in Section V.

II. NOTATIONS AND PRELIMINARIES

In the section, we introduce the notations mentioned in our paper. For the sake of brevity, the main notations are presented in Table I. $diag(\mathbf{x})$ is on behalf of diagonal matrix in which the diagonal is $\mathbf{x}$. $diag(\mathbf{X})$ is a vector which is consisted of $\mathbf{X}$'s diagonal elements. $\bar{\mathcal{X}}$ is denoted as the result of FFT along the 3_{th} dimension of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, i.e., $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, \cdot, 3)$. Naturally, we could restore $\mathcal{X}$ with the inverse FFT, i.e., $\mathcal{X} = \text{ifft}(\bar{\mathcal{X}}, \cdot, 3)$. Meanwhile, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ could be transformed into a $n_1 n_3 \times n_2$ matrix through the unfold operator, i.e., $\text{unfold}(\mathcal{X}) = [\mathcal{X}^{(1)}; \mathcal{X}^{(2)}; \dots; \mathcal{X}^{(n_3)}]$. As for the block circulant matrix, it takes the slices of the tensor to construct an $n_3 \times n_3$ square matrix, which is defined as

$$\text{bcirc}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \dots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \dots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \mathcal{X}^{(n_3-2)} & \dots & \mathcal{X}^{(1)} \end{bmatrix} \quad (1)$$

The following part provides some definitions involved in the paper.

TABLE I
THE MAIN NOTATIONS MENTIONED IN THE PAPER

Notations	Description
$x, \mathbf{x}, \mathbf{X}, \mathcal{X}$	scalar, vector, matrix, tensor
$\mathcal{X}^T$	the conjugate transpose of $\mathcal{X}$
$\mathcal{I}$	the identity tensor
$\mathcal{X}^{(i)}$	the i_{th} frontal slice of $\mathcal{X}$
$\ \mathcal{X}\ _0, \ \mathcal{X}\ _1, \ \mathcal{X}\ _{2,1}$	the $l_0, l_1, l_{2,1}$ norm of $\mathcal{X}$
$\ \mathcal{X}\ _*$	the tensor nuclear norm of $\mathcal{X}$
$\mathcal{A} * \mathcal{B}$	the tensor product of $\mathcal{A}$ and $\mathcal{B}$
$\text{unfold}(\mathcal{X})$	unfold $\mathcal{X}$ along mode-3
$\text{rank}_a(\mathcal{X})$	the tensor average rank of $\mathcal{X}$
$\text{rank}_t(\mathcal{X})$	the tensor tubal rank of $\mathcal{X}$
$\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, \cdot, 3)$	the discrete Fourier transform of $\mathcal{X}$ along mode-3

Definition 1 (Tensor product [46]): For two 3-dimensional tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the tensor product $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ is

$$\mathcal{C} = \text{fold}(\text{bcirc}(\mathcal{A}) \cdot \text{unfold}(\mathcal{B})). \quad (2)$$

Definition 2 (Some special tensors):

Conjugate transpose tensor [46]: $\mathcal{A}^T \in \mathbb{R}^{n_2 \times n_1 \times n_3}$ denoted as the conjugate transpose of $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is acquired by slicing along the third dimension with conjugate transposing each slice and then assembling each slice after reversing the order of previous arrangement.

F-diagonal tensor [46]: If each slice of a tensor $\mathcal{S}$ is a diagonal matrix, the tensor $\mathcal{S}$ is named f-diagonal.

Identity tensor [46]: The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_2}$ is the one whose first frontal slice is an $n_1 \times n_1$ identity matrix and other frontal slices are zeros.

Orthogonal Tensor [46]: A tensor $\mathcal{Q} \in \mathbb{R}^{n_1 \times n_1 \times n_2}$ is orthogonal if it satisfies the equation $\mathcal{Q} * \mathcal{Q}^T = \mathcal{Q}^T * \mathcal{Q} = \mathcal{I}$.

Definition 3 (t-SVD [46], [47]): A tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ has the factorization

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \quad (3)$$

where $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ should be a f-diagonal tensor; $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors.

Definition 4 (Tensor tubal rank [48]): For a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its tensor tubal rank, denoted as $\text{rank}_t(\mathcal{X})$, is defined as the number of nonzero singular tubes of $\mathcal{S}$ with $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. rank_t is

$$\text{rank}_t(\mathcal{X}) = \#\{i, \mathcal{S}(i, i, :) \neq \mathbf{0}\}. \quad (4)$$

Definition 5 (Tensor Nuclear Norm (TNN) [47]): Firstly, we introduce the property induced by inverse DFT, formulated as

$$\mathcal{S}(i, i, 1) = \frac{1}{n_3} \sum_{j=1}^{n_3} \bar{\mathcal{S}}(i, i, j). \quad (5)$$

Second, $\mathcal{S}$ is f-diagonal, which means $\#\{i, \mathcal{S}(i, i, :) \neq \mathbf{0}\}$ is equal to $\#\{i, \mathcal{S}(i, i, 1) \neq \mathbf{0}\}$. With these, the TNN can be defined as

$$\|\mathcal{X}\|_{\oplus} = \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \bar{\mathcal{S}}(i, i, j), \quad (6)$$

where $m = \max(n_1, n_2)$.

Definition 6 (Tensor Average Rank [47]): The tensor average rank of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted as $\text{rank}_a(\mathcal{X})$ is defined as

$$\text{rank}_a(\mathcal{X}) = \frac{1}{n_3} \text{rank}(\text{bcirc}(\mathcal{X})). \quad (7)$$

Definition 7 (Tensor Spectral Norm [47]): For a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor spectral norm is defined as

$$\|\mathcal{X}\| = \|\text{bcirc}(\mathcal{X})\|, \quad (8)$$

where $\|\text{bcirc}(\mathcal{X})\|$ is the spectral norm for matrix $\text{bcirc}(\mathcal{X})$.

III. THE PROPOSED GNLTR MODEL AND THE NONCONVEX DICTIONARY CONSTRUCTION METHOD

In this section, a unified nonconvex surrogate of tensor tubal rank, a unified anomaly detection nonconvex framework, and a nonconvex dictionary construction method are introduced.

A. Unified Nonconvex Surrogate

Various kinds of nonconvex surrogates such as the weighted tensor nuclear norm [40], truncated nuclear norm [39] for specific nonconvex optimization problems have been proposed in previous studies and have been broadly used in low-latitude spaces. Therefore, we integrate these nonconvex surrogates like ETP [49], L_p [50] and SCAD [51] into a unified nonconvex surrogate structure, which applies different nonconvex penalty functions for different nonconvex optimization requirements.

Considering the structure of TNN, we can define the unified nonconvex surrogate for tensor tubal rank, i.e.,

$$\Phi(\mathcal{X}) = \frac{1}{n_3} \sum_{i=1}^m \sum_{j=1}^{n_3} \phi(\bar{\mathcal{S}}(i, i, j)), \quad (9)$$

where $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $m = \max(n_1, n_2)$, and $[\mathcal{U}, \mathcal{S}, \mathcal{V}] = \text{t-SVD}(\mathcal{X})$. Besides, some simple restrictions on the penalty function ϕ are needed here in order to satisfy the tensor tubal rank property. ϕ should be continuous, concave, and monotonically nondecreasing, which also meets

$$\begin{cases} \phi(0) = 0 \\ \lim_{x \rightarrow \infty} \frac{\phi(x)}{x} = 0. \end{cases} \quad (10)$$

What should be notable is that these requirements are simple and easy to meet. Based on the properties of ϕ , the inequality, i.e.

$$\partial\phi(y) \leq \frac{\phi(x) - \phi(y)}{x - y} \leq \partial\phi(x) \quad (11)$$

holds true when $0 < x < y$ and $x, y \in \mathbb{R}$. Table II summarizes some popular nonconvex penalty functions meeting the above restrictions. In these nonconvex penalty functions, λ is the scale parameter; γ is the shape parameter; x is the independent variable. The above two variables λ and γ are greater than 0, with $x \geq 0$.

Moreover, we derive the properties of $\Phi(\mathcal{X})$ with these constraints of ϕ . $\Phi(\mathcal{X})$ is a continuous, monotonically non-decreasing, and concave function. As stated in Section I, we

mentioned that the unified nonconvex surrogate with appropriate parameters is a tighter nonconvex envelope with respect to the tensor nuclear norm, and now we provide the proof. On the set $\{\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3} \mid \|\mathcal{X}\| \leq 1\}$, we can choose the appropriate parameters λ, γ with one parameter fixed and another parameter solved by the equation $\phi(1)_{\lambda, \gamma} = 1$. Then, we hold $x \leq \phi_{\lambda, \gamma}(x) \leq 1$, which leads to

$$\|\mathcal{X}\|_{\oplus} \leq \Phi(\mathcal{X}) \leq \text{rank}_a(\mathcal{X}). \quad (12)$$

In addition, the inequality, i.e. $\text{rank}_a(\mathcal{X}) \leq \text{rank}_t(\mathcal{X})$, holds true for the tensor average rank and tensor tubal rank. Therefore, we claim that $\Phi(\mathcal{X})$ approximates tensor tubal rank better than $\|\mathcal{X}\|_{\oplus}$ with

$$\|\mathcal{X}\|_{\oplus} \leq \Phi(\mathcal{X}) \leq \text{rank}_a(\mathcal{X}) \leq \text{rank}_t(\mathcal{X}). \quad (13)$$

B. Unified Anomaly Detection Nonconvex Framework

1) *The Proposed GNLTR Model:* Let HSI data be presented by a 3-dimensional tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, where n_3 indicates the amount of spectral bands and n_1, n_2 show the spatial sizes of each frontal slice. We decompose $\mathcal{X}$ into backgrounds $\mathcal{L} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and anomalies $\mathcal{E} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The anomaly detection problem is replaced with

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{E}} \quad & \text{rank}_t(\mathcal{L}) + \alpha \|\mathcal{E}\|_{2,1} \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L} + \mathcal{E}, \end{aligned} \quad (14)$$

where α serves as a regularization parameter to balance the weight between backgrounds and anomalies, rank_t imposes restrictions on the low rank of backgrounds, and the $L_{2,1}$ norm is utilized for the spectral and spatial sparsity of anomalies. The $L_{2,1}$ norm is calculated by

$$\|\mathcal{E}\|_{2,1} = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sqrt{\sum_{t=1}^{n_3} \mathcal{E}(i, j, t)^2}. \quad (15)$$

However, calculating rank_t is an NP-hard and nonconvex problem. To make the optimization problem computable, we take the unified nonconvex surrogate as a nonconvex alternative to rank_t , which could be reformulated as

$$\begin{aligned} \min_{\mathcal{L}, \mathcal{E}} \quad & \Phi(\mathcal{L}) + \alpha \|\mathcal{E}\|_{2,1} \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{L} + \mathcal{E}, \end{aligned} \quad (16)$$

where $\Phi(\mathcal{L})$ is a unified nonconvex surrogate with ample nonconvex penalty functions.

Previous studies [23], [25], [40] have demonstrated that a reliable and well-designed dictionary has an important impact in improving the detection performance and robust representations of HSI data. Consequently, a comprehensive dictionary $\mathcal{D} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ extracted from the backgrounds without anomalies and a corresponding representation coefficient tensor $\mathcal{Z} \in \mathbb{R}^{n_4 \times n_2 \times n_3}$ imposed low-rank restrictions are inserted into the model to take the place of the background tensor. The renewed framework is

$$\begin{aligned} \min_{\mathcal{Z}, \mathcal{E}} \quad & \Phi(\mathcal{Z}) + \alpha \|\mathcal{E}\|_{2,1} \\ \text{s.t.} \quad & \mathcal{X} = \mathcal{D} * \mathcal{Z} + \mathcal{E}. \end{aligned} \quad (17)$$

TABLE II
SOME POPULAR NONCONVEX PENALTY FUNCTIONS

Name	Formula	Supergradient	Name	Formula	Supergradient
Capped L_1 [52]	$\begin{cases} \lambda x, & x < \gamma, \\ \lambda \gamma, & x \geq \gamma. \end{cases}$	$\begin{cases} \lambda, & x < \gamma, \\ [0, \lambda], & x = \lambda, \\ 0, & x \geq \lambda. \end{cases}$	ETP [49]	$\frac{\lambda}{1-e^{-\gamma}}(1 - e^{-\gamma x})$	$\frac{\lambda \gamma}{1-e^{-\gamma}} e^{-\gamma x}$
Geman [53]	$\frac{\lambda x}{\gamma + x}$	$\frac{\lambda \gamma}{(x + \gamma)^2}$	Laplace [54]	$\lambda(1 - e^{-\frac{x}{\lambda}})$	$\frac{\lambda}{\gamma} e^{-\frac{x}{\gamma}}$
Logarithm [55]	$\frac{\lambda}{\log(\gamma+1)} \log(\gamma x + 1)$	$\frac{\lambda \gamma}{(\gamma x + 1) \log(\gamma+1)}$	L_p [50]	$\lambda x^p, 0 < p < 1$	$\begin{cases} \infty, & x = 0, \\ \lambda p x^{p-1} & x > 0. \end{cases}$
MCP [56]	$\begin{cases} \lambda x - \frac{x^2}{2\gamma}, & x \leq \gamma \lambda \\ \frac{1}{2} \gamma \lambda^2, & x > \gamma \lambda \end{cases}$	$\begin{cases} \lambda - \frac{x}{\gamma}, & x < \gamma \lambda, \\ 0, & x \geq \gamma \lambda. \end{cases}$	SCAD [51]	$\begin{cases} \lambda x, & x < \lambda, \\ \frac{-x^2 + 2\lambda \gamma x - \lambda^2}{2(\gamma - 1)}, & \lambda \leq x < \gamma \lambda \\ \frac{\lambda^2(\gamma + 1)}{2}, & x \geq \gamma \lambda \end{cases}$	$\begin{cases} \lambda, & x \leq \lambda, \\ \frac{\gamma \lambda - x}{\gamma - 1}, & \lambda < x \leq \gamma \lambda, \\ 0, & x > \gamma \lambda \end{cases}$

2) *Optimization Solver*: To figure out the optimization problem, ADMM offers a clear, straightforward, and efficient method. Firstly, the tensor $\mathcal{Z}$ is expected to be decoupled from the restriction $\mathcal{X} = \mathcal{D} * \mathcal{Z} + \mathcal{E}$ to reduce the complexity of the following optimization formulas. We introduced an auxiliary variable $\mathcal{M}$ to replace $\mathcal{Z}$ and Eq. (17) is rewritten as

$$\begin{aligned} \min_{\mathcal{Z}, \mathcal{M}, \mathcal{E}} \quad & \Phi(\mathcal{Z}) + \alpha \|\mathcal{E}\|_{2,1} \\ \text{s.t.} \quad & \mathcal{Z} = \mathcal{M}, \quad \mathcal{X} = \mathcal{D} * \mathcal{M} + \mathcal{E}. \end{aligned} \quad (18)$$

Then, we transform Eq. (18) into the augmented Lagrangian function, i.e.

$$\begin{aligned} \min_{\mathcal{Z}, \mathcal{M}, \mathcal{E}} \quad & \Phi(\mathcal{Z}) + \alpha \|\mathcal{E}\|_{2,1} + \langle \beta_1, \mathcal{Z} - \mathcal{M} \rangle + \frac{\rho}{2} \|\mathcal{Z} - \mathcal{M}\|_F^2 \\ & + \langle \beta_2, \mathcal{X} - \mathcal{D} * \mathcal{M} - \mathcal{E} \rangle + \frac{\rho}{2} \|\mathcal{X} - \mathcal{D} * \mathcal{M} - \mathcal{E}\|_F^2, \end{aligned} \quad (19)$$

where ρ works as a penalty parameter; β_1 and β_2 are lagrangian multipliers. Eq. (19) is an unconstrained problem with six unknown variables. By fixing other unknowns, we update one unknown each time. The following are the optimization equations for each unknown variable.

- **Update $\mathcal{Z}$** : After extracting the items related to $\mathcal{Z}$, the formula is

$$\mathcal{Z}^{k+1} = \arg \min_{\mathcal{Z}} \Phi(\mathcal{Z}) + \frac{\rho^k}{2} \|\mathcal{Z} - \mathcal{M}^k\|_F^2 + \frac{\beta_1^k}{\rho^k} \|\mathcal{Z}\|_F^2 \quad (20)$$

where $\mathcal{Z}^{k+1}$ is the result of the $(k+1)_{th}$ iteration of $\mathcal{Z}$ and the same for others. Eq. (20) can be transformed to

$$\mathcal{Z}^{k+1} = \arg \min_{\mathcal{Z}} \sum_{j=1}^{n_3} \left(\sum_{i=1}^m \phi(\bar{\mathcal{S}}^k(i, i, j)) \right) + \frac{\rho^k}{2} \|\mathcal{Z}^{(j)} - \mathcal{H}^{(j)}\|_F^2 \quad (21)$$

where $\mathcal{H} = \mathcal{M}^k - \frac{\beta_1^k}{\rho^k}$, $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $m = \min(n_1, n_2)$, $\mathcal{Z} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, and $\mathcal{Z}^{(j)}$ stands for the j_{th} frontal slice of $\mathcal{Z}$. Moreover, we can derive the inequality, i.e.

$$\begin{aligned} \phi(\bar{\mathcal{S}}(i, i, j)) &\leq \phi(\bar{\mathcal{S}}^k(i, i, j)) \\ &\quad - \partial \phi(\bar{\mathcal{S}}^k(i, i, j))(\bar{\mathcal{S}}(i, i, j) - \bar{\mathcal{S}}^k(i, i, j)), \end{aligned} \quad (22)$$

through Eq. (11). With Eq. (22), the optimization problem Eq. (21) can be relaxed to

$$\begin{aligned} \mathcal{Z}^{k+1} = \arg \min_{\mathcal{Z}} \sum_{j=1}^{n_3} \left(\sum_{i=1}^m \partial \phi(\bar{\mathcal{S}}^k(i, i, j)) \bar{\mathcal{S}}(i, i, j) \right) \\ + \frac{\rho^k}{2} \|\mathcal{Z}^{(j)} - \mathcal{H}^{(j)}\|_F^2. \end{aligned} \quad (23)$$

It is obvious that Eq.(23) is composed of separated n_3 problems and the optimal solution could be obtained by solving the sub-problems. The j_{th} sub-problem is

$$\begin{aligned} \mathcal{Z}^{k+1(j)} = \arg \min_{\mathcal{Z}^{(j)}} \left(\sum_{i=1}^m \partial \phi(\bar{\mathcal{S}}^k(i, i, j)) \bar{\mathcal{S}}(i, i, j) \right) \\ + \frac{\rho^k}{2} \|\mathcal{Z}^{(j)} - \mathcal{H}^{(j)}\|_F^2. \end{aligned} \quad (24)$$

The optimal solution of the above equation can be obtained with the generalized weighted singular value thresholding (GWSVT) [57]–[59], i.e.

$$\bar{\mathcal{Z}}^{k+1(j)} = \bar{\mathbf{U}} * \bar{\mathbf{S}}_* * \bar{\mathbf{V}}^T, \quad (25)$$

where $[\bar{\mathbf{U}}, \bar{\mathbf{S}}, \bar{\mathbf{V}}] = \text{SVD}(\bar{\mathcal{H}}^{(j)})$, $\mathbf{s}_d = \text{diag}(\bar{\mathbf{S}})$, $\mathbf{s} = \text{diag}(\bar{\mathcal{S}}^{k(j)})$, $\bar{\mathbf{S}}_* = \text{diag}(\max(\mathbf{s}_d - \frac{\partial \phi(\mathbf{s})}{\rho^k}, 0))$, and $\mathcal{Z}^{k+1} = \text{ifft}(\bar{\mathcal{Z}}^{k+1}, \cdot, 3)$. The whole process of updating $\mathcal{Z}$ is shown in Algorithm 1.

Algorithm 1 Unified nonconvex surrogate tensor singular value thresholding

Input: Tensor $\mathcal{Z}^k$, Auxiliary variable $\mathcal{M}^k$, penalty parameter ρ^k , lagrangian multiplier β_1^k

Initialize: $[\mathcal{U}, \mathcal{S}^k, \mathcal{V}] = \text{t-SVD}(\mathcal{Z}^k)$, $\bar{\mathcal{S}}^k = \text{fft}(\mathcal{S}^k, \cdot, 3)$ $\mathcal{W} = \mathcal{M}^k - \frac{\beta_1^k}{\rho^k}$, $\bar{\mathcal{W}} = \text{fft}(\mathcal{W}, \cdot, 3)$;

- 1: **for** $i = 1$ to $\lceil \frac{(n_3+1)}{2} \rceil$ **do**
- 2: $[\bar{\mathbf{U}}, \bar{\mathbf{S}}, \bar{\mathbf{V}}] = \text{SVD}(\bar{\mathcal{W}}^{(i)})$;
- 3: $\mathbf{s}_d = \text{diag}(\bar{\mathbf{S}})$;
- 4: $\mathbf{s} = \text{diag}(\bar{\mathcal{S}}^{k(i)})$;
- 5: $\bar{\mathbf{S}}_* = \text{diag}(\max(\mathbf{s}_d - \frac{\partial \phi(\mathbf{s})}{\rho^k}, 0))$;
- 6: $\bar{\mathcal{Z}}^{(i)} = \bar{\mathbf{U}} * \bar{\mathbf{S}}_* * \bar{\mathbf{V}}^T$
- 7: **end for**
- 8: **for** $i = \lceil \frac{(n_3+1)}{2} \rceil + 1$ to n_3 **do**
- 9: $\bar{\mathcal{Z}}^{(i)} = \text{conj}(\bar{\mathcal{Z}}^{(n_3-i+2)})$;
- 10: **end for**
- 11: $\mathcal{Z} = \text{ifft}(\bar{\mathcal{Z}}, \cdot, 3)$,

Output: $\mathcal{Z}$

- **Update $\mathcal{E}$** : It is updated by

$$\mathcal{E}^{k+1} = \arg \min_{\mathcal{E}} \alpha \|\mathcal{E}\|_{2,1} + \frac{\rho^k}{2} \|\mathcal{E} - \mathcal{P}\|_F^2, \quad (26)$$

where $\mathcal{P} = \mathcal{X} - \mathcal{D} * \mathcal{M}^k + \frac{\beta_k^k}{\rho^k}$. For the sub-problem, its closed-form solution is

$$\mathcal{E}(:, :, t) = \begin{cases} \frac{\|\mathcal{P}(:, :, t)\|_2 - \frac{\alpha}{\rho^k}}{\|\mathcal{P}(:, :, t)\|_2} \mathcal{P}(:, :, t), & \frac{\alpha}{\rho^k} < \|\mathcal{P}(:, :, t)\|_2 \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

- **Update $\mathcal{M}$:** The function with respect to $\mathcal{M}$ is

$$\mathcal{M}^{k+1} = \arg \min_{\mathcal{M}} \frac{\rho^k}{2} \|\mathcal{Z}^{k+1} - \mathcal{M} + \frac{\beta_1^k}{\rho^k}\|_F^2 + \frac{\rho^k}{2} \|\mathcal{X} - \mathcal{D} * \mathcal{M} - \mathcal{E}^{k+1} + \frac{\beta_2^k}{\rho^k}\|_F^2. \quad (28)$$

The optimization problem comprises two squares of tensor F-norm. We set the differential coefficient of Eq. (28) to be $\mathbf{0}$, i.e.,

$$(\mathcal{D}^T * \mathcal{D} + \mathcal{I})\mathcal{M} - (\mathcal{Z}^{k+1} + \frac{\beta_1^k}{\rho^k} + \mathcal{D}^T * (\mathcal{X} - \mathcal{E}^{k+1} + \frac{\beta_2^k}{\rho^k})) = \mathbf{0}. \quad (29)$$

The above equation is linear, which indicates that $\mathcal{M}$ achieves a minimum at the point where the differential coefficient is $\mathbf{0}$, i.e.

$$\mathcal{M}^{k+1} = (\mathcal{D}^T * \mathcal{D} + \mathcal{I})^{-1} * (\mathcal{Z}^{k+1} + \frac{\beta_1^k}{\rho^k} + \mathcal{D}^T * (\mathcal{X} - \mathcal{E}^{k+1} + \frac{\beta_2^k}{\rho^k})). \quad (30)$$

- **Update β_1, β_2, ρ :** These variables are updated by

$$\beta_1^{k+1} = \beta_1^k + \rho^k (\mathcal{Z}^{k+1} - \mathcal{M}^{k+1}), \quad (31)$$

$$\beta_2^{k+1} = \beta_2^k + \rho^k (\mathcal{X} - \mathcal{D} * \mathcal{M}^{k+1} - \mathcal{E}^{k+1}), \quad (32)$$

$$\rho^{k+1} = \max(r\rho^k, \rho_{max}), \quad (33)$$

where the growth rate r is usually set to 1.1. It has been demonstrated that the update strategy of ρ can accelerate the ADMM convergence.

In [60], theoretical proof of ADMM convergence with more than two block variables is tough. Therefore, we could consider ADMM as converged when the changes in all variables are less than a small value $\varepsilon = e^{-6}$. The convergence condition is

$$\max \left(\begin{aligned} &\|\mathcal{Z}^{k+1} - \mathcal{Z}^k\|_\infty, \|\mathcal{E}^{k+1} - \mathcal{E}^k\|_\infty \\ &\|\mathcal{M}^{k+1} - \mathcal{M}^k\|_\infty, \|\mathcal{M}^{k+1} - \mathcal{Z}^{k+1}\|_\infty \\ &\|\mathcal{X} - \mathcal{D} * \mathcal{M}^{k+1} - \mathcal{E}^k\|_\infty \end{aligned} \right) \leq \varepsilon \quad (34)$$

Finally, we can get the anomaly detection map $\mathbf{T}$ by

$$\mathbf{T}(i, j) = \sqrt{\sum_{t=1}^{n_3} |\mathcal{E}(i, j, t)|^2}. \quad (35)$$

where $\mathcal{E} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathbf{T} \in \mathbb{R}^{n_1 \times n_2}$. We render all the procedures in Algorithm 2.

C. Dictionary Construction

In the previous section, a comprehensive dictionary $\mathcal{D}$ and a corresponding representation coefficient tensor $\mathcal{Z}$ are utilized for representing the backgrounds. The dictionary should consist of pure background parts so as to obtain a better representation. Here, we provide a nonconvex, feasible and

Algorithm 2 GNLTR for Hyperspectral Anomaly Detection

Input: HSI data $\mathcal{X}$, dictionary $\mathcal{D}$, regularization parameter α

Initialize: $\mathcal{Z}^0, \mathcal{M}^0, \mathcal{E}^0, \beta_1^0, \beta_2^0$ initialized to $\mathbf{0}$; $\mu = 1.1, \rho^0 = e^{-5}, \rho_{max} = e^8, \varepsilon = e^{-6}, max_iter = 100, k = 0$;

- 1: **while** not converged and $k \leq max_iter$ **do**
 - 2: Update $\mathcal{Z}^{k+1}$ via Algorithm 1;
 - 3: Update $\mathcal{E}^{k+1}$ via Eq. (27);
 - 4: Update $\mathcal{M}^{k+1}$ via Eq. (30);
 - 5: Update $\beta_1^{k+1}, \beta_2^{k+1}, \rho^{k+1}$ via Eqs. (31), (32), (33), respectively;
 - 6: Check the converge condition Eq. (34);
 - 7: $k = k + 1$;
 - 8: **end while**
 - 9: Obtain the optimal coefficient tensor $\mathcal{Z}^*$ and anomalies $\mathcal{E}^*$;
 - 10: Compute the backgrounds $\mathcal{L}$ via $\mathcal{L} = \mathcal{D} * \mathcal{Z}^*$;
 - 11: Compute the detection map $\mathbf{T}$ by Eq. (35);
- Output:** backgrounds $\mathcal{L}$, anomalies $\mathcal{E}^*$ and detection map $\mathbf{T}$.

effective dictionary construction technique that can effectively separate backgrounds and anomalies, i.e.,

$$\begin{aligned} &\min_{\mathcal{D}, \mathcal{L}} \Phi(\mathcal{D}) + \theta \|\mathcal{L}\|_{2,1} \\ &s.t. \quad \mathcal{X} = \mathcal{D} + \mathcal{L}. \end{aligned} \quad (36)$$

where $\mathcal{L}$ is the anomalous part. To solve this optimization problem, we also exploit ADMM. The augmented Lagrangian function transformed by Eq. (36) is

$$\begin{aligned} &\min_{\mathcal{D}, \mathcal{L}} \Phi(\mathcal{D}) + \theta \|\mathcal{L}\|_{2,1} + \langle \beta, \mathcal{X} - \mathcal{D} - \mathcal{L} \rangle \\ &\quad + \frac{\rho}{2} \|\mathcal{X} - \mathcal{D} - \mathcal{L}\|_F^2, \end{aligned} \quad (37)$$

where $\Phi(\mathcal{D})$ is the nonconvex alternative to $rank_t(\mathcal{D})$; β is a Lagrangian multiplier; ρ is a penalty parameter. ADMM is a good choice to address the optimisation Eq.(37). The sub-problems are similar to Eq. (19), which means we could use Algorithm 2 to handle the sub-problems with a few modifications. The details are as follows:

- **Update $\mathcal{D}$:** The dictionary is updated by

$$\mathcal{D}^{k+1} = \arg \min_{\mathcal{D}} \Phi(\mathcal{D}) + \frac{\rho^k}{2} \|\mathcal{X} - \mathcal{D} - \mathcal{L}^k + \frac{\beta^k}{\rho^k}\|_F^2. \quad (38)$$

The equation has a similar structure to Eq.(20) which can be solved by Algorithm 1.

- **Update $\mathcal{L}$:** The function related to $\mathcal{L}$ is

$$\mathcal{L}^{k+1} = \arg \min_{\mathcal{L}} \theta \|\mathcal{L}\|_{2,1} + \frac{\rho^k}{2} \|\mathcal{L} - \mathcal{P}\|_F^2 \quad (39)$$

where $\mathcal{P} = \mathcal{X} - \mathcal{D}^{k+1} + \frac{\beta^k}{\rho^k}$. We have noted that Eq.(39) has same composition as Eq.(26), thus we can obtain a closed-form solution by replacing the symbols of Eq.(27).

- **Update β, ρ :** the two variables could be updated by

$$\beta^{k+1} = \beta^k + \rho^k (\mathcal{X} - \mathcal{D}^{k+1} - \mathcal{L}^{k+1}), \quad (40)$$

$$\rho^{k+1} = \max(r\rho^k, \rho_{max}). \quad (41)$$

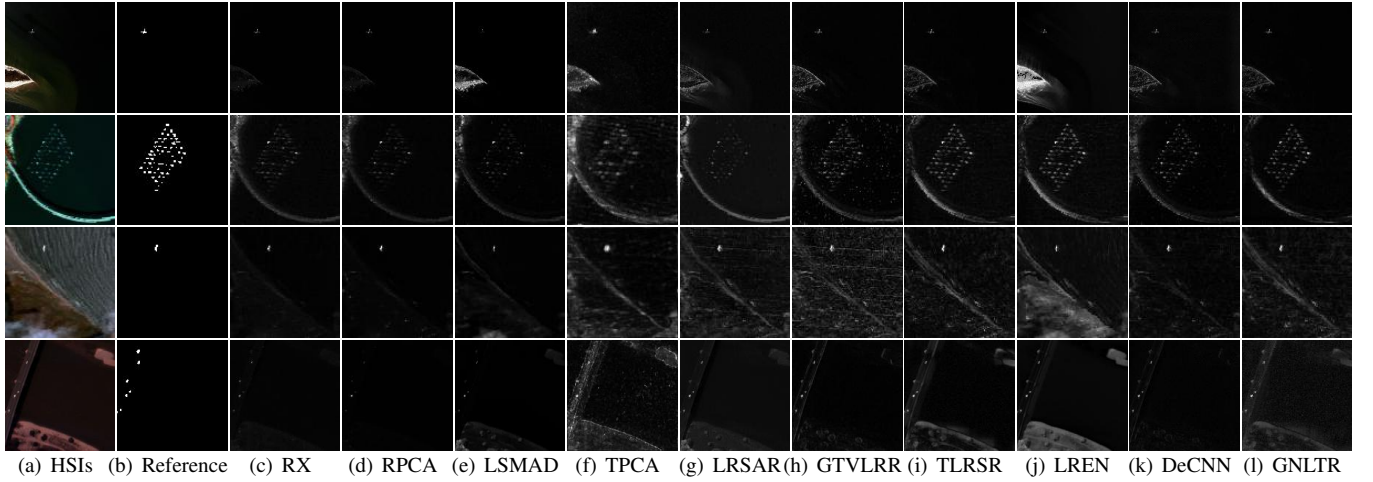


Fig. 2. Pseudo-color images, reference maps, detection maps on the Beach data. The top to bottom images are Beach 1 to 4.

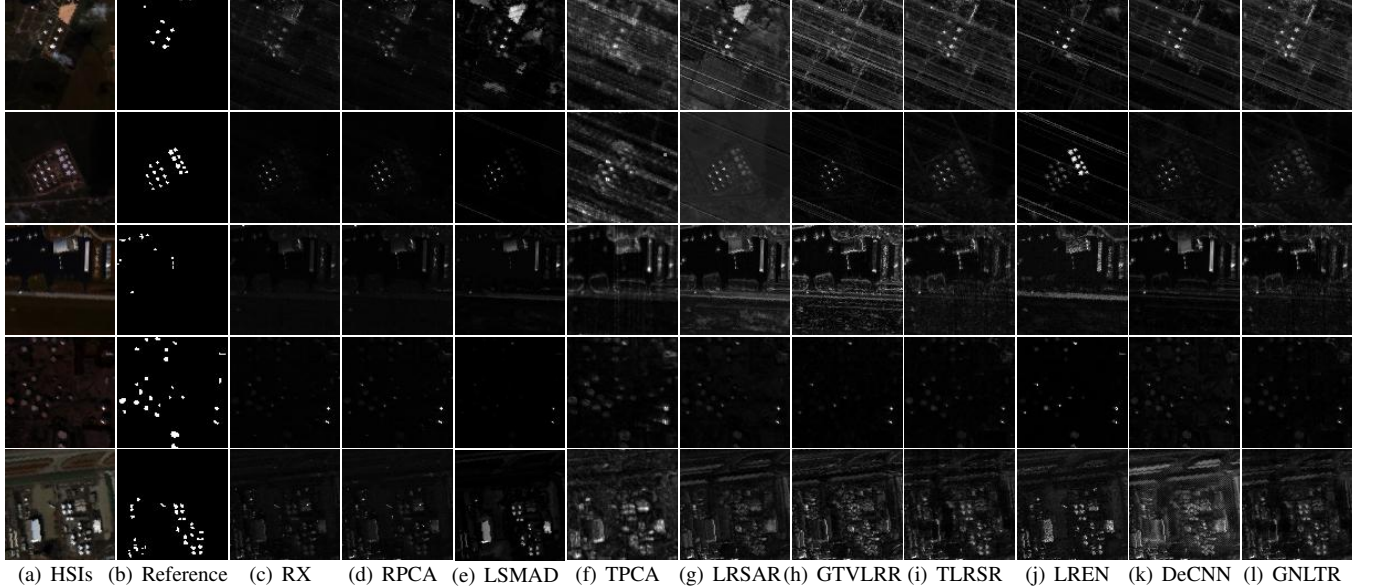


Fig. 3. Pseudo-color images, reference maps, detection maps on the Urban data. The top to bottom images are Urban 1 to 5.

Notably, the dictionary $\mathcal{D}$ is a 3-dimensional tensor that preserves the intrinsic spatial-spectral architecture of the HSI data $\mathcal{X}$. Also, the $L_{2,1}$ norm plays an important role in eliminating the impact of anomalies and the nonconvex surrogate $\Phi(\cdot)$ brings a purer dictionary than the convex surrogate due to the closer approximation to $rank_t$.

IV. EXPERIMENTAL RESULTS

In the section, we introduce several real-world hyperspectral data sets to assess the detection performance and effectiveness of our proposed GNLTR model. All the experimental results were obtained by a computer with an Intel Core i5-10400 2.90-GHz CPU plus 16-GB memory.

A. Hyperspectral Data Sets Description

The testing data sets include Beach, Urban, HYDICE-Urban, and San Diego. Some details about those datasets are

summarized as follows:

- In the Beach data set, ships, vehicles, flights, and man-made objects are labeled as anomalies. Meanwhile, beaches and water are considered as backgrounds. Two images of the data have an image size of 100×100 . The other two images are 150×150 . Fig. 2 (a) displays the pseudo-color images. Fig.2 (b) shows reference detection maps.
- Urban data set contains five 100×100 images derived from AVIRIS. Roofs and other objects that are significantly different from their surroundings will be viewed as anomalies. We present the detailed images in Fig. 3 (a) and (b).
- The HYDICE-Urban with $80 \times 100 \times 175$ pixels is urban scenes originated from the hyper-spectral digital imagery collection experiment (HYDICE). Roofs and vehicles are

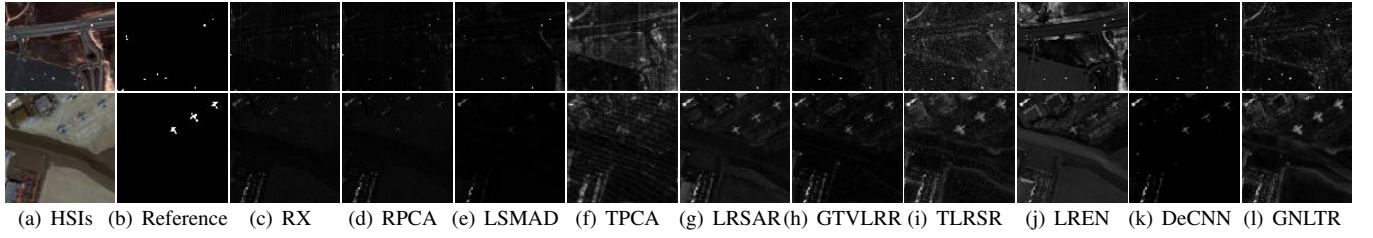


Fig. 4. Pseudo-color images, reference maps, detection maps on HYDICE-urban (top) and San Diego (bottom).

TABLE III

THE PERFORMANCES OF DIFFERENT NONCONVEX PENALTY FUNCTIONS FOR DICTIONARY CONSTRUCTION AND L_p FOR RESULT SOLVING ON AUC ON SEVERAL DATA SETS

Functions	Capped L_1	ETP	Geman	Laplace	Logarithm	L_p	MCP	SCAD
Beach-2	0.9753	0.9622	0.9640	0.9622	0.9609	0.9648	0.9638	0.9584
Urban-3	0.9918	0.9919	0.9872	0.9879	0.9907	0.9914	0.992	0.9922
HYDICE-Urban	0.9951	0.9956	0.9949	0.9962	0.9948	0.9963	0.9961	0.9948
San Diego	0.9954	0.9956	0.9940	0.9949	0.9955	0.9957	0.9957	0.9954

TABLE IV

THE PERFORMANCES OF DIFFERENT NONCONVEX PENALTY FUNCTIONS FOR RESULT SOLVING AND L_p FOR DICTIONARY CONSTRUCTION ON AUC IN SEVERAL DATA SETS

Functions	Capped L_1	ETP	Geman	Laplace	Logarithm	L_p	MCP	SCAD
Beach-2	0.9626	0.9629	0.9695	0.9695	0.9658	0.9648	0.9631	0.9626
Urban-3	0.9907	0.9912	0.9911	0.9911	0.9913	0.9914	0.9913	0.9913
HYDICE-Urban	0.9960	0.9961	0.9959	0.9959	0.9963	0.9963	0.9958	0.9960
San Diego	0.9957	0.9957	0.9958	0.9958	0.9958	0.9957	0.9957	0.9957

considered as anomalies, which contain 21 pixels in the different parts of the detection map. Fig. 4 (a) and (b) exhibit the sample images and related reference maps.

- San Diego is with $100 \times 100 \times 189$ pixels. Backgrounds are mostly made of normal things like grass, and roofs. As to the anomalies, three airplanes are naturally marked as the main components of the exceptions with details given in Fig. 4 (a) and (b).

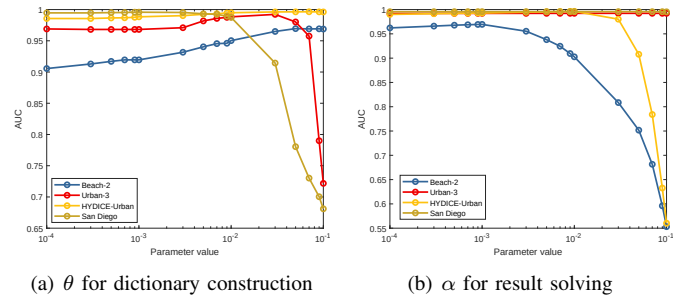


Fig. 5. Hyperparameters tuning on four data sets.

B. Parameter Tuning

For the proposed GNLTR model, four variables are supposed to be optimized to gain satisfactory detection results. The first two hyper-parameters θ , α involved in Eqs. (17) and (36) are selected in $\{0.0005, 0.001, 0.01, 0.03, 0.05, 0.07, 0.1\}$. The other two are in connection with the nonconvex surrogate selection, which are chosen in Table II. One of these parameters is adjusted while others are fixed. Fig. 5 demonstrates the robustness of the proposed model for θ and α , which means that GNLTR is insensitive to the parameters over a relatively large range on most data sets. Table III gives the performance of different nonconvex penalty functions used to build dictionary and L_p for result solving on AUCs with appropriate parameters on some data sets. Meanwhile, Table IV provides the performances of different nonconvex penalty functions for result solving and L_p for dictionary construction. Notably, the AUCs of different nonconvex penalty functions have little variation, which implies that our unified nonconvex framework is well structured with excellent performance.

C. Detection Performance

To examine the results, some baselines are compared with the proposed GNLTR, which include classic RX, four LRaSMD models (i.e., RPCA [20], LSMAD [23], LRSAR [25], GTVLRR [28]), two tensor representation models (i.e., TPCA [38], PCA-TLRSR [40]), and two deep learning-based HAD methods (i.e., LREN [41], DeCNN-AD [42]). In addition, we optimize the parameters in LSMAD, LARSR, and GTVLRR to achieve first-class detection accuracy. For LSMAD, the parameter $rank$ varies from 1 to 30. In terms of LARSR, two hyper-parameters λ and β are respectively selected from $\{0.001, 0.01, 0.05, 0.1, 0.5, 1, 2\}$. As to GTVLRR, there are three regularization parameters with the set point, i.e., $\beta = 0.2$, $\gamma = 0.05$ and λ varies from 0.05 to 1. For PCA-TLRSR, we get optimal values for two parameters according to the suggestion in [40]. Considering the parameters used in [41], we set the hidden nodes = 9, $k = 7$, and λ from

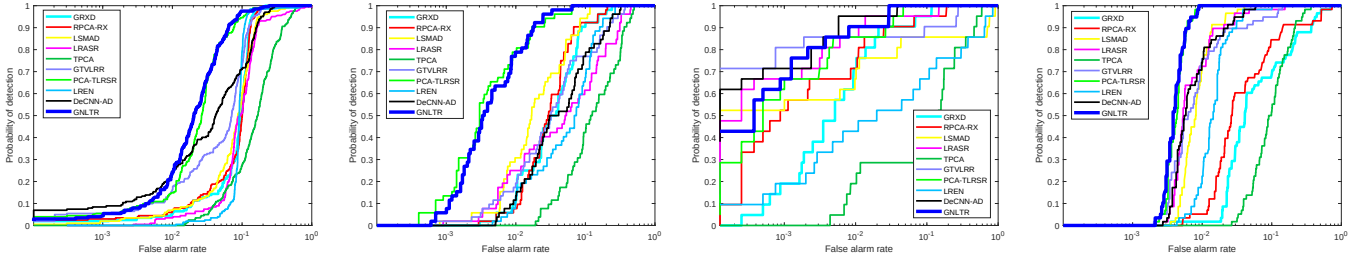


Fig. 6. ROC curves of (P_d, P_f) on different data sets (left to right) Beach-2, Urban-3, HYDICE-Urban, San Diego.

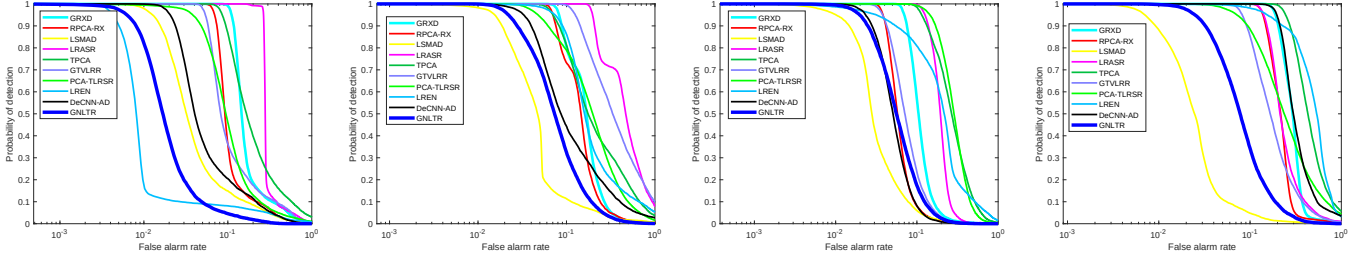


Fig. 7. ROC curves of (P_f, τ) on different data sets (left to right) Beach-2, Urban-3, HYDICE-Urban, San Diego.

10^{-4} to 1 for LREN. With regard to DeCNN-AD, K varies from 1 to 20, the values of β and λ change from 10^{-4} to 1. Moreover, the number of clusters is set to 15 and each cluster as atoms for the dictionary owns 20 pixels in LRASR and GTVLRR. As for our model, the parameters are programmed to the suitable values, which has been discussed in parameter tuning. In addition, we adopt principal component analysis to pre-process the original data to improve the effectiveness and reduce the computation time.

To qualitatively and quantitatively assess the detection performance, we provide visual observation and three evaluation indicators of the results. Before introducing the specific metrics, we briefly explain the meaning of P_d , P_f , τ , which stand for probability of detection, false alarm rate and threshold value, respectively. The first indicator is the receiver operating characteristic curve (ROC) in Fig.6 and Fig.7. The ROC of (P_d, P_f) depicts the varying correlation between P_d and P_f . A better model is considered to own a higher P_d for a given P_f . The relations between P_f and τ can be represented by the ROC of (P_f, τ) . A lower P_f indicates a lower likelihood that a method misclassifies a pixel for a given τ . The second one is the area under the ROC (AUC) in Table V and VI. For AUC of (P_d, P_f) , the closer the value of the method is to 1, the more accurate the method's detection is. In contrast, the lower AUC of (P_f, τ) means the method has a lower probability of false detection in general. The third metric is the separability boxplot which counts the detection results at the pixel locations of the anomaly and background categories with a statistical range from 10% to 90%. A more separable method owns a larger gap between the blue background box and the red anomaly box. With these evaluation indicators, the detection performance is as follows:

1) **Beach:** Four scenes of Beach are displayed in Fig. 2 with the correlative AUC scores in Table V and Table VI. Here, we employ Beach-2 to demonstrate the outstanding

result of GNLTR. The first three methods and LRASR rarely detect the anomalies. TPCA is poor in remaining the shape of anomalies. GTVLRR, PCA-TLRSR, LREN, DeCNN-AD and GNLTR successfully remain anomalies' forms. But GTVLRR and DeCNN-AD don't fully detect the anomalies. At the same time, GNLTR views the bay more as a background than an anomaly, which is better aligned with the detection map. In term of the Beach-2's ROC of (P_d, P_f) which is shown in Fig. 6(a), we find that DeCNN-AD get the first place at the beginning and GNLTR is in the lead with PCA-TLRSR in the second place when $P_f > 0.01$. In ROC of (P_f, τ) of Beach-2 presented in Fig.7(a), GNLTR gains the lowest value after $P_f > 0.05$. For AUCs of (P_d, P_f) , GNLTR is the best on Beach-2 and the second on Beach-1 and Beach-4. Moreover, for AUCs of (P_f, τ) , GNLTR obtains the first position on Beach-2. As a result, GNLTR has a noteworthy improvement in the average scores on (P_d, P_f) and a quality result on the scores of (P_f, τ) . In the separability map of Beach-2 (Fig.8(a)), GNLTR has the largest gap with narrow widths. Additionally, the median of backgrounds is close to 0, which indicates our method has a strong ability to suppress backgrounds.

2) **Urban:** Five urban scenarios make up this data, which are observed in Fig. 3. In these scenes, Urban-3 is a pretty good option as a representative of Urban for its typical backgrounds and anomalies. RX, RPCA, LSMAD, and TPCA do not own the capability to fully detect anomalies. The maps of LRASR, GTVLRR, LREN and DeCNN-AD make mistakes in treating a great deal of the background as anomalies. PCA-TLRSR and GNLTR successfully suppress the backgrounds better while detecting anomalies. For the relevant ROCs of (P_d, P_f) shown in Fig.6(b), PCA-TLRSR gets the first place when $P_f < 0.007$. GNLTR overtakes PCA-TLRSR and keeps the first when $P_f > 0.007$. GNLTR is only behind LSMAD in ROCs of (P_f, τ) in Fig.7(b). Besides, the related AUCs of (P_d, P_f) are available in Table V. GNLTR occupies the first position on Urban-3

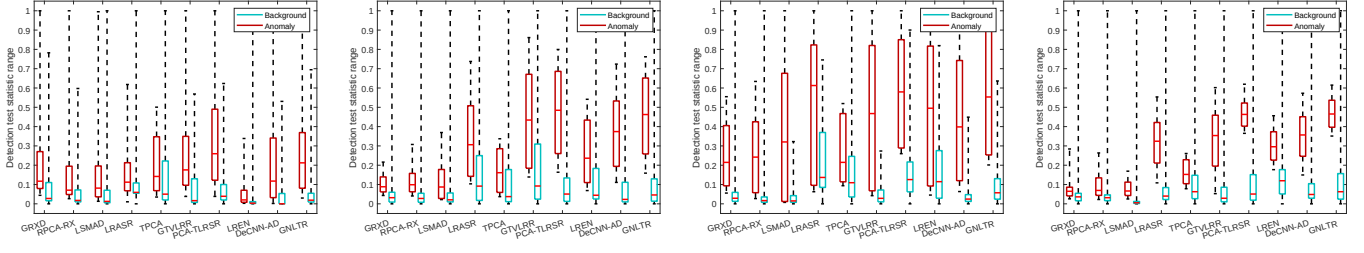


Fig. 8. Separability maps for different data sets (left to right) Beach-2, Urban-3, HYDICE-Urban, San Diego.

TABLE V
AUC VALUES OF (P_d, P_f) ON ALL TESTING DATA SETS

HSI data	RX	RPCA	LSMAD	TPCA	LRASR	GTVLRR	PCA-TLRSR	LREN	DeCNN-AD	GNLTR
Beach-1	0.9804	0.9761	0.8808	0.9890	0.9260	0.9773	0.9875	<u>0.9115</u>	<u>0.9737</u>	<u>0.9880</u>
Beach-2	0.9106	0.9097	0.9033	0.8061	0.8812	0.9306	<u>0.9638</u>	<u>0.9039</u>	<u>0.9411</u>	0.9695
Beach-3	0.9999	<u>0.9995</u>	0.9296	0.9982	0.9987	0.9954	<u>0.9993</u>	<u>0.9011</u>	<u>0.9995</u>	0.9993
Beach-4	0.9539	0.9599	0.8959	0.9583	0.9552	0.9815	0.9643	<u>0.8919</u>	<u>0.9739</u>	0.9795
Average	0.9612	0.9613	0.9024	0.9379	0.9403	0.9712	<u>0.9787</u>	<u>0.9021</u>	<u>0.9721</u>	0.9841
Urban-1	0.9907	0.9922	0.9612	0.9390	0.9685	0.8465	0.9949	<u>0.9786</u>	<u>0.9914</u>	<u>0.9947</u>
Urban-2	0.9946	0.9957	0.9783	0.9409	0.9974	0.8995	0.9983	<u>0.9981</u>	<u>0.9638</u>	0.9979
Urban-3	0.9513	0.9577	0.9697	0.8224	0.9336	0.9328	<u>0.9919</u>	<u>0.9209</u>	<u>0.9799</u>	0.9923
Urban-4	0.9887	0.9871	0.9902	0.9617	0.9932	0.9373	<u>0.9884</u>	<u>0.9905</u>	<u>0.9916</u>	0.9887
Urban-5	0.9692	0.9658	0.9115	0.9370	0.9115	0.9391	0.9837	<u>0.9554</u>	<u>0.9571</u>	<u>0.9830</u>
Average	0.9789	0.9797	0.9622	0.9202	0.9608	0.9110	0.9914	<u>0.9687</u>	<u>0.9678</u>	<u>0.9913</u>
HYDICE-Urban	0.9857	0.9842	0.8818	0.8242	0.9926	0.9746	0.9941	<u>0.9539</u>	0.9976	<u>0.9963</u>
San Diego	0.8886	0.9165	0.9893	0.8849	0.9920	0.9869	<u>0.9957</u>	<u>0.9892</u>	<u>0.9913</u>	0.9958

TABLE VI
AUC VALUES OF (P_f, τ) ON ALL TESTING DATA SETS

HSI data	RX	RPCA	LSMAD	TPCA	LRASR	GTVLRR	PCA-TLRSR	LREN	DeCNN-AD	GNLTR
Beach-1	0.0113	0.0096	0.0248	0.0491	0.0819	0.0331	0.0175	<u>0.0159</u>	<u>0.0323</u>	0.0225
Beach-2	0.0803	0.0569	0.0312	0.1079	0.1367	0.0577	0.0569	<u>0.0192</u>	<u>0.0353</u>	0.0119
Beach-3	0.0473	<u>0.0292</u>	0.0254	0.0857	0.1273	0.1187	0.0701	<u>0.1155</u>	<u>0.0445</u>	0.0906
Beach-4	0.0296	0.0065	<u>0.0137</u>	0.0968	0.0714	0.0283	0.0544	<u>0.0706</u>	<u>0.0246</u>	0.0440
Average	0.0422	<u>0.0244</u>	0.0238	0.0849	0.1043	0.0594	0.0497	<u>0.0553</u>	<u>0.0342</u>	0.0423
Urban-1	0.0932	0.0549	0.0549	0.1733	0.2543	0.1838	0.1334	<u>0.0588</u>	<u>0.0782</u>	0.1183
Urban-2	0.0193	0.0188	0.0054	0.1190	0.1357	0.0280	0.0434	<u>0.0258</u>	<u>0.0477</u>	0.0156
Urban-3	0.0501	0.0454	<u>0.0312</u>	0.0838	0.1590	0.1546	0.0715	<u>0.0771</u>	<u>0.0521</u>	0.0274
Urban-4	0.0185	0.0181	0.0005	0.0399	0.0319	0.0114	0.0168	<u>0.0045</u>	<u>0.0314</u>	<u>0.0032</u>
Urban-5	0.0734	0.0610	0.0328	0.1678	0.1242	0.1239	0.0882	<u>0.0883</u>	<u>0.1948</u>	<u>0.0382</u>
Average	0.0509	<u>0.0396</u>	0.0250	0.1168	0.1410	0.1003	0.0707	<u>0.0509</u>	<u>0.0808</u>	0.0405
HYDICE-Urban	0.0615	0.0349	0.0220	0.1610	0.1052	0.0442	0.1617	<u>0.1347</u>	<u>0.0309</u>	0.0364
San Diego	0.0670	0.0510	0.0087	0.1128	0.0629	0.0543	0.0748	<u>0.1206</u>	<u>0.0828</u>	<u>0.0263</u>

followed by PCA-TLRSR and the second place on Urban-1. GNLTR obtains the suboptimal effect in the average scores, whose value is only 0.01 less than PCA-TLRSR. For AUC values of (P_f, τ) in Table VI, GNLTR is the first on Urban-3 and second on Urban-2, Urban-3. The average value is also excellent. In Fig.8(b), the map shows PCA-TLRSR has the largest gap, tightly followed by GNLTR. What's more, the two boxes of the GNLTR have a smaller width relative to the PCA-TLRSR.

3) **HYDICE-Urban**: Fig. 4 gives us a view of the detection results. Unfortunately, RX, RPCA, and TPCA ignore the anomalies. The remaining methods show the ability to detect anomalies. But DeCNN-AD and GNLTR perform better among all compared models because of the clearest anomalies' forms and a relatively clean background. In Fig.6(c), Fig.7(c), Table V and Table VI, we provide the ROC curves and AUC values. Although, for ROCs of (P_d, P_f) , GNLTR is at the middle place when $P_f < 0.012$, it gains the highest P_d in the

following part. In regard to ROCs of (P_f, τ) , GNLTR gets the second place when $P_f < 0.05$ and the fourth in the remain part. For AUCs of (P_d, P_f) , GNLTR leaves all comparison methods far behind with the second highest value of 0.9963 except for DeCNN-AD. The GNLTR's score of (P_f, τ) obtains the fourth. In Fig.8(c), our model has the biggest gap and there is no overlap between the two boxes.

4) **San Diego**: Three high-value targets are distributed at the upper right part in the image and can be visually observed in Fig. 4. It is obvious that RX, RPCA, LSMAD, and TPCA fail to discover the anomalies. At the same time, the maps of LRASR, GTVLRR, LREN and DeCNN-AD lose the clear form of planes. On the contrary, PCA-TLRSR and GNLTR acquire a satisfactory structure. Considering the ROC curves in Fig. 6(d), PCA-TLRSR and GNLTR have a larger area than other compared methods. The P_d values of GNLTR are tightly followed by PCA-TLRSR. The P_d of GNLTR achieves 1 at FAR= 0.007. In Fig.7(d), GNLTR is the second one overall.

TABLE VII
TIME CONSUMPTION (S) OF GNLTR AND BASELINES ON SOME DATA SETS

HSIs	RX	RPCA	LSMAD	TPCA	LRASR	GTVLRR	PCA-TLRSR	LREN	DeCNN-AD	GNLTR
Beach-2	0.068	2.608	8.581	16.280	14.612	108.887	7.381	190.996	62.021	4.003
Urban-3	0.075	2.786	8.275	17.245	31.040	122.447	6.483	228.201	28.130	6.321
HYDICE-Urban	0.047	1.948	5.801	10.930	22.055	102.337	6.038	148.185	150.800	2.264
San Diego	0.062	2.532	8.414	15.741	27.980	116.145	2.253	183.336	30.590	2.087

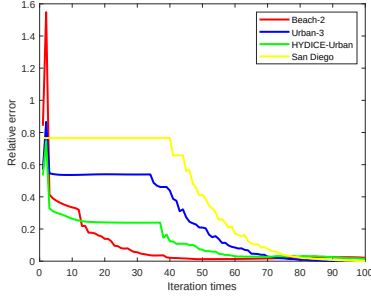


Fig. 9. Relative error for different iteration times on four data sets

Moreover, the AUCs of (P_d, P_f) in Table V indicate GNLTR has the highest values, tightly followed by PCA-TLRSR. The AUCs of (P_f, τ) in Table VI show the GNLTR is the second lowest value. For the separability map in Fig.8(d), the proposed method owns the biggest interval and the highest anomaly median.

D. Discussions

1) *Computational Cost*: Table VII shows the running time of GNLTR and all compared methods on some data sets. Due to the superior architecture, GNLTR's time consumption is excellently controlled, which has a lower cost than LSMAD, TPCA, LRASR, GTVLRR, PCA-TLRSR, LREN and DeCNN-AD. As to the computational complexity of Algorithm 2, there are five parts. Firstly, updating $\mathcal{Z}$ via Algorithm 1 takes about $O(n_1^3 n_3)$. For $\mathcal{M}$ and $\mathcal{E}$, they cost $O(n_1 n_2^3 n_3 + n_2(n_1 + n_3)n_3 \log(n_3))$, respectively. With regard to β_1, β_2 and ρ , they could be finished in $O(n_1 n_2^2 n_3)$ time. In summary, the total complexity is $O((n_1 + 1)n_2^3 n_3 + n_2(n_1 + n_2)n_3 \log(n_3) + n_1 n_2^2 n_3)$.

2) *Convergence Analysis*: In Fig.9, we provide the numerical convergence results on four data sets in which the relative errors are the max part of Eq.(34). For Beach-2, Urban-3 and HsYDICE-Urban, their relative errors hit the maximum value at times = 1 and then monotonically decline. Although the error of San Diego keeps stable at the beginning, it decreases after times = 40. It is obvious that all the errors finally approaching 0 when times > 70.

V. CONCLUSION

In this paper, we have proposed a novel unified anomaly detection nonconvex framework for hyperspectral anomaly detection. Tensor has been integrated into the framework to preserve the intrinsic 3-dimensional property of HSI. A unified nonconvex surrogate in the model has a better approximation than the tensor nuclear norm in terms of the tensor tubal rank,

which means it denotes the low rankness of the backgrounds more effectively. The $L_{2,1}$ norm regularization is an excellent choice to limit the sparsity of the anomaly component. ADMM is employed as the optimization solver. In addition, we extract a nonconvex dictionary construction method for anomaly detection. Experiments have proven the outstanding performance of our model in contrast with other existing anomaly detection methods. For future work, the liner TV-norm regularization can be adopted to denote the piecewise-smooth characteristic of backgrounds.

REFERENCES

- [1] P. W. Yuen and M. Richardson, "An introduction to hyperspectral imaging and its application for security, surveillance and target acquisition," *Imaging Sci. J.*, vol. 58, no. 5, pp. 241–253, 2010.
- [2] A. Koz, "Ground-based hyperspectral image surveillance systems for explosive detection: Part I—state of the art and challenges," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 12, pp. 4746–4753, 2019.
- [3] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2706–2717, 1999.
- [4] Q. Zhang, H. Wang, R. J. Plemmons, and V. P. Pauca, "Tensor methods for hyperspectral data analysis: a space object material identification study," *J. Opt. Soc. Amer. A*, vol. 25, no. 12, pp. 3001–3012, 2008.
- [5] H.-j. Wang, W.-j. Fan, Y.-k. Cui, L. Zhou, B.-y. Yan, D.-h. Wu, X.-r. Xu *et al.*, "Hyperspectral remote sensing monitoring of grassland degradation," *Spectrosc. Spect. Anal.*, vol. 30, no. 10, pp. 2734–2738, 2010.
- [6] Z. Wang, Y. Ma, Y. Zhang, and J. Shang, "Review of remote sensing applications in grassland monitoring," *Remote Sens.*, vol. 14, no. 12, p. 2903, 2022.
- [7] H. Wan, C. Wang, Y. Li, Q. Wang, J. Li, and X. Liu, "Monitoring an invasive plant using hyperspectral remote sensing data," *Trans. Chin. Soc. Agric. Eng.*, vol. 26, no. 1, pp. 59–63, 2010.
- [8] L. Papp, B. Van Leeuwen, P. Szilassi, Z. Tobak, J. Szatmári, M. Árvai, J. Mészáros, and L. Pásztor, "Monitoring invasive plant species using hyperspectral remote sensing data," *Land*, vol. 10, no. 1, p. 29, 2021.
- [9] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, 2019.
- [10] G. Zhang, W. Xu, W. Zhao, C. Huang, E. N. Yk, Y. Chen, and J. Su, "A multiscale attention network for remote sensing scene images classification," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9530–9545, 2021.
- [11] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Survey.*, vol. 41, no. 3, pp. 1–58, 2009.
- [12] 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.
- [13] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, "Analysis and optimizations of global and local versions of the rx algorithm for anomaly detection in hyperspectral data," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, 2013.
- [14] W. Li and Q. Du, "Decision fusion for dual-window-based hyperspectral anomaly detector," *J. Appl. Remote Sens.*, vol. 9, no. 1, pp. 097297–097297, 2015.
- [15] W.-M. Liu and C.-I. Chang, "Multiple-window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 644–658, 2013.

- [16] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerospace Conf.*, vol. 3. IEEE, 2004.
- [17] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-rxd and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, 2014.
- [18] 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.
- [19] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel rx algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, 2016.
- [20] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, "Low-rank decomposition-based anomaly detection," in *Proc. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imagery XIX*, vol. 8743. SPIE, 2013, pp. 171–177.
- [21] 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.
- [22] M. Rahmani and G. K. Atia, "High dimensional low rank plus sparse matrix decomposition," *IEEE Trans. Signal Process.*, vol. 65, no. 8, pp. 2004–2019, 2017.
- [23] 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.
- [24] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, p. 289, 2016.
- [25] 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.
- [26] L. Zhang and C. Zhao, "A spectral-spatial method based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Int. J. Remote Sens.*, vol. 38, no. 14, pp. 4047–4068, 2017.
- [27] L. Zhu, G. Wen, and S. Qiu, "Low-rank and sparse matrix decomposition with cluster weighting for hyperspectral anomaly detection," *Remote Sens.*, vol. 10, no. 5, p. 707, 2018.
- [28] 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.
- [29] 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.*, 2023.
- [30] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [31] K. Zhang, M. Wang, S. Yang, and L. Jiao, "Spatial-spectral-graph-regularized low-rank tensor decomposition for multispectral and hyperspectral image fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1030–1040, 2018.
- [32] 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.
- [33] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, 2017.
- [34] Y. Chang, L. Yan, X.-L. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. on Cybern.*, vol. 50, no. 11, pp. 4558–4572, 2020.
- [35] Y. Chen, S. Wang, C. Peng, Z. Hua, and Y. Zhou, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 30, pp. 4022–4035, 2021.
- [36] Y. Chen, X. Xiao, C. Peng, G. Lu, and Y. Zhou, "Low-rank tensor graph learning for multi-view subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 92–104, 2021.
- [37] S. Li, W. Wang, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Low-rank tensor decomposition based anomaly detection for hyperspectral imagery," in *Proc. Int. Conf. Image Process.* IEEE, 2015, pp. 4525–4529.
- [38] Z. Chen, B. Yang, and B. Wang, "A preprocessing method for hyperspectral target detection based on tensor principal component analysis," *Remote Sens.*, vol. 10, no. 7, p. 1033, 2018.
- [39] 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.
- [40] 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.
- [41] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "Lren: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 5, 2021, pp. 4139–4146.
- [42] 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.
- [43] 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.
- [44] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "Bs 3 lnet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, 2023.
- [45] H. Wang, F. Zhang, J. Wang, T. Huang, J. Huang, and X. Liu, "Generalized nonconvex approach for low-tubal-rank tensor recovery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 8, pp. 3305–3319, 2021.
- [46] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.
- [47] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, 2019.
- [48] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
- [49] C. Gao, N. Wang, Q. Yu, and Z. Zhang, "A feasible nonconvex relaxation approach to feature selection," in *Proc. AAAI Conf. Artif. Intell.*, vol. 25, no. 1, 2011, pp. 356–361.
- [50] L. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," *Technometrics*, vol. 35, no. 2, pp. 109–135, 1993.
- [51] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," *J. Amer. Stat. Assoc.*, vol. 96, no. 456, pp. 1348–1360, 2001.
- [52] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *J. Mach. Learn. Res.*, vol. 11, no. 3, 2010.
- [53] D. Geman and G. Reynolds, "Constrained restoration and the recovery of discontinuities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 3, pp. 367–383, 1992.
- [54] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic l_0 -minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, 2008.
- [55] J. H. Friedman, "Fast sparse regression and classification," *Int. J. Forecasting*, vol. 28, no. 3, pp. 722–738, 2012.
- [56] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," *Ann. Statist.*, 2010.
- [57] K. Chen, H. Dong, and K.-S. Chan, "Reduced rank regression via adaptive nuclear norm penalization," *Biometrika*, vol. 100, no. 4, pp. 901–920, 2013.
- [58] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," in *Proc. AAAI Conf. Artif. Intell.*, vol. 29, no. 1, 2015.
- [59] S. Gaïffas and G. Lecué, "Weighted algorithms for compressed sensing and matrix completion," *arXiv preprint arXiv:1107.1638*, 2011.
- [60] C. Chen, B. He, Y. Ye, and X. Yuan, "The direct extension of admm for multi-block convex minimization problems is not necessarily convergent," *Math. Program.*, vol. 155, no. 1–2, pp. 57–79, 2016.