SIAM J. Imaging Sciences
Vol. 19, No. 2, pp. 807–838

Spectral-Spatial Extraction through Layered Tensor Decomposition for Hyperspectral Anomaly Detection*

Quan Yu†, Yu-Hong Dai‡, and Minru Bai§

Abstract. Low rank tensor representation (LRTR) methods are very useful for hyperspectral anomaly detection (HAD). However, existing LRTR methods often overlook spectral anomaly and rely on computationally expensive large-scale matrix singular value decomposition. To overcome these limitations, we propose the following highly efficient layered tensor decomposition (LTD) framework that simultaneously optimizes two key components within a unified model: layer 1, which reduces spectral redundancy and extracts spectral anomaly, and layer 2, which captures spatial low rank features and extracts spatial anomaly. The resulting spectral and spatial anomaly maps are then integrated to achieve a robust final detection result. An iterative algorithm based on proximal alternating minimization is developed to solve the proposed LTD model, with convergence guarantees provided. Moreover, we introduce a rank reduction strategy with validation mechanism that adaptively reduces data size while preventing excessive reduction. Theoretically, we rigorously establish the equivalence between the tensor tubal rank and tensor group sparsity regularization (TGSR) and, under mild conditions, demonstrate that the relaxed formulation of TGSR shares the same global minimizers and optimal values as its original counterpart. Experimental results on the Airport-Beach-Urban and MVTec datasets demonstrate that our approach outperforms state-of-the-art methods.

Key words. anomaly detection, layered tensor decomposition, low rank representation, group sparsity regularization

MSC codes. 15A69, 47A80, 65K05

DOI. 10.1137/25M1755679

1. Introduction.

Hyperspectral images (HSIs) integrate 2D spatial information with detailed spectral data, capturing the fundamental characteristics of objects. This integration offers substantial advantages in applications such as denoising [44], classification [17], image fusion [6, 14], and anomaly detection [50]. Recently, hyperspectral anomaly detection (HAD) has garnered significant attention due to its critical role in public safety and defense applications. The objective of HAD is to detect and isolate anomalous objects from their background.

---

[*] Received by the editors April 29, 2025; accepted for publication (in revised form) November 19, 2025; published electronically April 27, 2026.

https://doi.org/10.1137/25M1755679

Funding: The second author was supported by National Natural Science Foundation of China grants 12021001 and 92473208. The third author was support by National Natural Science Foundation of China grants 11971159 and 12071399 and by the Hunan Provincial Key Laboratory of Intelligent Information Processing and Applied Mathematics.

[†] School of Mathematics, Hunan University, Changsha, 410082 China (quanyu@hnu.edu.cn).

[‡] State Key Laboratory of Mathematical Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, 100190 China, and School of Mathematical Sciences, University of Chinese Academy of Sciences, Beijing, 100049 China (dyh@lsec.cc.ac.cn).

[§] Corresponding author. School of Mathematics, Hunan University, Changsha, 410082 China (minru-bai@hnu.edu.cn).

To address this problem, a variety of approaches have been developed, including statistic based methods, deep learning based methods, and low rank representation based methods.

One of the most commonly used statistical approaches for anomaly detection is the Reed-Xiaoli (RX) detector [35], which models background pixels using a multivariate Gaussian distribution. The RX detector calculates the mean and covariance matrix of the background and then determines the Mahalanobis distance between a test pixel and the background mean to identify anomalies based on a predefined threshold. Although the RX detector is simple and effective, its exclusive reliance on the covariance matrix limits its ability to capture finer relationships within HSI data. To address this limitation, several variants of the RX detector have been introduced, such as kernel RX [25, 53], subspace RX [39], and local RX [33]. However, despite these enhancements, these methods still rely on manually crafted distribution models, which may struggle to capture the complexity and variability of real-world backgrounds.

Recent studies have highlighted deep learning based methods for effective HAD, which utilize deep networks to mine and interpret higher order information contained in HSI. These methods can be categorized into unsupervised, supervised, and self-supervised learning. Unsupervised learning methods, such as autoencoder (AE) detector [3] and robust graph AE (RGAE) detector [9], learn background representations or anomalies by analyzing the inherent structure of the data without requiring labeled data. Supervised learning methods, such as convolutional neural network based detection (CNND) [28], use labeled data to classify each pixel as background or anomalous. Self-supervised learning methods, like blind-spot self-supervised learning network (BS^{3}LNet) [13] and pixel-shuffle downsampling blind-spot reconstruction network (PDBSNet) [46], learn to model the distribution of normal data by reconstructing masked or rearranged pixels from their context. Anomalies are then detected as data points that result in a high reconstruction error, as they deviate from the learned normal patterns. Building upon this, the nonlocal and local feature-coupled self-supervised network (NL2Net) [45] refines the reconstruction process by creating a more comprehensive context model that integrates both local and nonlocal features. To push performance further, the work in [37] synergistically integrates a deep denoiser prior with a nonlocal framework, harnessing the complementary strengths of deep learning for feature extraction and nonlocal low rank for structural modeling. However, these methods often demand extensive training times and involve intricate tuning of network architectures and training parameters.

Beyond the statistical and deep learning methods mentioned above, recent research has focused on the low rank representation (LRR) of the background and the sparsity of anomaly. With the maturation of matrix theory, many researchers first convert HSI into a two-dimensional matrix and then use LRR for anomaly detection. A widely adopted approach is robust principal component analysis (RPCA), which decomposes HSI into a low rank matrix and a sparse anomaly matrix [7]. Subsequently, a series of improved methods have been proposed, such as those addressing noise [41] and incorporating Mahalanobis distance metrics [52]. The above methods struggle to distinguish weak anomalies. To address this limitation, a variant based on RPCA was introduced, which models the low rank background component via a representation learned from a fixed, predefined dictionary [36, 55]. Additionally, Ma et al. [31] enhanced the robustness and detection performance by employing a learned background dictionary. In a parallel line of research, other frameworks integrate more sophisticated priors to improve performance; for instance, HADDNLP [38] utilizes a double nonlocal prior to better characterize the background structure. HSI can be treated as a third order tensor with one spectral and two spatial dimensions. Matrix based methods often overlook this multidimensional structure, potentially degrading detection performance. Consequently, methods based on LRTR of the background and anomaly sparsity have rapidly developed for HAD. Chen, Yang, and Wang [8] extended matrix RPCA to tensors, proposing tensor principal component analysis (TPCA) to effectively distinguish between the background and targets. Similarly, many methods construct a tensor background dictionary to enhance representation quality [47, 50]. Additionally, some regularizations, such as total variation regularization [12], are considered to improve anomaly detection quality.

Until now, the LRTR based HAD method has faced two main drawbacks. First, due to the redundancy in the spectral dimension of HSI, some algorithms perform orthogonal transformations such as TPCA and maximum noise fraction (MNF) transformation [16, 47] to reduce the spectral dimension and computational load, followed by anomaly detection on the preprocessed data. This approach results in the loss of spectral domain anomaly and divides the original data processing into two stages, significantly compromising effectiveness. Second, extracting spatial domain anomaly often involves the singular value decomposition (SVD) of large matrices, leading to low computational efficiency.

To overcome the aforementioned limitations, we introduce an HAD model that employs layered tensor decomposition (LTD) to efficiently detect both spectral and spatial anomalies within a unified framework, as depicted in Figure 1. Our model comprises two integrated layers that operate synergistically. Layer 1 reduces spectral dimensionality, thereby lowering computational complexity, and simultaneously generates the coefficient tensor while identifying spectral anomaly often missed by existing models. Layer 2 captures the low rank spatial structure of the coefficient tensor, during which spatial anomaly is detected. Finally, we integrate the spectral and spatial anomalies to achieve a superior anomaly detection result. In layer 2, we demonstrate the equivalence between tensor tubal rank and tensor group sparsity regularization (TGSR). Additionally, we show that TGSR and its relaxed formulation share the same global minimizers and optimal values under certain mild conditions. Leveraging these insights, the relaxed formulation of TGSR is employed to characterize the rank of the coefficient tensor. This regularization decomposes the coefficient tensor into two smaller tensors, whose sizes are directly related to the rank of the coefficient tensor. Building on the aforementioned decompositions and the established theory related to group sparse support sets, we developed a rank reduction strategy with validation mechanism. This strategy adaptively and dynamically reduces the size of these two tensors, thereby improving computational efficiency and eliminating the reliance on large-scale matrix SVD. Additionally, it ensures that the reduction process is appropriate, preventing excessively low ranks that could compromise computational performance. In summary, our main contributions include the following:

1. We propose a novel layered tensor decomposition (LTD) framework that holistically decomposes large-scale HSI data into a product of multiple smaller tensors, substantially enhancing algorithmic performance.
2. We present a unified framework that leverages LTD and, to the best of our knowledge, represents the first LRR based method to simultaneously detect spectral and spatial anomalies within a single optimization process. By strategically fusing the resulting anomaly maps, our approach yields exceptionally robust detection performance.

(3) We demonstrate the equivalence between tensor tubal rank and tensor group sparsity regularization (TGSR). Additionally, we show that TGSR and its relaxed formulation share the same global minimizers and optimal values under certain mild conditions. Furthermore, we develop theories related to group sparse support sets.

(4) We propose a rank reduction strategy with validation mechanism that adaptively and dynamically reduces tensor size. This approach ensures that the reduction process is appropriate, correcting any erroneous reductions to prevent excessively low ranks that could compromise performance.

(5) We propose a proximal alternating minimization algorithm to solve the proposed model and provide a comprehensive convergence analysis.

Experimental results on Airport-Beach-Urban and MVTec databases demonstrate the advantages of our method.

2. Preliminaries.

In this section, we present the notation and preliminaries utilized throughout this article. The sets of real numbers and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. Some basic notation is summarized in Table 1.

Next, we introduce the t-product between two third order tensors. This, in turn, allows us to propose a new tensor decomposition framework known as (skinny) T-SVD.

Definition 1 (t-product [22]). *Let $\mathcal{X} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2\times n_4\times n_3}$ be two third order tensors. Their* t*-product is defined as $\mathcal{Z} = \mathcal{X} * \mathcal{Y} \in \mathbb{R}^{n_1\times n_4\times n_3}$ and is computed via*

$$\mathcal{Z}(i,j,:) = \sum_{k=1}^{n_2} \mathcal{X}(i,k,:) \bullet \mathcal{Y}(k,j,:),$$

where $\bullet$ denotes the circular convolution between two tubal fiber vectors.

Table 1
Notation.

Notation	Description
$[n]$	the set $\{1,2,\ldots,n\}$
x, $\boldsymbol{x}$, X, $\mathcal{X}$	scalar, vector, matrix, tensor
$\mathcal{X}(:,:,k)/X^{(k)}$	kth frontal slice of a third order tensor $\mathcal{X}$
$\mathcal{X}(i,j,k)/\mathcal{X}_{ijk}$	the (i,j,k)th index value of a third order tensor $\mathcal{X}$
$\lVert\mathcal{X}\rVert$	the Frobenius norm of $\mathcal{X}$, which is defined as $\sqrt{\sum_{ijk}\lvert\mathcal{X}_{ijk}\rvert^2}$
$\langle\mathcal{X},\mathcal{Y}\rangle$	the inner product of two same-sized third order tensors $\mathcal{X}$ and $\mathcal{Y}$, which is defined as $\sum_{ijk}\mathcal{X}_{ijk}\mathcal{Y}_{ijk}$
$X_{(3)}$	the mode-3 unfolding of a third order tensor $\mathcal{X}$ (Definition 3)
$\times_3$	the mode-3 tensor-matrix product (Definition 4)
$\mathcal{X}^T$	the conjugate transpose of a third order tensor $\mathcal{X}$ (Definition 6)
$\mathcal{X} * \mathcal{Y}$	tensor product between two third order tensors $\mathcal{X}$ and $\mathcal{Y}$ (Definition 1)
$\lVert X\rVert_2$	the spectral norm of a matrix X, which is the largest singular value
$\delta_{\mathbb{X}}(\mathcal{X})$	the indicator function of a subset $\mathbb{X}$, which is 0 if $\mathcal{X}\in\mathbb{X}$ and ∞ if $\mathcal{X}\notin\mathbb{X}$
$\mathcal{X}\geq 0$	the tensor $\mathcal{X}$ with all elements being nonnegative
$\mathcal{I}$	the identity tensor (Definition 7)
$\Gamma(\mathcal{X})$	the group sparse support set of $\mathcal{X}$, which is defined as $\{j \mid \lVert\mathcal{X}(:,j,:)\rVert \neq 0\}$

Theorem 2.1 (T-SVD and skinny T-SVD [22, 30]). *Let $\mathcal{X}$ be an $n_1 \times n_2 \times n_3$ real-valued tensor. Then it can be factorized as $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor. Then the skinny T-SVD of $\mathcal{X}$ is $\mathcal{X} = \mathcal{U}_1 * \mathcal{S}_1 * \mathcal{V}_1^T$, where $\mathcal{U}_1 = \mathcal{U}(:, 1:r, :)$, $\mathcal{S}_1 = \mathcal{S}(1:r, 1:r, :)$, and $\mathcal{V}_1 = \mathcal{V}(:, 1:r, :)$ in which r denotes the tensor tubal rank of $\mathcal{X}$ (see Definition 2).*

Definition 2 (tensor tubal rank [21]). *Suppose that $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with T-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. The tensor tubal rank is defined as $\operatorname{rank}_t(\mathcal{X}) = \#\{i : \mathcal{S}(i,i,:) \neq 0\}$.*

Some related concepts, including the f-diagonal tensor and the conjugate transpose, among others, are elucidated in Appendix A.

Definition 3 (mode-3 unfolding [23]). *Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its mode-3 unfolding $X_{(3)}$ is an $n_3 \times (n_1 n_2)$ matrix, which satisfies $X_{(3)}(k, (i-1)n_2 + j) = \mathcal{X}(i,j,k)$ for $i \in [n_1]$, $j \in [n_2]$, and $k \in [n_3]$.*

Definition 4 (mode-3 tensor-matrix product [23]). *Given a third order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and a matrix $M \in \mathbb{R}^{p \times n_3}$, the mode-3 product of $\mathcal{X}$ and M, denoted as $\mathcal{X} \times_3 M \in \mathbb{R}^{n_1 \times n_2 \times p}$, is defined by $(\mathcal{X} \times_3 M)_{ijk} = \sum_{l=1}^{n_3} \mathcal{X}_{ijl} M_{kl}$, where $i \in [n_1]$, $j \in [n_2]$, and $k \in [p]$.*

3. Proposed anomaly detection method.

In this section, we introduce our proposed anomaly detection method, which is composed of a layered tensor decomposition (LTD) framework and a subsequent spectral-spatial fusion stage. The LTD framework systematically decomposes the detection task into two complementary layers operating within a unified model: layer 1 to extract spectral anomaly and layer 2 to extract spatial anomaly. The fusion stage then integrates the outputs from these layers to produce the final detection result. The overall architecture of our method is illustrated in Figure 1.

3.1. An LTD framework.

In this subsection, we formulate the objective function of the proposed LTD model. The function integrates two key components that operate concurrently: layer 1, which addresses spectral redundancy and extracts spectral anomaly, and layer 2, which captures spatial low rank features and identifies spatial anomaly. The mathematical formulation of each component is detailed below.

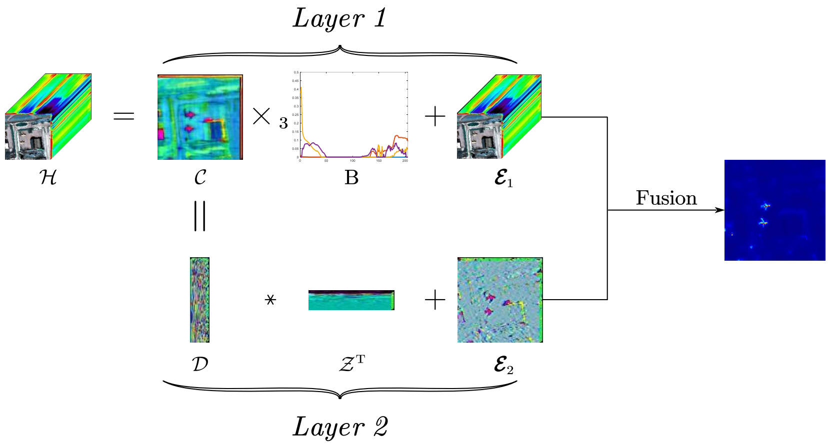


Figure 1. *Flowchart of the proposed anomaly detection method.*

Layer 1: Spectral anomaly extraction. Hyperspectral images (HSIs) are characterized by high correlation between spectral bands, leading to significant data redundancy [24]. A common preprocessing step is therefore dimensionality reduction [16, 47]. Building on this principle, the first layer of our model aims to represent the HSI by separating its low-dimensional background from spectral anomaly. We formulate this decomposition as

$$\mathcal{H} = \mathcal{C} \times_3 B + \mathcal{E}_1,$$

where $\mathcal{H} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is the input HSI. The model represents $\mathcal{H}$ as a combination of three components: a spectral dictionary $B \in \mathbb{R}^{n_3 \times b}$, whose columns represent the primary spectral signatures of the background; a coefficient tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times b}$, which encodes the spatial abundance of each background signature; and a spectral anomaly tensor $\mathcal{E}_1 \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, which captures pixels that deviate from the background model.

To ensure a physically meaningful and stable decomposition, we incorporate several constraints. First, since spectral signatures are inherently nonnegative, we impose a nonnegativity constraint on the dictionary, $B \geq 0$. Second, to eliminate the inherent scale ambiguity between B and $\mathcal{C}$, we enforce a unit-norm constraint on each spectral tube of the coefficient tensor, i.e., $\|\mathcal{C}(i,j,:)\| = 1$. With these foundational constraints established, we can now formulate a complete optimization problem that also accounts for practical factors like noise and the sparse nature of anomalies. We account for the presence of Gaussian noise in the data by using a Frobenius norm fidelity term. Furthermore, to enforce the prior knowledge that anomalies are spatially sparse yet spectrally dense, we penalize the anomaly tensor $\mathcal{E}_1$ using the joint sparsity inducing $\ell_{1,1,\phi}$-norm. Integrating all these components leads to the objective function

$$\tag{3.1} \begin{aligned} \min_{\mathcal{C},B,\mathcal{E}_1} \quad & \frac{\lambda_1}{2} \|B\|^2 + \lambda_2 \|\mathcal{E}_1\|_{11\phi} + \frac{\lambda_3}{2} \|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1\|^2 \\ \text{s.t.} \quad & B \geq 0, \ \|\mathcal{C}(i,j,:)\| = 1 \ \forall i \in [n_1], \ j \in [n_2], \end{aligned}$$

where $\|\mathcal{E}_1\|_{11\phi} := \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \phi(\|\mathcal{E}_1(i,j,:)\|)$ with function $\phi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$. The parameter $\lambda_1 > 0$ constrains the size of the entries in B to prevent excessively large values, which could induce instability in the results.

Remark 3.1. In [50], it is demonstrated that $\phi^{CapL1}(\cdot) := \min\{\cdot, 1\}$ performs well in anomaly detection, both in terms of computational efficiency and effectiveness. Therefore, in this paper, we set $\phi = \phi^{CapL1}$.

Layer 2: Spatial anomaly extraction. The coefficient tensor $\mathcal{C}$, produced by the spectral decomposition in the first layer, serves as the input for the second layer of our model. While the spectral dimension has been compressed, $\mathcal{C}$ inherits the strong spatial self-similarity of the background from the original HSI, resulting in a spatially low rank structure. The tensor tubal rank is a powerful tool for quantifying this property [51]. Accordingly, we model the spatial anomaly extraction task as a decomposition of $\mathcal{C}$ into a low rank background component $\mathcal{L}$ and a sparse spatial anomaly component $\mathcal{E}_2$. This leads to the following optimization problem:

$$\tag{3.2} \min_{\mathcal{L},\mathcal{E}_2} \ \lambda_4 \operatorname{rank}_t(\mathcal{L}) + \lambda_5 \|\mathcal{E}_2\|_{11\phi} \quad \text{s.t.} \quad \mathcal{C} = \mathcal{L} + \mathcal{E}_2.$$

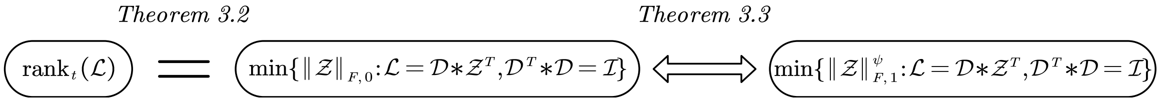


Figure 2. *Relationships between tensor tubal rank and group sparsity regularization.*

Driven by the group sparsity structure evident in the coding of low rank matrices [10, 43] and low rank tensors [11], we represent $\mathcal{L}$ within a subspace $\mathcal{D}$ and limit its rank by imposing group sparsity on its coefficients $\mathcal{Z}$. To justify the validity of this representation, we investigate the relationship between tensor tubal rank and group sparsity regularization.

Theorem 3.2. *For any tensor $\mathcal{L}\in\mathbb{R}^{n_1\times n_2\times b}$, one has*

$$\mathrm{rank}_t(\mathcal{L})=\min\left\{\|\mathcal{Z}\|_{F,0}:\mathcal{L}=\mathcal{D}*\mathcal{Z}^T,\mathcal{D}^T*\mathcal{D}=\mathcal{I},\mathcal{D}\in\mathbb{R}^{n_1\times r\times b},\mathcal{Z}\in\mathbb{R}^{n_2\times r\times b}\right\},$$

with $\|\mathcal{Z}\|_{F,0}=\sum_{j=1}^{r}\|\mathcal{Z}(:,j,:)\|^0$, for any r that satisfies $\mathrm{rank}_t(\mathcal{L})\le r\le\min\{n_1,n_2\}$.

Proof. See Appendix B for the proof. ■

Theorem 3.3. *There exists a $\bar{\nu}>0$. For any function ψ that satisfies $\psi(0)=0$, $x/\nu\le\psi(x)<1$ for $x\in(0,\nu)$ and $\psi(x)=1$ for $x\ge\nu$, where $0<\nu<\bar{\nu}$, the following two problems share the same global minimizers and optimal values:*

$$\begin{cases}(P_0)\ \ \min\left\{\|\mathcal{Z}\|_{F,0}:\mathcal{L}=\mathcal{D}*\mathcal{Z}^T,\mathcal{D}^T*\mathcal{D}=\mathcal{I}\right\};\\(P_\psi)\ \ \min\left\{\|\mathcal{Z}\|_{F,1}^{\psi}:\mathcal{L}=\mathcal{D}*\mathcal{Z}^T,\mathcal{D}^T*\mathcal{D}=\mathcal{I}\right\}.\end{cases}$$

Here $\|\mathcal{Z}\|_{F,1}^{\psi}=\sum_{j=1}^{r}\psi(\|\mathcal{Z}(:,j,:)\|)$.

Proof. See Appendix C for the proof. ■

Remark 3.4. The relationships between tensor tubal rank and group sparsity regularization are shown in Figure 2. Theorem 3.2 establishes that the tensor tubal rank $\mathrm{rank}_t(\mathcal{L})$ is equivalent to the optimal values of problem (P_0). However, due to the nonconvex and nonsmooth nature of $\|\mathcal{Z}\|_{F,0}$, algorithms based on it exhibit unstable numerical performance. Consequently, we demonstrate in Theorem 3.3 that problem (P_0) and problem (P_ψ) have the same global minimizers and optimal values.

Optimizing $\|\mathcal{Z}\|_{F,1}^{\psi}$ is significantly more efficient than optimizing tensor nuclear norm, as it eliminates the need for large-scale SVD, and the dimensionality of $\mathcal{Z}$ is considerably smaller than that of $\mathcal{L}$. Furthermore, in algorithms based on $\|\mathcal{Z}\|_{F,1}^{\psi}$, the size of r can adaptively decrease until it approximates $\mathrm{rank}_t(\mathcal{L})$, thereby obviating the need for additional techniques to adjust r. This is in contrast to tensor decomposition methods discussed in [54], which require adjusting r. By leveraging the ability to adaptively adjust r and the fact that $\mathrm{rank}_t(\mathcal{L})$ is much smaller, models based on $\|\mathcal{Z}\|_{F,1}^{\psi}$ exhibit not only highefficiency but also superior performance.

Motivated by Theorems 3.2 and 3.3, we transform problem (3.2) into the formulation

$$\min_{\mathcal{D},\mathcal{Z},\mathcal{E}_2}\ \lambda_4\|\mathcal{Z}\|_{F,1}^{\psi}+\lambda_5\|\mathcal{E}_2\|_{11\phi}\quad\text{s.t.}\quad\mathcal{C}=\mathcal{D}*\mathcal{Z}^T+\mathcal{E}_2,\ \mathcal{D}^T*\mathcal{D}=\mathcal{I},\tag{3.3}$$

where $\mathcal{E}_2 \in \mathbb{R}^{n_1 \times n_2 \times b}$, $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times b}$, $\mathcal{Z} \in \mathbb{R}^{n_2 \times r \times b}$. In this paper, we choose $\psi(x) = \min\{x^p/\nu^p, 1\}$ for some $0 < p < 1$ and $\nu > 0$.

LTD framework. Integrating the spectral anomaly extraction from layer 1 and the spatial anomaly extraction from layer 2, the proposed LTD model can be formulated as

$$\begin{aligned}(3.4)\quad &\min_{\mathcal{C},B,\mathcal{E}_1,\mathcal{D},\mathcal{Z},\mathcal{E}_2} && \frac{\lambda_1}{2}\|B\|^2 + \lambda_2\|\mathcal{E}_1\|_{11\phi} + \frac{\lambda_3}{2}\|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1\|^2 + \lambda_4\|\mathcal{Z}\|_{F,1}^{\psi} + \lambda_5\|\mathcal{E}_2\|_{11\phi} \\ &\text{s.t.} && \mathcal{C} = \mathcal{D} * \mathcal{Z}^T + \mathcal{E}_2,\ B \in \mathbb{B},\ \mathcal{C} \in \mathbb{C},\ \mathcal{D} \in \mathbb{D},\end{aligned}$$

where $\mathbb{B} = \{B \mid B \geq 0\}$, $\mathbb{C} = \{\mathcal{C} \mid \|\mathcal{C}(i,j,:)\| = 1\}$, and $\mathbb{D} = \left\{\mathcal{D} \mid \mathcal{D}^T * \mathcal{D} = \mathcal{I}\right\}$.

Remark 3.5. The fundamental novelty of our LTD framework lies in two key areas: (1) a unified framework for enhanced performance, and (2) specific architectural innovations for superior efficiency.

First, regarding performance, LTD is the first low rank representation based framework to unify the detection of both spectral ($\mathcal{E}_1$) and spatial ($\mathcal{E}_2$) anomalies within a single optimization process. This unified design enables mutual refinement between the two layers, leading to demonstrably superior accuracy and robustness, as validated by our ablation study (see subsection 5.2.4). Second, for efficiency, the framework incorporates the following two key innovations: (a) Its layered decomposition structurally breaks down the large input tensor into smaller, computationally tractable ones, inherently reducing computational complexity. (b) Leveraging our new theoretical results (Theorems 3.2 and 3.3), we replace the computationally expensive nuclear norm with a far more efficient group sparsity regularization to handle spatial low-rankness. Our adaptive algorithm ensures that complex operations are confined to small scale tensors, making this approach significantly faster than conventional methods.

3.2. Spectral-spatial fusion.

To effectively highlight anomalous targets, we design an adaptive fusion architecture that integrates two complementary detection maps, as illustrated in Figure 3. First, we compute the spectral anomaly detection map, given by $T_1 = \sqrt{\sum_{k=1}^{n_3} |\mathcal{E}_1^*(:,:,k)|^2}$, and the spatial anomaly detection map, given by $T_2 = \sqrt{\sum_{k=1}^{b} |\mathcal{E}_2^*(:,:,k)|^2}$. Then, we obtain the final detection result as

$$(3.5)\qquad T = \mathrm{IGF}\left(T_1 \circ T_2\right) \quad \text{or} \quad T = \mathrm{IGF}(\mathrm{IGF}(\mathrm{IGF}\left(T_1 \circ T_2\right), T_1), T_2),$$

where $\circ$ denotes the Hadamard product, and IGF is a guided image filter [15]. Notably, spectral-spatial fusion is a modular postprocessing stage, independent of the core LTD model, which allows for the flexible design of optimal fusion strategies tailored to different datasets.

Remark 3.6. From Figure 3, it can be seen that T_1 and T_2 have complementary characteristics. Specifically, the noise points and anomalous targets are complementary in their positions in T_1 and T_2. Therefore, we take the Hadamard product of T_1 and T_2. Although $T_1 \circ T_2$ still contains slight noise, considering that the guided image filter can effectively remove noise and enhance details, we apply the guided image filter to $T_1 \circ T_2$, resulting in an image with reduced noise and clearer anomalous targets.

Remark 3.7. To clarify the notation and motivation behind the two fusion strategies in (3.5), we provide the following detailed explanation. The operator $\mathrm{IGF}(I, G)$ represents the

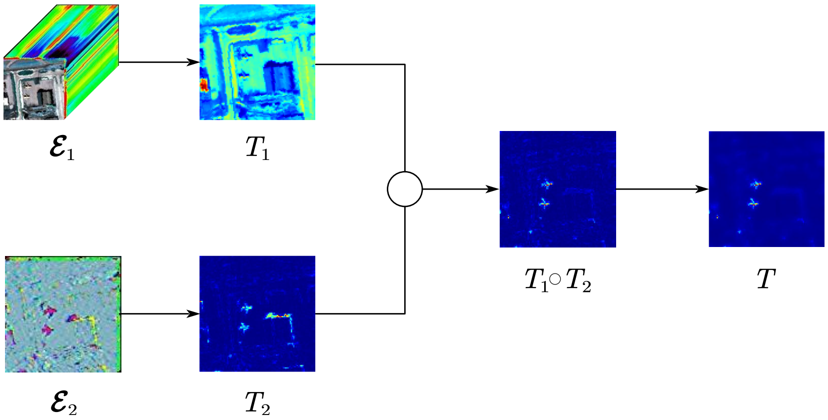


Figure 3. *Flow chart of spectral-spatial fusion.*

guided image filter, an edge-preserving smoothing filter designed to transfer structural information from a guidance image G to an input image I. For the self-guided case, $\mathrm{IGF}(I,I)$, we adopt the shorthand $\mathrm{IGF}(I)$. This configuration effectively suppresses noise while preserving the dominant structures within the input image.

With this clarified, we can see that the two fusion strategies serve distinct purposes.

- **Direct fusion**: The first strategy, $T = \mathrm{IGF}\,(T_1 \circ T_2)$, performs self-guided smoothing on the combined anomaly map. It is a straightforward and efficient method for reducing noise while retaining the strong anomalous features that are jointly present in both the spectral T_1 and spatial T_2 maps.
- **Cascaded fusion**: The second strategy, $T = \mathrm{IGF}(\mathrm{IGF}(\mathrm{IGF}\,(T_1 \circ T_2)\,, T_1), T_2)$, is a more sophisticated, multistage refinement process. It first applies self-guided filtering to the combined map $T_1 \circ T_2$ for initial smoothing. The result is then sequentially refined, first using the spectral map T_1 as a guide to preserve spectrally distinct anomalies, and subsequently using the spatial map T_2 to reinforce spatially prominent structures. This hierarchical process is designed to better preserve subtle targets that may be distinct to only one of the two domains.

We employ an adaptive fusion methodology, where the choice between the direct and cascaded fusion strategies is determined individually for each dataset to ensure optimal performance across diverse scenarios.

4. Algorithm description and theoretical analysis.

4.1. PAM algorithm. Using the half-quadratic splitting technique, we reformulate (3.4) as an unconstrained problem as follows:

$$\begin{aligned} &\min F\left(\mathcal{C}, B, \mathcal{E}_1, \mathcal{D}, \mathcal{Z}, \mathcal{E}_2\right) \\ &= \frac{\lambda_1}{2}\left\|B\right\|^2 + \lambda_2 \left\|\mathcal{E}_1\right\|_{11\phi} + \frac{\lambda_3}{2}\left\|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1\right\|^2 \\ &\quad + \lambda_4 \left\|\mathcal{Z}\right\|_{F,1}^{\psi} + \lambda_5 \left\|\mathcal{E}_2\right\|_{11\phi} + \frac{\lambda_6}{2}\left\|\mathcal{C} - \mathcal{D} * \mathcal{Z}^T - \mathcal{E}_2\right\|^2 + \delta_{\mathbb{B}}\left(B\right) + \delta_{\mathbb{C}}\left(\mathcal{C}\right) + \delta_{\mathbb{D}}\left(\mathcal{D}\right). \end{aligned} \tag{4.1}$$

Within the framework of the proximal alternating minimization (PAM) algorithm [1], we alternatively update each variable as

$$\begin{cases} \mathcal{C}^{t+1} = \arg\min\limits_{\mathcal{C}} \ F\left(\mathcal{C}, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \dfrac{\rho_1}{2}\left\|\mathcal{C} - \mathcal{C}^t\right\|^2; \\ B^{t+1} = \arg\min\limits_{B} \ F\left(\mathcal{C}^{t+1}, B, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \dfrac{\rho_2}{2}\left\|B - B^t\right\|^2; \\ \mathcal{E}_1^{t+1} = \arg\min\limits_{\mathcal{E}_1} \ F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \dfrac{\rho_3}{2}\left\|\mathcal{E}_1 - \mathcal{E}_1^t\right\|^2; \\ \mathcal{D}^{t+1} = \arg\min\limits_{\mathcal{D}} \ F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \dfrac{\rho_4}{2}\left\|\mathcal{D} - \mathcal{D}^t\right\|^2; \\ \mathcal{Z}^{t+1} = \arg\min\limits_{\mathcal{Z}} \ F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}, \mathcal{E}_2^t\right) + \dfrac{\rho_5}{2}\left\|\mathcal{Z} - \mathcal{Z}^t\right\|^2; \\ \mathcal{E}_2^{t+1} = \arg\min\limits_{\mathcal{E}_2} \ F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2\right) + \dfrac{\rho_6}{2}\left\|\mathcal{E}_2 - \mathcal{E}_2^t\right\|^2, \end{cases}$$

where $\rho_i,\ i \in [6]$ are positive constants, and t denotes the iteration number. Next, we detail the updates for $\mathcal{C}$, B, $\mathcal{E}_1$, $\mathcal{D}$, $\mathcal{Z}$, and $\mathcal{E}_2$.

4.1.1. Computing $\mathcal{C}^{t+1}$ and B^{t+1}.

The subproblems of $\mathcal{C}$ and B are

$$\min_{\mathcal{C}} \ f\left(\mathcal{C}, B^t\right) + \delta_{\mathbb{C}}\left(\mathcal{C}\right) + \frac{\rho_1}{2}\|\mathcal{C} - \mathcal{C}^t\|^2; \tag{4.2}$$

$$\min_{B} \ f\left(\mathcal{C}^{t+1}, B\right) + \delta_{\mathbb{B}}\left(B\right) + \frac{\rho_2}{2}\|B - B^t\|^2, \tag{4.3}$$

where $f\left(\mathcal{C}, B\right) = \frac{\lambda_1}{2}\|B\|^2 + \frac{\lambda_3}{2}\|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1^t\|^2 + \frac{\lambda_6}{2}\|\mathcal{C} - \mathcal{D}^t * \mathcal{Z}^{t^T} - \mathcal{E}_2^t\|^2$. However, (4.2) and (4.3) may have no closed-form solutions; we update $\mathcal{C}$ by solving the subproblem

$$\begin{aligned} &\arg\min_{\mathcal{C}} \delta_{\mathbb{C}}\left(\mathcal{C}\right) + \left\langle \nabla_{\mathcal{C}} f\left(\mathcal{C}^t, B^t\right), \mathcal{C} - \mathcal{C}^t \right\rangle + \frac{l_{\mathcal{C}}^t}{2}\|\mathcal{C} - \mathcal{C}^t\|^2 + \frac{\rho_1}{2}\|\mathcal{C} - \mathcal{C}^t\|^2 \\ &\quad = \arg\max_{\mathcal{C}} \delta_{\mathbb{C}}\left(\mathcal{C}\right) + \langle \mathcal{C}, \hat{\mathcal{C}}^{t+1} \rangle, \end{aligned} \tag{4.4}$$

where $\nabla_{\mathcal{C}} f(\mathcal{C}^t, B^t) = \lambda_3(\mathcal{C}^t \times_3 B^t + \mathcal{E}_1^t - \mathcal{H}) \times_3 B^{t^T} + \lambda_6(\mathcal{C}^t - \mathcal{D}^t * \mathcal{Z}^{t^T} - \mathcal{E}_2^t)$, $l_{\mathcal{C}}^t = \lambda_3\|B^t\|_2^2 + \lambda_6$, and $\hat{\mathcal{C}}^{t+1} = (l_{\mathcal{C}}^t \mathcal{C}^t + \rho_1 \mathcal{C}^t - \nabla_{\mathcal{C}} f\left(\mathcal{C}^t, B^t\right))/(l_{\mathcal{C}}^t + \rho_1)$. By applying the Cauchy–Schwarz inequality, the optimal solution of (4.4) is given by

$$\mathcal{C}^{t+1}(i,j,:) = \hat{\mathcal{C}}^{t+1}(i,j,:)/\|\hat{\mathcal{C}}^{t+1}(i,j,:)\|, \quad i \in [n_1],\ j \in [n_2]. \tag{4.5}$$

Likewise, we can update B by solving the subproblem

$$\begin{aligned} &\arg\min_{B} \delta_{\mathbb{B}}\left(B\right) + \left\langle \nabla_B f\left(\mathcal{C}^{t+1}, B^t\right), B - B^t \right\rangle + \frac{l_B^t}{2}\left\|B - B^t\right\|^2 + \frac{\rho_2}{2}\left\|B - B^t\right\|^2 \\ &\quad = \arg\min_{B} \delta_{\mathbb{B}}\left(B\right) + \frac{l_B^t + \rho_2}{2}\|B - \hat{B}^{t+1}\|^2, \end{aligned} \tag{4.6}$$

where $\nabla_B f(\mathcal{C}^{t+1}, B^t) = \lambda_1 B^t + \lambda_3(B^t C_{(3)}^{t+1} + {E_1}_{(3)}^t - H_{(3)}) C_{(3)}^{t+1^T}$, $l_B^t = \lambda_1 + \lambda_3\|C_{(3)}^{t+1}\|_2^2$, and $\hat{B}^{t+1} = (l_B^t B^t + \rho_2 B^t - \nabla_B f(\mathcal{C}^{t+1}, B^t))/(l_B^t + \rho_2)$. It is not difficult to see that the optimal solution of (4.6) is given by

$$B^{t+1} = \max\{0, \hat{B}^{t+1}\}. \tag{4.7}$$

4.1.2. Computing $\mathcal{E}_1^{t+1}$. The subproblem of $\mathcal{E}_1$ is

$$\begin{aligned}&\arg\min_{\mathcal{E}_1}\ \lambda_2\|\mathcal{E}_1\|_{11\phi}+\frac{\lambda_3}{2}\|\mathcal{H}-\mathcal{C}^{t+1}\times_3 B^{t+1}-\mathcal{E}_1\|^2+\frac{\rho_3}{2}\|\mathcal{E}_1-\mathcal{E}_1^t\|^2\\ =&\arg\min_{\mathcal{E}_1}\ \hat{\lambda}_2\|\mathcal{E}_1\|_{11\phi}+\frac{1}{2}\|\mathcal{E}_1-\hat{\mathcal{E}}_1^{t+1}\|^2,\end{aligned}\tag{4.8}$$

where $\hat{\lambda}_2=\lambda_2/(\lambda_3+\rho_3)$ and $\hat{\mathcal{E}}_1^{t+1}=(\lambda_3(\mathcal{H}-\mathcal{C}^{t+1}\times_3 B^{t+1})+\rho_3\mathcal{E}_1^t)/(\lambda_3+\rho_3)$. By [34, Appendix A.1], the optimal solution of (4.8) is given by

$$\mathcal{E}_1^{t+1}(i,j,:)=\begin{cases}\max\{0,\ 1-\hat{\lambda}_2/e_1\}\hat{\mathcal{E}}_1^{t+1}(i,j,:), & e_1\le 1+\hat{\lambda}_2;\\ \hat{\mathcal{E}}_1^{t+1}(i,j,:), & e_1>1+\hat{\lambda}_2,\end{cases}\quad i\in[n_1],\ j\in[n_2],\tag{4.9}$$

where $e_1=\|\hat{\mathcal{E}}_1^{t+1}(i,j,:)\|$.

4.1.3. Computing $\mathcal{D}^{t+1}$. The subproblem of $\mathcal{D}$ is

$$\begin{aligned}&\arg\min_{\mathcal{D}}\frac{\lambda_6}{2}\|\mathcal{C}^{t+1}-\mathcal{D}*\mathcal{Z}^{t^T}-\mathcal{E}_2^t\|^2+\frac{\rho_4}{2}\left\|\mathcal{D}-\mathcal{D}^t\right\|^2+\delta_{\mathbb{D}}(\mathcal{D})\\ &=\arg\max_{\mathcal{D}}\left\langle\mathcal{D},\lambda_6\left(\mathcal{C}^{t+1}-\mathcal{E}_2^t\right)*\mathcal{Z}^t+\rho_4\mathcal{D}^t\right\rangle+\delta_{\mathbb{D}}(\mathcal{D}).\end{aligned}\tag{4.10}$$

To solve the above problem, we introduce the following lemma.

Lemma 4.1. *Suppose a tensor $\mathcal{G}\in\mathbb{R}^{n_1\times n_2\times b}$ with tubal rank r has the skinny* T-SVD *$\mathcal{G}=\mathcal{U}*\mathcal{S}*\mathcal{V}^T$, where $\mathcal{U}\in\mathbb{R}^{n_1\times r\times b}$ and $\mathcal{V}\in\mathbb{R}^{n_2\times r\times b}$. Then,*

$$\arg\max_{\mathcal{D}^T*\mathcal{D}=\mathcal{I}}\ \langle\mathcal{D},\mathcal{G}\rangle=\mathcal{U}*\mathcal{V}^T.\tag{4.11}$$

Proof. See Appendix D for the proof. ■

By using Lemma 4.1, the optimal solution of (4.10) is given by

$$\mathcal{D}^{t+1}=\mathcal{U}_{\mathcal{D}}*\mathcal{V}_{\mathcal{D}}^T,\tag{4.12}$$

where $\mathcal{U}_{\mathcal{D}}$ and $\mathcal{V}_{\mathcal{D}}$ are obtained by the skinny T-SVD decomposition as follows: $\lambda_6(\mathcal{C}^{t+1}-\mathcal{E}_2^t)*\mathcal{Z}^t+\rho_4\mathcal{D}^t=\mathcal{U}_{\mathcal{D}}*\mathcal{S}_{\mathcal{D}}*\mathcal{V}_{\mathcal{D}}^T$.

4.1.4. Computing $\mathcal{Z}^{t+1}$. The subproblem of $\mathcal{Z}$ is

$$\begin{aligned}&\arg\min_{\mathcal{Z}}\lambda_4\|\mathcal{Z}\|_{F,1}^{\psi}+\frac{\lambda_6}{2}\left\|\mathcal{C}^{t+1}-\mathcal{D}^{t+1}*\mathcal{Z}^T-\mathcal{E}_2^t\right\|^2+\frac{\rho_5}{2}\left\|\mathcal{Z}-\mathcal{Z}^t\right\|^2\\ &=\arg\min_{\mathcal{Z}}\hat{\lambda}_4\|\mathcal{Z}\|_{F,1}^{\psi}+\frac{1}{2}\|\mathcal{Z}-\hat{\mathcal{Z}}^{t+1}\|^2,\end{aligned}\tag{4.13}$$

where $\hat{\lambda}_4=\lambda_4/(\lambda_6+\rho_5)$ and $\hat{\mathcal{Z}}^{t+1}=(\lambda_6(\mathcal{C}^{t+1}-\mathcal{E}_2^t)^T*\mathcal{D}^{t+1}+\rho_5\mathcal{Z}^t)/(\lambda_6+\rho_5)$. By [34, Appendix A] and [32], the optimal solution of (4.13) is given by

$$\mathcal{Z}^{t+1}(:,j,:)=\begin{cases}\mathrm{Prox}_{\hat{\lambda}_4}\psi\left(\hat{z}^{t+1}\right)\hat{\mathcal{Z}}^{t+1}(:,j,:)/\hat{z}^{t+1}, & \hat{z}^{t+1}\neq 0;\\ 0, & \hat{z}^{t+1}=0,\end{cases}\quad j\in[r],\tag{4.14}$$

where $\hat{z}^{t+1} = \|\hat{\mathcal{Z}}^{t+1}(:, j, :)\|$, and

$$\mathrm{Prox}_{\hat{\lambda}_4}\psi(z) = \begin{cases} u_1 & \text{if } \hat{\lambda}_4\psi(u_1) + \frac{1}{2}(u_1 - z)^2 \le \hat{\lambda}_4\psi(u_2) + \frac{1}{2}(u_2 - z)^2; \\ u_2 & \text{otherwise.} \end{cases}$$

Here $u_1 = \min\{\mathrm{prox}_{\hat{\lambda}_4}\|\cdot\|^p(z), v\}$, $u_2 = \max\{z, v\}$, and

$$\mathrm{prox}_{\hat{\lambda}_4}\|\cdot\|^p(z) = \begin{cases} 0 & \text{if } z < \pi_2; \\ \{0, \pi_1\} & \text{if } z = \pi_2; \\ \pi_\star & \text{if } z > \pi_2, \end{cases}$$

where $\pi_1 = (2\hat{\lambda}_4(1-p))^{1/(2-p)}$, $\pi_2 = \pi_1 + \hat{\lambda}_4 p \pi_1^{p-1}$, and $\pi_\star \in (\pi_1, z)$ is the solution of $g(\pi) = \pi + \hat{\lambda}_4 p \pi^{p-1} - z = 0$ with $\pi > 0$.

4.1.5. Computing $\mathcal{E}_2^{t+1}$.

The subproblem of $\mathcal{E}_2$ is

$$\begin{aligned} &\arg\min_{\mathcal{E}_2} \lambda_5 \|\mathcal{E}_2\|_{11\phi} + \frac{\lambda_6}{2}\|\mathcal{C}^{t+1} - \mathcal{D}^{t+1} * \mathcal{Z}^{t+1^T} - \mathcal{E}_2\|^2 + \frac{\rho_6}{2}\left\|\mathcal{E}_2 - \mathcal{E}_2^t\right\|^2 \\ &= \arg\min_{\mathcal{E}_2} \hat{\lambda}_5 \|\mathcal{E}_2\|_{11\phi} + \frac{1}{2}\|\mathcal{E}_2 - \hat{\mathcal{E}}_2^{t+1}\|^2, \end{aligned} \tag{4.15}$$

where $\hat{\lambda}_5 = \lambda_5/(\lambda_6 + \rho_6)$ and $\hat{\mathcal{E}}_2^{t+1} = (\lambda_6(\mathcal{C}^{t+1} - \mathcal{D}^{t+1} * \mathcal{Z}^{t+1^T}) + \rho_6\mathcal{E}_2^t)/(\lambda_6 + \rho_6)$. By [34, Appendix A.1], the optimal solution of (4.15) is given by

$$\mathcal{E}_2^{t+1}(i,j,:) = \begin{cases} \max\{0,\ 1 - \hat{\lambda}_5/e_2\}\hat{\mathcal{E}}_2^{t+1}(i,j,:), & e_2 \le 1 + \hat{\lambda}_5; \\ \hat{\mathcal{E}}_2^{t+1}(i,j,:), & e_2 > 1 + \hat{\lambda}_5, \end{cases} \quad i \in [n_1],\ j \in [n_2], \tag{4.16}$$

where $e_2 = \|\hat{\mathcal{E}}_2^{t+1}(i,j,:)\|$.

The entire procedure of the proposed method is summarized in Algorithm 4.1.

Algorithm 4.1. The PAM-based solver for the proposed LTD model.

```
  Input: HSI data $\mathcal{H}$, parameters $\{\lambda_i\}_{i=1}^6$ and $\{\rho_i\}_{i=1}^6$. Set $t \leftarrow 0$.
1 Update $\mathcal{C}^{t+1}$ via (4.5);
2 Update $B^{t+1}$ via (4.7);
3 Update $\mathcal{E}_1^{t+1}$ via (4.9);
4 Update $\mathcal{D}^{t+1}$ via (4.12);
5 Update $\mathcal{Z}^{t+1}$ via (4.14);
6 Update $\mathcal{E}_2^{t+1}$ via (4.16);
7 If a termination criterion is met, set $\mathcal{E}_1^* := \mathcal{E}_1^{t+1}$, $\mathcal{E}_2^* := \mathcal{E}_2^{t+1}$; else, set $t \leftarrow t+1$, return to
   step 1;
8 Compute the detection map $T$ by (3.5).
  Output: Detection map $T$.
```

4.2. Convergence analysis. In this subsection, we rigorously establish the global convergence guarantee of Algorithm 4.1. Prior to proving the convergence of the proposed method, we first introduce several pivotal lemmas that are essential for our subsequent analysis.

Lemma 4.2 (sufficient decrease condition). *Assume that $\mathcal{W}^t := (\mathcal{C}^t, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t)$ is generated by Algorithm 4.1. Then, $\mathcal{W}^t$ satisfies the following properties:*

(1) $F(\mathcal{W}^{t+1}) + \rho\|\mathcal{W}^{t+1} - \mathcal{W}^t\|^2 \le F(\mathcal{W}^t)$, *where* $\rho = \min_{i\in[6]}\{\rho_i/2\}$.

(2) $\lim_{t\to\infty} \|\mathcal{W}^{t+1} - \mathcal{W}^t\| = 0$.

(3) $\mathcal{W}^t$ *is bounded if either* $\mathcal{Z}^t$ *or* $\mathcal{E}_2^t$ *is bounded.*

Proof. (1) From the Lipschitz continuity of $\nabla_{\mathcal{C}} f(\mathcal{C}, B^t)$ about $\mathcal{C}$, it holds that

$$f\left(\mathcal{C}^{t+1}, B^t\right) \le f\left(\mathcal{C}^t, B^t\right) + \left\langle \nabla_{\mathcal{C}} f\left(\mathcal{C}^t, B^t\right), \mathcal{C}^{t+1} - \mathcal{C}^t\right\rangle + \frac{l_{\mathcal{C}}^t}{2}\|\mathcal{C}^{t+1} - \mathcal{C}^t\|^2. \tag{4.17}$$

Given that $\mathcal{C}^{t+1}$ is the minimizer of (4.4), we derive

$$\delta_{\mathbb{C}}\left(\mathcal{C}^{t+1}\right) + \left\langle \nabla_{\mathcal{C}} f\left(\mathcal{C}^t, B^t\right), \mathcal{C}^{t+1} - \mathcal{C}^t\right\rangle + \frac{l_{\mathcal{C}}^t}{2}\|\mathcal{C}^{t+1} - \mathcal{C}^t\|^2 + \Delta_{\mathcal{C}} \le \delta_{\mathbb{C}}\left(\mathcal{C}^t\right), \tag{4.18}$$

where $\Delta_{\mathcal{C}} = \frac{\rho_1}{2}\|\mathcal{C}^{t+1} - \mathcal{C}^t\|^2$. By summing (4.17) and (4.18), one has

$$F\left(\mathcal{C}^{t+1}, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \Delta_{\mathcal{C}} \le F\left(\mathcal{C}^t, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right). \tag{4.19}$$

Similarly, for B, let $\Delta_B = \frac{\rho_2}{2}\|B^{t+1} - B^t\|^2$. Then, we have

$$F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \Delta_B \le F\left(\mathcal{C}^{t+1}, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right). \tag{4.20}$$

Given that $\mathcal{E}_1^{t+1}$, $\mathcal{D}^{t+1}$, $\mathcal{Z}^{t+1}$, and $\mathcal{E}_2^{t+1}$ are the minimizers of (4.8), (4.10), (4.13), and (4.15), respectively, we can deduce that

$$\begin{aligned} &F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \Delta_{\mathcal{E}_1} \le F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right), \\ &F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^t, \mathcal{E}_2^t\right) + \Delta_{\mathcal{D}} \le F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t\right), \\ &F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^t\right) + \Delta_{\mathcal{Z}}+ \le F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^t, \mathcal{E}_2^t\right), \\ &F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^{t+1}\right) + \Delta_{\mathcal{E}_2} \le F\left(\mathcal{C}^{t+1}, B^{t+1}, \mathcal{E}_1^{t+1}, \mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^t\right), \end{aligned} \tag{4.21}$$

where $\Delta_{\mathcal{E}_1} = \frac{\rho_3}{2}\|\mathcal{E}_1^{t+1} - \mathcal{E}_1^t\|^2$, $\Delta_{\mathcal{D}} = \frac{\rho_4}{2}\|\mathcal{D}^{t+1} - \mathcal{D}^t\|^2$, $\Delta_{\mathcal{Z}} = \frac{\rho_5}{2}\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\|^2$, and $\Delta_{\mathcal{E}_2} = \frac{\rho_6}{2}\|\mathcal{E}_2^{t+1} - \mathcal{E}_2^t\|^2$. Combining (4.19), (4.20), and (4.21), we have

$$F\left(\mathcal{W}^{t+1}\right) + \rho\|\mathcal{W}^{t+1} - \mathcal{W}^t\|^2 \le F\left(\mathcal{W}^t\right), \tag{4.22}$$

where $\rho = \min_{i\in[6]}\{\rho_i/2\}$. Thus, we complete the proof of statement (1).

(2) Summing up (4.22) over $t = 0, 1, \ldots, \infty$, it gives

$$\rho\sum_{t=0}^{\infty}\|\mathcal{W}^{t+1} - \mathcal{W}^t\|^2 \le F\left(\mathcal{W}^0\right) - F\left(\mathcal{W}^\infty\right) < \infty, \tag{4.23}$$

where the last inequality uses $F\left(\mathcal{W}^\infty\right) \ge 0$. Thus, $\lim_{t\to\infty}\|\mathcal{W}^{t+1} - \mathcal{W}^t\| = 0$.

(3) According to (4.22), we have

$$F\left(\mathcal{W}^{t+1}\right) \leq F\left(\mathcal{W}^{t}\right) \leq \cdots \leq F\left(\mathcal{W}^{0}\right) < \infty.$$

Thus, $F(\mathcal{W}^t)$ is bounded. Combining this with

$$\begin{aligned} F\left(\mathcal{W}^{t}\right) = {} & \frac{\lambda_1}{2}\|B^t\|^2 + \lambda_2\|\mathcal{E}_1^t\|_{11\phi} + \frac{\lambda_3}{2}\|\mathcal{H} - \mathcal{C}^t \times_3 B^t - \mathcal{E}_1^t\|^2 + \lambda_4\|\mathcal{Z}^t\|_{F,1}^{\psi} \\ & + \lambda_5\|\mathcal{E}_2^t\|_{11\phi} + \frac{\lambda_6}{2}\|\mathcal{C}^t - \mathcal{D}^t * \mathcal{Z}^{t^T} - \mathcal{E}_2^t\|^2 + \delta_{\mathbb{B}}\left(B^t\right) + \delta_{\mathbb{C}}\left(\mathcal{C}^t\right) + \delta_{\mathbb{D}}\left(\mathcal{D}^t\right), \end{aligned} \tag{4.24}$$

we can deduce that each term in (4.24) is bounded. From the boundedness of $\|B^t\|$, $\delta_{\mathbb{C}}(\mathcal{C}^t)$ and $\delta_{\mathbb{D}}(\mathcal{D}^t)$, we can deduce that B^t, $\mathcal{C}^t$, and $\mathcal{D}^t$ are bounded. Combining this with the boundedness of $\|\mathcal{H} - \mathcal{C}^t \times_3 B^t - \mathcal{E}_1^t\|$ and $\|\mathcal{C}^t - \mathcal{D}^t * \mathcal{Z}^{t^T} - \mathcal{E}_2^t\|$ and the fact that either $\mathcal{Z}^t$ or $\mathcal{E}_2^t$ is bounded, we obtain that $\mathcal{Z}^t$, $\mathcal{E}_1^t$, and $\mathcal{E}_2^t$ are bounded. Therefore, we can conclude that $\mathcal{W}^t$ is bounded. ■

Lemma 4.3 (relative error condition). *Let the sequence $\{\mathcal{W}^t\}$ be generated by Algorithm 4.1. Suppose that either $\mathcal{Z}^t$ or $\mathcal{E}_2^t$ is bounded. Then, there exists $\varpi > 0$ such that $\|V^{t+1}\| \leq \varpi\|\mathcal{W}^{t+1} - \mathcal{W}^t\|$ for any $V^{t+1} \in \partial F(\mathcal{W}^{t+1})$.*

Proof. By the first order necessary optimality conditions in (4.4), (4.6), (4.8), (4.10), (4.13), and (4.15), we have

$$\begin{cases} 0 \in \partial\delta_{\mathbb{C}}(\mathcal{C}^{t+1}) + \nabla_{\mathcal{C}}G_1(\mathcal{C}^t, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t) + (l_{\mathcal{C}}^t + \rho_1)(\mathcal{C}^{t+1} - \mathcal{C}^t); \\ 0 \in \partial\delta_{\mathbb{B}}(\mathcal{B}^{t+1}) + \nabla_{\mathcal{B}}G_1(\mathcal{C}^{t+1}, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t) + (l_{\mathcal{B}}^t + \rho_2)(\mathcal{B}^{t+1} - B^t); \\ 0 \in \lambda_2\partial\|\mathcal{E}_1^{t+1}\|_{11\phi} + \lambda_3(-\mathcal{H} + \mathcal{C}^{t+1} \times_3 B^{t+1} + \mathcal{E}_1^{t+1}) + \rho_3(\mathcal{E}_1^{t+1} - \mathcal{E}_1^t); \\ 0 \in \partial\delta_{\mathbb{D}}(\mathcal{D}^{t+1}) + \nabla_{\mathcal{D}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^t, \mathcal{E}_2^t) + \rho_4(\mathcal{D}^{t+1} - \mathcal{D}^t); \\ 0 \in \lambda_4\partial\|\mathcal{Z}^{t+1}\|_{F,1}^{\psi} + \nabla_{\mathcal{Z}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^t) + \rho_5(\mathcal{Z}^{t+1} - \mathcal{Z}^t); \\ 0 \in \lambda_5\partial\|\mathcal{E}_2^{t+1}\|_{11\phi} + \lambda_6(-\mathcal{C}^{t+1} + \mathcal{D}^{t+1} * \mathcal{Z}^{{t+1}^T} + \mathcal{E}_2^{t+1}) + \rho_6(\mathcal{E}_2^{t+1} - \mathcal{E}_2^t), \end{cases} \tag{4.25}$$

where $G_1(\mathcal{C}, B, \mathcal{E}_1, \mathcal{D}, \mathcal{Z}, \mathcal{E}_2) = \frac{\lambda_1}{2}\|B\|^2 + \frac{\lambda_3}{2}\|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1\|^2 + \frac{\lambda_6}{2}\|\mathcal{C} - \mathcal{D} * \mathcal{Z}^T - \mathcal{E}_2\|^2$ and $G_2(\mathcal{D}, \mathcal{Z}, \mathcal{E}_2) = \frac{\lambda_6}{2}\|\mathcal{C}^{t+1} - \mathcal{D} * \mathcal{Z}^T - \mathcal{E}_2\|^2$. We denote V_i^{t+1} for $i \in [6]$ as follows:

$$\begin{cases} V_1^{t+1} = \nabla_{\mathcal{C}}G_1(\mathcal{W}^{t+1}) - \nabla_{\mathcal{C}}G_1(\mathcal{W}^t) - (l_{\mathcal{C}}^t + \rho_1)(\mathcal{C}^{t+1} - \mathcal{C}^t); \\ V_2^{t+1} = \nabla_{\mathcal{B}}G_1(\mathcal{W}^{t+1}) - \nabla_{\mathcal{B}}G_1(\mathcal{C}^{t+1}, B^t, \mathcal{E}_1^t, \mathcal{D}^t, \mathcal{Z}^t, \mathcal{E}_2^t) - (l_{\mathcal{B}}^t + \rho_2)(\mathcal{B}^{t+1} - B^t); \\ V_3^{t+1} = -\rho_3(\mathcal{E}_1^{t+1} - \mathcal{E}_1^t); \\ V_4^{t+1} = \nabla_{\mathcal{D}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^{t+1}) - \nabla_{\mathcal{D}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^t, \mathcal{E}_2^t) - \rho_4(\mathcal{D}^{t+1} - \mathcal{D}^t); \\ V_5^{t+1} = \nabla_{\mathcal{Z}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^{t+1}) - \nabla_{\mathcal{Z}}G_2(\mathcal{D}^{t+1}, \mathcal{Z}^{t+1}, \mathcal{E}_2^t) - \rho_5(\mathcal{Z}^{t+1} - \mathcal{Z}^t); \\ V_6^{t+1} = -\rho_6(\mathcal{E}_2^{t+1} - \mathcal{E}_2^t). \end{cases} \tag{4.26}$$

Combining (4.25) and (4.26) and recalling the definition of $F(\mathcal{W})$, one has

$$V^{t+1} = \left(V_1^{t+1}, V_2^{t+1}, V_3^{t+1}, V_4^{t+1}, V_5^{t+1}, V_6^{t+1}\right) \in \partial F\left(\mathcal{W}^{t+1}\right).$$

It is evident that G_1 and G_2 are Lipschitz continuous on any bounded set. From Lemma 4.2(3), we know that $\mathcal{W}^t$ is bounded. Hence, we obtain

$$\left\|V^{t+1}\right\| \leq \varpi\left\|\mathcal{W}^{t+1} - \mathcal{W}^t\right\|, \tag{4.27}$$

where $\varpi = \max_{i\in[6]}\{\rho_i + l_g\}$, and l_g is the maximum Lipschitz constant of G_1 and G_2. ■

Lemma 4.4. *The function $F(\mathcal{W})$ is a Kurdyka–Łojasiewicz (KL) function.*

Proof. According to [5], the Frobenius norm functions $\frac{\lambda_1}{2}\|B\|^2$, $\frac{\lambda_3}{2}\|\mathcal{H} - \mathcal{C} \times_3 B - \mathcal{E}_1\|^2$, and $\frac{\lambda_6}{2}\|\mathcal{C} - \mathcal{D} * \mathcal{Z}^T - \mathcal{E}_2\|^2$ are semialgebraic functions. The terms $\lambda_2\|\mathcal{E}_1\|_{11\phi}$, $\lambda_5\|\mathcal{E}_2\|_{11\phi}$, and $\lambda_4\|\mathcal{Z}\|_{F,1}^{\psi}$ are also semialgebraic functions because they are composed of semialgebraic operations, including the Frobenius norm, minimum, and finite summation. The functions $\delta_{\mathbb{B}}(B)$, $\delta_{\mathbb{C}}(\mathcal{C})$, and $\delta_{\mathbb{D}}(\mathcal{D})$ are semialgebraic because they are indicator functions with semialgebraic sets [5]. Hence, $F(\mathcal{W})$ is semialgebraic because it is a finite sum of semialgebraic functions. Additionally, since $F(\mathcal{W})$ is a proper continuous function, it follows from [5, Theorem 3] that $F(\mathcal{W})$ is a KL function. This completes the proof. ■

Finally, we provide a theoretical guarantee for the convergence of Algorithm 4.1.

Theorem 4.5. *Consider the sequence $\{\mathcal{W}^t\}$ obtained by Algorithm 4.1. Assuming that either $\mathcal{Z}^t$ or $\mathcal{E}_2^t$ is bounded, the sequence $\{\mathcal{W}^t\}$ converges to a critical point of $F(\mathcal{W})$.*

Proof. We begin by noting that Lemma 4.2(3) establishes the boundedness of the sequence $\{\mathcal{W}^t\}$. Given this boundedness, the Bolzano–Weierstrass theorem ensures the existence of a convergent subsequence. Furthermore, by exploiting the continuity of $F(\mathcal{W})$, along with the results derived from Lemmas 4.2, 4.3, and 4.4, we can rigorously establish the desired conclusion as articulated in [2, Theorem 2.9]. ■

4.3. Rank reduction strategy with validation mechanism.

In this subsection, we propose a rank reduction strategy to decrease r, which is related to the dimensions of $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times b}$ and $\mathcal{Z} \in \mathbb{R}^{n_2 \times r \times b}$, thereby reducing the complexity of Algorithm 4.1. Unlike most decomposition methods that set a small initial r^0, we set $r^0 = \min\{n_1, n_2\}$ and leverage the algorithm's inherent capability to adaptively reduce r^t. This approach is initially supported by Theorem 3.2. Furthermore, we present the following theorem to substantiate our proposed rank reduction strategy.

Theorem 4.6. *There exists $t^{\#} \in \mathbb{N}$ such that $\Gamma(\mathcal{Z}^t) = \Gamma(\mathcal{Z}^{t+1})$ for each $t \geq t^{\#}$, where $\Gamma(\mathcal{Z}) = \{j \mid \|\mathcal{Z}(:,j,:)\| \neq 0, j = 1, \ldots, r\}$.*

Proof. By Lemma 4.2, which states that $\lim_{t\to\infty} \|\mathcal{Z}^{t+1} - \mathcal{Z}^t\| = 0$, there exists $t^{\#} \in \mathbb{N}$ such that $\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\| < \min\{(2\hat{\lambda}_4(1-p))^{1/(2-p)}, \nu\}$ for each $t \geq t^{\#}$. Proving by contradiction, we assume that there exists $t \geq t^{\#}$ such that $\Gamma(\mathcal{Z}^t) \neq \Gamma(\mathcal{Z}^{t+1})$. Then there exists $j \in r$ such that either (a) $\mathcal{Z}^t(:,j,:) \neq 0$ and $\mathcal{Z}^{t+1}(:,j,:) = 0$, or (b) $\mathcal{Z}^t(:,j,:) = 0$ and $\mathcal{Z}^{t+1}(:,j,:) \neq 0$. Hence by (4.14), we have for both cases that

$$\left\|\mathcal{Z}^{t+1} - \mathcal{Z}^t\right\| \geq \left\|\mathcal{Z}^{t+1}(:,j,:) - \mathcal{Z}^t(:,j,:)\right\| \geq \min\{(2\hat{\lambda}_4(1-p))^{1/(2-p)}, \nu\},$$

which yields a contradiction. The proof is complete. ■

Based on the aforementioned theorem, when $t \geq t^{\#}$, we can remove the zero lateral slices of $\mathcal{Z}^t$ and the corresponding lateral slices in $\mathcal{D}^t$. Consequently, the dimensions of $\mathcal{D}^t$ and $\mathcal{Z}^t$ can be reduced to $\mathbb{R}^{n_1 \times |\Gamma(\mathcal{Z}^{\#})| \times b}$ and $\mathbb{R}^{n_2 \times |\Gamma(\mathcal{Z}^{\#})| \times b}$, respectively, when $t \geq t^{\#}$. However, when $t < t^{\#}$, the computational cost remains high due to the large value of $\min\{n_1, n_2\}$, and $t^{\#}$ is uncertain and potentially very large.

Therefore, we initiate the reduction of r from $t = 1$, as illustrated in step 15 of Algorithm 4.2. However, this approach introduces the following potential issue: when

Algorithm 4.2. Rank reduction with validation mechanism strategy for Algorithm 4.1.

```
Input: HSI data $\mathcal{H}$, parameters $\{\lambda_i\}_{i=1}^6$ and $\{\rho_i\}_{i=1}^6$. Set $\mathbb{S}=\emptyset$ and $t \leftarrow 0$.
9  Update $\mathcal{C}^{t+1}$, $B^{t+1}$, $\mathcal{E}_1^{t+1}$, $\mathcal{D}^{t+1}$, $\mathcal{Z}^{t+1}$, and $\mathcal{E}_2^{t+1}$ using steps 1–6 of Algorithm 4.1;
10 if $\mathbb{S} \neq \emptyset$ then
11     Compute $\mathcal{Z}_{sub}^{t+1} = \text{Prox}_{\hat{\lambda}_4}\psi(\lambda_6(\mathcal{C}^{t+1}-\mathcal{E}_2^{t+1})^T * \mathcal{D}_{sub}^t(:,\mathbb{S},:)/(\lambda_6+\rho_5))$;
12     Identify the top 5 lateral slices of $\mathcal{Z}_{sub}^{t+1}$ with the highest Frobenius norms, or all if fewer than 5, and denote this set as $\mathbb{S}_1$;
13     Let $\mathcal{Z}^{t+1} = [\mathcal{Z}^{t+1}, \mathcal{Z}_{sub}^{t+1}(:,\mathbb{S}_1,:)]$ and $\mathcal{D}^{t+1} = [\mathcal{D}^{t+1}, \mathcal{D}_{sub}^t(:,\mathbb{S}_1,:)]$;
14 end
15 Let $\mathbb{S} = \{j \mid \|\mathcal{Z}^{t+1}(:,j,:)\| = 0\}$, and define $\mathcal{D}_{sub}^{t+1} = \mathcal{D}^{t+1}(:,\mathbb{S},:)$. Then, remove the lateral slices at positions in $\mathbb{S}$ from $\mathcal{D}^{t+1}$ and $\mathcal{Z}^{t+1}$;
16 If a termination criterion is met, set $\mathcal{E}_1^* := \mathcal{E}_1^{t+1}$, $\mathcal{E}_2^* := \mathcal{E}_2^{t+1}$; else, set $t \leftarrow t+1$, return to step 9;
17 Compute the detection map $T$ by (3.5).
Output: Detection map $T$.
```

$\mathcal{Z}^{t+1}(:,j,:)=0$, it does not necessarily imply that $\mathcal{Z}^{t+2}(:,j,:)=0$ for $t+1<t^{\#}$. Consequently, removing $\mathcal{Z}^{t+1}(:,j,:)$ may be unjustified. To address this, we incorporate a verification step in steps 10–14 of Algorithm 4.2 to assess the validity of removing $\mathcal{Z}^{t+1}(:,j,:)$. If $\mathcal{Z}_{sub}^{t+1}(:,j,:)=0$, we deem it reasonable to remove $\mathcal{Z}^{t+1}(:,j,:)$; conversely, if $\mathcal{Z}_{sub}^{t+1}(:,j,:)\neq 0$, we consider it unreasonable to remove $\mathcal{Z}^{t+1}(:,j,:)$ and reintegrate $\mathcal{Z}_{sub}^{t+1}(:,j,:)$ back into $\mathcal{Z}^{t+1}$. This procedure is detailed in Algorithm 4.2.

Remark 4.7. In Algorithm 4.2, we see the following:

(1) Step 15 significantly reduces the value of r, thereby decreasing the algorithm's complexity. Meanwhile, steps 10–14 ensure optimal solution quality by providing a mechanism to correct any erroneous reductions in r. This interplay between the steps enhances the algorithm's efficiency while maintaining high-quality results.

(2) In step 15, if $|\mathbb{S}| = size(\mathcal{Z}^{t+1},2)$, executing this step would result in $\mathcal{D}^{t+1}$ and $\mathcal{Z}^{t+1}$ becoming empty. We consider this to be unreasonable, and therefore this step is not executed in such cases.

(3) In step 12, to prevent r from increasing too rapidly, we impose a limited maximum increase of 5 per step.

4.4. Complexity analysis. Here we analyze the computational complexity of Algorithm 4.2. At each iteration, updating $\mathcal{C}$ involves performing the tensor-matrix product and t-product, with a computational complexity of $\mathcal{O}(n_1 n_2(n_3+r^t)b)$ when the iterates $\mathcal{D}^t$ and $\mathcal{Z}^t$ contain r^t nonzero lateral slices. For updating B and $\mathcal{E}_1$, the primary computational cost is associated with the matrix (tensor-matrix) product, whose complexity is $\mathcal{O}(n_1 n_2 n_3 b)$. Updating $\mathcal{D}$ requires a computational cost of $\mathcal{O}((n_1+n_2+\log b)n_1 r^t b)$ to perform the (inverse) discrete Fourier transform, SVD, and matrix product calculations. The computational cost for updating $\mathcal{Z}$ and $\mathcal{E}_2$ is $\mathcal{O}(n_1 n_2 r^t b + n_1(n_2+r^t)b\log b)$ and $\mathcal{O}(n_1 n_2 r^t b + (n_1+n_2)r^t b \log b)$, respectively. In summary, the overall computational complexity of our proposed method is $\mathcal{O}(n_1 n_2(n_3 + r^t)b)$, assuming that $n_1 \approx n_2$ in most cases. It is important to note that r^t decreases with each iteration and eventually becomes much smaller than $\min\{n_1, n_2\}$. The computational complexity of a single SVD operation for the input HSI $\mathcal{H}$ is $\mathcal{O}(n_1n_2n_3\min\{n_1n_2, n_3\})$ in matrix based methods. Therefore, the proposed method offers a lower computational cost compared to other low rank approximation based HAD methods.

5. Experiments.

In this section, we present a series of experiments designed to assess the performance of the proposed algorithm. These experiments were conducted on a server equipped with 16 logical CPU cores and 16 GB of memory. All algorithms were implemented using MATLAB 2022a, and no preprocessing was applied to maintain fairness.

Stopping criteria: For the iterative procedures in Algorithms 4.1 and 4.2, we employ the following absolute error condition as the termination criterion:

$$\|w^{t+1} - w^t\| < 1e\text{-}2.$$

Here, w denotes the concatenated vector of all model variables, defined as $w := [\mathcal{C}(:); B(:); \mathcal{E}_1(:); \mathcal{D}(:); \mathcal{Z}(:); \mathcal{E}_2(:)]$. This criterion is supported by our convergence analysis, which demonstrates that $\|w^{t+1} - w^t\|$ tends to zero and provides a bound on the subgradient norm, thereby serving as a reliable indicator of convergence to a stationary point.

Parameter settings: The final hyperparameter values used to generate our main results are set as follows: $\rho_i = 1e\text{-}2, i \in [6]$, $b \in [2,6]$, $\lambda_1 = 1e\text{-}2$, $\lambda_2 = 5$, $\lambda_3 \in \{1e\text{-}2, 1e\text{-}1, 5e\text{-}1, 1\}$, $\lambda_4 = 5e\text{-}1$, $\lambda_5 = 1e\text{-}1$, $\lambda_6 = \lambda_3/10$. A detailed sensitivity analysis justifying these choices is presented in subsection 5.2.1.

For comparison, we employed several established methods in our experiments. These included RX [35], which is statistics based; TD-PEPTE [42], which is pixel-level topology entropy based; RPCA [41], LRASR [49], and SuperRPCA [29], which are matrix based; PTA [27], TPCA [8], and TLRSR [47], which are tensor based; and RGAE [9] and GAED [48], which are deep learning based. To evaluate detection performance, we utilized three widely recognized metrics in addition to visually observing the following resulting anomaly maps: the receiver operating characteristic (ROC) curve [19], the area under the ROC curve (AUC) value [20], and the separability map between anomalies and the background.

5.1. Datasets description.

In this section, the proposed method is evaluated on two distinct benchmarks: the Airport-Beach-Urban (ABU) hyperspectral database and the MVTec natural image database. The latter, although not hyperspectral, is included to demonstrate the generality of our framework as well as its broader applicability beyond the HSI domain. Details of each dataset are provided below.

(1) Airport-Beach-Urban dataset:[1] This dataset includes three distinct scenes: Airport, Beach, and Urban. Each of the Airport and Beach scenes contain four images, while the Urban scene includes five images. All images were captured using the Airborne Visible/Infrared Imaging Spectrometer sensor, producing images with approximately 200 spectral bands, except for one Beach scene, which was captured by the Reflective Optics System Imaging Spectrometer sensor, producing images with approximately 100 spectral bands. Detailed acquisition procedures are described in [18]. The sample images, sized at 100 × 100 or 150 × 150 pixels, include corresponding reference maps. These reference maps were manually labeled using the Environment for Visualizing

---

[1] http://xudongkang.weebly.com/data-sets.html.

Images software. Due to space limitations, we selected only two images from the Airport and Urban scene datasets for the experiments, as illustrated in Figure 4, which shows the selected image scenes and their corresponding ground-truth maps.

(2) MVTec dataset:[2] This dataset [4], developed by MVTec Software GmbH, is specifically designed for anomaly detection in a real industrial scenario. This dataset encompasses 15 different industrial products, which are categorized into texture and object classes. The texture class includes five categories: carpet, grid, leather, tile, and wood. The object class comprises ten categories: bottle, cable, capsule, hazelnut, metal nut, pill, screw, toothbrush, transistor, and zipper. Each product category contains multiple defect types, resulting in a total of 73 distinct defects within the dataset. The resolution of the images ranges from 700 × 700 pixels to 1024 × 1024 pixels. Due to space constraints, we only selected images of hazelnuts exhibiting all defect types (crack, cut, hole, and print) for our experiments and resized them to 512 × 512 pixels. Figure 5 presents the selected images along with their corresponding ground-truth maps.

5.2. Discussion.

5.2.1. Parameter analysis. For the proposed model LTD, three main types of parameters—the proximal parameters $\rho_i, i \in [6]$, the parameter b, and the regularization parameters $\lambda_i, i \in [6]$—affect the performance. Following [26], we set the proximal parameters $\rho_i = 1e\text{-}2, i \in [6]$ in all experiments. For the remaining parameters, we conducted a sensitivity analysis by varying one parameter at a time while holding the others fixed at a robust baseline configuration. The baseline configuration is as follows: $b \in [2,6]$, $\lambda_1 = 1e\text{-}2$, $\lambda_2 = 5$, $\lambda_3 \in \{1e\text{-}2, 1e\text{-}1, 5e\text{-}1, 1\}$, $\lambda_4 = 5e\text{-}1$, $\lambda_5 = 1e\text{-}1$, $\lambda_6 = \lambda_3/10$. The final settings derived from this analysis are detailed below.

First, we analyze the impact of the parameter b on the algorithm's performance and efficiency. This parameter controls the dimensions of tensor $\mathcal{C}$ and matrix B. We explored

(a) Airport1. (b) Airport2. (c) Urban1. (d) Urban2.

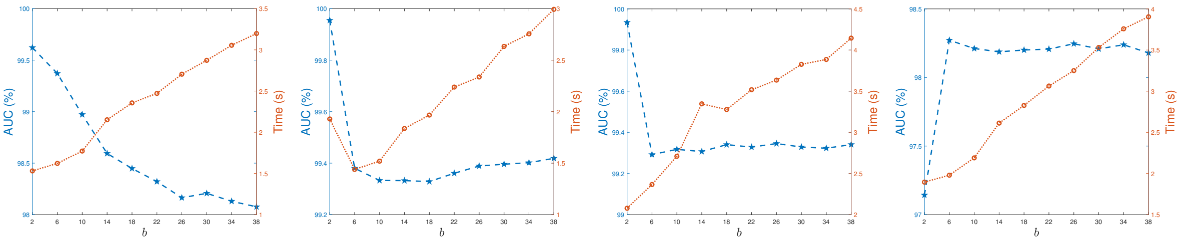
Figure 4. *Pseudocolor images and ground-truth maps of ABU dataset.*

(a) Crack. (b) Cut. (c) Hole. (d) Print.

Figure 5. *Color images and ground-truth maps of MVTec dataset.*

---

[2] https://www.mvtec.com/company/research/datasets/mvtec-ad.



(a) Airport1. (b) Airport2. (c) Urban1. (d) Urban2.

Figure 6. *AUC values (%) and corresponding running time(s) of LTD with different b for ABU dataset.*

values of b ranging from 2 to 38, with increments of 4. Figure 6 illustrates the AUC values and corresponding running times of LTD for these different values of b. As shown in the figure, the AUC values reach their optimum at very small values of b. As b increases, the AUC values tend to stabilize. Regarding running times, as b increases, the sizes of tensor $\mathcal{C}$ and matrix B also increase, resulting in a gradual increase in running times in most cases. Therefore, in our experiments, we set b within the range of $[2, 6]$ for all cases.

Next, we investigate the sensitivity of the model to each of the six regularization parameters $\lambda_i, i \in [6]$. From Figure 7, we observe that LTD is not sensitive to the parameter λ_1 when it is less than or equal to 1. Its primary function is to prevent B from becoming too large; therefore, we set $\lambda_1 = 1e\text{-}2$ in our experiments. For λ_2, we set it to 5, and for λ_5, we set it to $1e$-1, as these values yielded optimal results across all datasets. Regarding λ_3, we select the best values from $\{1e\text{-}2, 1e\text{-}1, 5e\text{-}1, 1\}$ based on the AUC values for different datasets and then fine-tuned them to achieve the optimal results. With respect to λ_4, we found that LTD's performance is not significantly affected when it is set to $1e$-1 or higher. Therefore, we set $\lambda_4 = 5e\text{-}1$ in our experiments. For the ABU dataset, we observed that the optimal AUC values for λ_3 and λ_6 exhibit a multiplicative relationship, with λ_6 being $\lambda_3/10$. Therefore, we set $\lambda_6 = \lambda_3/10$. Similarly, for the MVTec dataset, we set $\lambda_6 = \lambda_3/100$ to achieve the optimal AUC values.

5.2.2. Effects of rank reduction strategy with validation mechanism. In our model, the large tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_2 \times b}$ is decomposed into two smaller tensors, $\mathcal{D} \in \mathbb{R}^{n_1 \times r \times b}$ and $\mathcal{Z} \in \mathbb{R}^{n_2 \times r \times b}$, utilizing the t-product, with the dimensions of $\mathcal{D}$ and $\mathcal{Z}$ determined by the parameter r. By leveraging the principle of group sparsity, we propose the rank reduction strategy with validation mechanism in Algorithm 4.2 to adaptively adjust the value of r.

Initially, we depict the variation in r values as a function of iteration count for the MVTec dataset in Figure 8. From Figure 8, it is evident that r significantly decreases during the initial stages of the algorithm, iterating at a lower value to effectively find a suboptimal solution. However, as iterations progress, the smaller r becomes insufficient to achieve optimal results. Consequently, the algorithm dynamically adjusts r, incrementally restoring some of the previously reduced values to compensate for the excessive reduction. This adjustment gradually increases r and eventually stabilizes in the later stages, consistent with the description in Theorem 4.6. This process demonstrates the algorithm's effectiveness in adaptively adjusting r, thereby achieving a good balance between computational efficiency and decomposition accuracy.

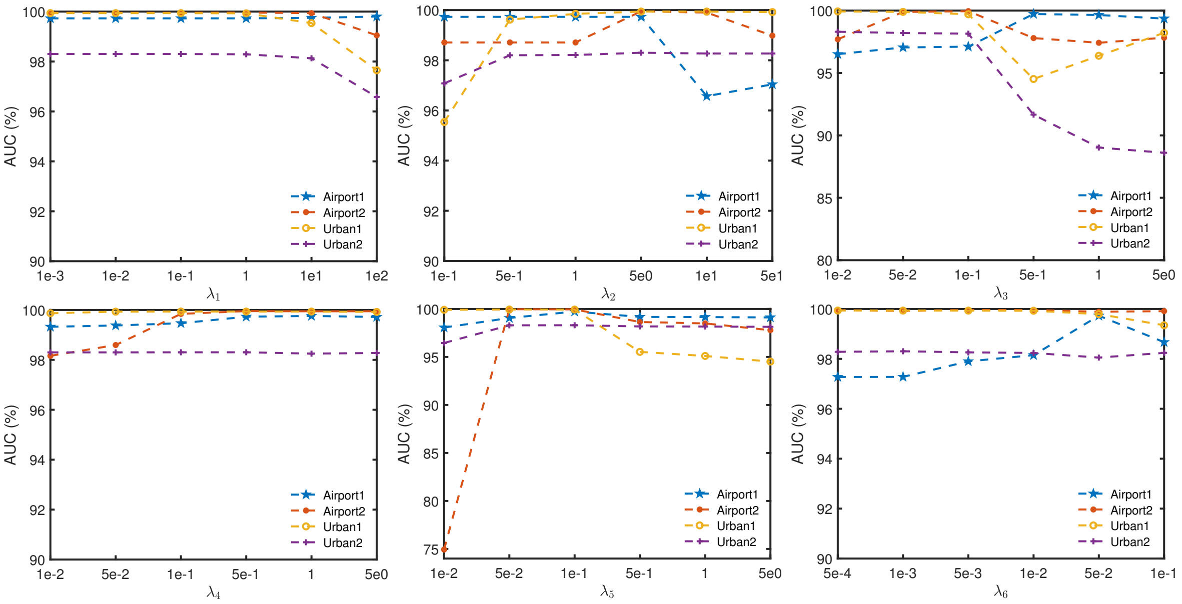


Figure 7. *AUC values (%) of LTD with different λ_i $(i \in [6])$ for ABU dataset.*

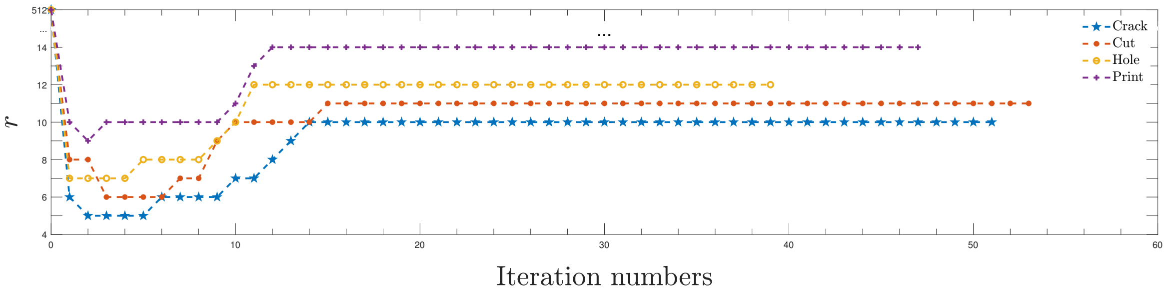
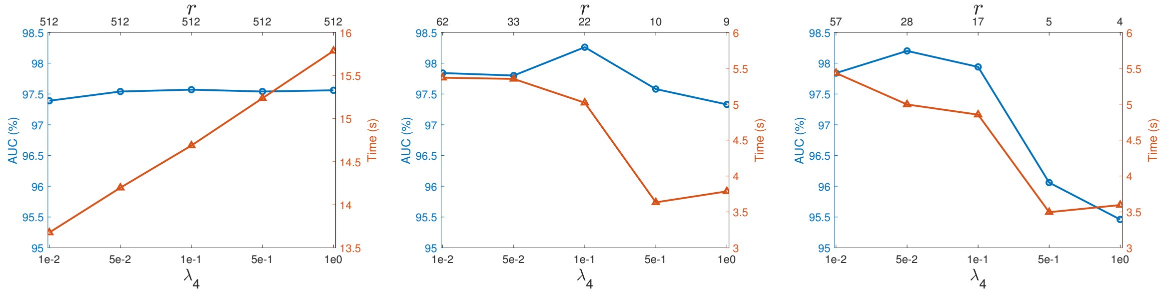


Figure 8. *Variation of r values across iterations for Algorithm 4.2.*



(a) (b) (c)

Figure 9. *Performance analysis of* (a) *Algorithm* 4.1, (b) *Algorithm* 4.2, *and* (c) *Algorithm* 4.2 *without validation mechanism on AUC values (%), running time(s), and r metrics across various values of λ_4 on the Crack dataset.*

Table 2

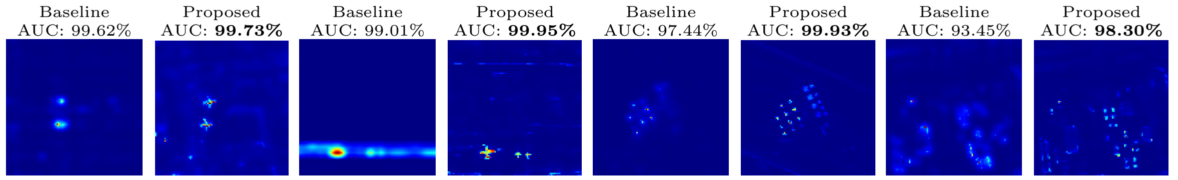
Comparison of AUC values (%) of different anomaly detection maps for ABU dataset.

	Airport1	Airport2	Urban1	Urban2
T_1	42.97	92.83	99.06	94.93
T_2	96.82	99.91	99.63	94.30
$T_1 \circ T_2$	97.58	99.95	99.92	98.29
T	99.73	99.95	99.93	98.30

Moreover, Figure 9 presents a detailed evaluation of AUC, computational time, and r metrics for Algorithm 4.1, Algorithm 4.2, and Algorithm 4.2 without validation mechanism (i.e., excluding steps 10–14) across a range of λ_4 values. From the detection results, the AUC value produced by Algorithm 4.2 without validation mechanism exhibits significant sensitivity to the parameter λ_4. This sensitivity arises because an excessively large λ_4 results in an overly small r, leading to suboptimal performance. This issue is prevalent among most algorithms that employ rank reduction operations. In contrast, our proposed Algorithm 4.2 incorporates a validation mechanism within its rank reduction strategy. This mechanism evaluates the reasonableness of the reduction in r. By leveraging this validation mechanism, it mitigates the issue of an excessively large λ_4, causing r to become too small, thereby achieving more stable results and reducing sensitivity to λ_4. Algorithm 4.1, which lacks a rank reduction operation, also demonstrates minimal sensitivity to the parameter λ_4. Regarding time consumption, Algorithm 4.1, due to the absence of a rank reduction operation, maintains r at a relatively large value of 512, resulting in substantial time consumption. Conversely, both Algorithm 4.2 and Algorithm 4.2 without validation mechanism incorporate a rank reduction operation, which keeps r at a smaller value, thereby significantly reducing time consumption compared to Algorithm 4.1. In summary, our proposed Algorithm 4.2, which integrates a validation mechanism within its rank reduction strategy, not only achieves high AUC values and exhibits reduced sensitivity to the parameter λ_4 but also significantly lowers time consumption.

5.2.3. Effects of spectral-spatial fusion. Figure 3 demonstrates that the anomalies in T_1 and T_2 exhibit complementary characteristics, suggesting that their fusion can produce a more effective anomaly detection map. To more intuitively illustrate the efficacy of fusing spectral anomaly detection map T_1 and spatial anomaly detection map T_2, Table 2 presents the AUC values for various anomaly detection maps across the ABU dataset. The table indicates that, irrespective of the dataset, the AUC values for the spectral-spatial fusion anomaly detection map $T_1 \circ T_2$ consistently surpass those of the spectral anomaly detection map T_1 and the spatial anomaly detection map T_2. Moreover, the results T derived from the guided image filter further enhance the performance of $T_1 \circ T_2$.

5.2.4. Effects of the unified framework. To validate the advantages of our proposed unified LTD framework, we conducted an ablation study, comparing it against a decoupled, two-stage baseline. This baseline first applies layer 1 for spectral feature extraction and then uses its resulting coefficient tensor as input to a subsequent layer 2 stage for spatial anomaly detection. The visual results and corresponding AUC values are presented in Figure 10.



(a) Airport1. (b) Airport2. (c) Urban1. (d) Urban2.

Figure 10. *Target detection results for ABU dataset. For each scene, the left map is the result of the baseline, and the right map is from our proposed (LTD) method. The corresponding AUC value is displayed above each map.*

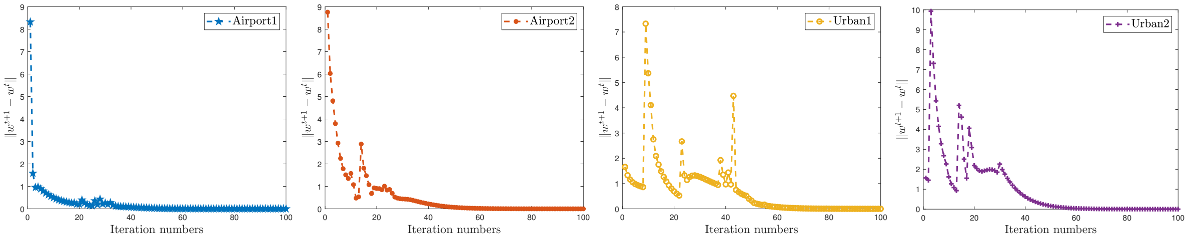


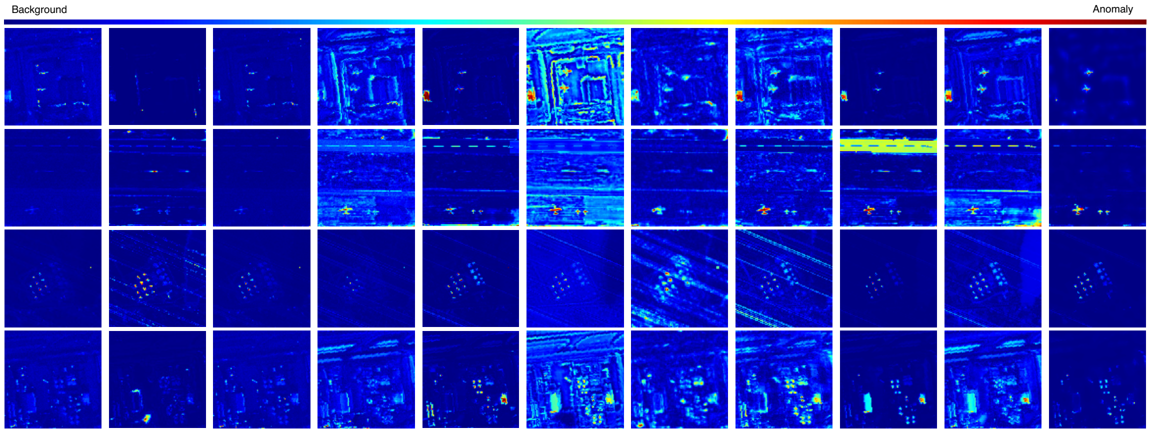
Figure 11. *Absolute changes in* $w := [\mathcal{C}(:); B(:); \mathcal{E}_1(:); \mathcal{D}(:); \mathcal{Z}(:); \mathcal{E}_2(:)]$ *over iterations across four different scenes from the ABU dataset.*

The results presented in Figure 10 provide comprehensive evidence of the effectiveness of the proposed unified framework. On the AUC metric, our LTD method consistently outperforms the baseline across all four scenes, with the most notable improvement observed in the Urban2 scene (98.30% vs. 93.45%). This numerical superiority is visually corroborated by the detection maps. The baseline method's lower AUC values can be attributed to two distinct failure modes: in the Airport2 scene, it suffers from severe false alarms, while in the other scenes, it exhibits poor sensitivity, failing to detect the true anomalies with sufficient intensity. In stark contrast, our proposed LTD method produces a much cleaner background while simultaneously highlighting the true anomalies with high contrast and integral shapes. This clear improvement in both background suppression and target enhancement explains the superior AUC performance and validates the effectiveness of our joint optimization strategy.

5.2.5. Convergence analysis. Figure 11 illustrates the convergence behavior of the proposed Algorithm 4.2 on four scenes from the ABU dataset. As the figure demonstrates, the absolute changes across all experiments decrease rapidly and stabilize near zero, typically within 60 iterations. The temporary oscillations observed in the curves correspond to the moments where our adaptive rank reduction strategy adjusts the tensor sizes. Following each adjustment, the algorithm quickly resumes its downward convergence trend. This analysis empirically verifies the robust convergence of the proposed algorithm.

5.3. Detection performance.

5.3.1. ABU dataset. For the ABU dataset, the reference map and detection results are shown in Figures 4 and 12. The proposed method, LTD, effectively detects all anomalies while



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Figure 12. *Target detection results by different methods for ABU dataset.* (a) *RX.* (b) *TD-PEPTE.* (c) *RPCA.* (d) *LRASR.* (e) *SuperRPCA.* (f) *PTA.* (g) *TPCA.* (h) *TLRSR.* (i) *RGAE.* (j) *GAED.* (k) *LTD.*

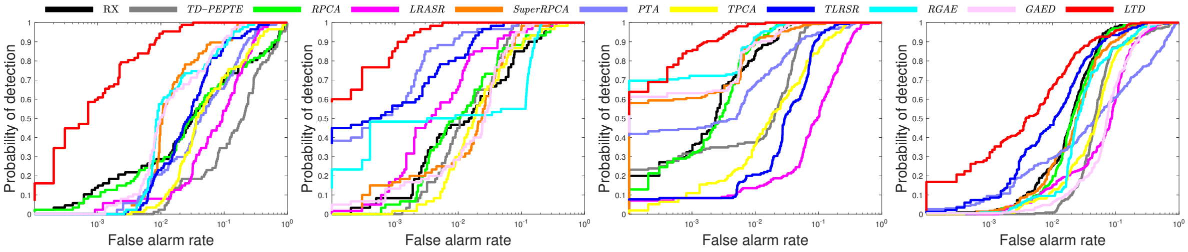
Table 3

Comparison of AUC values (%) and running time(s) of different methods for ABU dataset.

Algorithm	Index	RX	TD-PEPTE	RPCA	LRASR	SuperRPCA	PTA	TPCA	TLRSR	RGAE	GAED	LTD
Airport1	AUC	84.04	73.45	84.28	87.70	97.17	90.96	88.11	94.58	96.40	96.77	**99.73**
	Time	0.05	7.97	4.62	20.06	8.97	15.77	18.63	2.50	51.88	41.10	1.66
Airport2	AUC	95.26	97.69	96.27	97.95	96.71	99.55	95.26	99.49	93.25	96.81	**99.95**
	Time	0.06	7.51	1.93	19.45	8.19	14.68	17.93	2.28	50.00	39.15	1.89
Urban1	AUC	99.46	97.84	99.57	87.03	99.46	97.70	95.45	95.42	99.73	99.59	**99.93**
	Time	0.05	7.64	3.98	22.87	8.89	16.69	17.98	2.40	53.78	39.84	2.13
Urban2	AUC	96.92	91.76	96.58	89.13	95.97	82.58	93.00	97.11	94.90	90.18	**98.30**
	Time	0.06	7.78	3.85	20.30	9.18	15.99	19.32	2.49	42.26	37.99	1.89

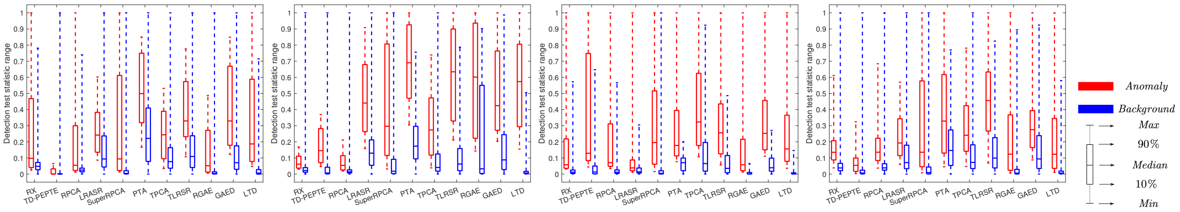
including fewer background pixels compared to other methods. Specifically, although the anomaly detection maps produced by RX, TD-PEPTE, RPCA, LRASR, and TPCA contain few background pixels, they fail to detect many anomalies or detect them weakly. Conversely, the anomaly detection maps generated by SuperRPCA, PTA, TLRSR, RGAE, and GAED can identify most anomalies but include a significant amount of background pixels, such as the red area in the lower left corner of Airport1 and the dashed lines at the top of Airport2. In contrast, the anomaly detection map produced by our proposed LTD method successfully detects nearly all anomalies while containing minimal background pixels.

Table 3 provides a detailed comparison of the AUC values and running times for each method. On the AUC metric, the proposed LTD method achieves the best results, consistent with the observations in Figure 12. Regarding running times, our method is the second fastest. Although RX exhibits shorter running times, its AUC values are unsatisfactory in most scenarios. In contrast, other methods have longer running times compared to ours, with the deep learning based method RGAE taking approximately 50 s, which is about 25 times the running time of our proposed method. Moreover, the AUC values of these methods are not as high as those achieved by the LTD method.



(a) Airport1. (b) Airport2. (c) Urban1. (d) Urban2.

Figure 13. *ROC curves obtained by different methods for ABU dataset.*



(a) Airport1. (b) Airport2. (c) Urban1. (d) Urban2.

Figure 14. *Separability maps of different methods for ABU dataset.*

Figure 13 provides a detailed comparison of the ROC curves for each method. It shows that our proposed LTD method has a higher detection probability than other methods for most false alarm rates. For the Airport dataset in particular, our method maintains a higher detection probability across all false alarm rates.

To further evaluate the performance of LTD, we utilize normalized background-anomaly separation maps. A larger distance between the anomaly and background boxes indicates superior separation performance, while shorter background boxes reflect more effective suppression of background information. As illustrated in Figure 14, the LTD method consistently demonstrates both shorter background boxes and greater separation between anomaly and background boxes across all scenarios. In contrast, some comparative algorithms only exhibit shorter background boxes for specific datasets. For instance, the RX algorithm achieves shorter background boxes solely in the Urban1 dataset. Similarly, certain algorithms achieve greater separation between anomaly and background boxes only in select datasets. For example, the PTA algorithm shows greater separation exclusively in the Airport2 dataset, while in other datasets, the anomaly and background boxes remain closely spaced.

5.3.2. MVTec dataset. For the MVTec dataset, the detection results of the compared methods are illustrated in Figures 5 and 15. The corresponding AUC values and running times are provided in Table 4. One can see that the detection maps produced by the proposed LTD method closely resemble the reference maps, outperforming other methods. Notably, the detection map contains less background noise compared to the RX method, demonstrating the effectiveness of the proposed LTD method in suppressing background. In addition, as shown in Table 4, the AUC values of our proposed LTD method show significant improvements

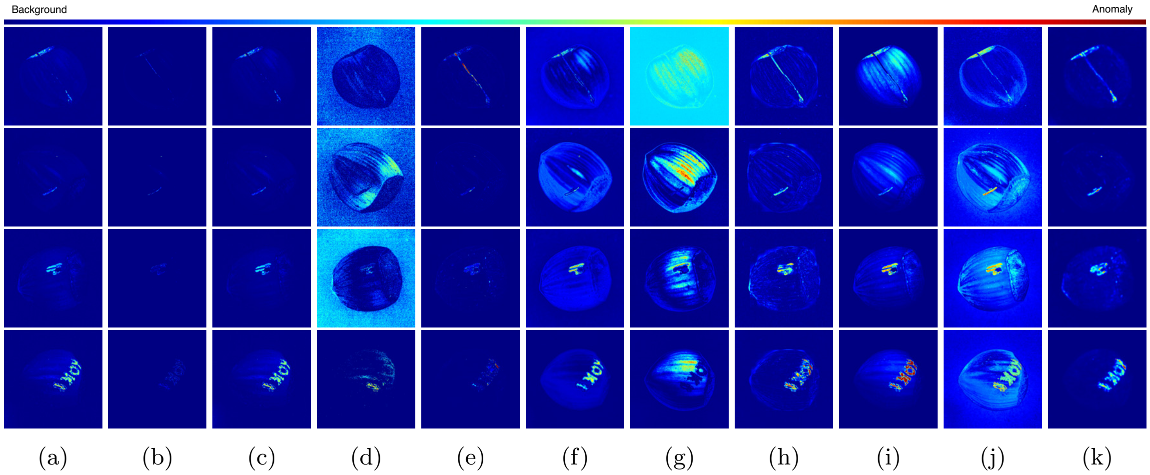


Figure 15. *Target detection results by different methods for MVTec dataset.* (a) *RX.* (b) *TD-PEPTE.* (c) *RPCA.* (d) *LRASR.* (e) *SuperRPCA.* (f) *PTA.* (g) *TPCA.* (h) *TLRSR.* (i) *RGAE.* (j) *GAED.* (k) *LTD.*

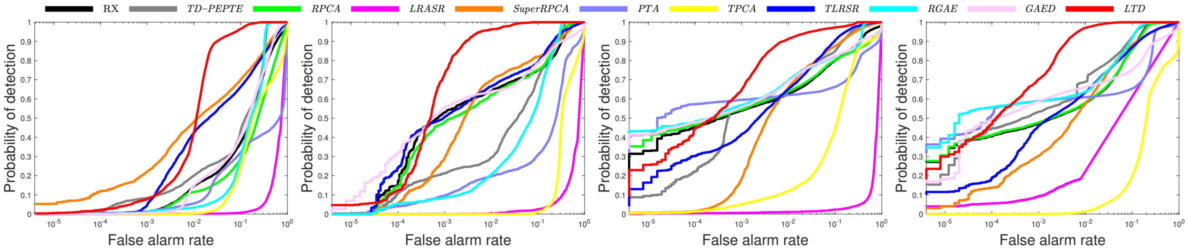
Table 4
Comparison of AUC values (%) and running time(s) of different methods for MVTec dataset.

Algorithm	Index	RX	TD-PEPTE	RPCA	LRASR	SuperRPCA	PTA	TPCA	TLRSR	RGAE	GAED	LTD
Crack	AUC	77.23	84.20	71.12	29.98	87.88	45.66	65.75	83.34	82.03	78.37	**98.26**
	Time	0.05	7.90	0.40	977.33	8.79	16.04	14.33	29.41	818.99	669.92	4.99
Cut	AUC	89.46	91.50	91.50	33.48	93.11	72.30	56.40	91.11	88.54	86.02	**99.76**
	Time	0.05	8.35	0.40	991.03	7.91	16.81	13.96	32.88	837.77	662.67	4.75
Hole	AUC	91.26	96.24	88.50	22.93	95.41	78.77	82.01	95.75	93.88	90.34	**99.25**
	Time	0.05	7.89	0.40	979.06	7.54	16.26	14.62	46.29	814.96	677.04	2.62
Print	AUC	95.70	97.79	95.60	58.68	95.14	89.03	68.48	94.21	97.05	87.42	**99.87**
	Time	0.04	8.13	0.40	982.01	7.48	15.31	14.00	32.08	978.39	684.22	3.44

over other methods. Specifically, the AUC values are higher by 11.81%, 7.14%, 3.13%, and 2.13% across four datasets compared to the second-highest values. Regarding running times, our method, which employs a rank reduction strategy with validation mechanism, achieves an average running time of approximately 4 seconds, significantly lower than that of deep learning based methods.

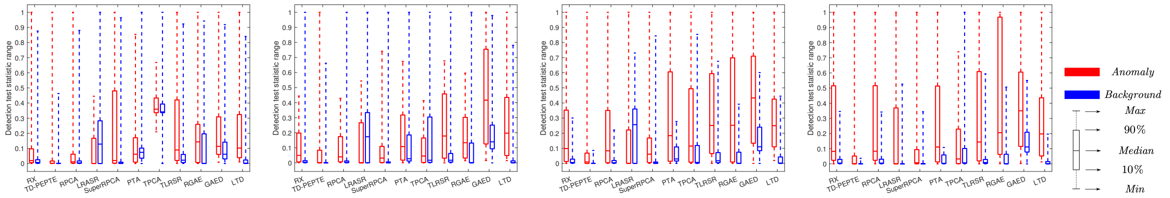
Figure 16 presents a detailed comparison of the ROC curves for each method. The detection probabilities of the proposed LTD method approach 1 when the false alarm rate is approximately 1e-1. Notably, our LTD method outperforms others in detection probability when the false alarm rate exceeds 1e-2. Figure 17 further compares the methods through box plots, confirming that our approach achieves the superior performance, aligning with the results in Figure 16.

The anomaly detection results from the MVTec dataset are also consistent with those from the ABU dataset, and all these demonstrate that our method achieves superior anomaly detection accuracy and operates with enhanced efficiency.



(a) Crack. (b) Cut. (c) Hole. (d) Print.

Figure 16. *ROC curves obtained by different methods for MVTec dataset.*



(a) Crack. (b) Cut. (c) Hole. (d) Print.

Figure 17. *Separability maps of different methods for MVTec dataset.*

6. Conclusion.

In this paper, we propose a novel LTD model for HAD that integrates spectral and spatial analysis into a unified framework. Within this framework, one component (layer 1) addresses spectral redundancy and extracts spectral anomaly, while the other component (layer 2) models the low rank spatial structure and identifies spatial anomaly. A key innovation in layer 2 is the substitution of the tensor tubal rank with group sparsity regularization, which iteratively reduces data size by eliminating groups of all-zero elements, further improving algorithmic efficiency. By integrating spectral anomaly from layer 1 with spatial anomaly from layer 2, we derive a spectral-spatial anomaly, leveraging the complementarity between the spectral and spatial domains to boost detection performance. Lastly, a guided image filter is applied to the spectral-spatial anomaly, yielding clearer target detection with reduced noise. We also introduce a PAM algorithm to solve the LTD model, proving that the iterative sequence converges to a critical point. Experimental results demonstrate that our method not only delivers superior performance but also operates faster than most state-of-the-art HAD methods.

Appendix A. Preliminaries of tensors.

In the appendix, we present a comprehensive set of definitions for the tensor concepts that are utilized consistently throughout this paper.

Definition 5 (f-diagonal tensor [22]). *If every frontal slice of a tensor constitutes a diagonal matrix, then the tensor is termed f-diagonal.*

Definition 6 (conjugate transpose [22]). *The conjugate transpose of a tensor, denoted by $\mathcal{X}^T$, for $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, is the tensor obtained by taking the conjugate transpose of each frontal slice and then reversing the order of transposed frontal slices 2 through n_3.*

Definition 7 (identity tensor [22]). *The identity tensor, denoted by $\mathcal{I} \in \mathbb{R}^{n\times n\times n_3}$, is the tensor whose first frontal slice is the identity matrix, with all other frontal slices being zeros.*

Definition 8 (orthogonal tensor [22]). *A tensor $\mathcal{P} \in \mathbb{R}^{n\times n\times n_3}$ is orthogonal if it fulfills the condition $\mathcal{P}^T * \mathcal{P} = \mathcal{P} * \mathcal{P}^T = \mathcal{I}$.*

Appendix B. Proof of Theorem 3.2. Initially, let us delineate the notation to be utilized.

- The symbol $\mathcal{O}$ denotes the tensor wherein all elements are zero;
- for the given tensors $\mathcal{X} \in \mathbb{R}^{n_1\times r_1\times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_1\times r_2\times n_3}$, the tensor $\mathcal{Z} = [\mathcal{X}, \mathcal{Y}] \in \mathbb{R}^{n_1\times(r_1+r_2)\times n_3}$ is defined as the concatenation of $\mathcal{X}$ and $\mathcal{Y}$ along the lateral slices, i.e., $\mathcal{Z}(:, 1 : r_1, :) = \mathcal{X}$ and $\mathcal{Z}(:, r_1 + 1 : r_1 + r_2, :) = \mathcal{Y}$.

Proof. Let $\hat{r} := \|\mathcal{Z}\|_{F,0}$. Without loss of generality, assume that $\mathcal{Z}$ can be partitioned as $\mathcal{Z} = [\mathcal{Z}_1, \mathcal{O}]$, where $\mathcal{Z}_1 \in \mathbb{R}^{n_2\times\hat{r}\times b}$ and $\mathcal{O} \in \mathbb{R}^{n_2\times(r-\hat{r})\times b}$. Correspondingly, tensor $\mathcal{D}$ is divided into $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2]$, with $\mathcal{D}_1 \in \mathbb{R}^{n_1\times\hat{r}\times b}$ and $\mathcal{D}_2 \in \mathbb{R}^{n_1\times(r-\hat{r})\times b}$. Consequently, $\mathcal{D} * \mathcal{Z}^T = \mathcal{D}_1 * \mathcal{Z}_1^T$. Combining this with $\mathcal{L} = \mathcal{D} * \mathcal{Z}^T$, one has

$$\operatorname{rank}_t(\mathcal{L}) = \operatorname{rank}_t\left(\mathcal{D} * \mathcal{Z}^T\right) = \operatorname{rank}_t\left(\mathcal{D}_1 * \mathcal{Z}_1^T\right) \le \operatorname{rank}_t(\mathcal{Z}_1) \le \hat{r} = \|\mathcal{Z}\|_{F,0},$$

where the first inequality follows from [54, Lemma 2]. When $\mathcal{D} = \mathcal{U}(:, 1 : r, :)$ and $\mathcal{Z} = \mathcal{V} * \mathcal{S}(1 : r, :, :)^T$, equality is achieved. The tensors $\mathcal{U}$, $\mathcal{S}$, and $\mathcal{V}$ are derived from the T-SVD of $\mathcal{L}$, represented as $\mathcal{L} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$. ■

Appendix C. Proof of Theorem 3.3. Before proving the theorem, we introduce some notation. We denote

$$\Omega = \left\{\mathcal{Z} \in \mathbb{R}^{n_2\times r\times b} \mid \exists\, \mathcal{D} \text{ s.t. } \mathcal{L} = \mathcal{D} * \mathcal{Z}^T, \mathcal{D}^T * \mathcal{D} = \mathcal{I}\right\}.$$

We also denote the group support set of $\mathcal{Z}$ as

$$\Gamma(\mathcal{Z}) := \{j \mid \|\mathcal{Z}(:, j, :)\| \ne 0, j = 1, \dots, r\} = \Gamma_1(\mathcal{Z}) \cup \Gamma_2(\mathcal{Z}),$$

$$\Gamma_1(\mathcal{Z}) := \{j \mid \|\mathcal{Z}(:, j, :)\| < \nu, j \in \Gamma(\mathcal{Z})\}, \Gamma_2(\mathcal{Z}) := \{j \mid \|\mathcal{Z}(:, j, :)\| \ge \nu, j \in \Gamma(\mathcal{Z})\}.$$

For an integer s with $0 \le s \le r$, denote $Q^s = \left\{\mathcal{Z} \in \mathbb{R}^{n_2\times r\times b} \mid \|\mathcal{Z}\|_{F,0} \le s\right\}$ and $\operatorname{dist}(\Omega, Q^s) = \inf_{\mathcal{Z}\in\Omega} \operatorname{dist}(\mathcal{Z}, Q^s)$. Here $\operatorname{dist}(\mathcal{Z}, Q^s) = \inf_{\tilde{\mathcal{Z}}\in Q^s} \|\mathcal{Z} - \tilde{\mathcal{Z}}\|$.

For simplicity, we denote

$$\begin{cases} (P_0) \ \min\left\{\|\mathcal{Z}\|_{F,0} : \mathcal{L} = \mathcal{D} * \mathcal{Z}^T, \mathcal{D}^T * \mathcal{D} = \mathcal{I}\right\}; \\ (P_\psi) \ \min\left\{\|\mathcal{Z}\|_{F,1}^{\psi} : \mathcal{L} = \mathcal{D} * \mathcal{Z}^T, \mathcal{D}^T * \mathcal{D} = \mathcal{I}\right\}. \end{cases}$$

Obviously, the optimal value of (P_0) is a positive integer, which we denote as N. Then for any $\mathcal{Z} \in \Omega$, one has $\|\mathcal{Z}\|_{F,0} \ge N$, which implies that $\operatorname{dist}(\Omega, Q^{N-n}) > 0$ for all $n = 1, 2, \dots, N$. Up to this point, we can construct $\bar{\nu}$ as follows:

$$\bar{\nu} = \min\left\{\frac{1}{n} \operatorname{dist}\left(\Omega, Q^{N-n}\right) : n = 1, \dots, N\right\}.$$

Proof. (i) Let $(\mathcal{D}^*, \mathcal{Z}^*)$ be a global minimizer of problem (P_0) with $\|\mathcal{Z}^*\|_{F,0} = N$. We demonstrate that $(\mathcal{D}^*, \mathcal{Z}^*)$ is also a global minimizer of problem (P_ψ) for any $0 < \nu < \bar{\nu}$. Given that the global optimality of (P_0) implies $\|\mathcal{Z}\|_{F,0} \geq N$ for $\mathcal{Z} \in \Omega$, we establish this conclusion by considering two cases.

Case 1. $\|\mathcal{Z}\|_{F,0} = N$. It is easy to see that for any $j \in \Gamma(\mathcal{Z})$,

$$\begin{aligned}\|\mathcal{Z}(:,j,:)\| &\geq \min\{\|\mathcal{Z}(:,l,:)\| > 0 : l = 1, \ldots, r\}\\ &= \operatorname{dist}\left(\mathcal{Z}, Q^{N-1}\right) \geq \operatorname{dist}\left(\Omega, Q^{N-1}\right) \geq \bar{\nu} > \nu,\end{aligned}$$

which means that $\|\mathcal{Z}\|_{F,1}^{\psi} = N = \|\mathcal{Z}^*\|_{F,1}^{\psi}$.

Case 2. $\|\mathcal{Z}\|_{F,0} = M > N$. Without loss of generality, assume $\|\mathcal{Z}(:,1,:)\|, \ldots, \|\mathcal{Z}(:,j_1,:)\| \in (0,\nu)$, $\|\mathcal{Z}(:,j_1+1,:)\|, \ldots, \|\mathcal{Z}(:,j_2,:)\| \in [\nu, +\infty)$ and $\|\mathcal{Z}(:,j_2+1,:)\| = \cdots = \|\mathcal{Z}(:,r,:)\| = 0$. If $M' = \Gamma_2(\mathcal{Z}) \geq N$, from $\phi(x) > 0$ for $x > 0$, one has $\|\mathcal{Z}\|_{F,1}^{\psi} > N$. Now assume $M' < N$; we know that

$$\frac{1}{N - M'} \operatorname{dist}\left(\Omega, Q^{M'}\right) \geq \bar{\nu}.$$

From the above cases, together with

$$\begin{aligned}\|\mathcal{Z}(:,1,:)\| + \cdots + \|\mathcal{Z}(:,j_1,:)\| &\geq \sqrt{\|\mathcal{Z}(:,1,:)\|^2 + \cdots + \|\mathcal{Z}(:,j_1,:)\|^2}\\ &\geq \operatorname{dist}\left(\mathcal{Z}, Q^{M'}\right) \geq \operatorname{dist}\left(\Omega, Q^{M'}\right),\end{aligned}$$

we get

$$\begin{aligned}\|\mathcal{Z}\|_{F,1}^{\psi} &= \psi\left(\|\mathcal{Z}(:,1,:)\|\right) + \cdots + \psi\left(\|\mathcal{Z}(:,j_1,:)\|\right) + \cdots + \psi\left(\|\mathcal{Z}(:,j_2,:)\|\right)\\ &\geq \frac{1}{\nu}\left(\|\mathcal{Z}(:,1,:)\| + \cdots + \|\mathcal{Z}(:,j_1,:)\|\right) + M'\\ &\geq \frac{1}{\nu} \operatorname{dist}\left(\Omega, Q^{M'}\right) + M'\\ &\geq \frac{1}{\nu}\left(N - M'\right)\bar{\nu} + M'\\ &> \frac{1}{\bar{\nu}}\left(N - M'\right)\bar{\nu} + M' = N.\end{aligned} \tag{C.1}$$

The aforementioned two cases indicate that $\|\mathcal{Z}\|_{F,1}^{\psi} \geq N = \|\mathcal{Z}^*\|_{F,1}^{\psi}$ for all $\mathcal{Z} \in \Omega$. Hence, $(\mathcal{D}^*, \mathcal{Z}^*)$ is also a global minimizer of problem (P_ψ). Furthermore, for each global minimizer $(\mathcal{D}^*, \mathcal{Z}^*)$ of problem (P_ψ), we have $\|\mathcal{Z}^*\|_{F,0} = \|\mathcal{Z}^*\|_{F,1}^{\psi}$.

(ii) Let $(\mathcal{D}^\#, \mathcal{Z}^\#)$ be a global minimizer of problem (P_ψ) for $0 < \nu < \bar{\nu}$. Assume on the contrary that $(\mathcal{D}^\#, \mathcal{Z}^\#)$ is not a solution of problem (P_0). Let $(\mathcal{D}^*, \mathcal{Z}^*)$ be a global minimizer of problem (P_0), that is, $\|\mathcal{Z}^*\|_{F,0} = N$. By $\psi(x) \leq |x|^0$, we have $\|\mathcal{Z}^*\|_{F,1}^{\psi} \leq \|\mathcal{Z}^*\|_{F,0}$. Similarly to the proof of Case 2 above, we will obtain $\|\mathcal{Z}^\#\|_{F,1}^{\psi} > N = \|\mathcal{Z}^*\|_{F,0} \geq \|\mathcal{Z}^*\|_{F,1}^{\psi}$ for any $0 < \nu < \bar{\nu}$. This contradicts the global optimality of $(\mathcal{D}^\#, \mathcal{Z}^\#)$ for problem (P_ψ). Hence $(\mathcal{D}^\#, \mathcal{Z}^\#)$ is a global minimizer of problem (P_0).

Therefore, whenever $0 < \nu < \bar{\nu}$, problems (P_0) and (P_ψ) have the same global minimizers and optimal values. ■

Appendix D. Proof of Lemma 4.1.

Lemma D.1 ([22]). *Let* $\mathcal{X} \in \mathbb{R}^{n_1 \times r \times n_3}$ *and* $\mathcal{Y} \in \mathbb{R}^{r \times n_2 \times n_3}$ *be arbitrary tensors, and let* $\mathcal{Z} = \mathcal{X} * \mathcal{Y}$. *Then,*

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} \Longleftrightarrow \bar{Z}^{(k)} = \bar{X}^{(k)} \bar{Y}^{(k)} \text{ for all } k \in [n_3],$$

where $\bar{\mathcal{Z}}$ *denotes the fast Fourier transform of* $\mathcal{Z}$ *along its third mode (and this holds similarly for* $\bar{\mathcal{X}}$ *and* $\bar{\mathcal{Y}}$*).*

Proof. Given that $F_b \in \mathbb{C}^{b \times b}$ is the discrete Fourier transform matrix, and from the property $F_b^T F_b = bI$, we have

$$b \langle \mathcal{D}, \mathcal{G} \rangle = \left\langle \mathcal{D} \times_3 \left(F_b^T F_b \right), \mathcal{G} \right\rangle = \langle \mathcal{D} \times_3 F_b, \mathcal{G} \times_3 F_b \rangle = \sum_{k=1}^{b} \langle \bar{D}^{(k)}, \bar{G}^{(k)} \rangle.$$

We obtain from the above equality that

$$\underset{\mathcal{D}^T * \mathcal{D} = \mathcal{I}}{\arg\max} \ \langle \mathcal{D}, \mathcal{G} \rangle = \underset{\mathcal{D}^T * \mathcal{D} = \mathcal{I}}{\arg\max} \sum_{k=1}^{b} \langle \bar{D}^{(k)}, \bar{G}^{(k)} \rangle = \underset{\bar{D}^{(k)T} \bar{D}^{(k)} = I}{\arg\max} \sum_{k=1}^{b} \langle \bar{D}^{(k)}, \bar{G}^{(k)} \rangle,$$

where the last equality follows from Lemma D.1. By the orthogonal Procrustes problem [40], we obtain

$$\underset{\bar{D}^{(k)T} \bar{D}^{(k)} = I}{\arg\max} \ \langle \bar{D}^{(k)}, \bar{G}^{(k)} \rangle = \bar{U}^{(k)} \bar{V}^{(k)^T},$$

where $\bar{U}^{(k)}$ and $\bar{V}^{(k)}$ are obtained by the skinny SVD as follows: $\bar{G}^{(k)} = \bar{U}^{(k)} \bar{S}^{(k)} \bar{V}^{(k)^T}$. This completes the proof of this statement. ■

Data availability statement. The data that support the findings of this study are available upon reasonable request from the authors.

REFERENCES

[1] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457, https://doi.org/10.1287/moor.1100.0449.

[2] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: Proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129, https://doi.org/10.1007/s10107-011-0484-9.

[3] E. Bati, A. Çalişkan, A. Koz, and A. A. Alatan, *Hyperspectral anomaly detection method based on auto-encoder*, in Image and Signal Processing for Remote Sensing XXI, L. Bruzzone, ed., SPIE, 2015, https://doi.org/10.1117/12.2195180.

[4] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, *MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection*, in Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2019, pp. 9584–9592, https://doi.org/10.1109/CVPR.2019.00982.

[5] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494, https://doi.org/10.1007/s10107-013-0701-9.

[6] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb, *Blind image fusion for hyperspectral imaging with the directional total variation*, Inverse Problems, 34 (2018), 044003, https://doi.org/10.1088/1361-6420/aaaf63.

[7] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, *Low-rank decomposition-based anomaly detection*, in Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XIX, S. S. Shen and P. E. Lewis, eds., SPIE, 2013, https://doi.org/10.1117/12.2015652.

[8] Z. Chen, B. Yang, and B. Wang, *A preprocessing method for hyperspectral target detection based on tensor principal component analysis*, Remote Sens., 10 (2018), 1033, https://doi.org/10.3390/rs10071033.

[9] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, *Hyperspectral anomaly detection with robust graph autoencoders*, IEEE Trans. Geosci. Remote Sens., 60 (2022), pp. 1–14, https://doi.org/10.1109/TGRS.2021.3097097.

[10] J. Fan, L. Ding, Y. Chen, and M. Udell, *Factor group-sparse regularization for efficient low-rank matrix recovery*, in Proceedings of the 33rd International Conference on Neural Information Processing Systems, Adv. Neural Inform. Process. Syst. 32, Curran Associates, Red Hook, NY, 2019, https://papers.neurips.cc/paper_files/paper/2019/hash/0fc170ecbb8ff1afb2c6de48ea5343e7-Abstract.html.

[11] J. Fan, L. Ding, C. Yang, Z. Zhang, and M. Udell, *Euclidean-norm-induced Schatten-p quasi-norm regularization for low-rank tensor completion and tensor robust principal component analysis*, Trans. Mach. Learn. Res., 2023 (2023), https://openreview.net/forum?id=Grhi800jVz.

[12] 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., 61 (2023), pp. 1–14, https://doi.org/10.1109/TGRS.2023.3274661.

[13] 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., 61 (2023), pp. 1–18, https://doi.org/10.1109/TGRS.2023.3246565.

[14] X. Han, J. Yu, J.-H. Xue, and W. Sun, *Hyperspectral and multispectral image fusion using optimized twin dictionaries*, IEEE Trans. Image Process., 29 (2020), pp. 4709–4720, https://doi.org/10.1109/TIP.2020.2968773.

[15] K. He, J. Sun, and X. Tang, *Guided image filtering,* in Computer Vision – ECCV 2010, Lecture Notes in Comput. Sci. 6311, K. Daniilidis, P. Maragos, and N. Paragios, eds., Springer, Berlin, Heidelberg, 2010, pp. 1–14, https://doi.org/10.1007/978-3-642-15549-9_1.

[16] 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., 61 (2023), pp. 1–17, https://doi.org/10.1109/TGRS.2023.3270667.

[17] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, *More diverse means better: Multimodal deep learning meets remote-sensing imagery classification*, IEEE Trans. Geosci. Remote Sens., 59 (2021), pp. 4340–4354, https://doi.org/10.1109/TGRS.2020.3016820.

[18] 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., 55 (2017), pp. 5600–5611, https://doi.org/10.1109/TGRS.2017.2710145.

[19] J. Kerekes, *Receiver operating characteristic curve confidence intervals and regions*, IEEE Geosci. Remote Sens. Lett., 5 (2008), pp. 251–255, https://doi.org/10.1109/LGRS.2008.915928.

[20] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, *Anomaly detection in hyperspectral images based on an adaptive support vector method*, IEEE Geosci. Remote Sens. Lett., 8 (2011), pp. 646–650, https://doi.org/10.1109/LGRS.2010.2098842.

[21] 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., 34 (2013), pp. 148–172, https://doi.org/10.1137/110837711.

[22] M. E. Kilmer and C. D. Martin, *Factorization strategies for third-order tensors*, Linear Algebra Appl., 435 (2011), pp. 641–658, https://doi.org/10.1016/j.laa.2010.09.020.

[23] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.

[24] O. Kuybeda, D. Malah, and M. Barzohar, *Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors*, IEEE Trans. Signal Process., 55 (2007), pp. 5579–5592, https://doi.org/10.1109/TSP.2007.901645.

[25] H. Kwon and N. Nasrabadi, *Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery*, IEEE Trans. Geosci. Remote Sens., 43 (2005), pp. 388–397, https://doi.org/10.1109/TGRS.2004.841487.

[26] B.-Z. Li, X.-L. Zhao, T.-Y. Ji, X.-J. Zhang, and T.-Z. Huang, *Nonlinear transform induced tensor nuclear norm for tensor completion*, J. Sci. Comput., 92 (2022), 83, https://doi.org/10.1007/s10915-022-01937-1.

[27] 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., 33 (2022), pp. 1037–1050, https://doi.org/10.1109/TNNLS.2020.3038659.

[28] W. Li, G. Wu, and Q. Du, *Transferred deep learning for anomaly detection in hyperspectral imagery*, IEEE Geosci. Remote Sens. Lett., 14 (2017), pp. 597–601, https://doi.org/10.1109/LGRS.2017.2657818.

[29] J.-T. Lin and C.-H. Lin, *SuperRPCA: A collaborative superpixel representation prior-aided RPCA for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 62 (2024), pp. 1–16, https://doi.org/10.1109/TGRS.2024.3459956.

[30] 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., 42 (2020), pp. 925–938, https://doi.org/10.1109/TPAMI.2019.2891760.

[31] X. Ma, X. Zhang, N. Huyan, J. Gu, X. Tang, and L. Jiao, *Background representation learning with structural constraint for hyperspectral anomaly detection*, IEEE Geosci. Remote Sens. Lett., 19 (2022), pp. 1–5, https://doi.org/10.1109/LGRS.2021.3073176.

[32] G. Marjanovic and V. Solo, *On ℓ_q optimization and matrix completion*, IEEE Trans. Signal Process., 60 (2012), pp. 5714–5724, https://doi.org/10.1109/TSP.2012.2212015.

[33] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, *Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data*, IEEE J. Selected Topics Appl. Earth Observations Remote Sens., 6 (2013), pp. 801–814, https://doi.org/10.1109/JSTARS.2013.2238609.

[34] L. Pan and X. Chen, *Group sparse optimization for images recovery using capped folded concave functions*, SIAM J. Imaging Sci., 14 (2021), pp. 1–25, https://doi.org/10.1137/19M1304799.

[35] I. Reed and X. Yu, *Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution*, IEEE Trans. Acoustics Speech Signal Process., 38 (1990), pp. 1760–1770, https://doi.org/10.1109/29.60107.

[36] 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., 62 (2024), pp. 1–15, https://doi.org/10.1109/TGRS.2023.3349128.

[37] L. Ren, D. Wang, L. Gao, M. Wang, and M. Huang, *Nonlocal and deep priors for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote Sens., 63 (2025), pp. 1–15, https://doi.org/10.1109/TGRS.2025.3593019.

[38] L. Ren, D. Wang, L. Gao, M. Wang, M. Huang, and H. Zhang, *HADDNLP: Hyperspectral anomaly detection via double nonlocal priors*, Pattern Recog., 172 (2026), 112535, https://doi.org/10.1016/j.patcog.2025.112535.

[39] A. Schaum, *Joint subspace detection of hyperspectral targets*, in 2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720), Vol. 3, 2004, 1824, https://doi.org/10.1109/AERO.2004.1367963.

[40] P. H. Schönemann, *A generalized solution of the orthogonal Procrustes problem*, Psychometrika, 31 (1966), pp. 1–10, https://doi.org/10.1007/BF02289451.

[41] 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., 8 (2014), 083641, https://doi.org/10.1117/1.JRS.8.083641.

[42] X. Sun, L. Zhuang, L. Gao, H. Gao, X. Sun, and B. Zhang, *A parameter-free topological disassembly-guided method for hyperspectral target detection*, IEEE J. Sel. Top. Appl., 18 (2025), pp. 17875–17888, https://doi.org/10.1109/JSTARS.2025.3587030.

[43] T. Tao, Y. Qian, and S. Pan, *Column $\ell_{2,0}$-norm regularized factorization model of low-rank matrix recovery and its computation*, SIAM J. Optim., 32 (2022), pp. 959–988, https://doi.org/10.1137/20M136205X.

[44] Z. Tu, J. Lu, H. Zhu, H. Pan, W. Hu, Q. Jiang, and Z. Lu, *A new nonconvex low-rank tensor approximation method with applications to hyperspectral images denoising*, Inverse Problems, 39 (2023), 065003, https://doi.org/10.1088/1361-6420/acc88a.

[45] 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. Top. Appl., 18 (2025), pp. 6981–6993, https://doi.org/10.1109/JSTARS.2025.3542457.

[46] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, *PDBSNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection*, IEEE Trans. Geosci. Remote, 61 (2023), pp. 1–14, https://doi.org/10.1109/TGRS.2023.3276175.

[47] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, *Learning tensor low-rank representation for hyperspectral anomaly detection*, IEEE Trans. Cybernet., 53 (2023), pp. 679–691, https://doi.org/10.1109/TCYB.2022.3175771.

[48] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, *Hyperspectral anomaly detection with guided autoencoder*, IEEE Trans. Geosci. Remote Sens., 60 (2022), pp. 1–18, https://doi.org/10.1109/TGRS.2022.3207165.

[49] 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., 54 (2016), pp. 1990–2000, https://doi.org/10.1109/TGRS.2015.2493201.

[50] Q. Yu and M. Bai, *Generalized nonconvex hyperspectral anomaly detection via background representation learning with dictionary constraint*, SIAM J. Imaging Sci., 17 (2024), pp. 917–950, https://doi.org/10.1137/23M157363X.

[51] Q. Yu and X. Zhang, *T-product factorization based method for matrix and tensor completion problems*, Comput. Optim. Appl., 84 (2023), pp. 761–788, https://doi.org/10.1007/s10589-022-00439-y.

[52] 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., 54 (2016), pp. 1376–1389, https://doi.org/10.1109/TGRS.2015.2479299.

[53] 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., 54 (2016), pp. 6497–6504, https://doi.org/10.1109/TGRS.2016.2585495.

[54] P. Zhou, C. Lu, Z. Lin, and C. Zhang, *Tensor factorization for low-rank tensor completion*, IEEE Trans. Image Process., 27 (2018), pp. 1152–1163, https://doi.org/10.1109/TIP.2017.2762595.

[55] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, *Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations*, IEEE Trans. Geosci. Remote Sens., (2022), pp. 1–17, https://doi.org/10.1109/TGRS.2020.3040221.