Neurocomputing 686 (2026) 133709

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Robust asymmetric encoder-decoder nonnegative matrix factorization for hyperspectral anomaly detection

Shirin Moradi [a], Amjad Seyedi [b], Wafa Barkhoda [a,c], Fardin Akhlaghian Tab [a,*]

[a] Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran
[b] Department of Mathematics and Operational Research, University of Mons, Mons, Belgium
[c] Faculty of Information Technology, Kermanshah University of Technology, Kermanshah, Iran

ARTICLE INFO

Communicated by Z. Wang

Keywords:
Hyperspectral analysis
Anomaly detection
Nonnegative matrix factorization
Robustness
Encoder-decoder structure

ABSTRACT

Hyperspectral anomaly detection remains a challenging task due to the limitations of traditional low-rank matrix factorization methods, which often fail to effectively isolate anomalies. Most approaches model hyperspectral images as a combination of low-rank background and high-rank anomaly components, but do not explicitly learn the anomaly, leading to contamination of the background representation and suboptimal detection. Furthermore, existing factorization-based models often lack an encoder, restricting their ability to enforce mutual consistency between latent features and reconstruction, especially in noisy environments. To overcome these challenges, we propose Robust Asymmetric Encoder–Decoder Nonnegative Matrix Factorization (REDNMF), a novel optimization framework that jointly learns a sparse anomaly matrix and low-rank background components. REDNMF incorporates an explicitly modeled anomaly matrix, improving robustness and detection accuracy. Unlike traditional decoder-only schemes, our asymmetric encoder–decoder architecture enables flexible and discriminative feature learning, enhancing both reconstruction fidelity and anomaly-background separation. Additionally, REDNMF enforces pixel-wise sparsity on the anomaly matrix using the $L_{2,1}$ norm, promoting structured sparsity that aligns with the hyperspectral image geometry. This design helps localize anomalies more precisely and reduces sensitivity to noise and modeling errors. All components are integrated into a unified framework and optimized jointly. Extensive experiments on both real-world and benchmark hyperspectral datasets demonstrate that REDNMF consistently outperforms state-of-the-art methods in detection accuracy and robustness.

1. Introduction

Hyperspectral image (HSI), a key area within remote sensing technology, has garnered significant attention in recent years due to ongoing technological innovations and its wide-ranging applications across various fields [1,2]. These images encompass a broad spectrum of information, allowing for the precise differentiation of numerous ground objects [3]. HSI is employed in multiple vision tasks, such as target recognition [4], anomaly detection [5], change detection [6], spectral unmixing [7], and spectral analysis [8,9]. Typically, HSIs are organized as 3D data cubes, comprising two spatial dimensions and one spectral dimension. Each pixel within the spatial dimensions contains a spectrum of values across different wavelengths, forming a spectral data vector. This tensor structure enables detailed analysis of each pixel's spectral properties and supports efficient data storage and processing. By utilizing advanced algorithms like deep learning and clustering approaches, these techniques can extract complex patterns and characteristics from vast spectral data. Building on these extracted features, computer vision algorithms automate the interpretation of HSI data, enabling more precise and efficient detection, classification, and segmentation operations [10].

Real-world issues often include data that deviate from expected model assumptions, known as outliers or anomalies [11]. These anomalies, which are often rare pixels with distinct spectral signatures compared to their surroundings, are vital in different fields such as security [12], environmental monitoring [13], search and rescue operations [14], and identifying invasive species [15] or man-made objects in natural environments [16,17]. The extensive spectral information in the HSI data makes them particularly valuable for anomaly detection, but challenging [18]. In hyperspectral imaging, anomalies are identified as individual pixels or clusters of pixels whose spectral characteristics deviate markedly from those of the surrounding background, causing them to

* Corresponding author.
Email addresses: shirin.moradi@uok.ac.ir (S. Moradi), seyedamjad.seyedi@umons.ac.be (A. Seyedi), barkhoda@kut.ac.ir (W. Barkhoda), f.akhlaghian@uok.ac.ir (F. Akhlaghian Tab).

https://doi.org/10.1016/j.neucom.2026.133709
Received 13 October 2025; Received in revised form 9 April 2026; Accepted 15 April 2026
Available online 20 April 2026
0925-2312/

appear as distinct entities within the dataset. Anomalies in hyperspectral data are not inherently negative or undesirable; for example, airplanes at airports may be considered anomalies in certain datasets, yet they are anticipated and highly relevant in that specific context. However, accurately detecting anomalies in HSI data is challenging due to its high dimensionality and the difficulty of distinguishing true anomalies from noise and outliers. These challenges are further compounded by the large pixel sizes, which can obscure subtle spectral variations. In numerous instances, a single pixel can encompass a large region, and an anomaly may only occupy a small portion of that pixel, thus not fully representing the pixel's overall spectral signature.

Traditional hyperspectral anomaly detection methods mainly rely on statistical differences between background and anomalous pixels [19]. A representative approach is the RX detector proposed by Reed and Xiaoli [20], which measures the Mahalanobis distance between a test pixel and the estimated background using either global (GRX) or local (LRX) statistics [21]. However, complex and spatially varying backgrounds often violate Gaussian assumptions [22], leading to the development of kernel-based RX methods that better model nonlinear data distributions [23,24]. To reduce dependence on explicit background modeling, reconstruction-based methods such as dictionary learning identify anomalies through reconstruction errors [25,26], but their performance is highly sensitive to basis quality and degrades in the presence of noise, complex spatial–spectral structures, and high dimensionality [27]. Matrix factorization techniques, particularly Nonnegative Matrix Factorization (NMF), address these challenges by producing interpretable, part-based representations through nonnegativity constraints [28,29]. NMF decomposes the data into two nonnegative matrices by minimizing reconstruction error [30], making it well suited for hyperspectral data with inherently nonnegative reflectance values [31]. Nevertheless, basic NMF is sensitive to noise and outliers [32,33], motivating robust variants such as HCNMF [34], $L_{2,1}$-NMF [33], Manhattan NMF [35,36], Huber-based NMF [37], Cauchy NMF [38], correntropy-based NMF [39], and Robust NMF (RNMF), which explicitly separates sparse noise from informative components [40]. These NMF-based frameworks have been successfully applied to anomaly detection in documents [17], graphs [41], low-SNR audio [42], and hyperspectral imagery, with recent methods incorporating spatial constraints and robust background reconstruction, such as DVAD [43], SC-CNMF [44], and robust NMF-based background modeling [45]. Despite these advances, achieving robust and reliable anomaly detection in hyperspectral images under complex backgrounds and severe noise remains an open challenge.

Autoencoders are fundamental self-representation models in unsupervised machine learning, designed to facilitate feature learning through the complementary functions of an encoder and a decoder. The encoder's role is to compress the input data, identifying latent features that encapsulate its core characteristics within a reduced-dimensional space. Conversely, the decoder reconstructs the input from this compressed representation, ensuring that the key information is preserved and refined. This synergy between the encoder and decoder allows autoencoders to produce compact, meaningful representations that effectively highlight the most critical aspects of the input data. Consequently, autoencoders are widely applied in tasks such as data compression, noise reduction, anomaly detection, and data representation [46–48]. To realize this encoding-decoding function in the NMF method, the Symmetric Encoder–Decoder NMF model is proposed that uses the transposed decoder weights as the encoder weights, making it equivalent to the Projective NMF [49]. This symmetric Encoder–Decoder model features an implicit orthogonality constraint, making it suitable for data representation and clustering tasks. This paper, in contrast, proposes an Asymmetric Encoder–Decoder factorization model with independent encoder and decoder weight matrices. This design prioritizes flexibility over orthogonality, making it more effective for data reconstruction and denoising tasks.

Despite considerable progress, several critical challenges remain in applying NMF-based low-rank models to hyperspectral anomaly detection. Most existing approaches model hyperspectral images as the superposition of a low-rank background component and a sparse anomaly-related residual. Typically, the hyperspectral data are decomposed into spatial and spectral latent factors, and anomalies are inferred from the reconstruction residual between the observed data and the low-rank approximation. However, in these methods, the anomaly component is not explicitly modeled or learned during the factorization process. As a result, anomalous pixels can contaminate the low-rank background estimation, causing distortions in the learned background subspace and leading to degraded detection performance, especially when anomalies are strong or spatially clustered. This limitation highlights the need for a factorization framework that can explicitly disentangle anomalies from the background during the decomposition process. Another limitation of existing NMF-based hyperspectral anomaly detection methods lies in their model architecture. Most prior approaches follow a decoder-only paradigm, where the hyperspectral image is reconstructed directly from learned basis matrices without a corresponding encoder or forward mapping from the original data to the latent space. The absence of such a self-representation mechanism restricts the model's ability to enforce consistency between the latent representation and the reconstructed data, limiting robustness under complex scenes or noisy conditions. Introducing an encoder–decoder structure can provide mutual constraints between encoding and reconstruction, thereby improving the stability and interpretability of the learned representations. Furthermore, anomalies in hyperspectral images are inherently sparse and typically occupy only a small fraction of the scene. Many existing methods impose element-wise sparsity constraints that do not fully exploit the spatial structure of hyperspectral data. Incorporating structured sparsity at the pixel level is therefore crucial for accurately isolating anomalous pixels while preserving the low-rank nature of the background. Explicitly enforcing pixel-wise sparsity can enhance anomaly localization and improve robustness against noise and modeling errors.

Motivated by these observations, this paper proposes Robust Asymmetric Encoder–Decoder NMF (REDNMF), a unified optimization framework for hyperspectral anomaly detection that explicitly addresses the limitations of existing low-rank methods. Unlike traditional approaches that treat reconstruction residuals as anomaly indicators, REDNMF introduces a learnable sparse anomaly matrix that is jointly optimized with the background components. This explicit modeling prevents anomalous pixels from influencing the background estimation and enables more accurate and noise-resilient anomaly isolation. To further enhance representation capability, REDNMF integrates an asymmetric encoder–decoder self-representation mechanism. In contrast to conventional decoder-only or symmetric factorization models, the asymmetric design decouples the encoder and decoder pathways, allowing more flexible and specialized feature learning for hyperspectral data. This structure enforces consistency between the latent representation and the reconstruction, resulting in improved background reconstruction and clearer separation between anomalies and normal background components. In addition, REDNMF incorporates pixel-wise sparsity regularization through the $L_{2,1}$ norm, which exploits the column-wise organization of hyperspectral images. This constraint promotes sparsity at the pixel level rather than at individual spectral elements, aligning with the physical characteristics of hyperspectral anomalies and improving detection interpretability and localization accuracy. All components—robust factorization, asymmetric self-representation, and structured sparsity—are jointly optimized within a single framework. Extensive experiments conducted on multiple real-world and benchmark hyperspectral datasets demonstrate that REDNMF consistently outperforms several state-of-the-art anomaly detection methods, achieving superior detection accuracy and robustness. In summary, the main contributions of this work are:

- A robust factorization model with a learnable anomaly matrix, explicitly isolating anomalies thereby improving robustness to noise and large deviations.
- An asymmetric encoder–decoder mechanism that enhances self-representation and provides improved reconstruction and anomaly-background separation.
- A pixel-wise sparsity regularization using the $L_{2,1}$ norm, enabling accurate and interpretable anomaly detection aligned with the hyperspectral data structure.
- Extensive experiments demonstrate the superior performance of REDNMF over existing methods across diverse hyperspectral datasets.

This paper is organized as follows: Section 2 reviews the related work in the field. Section 3 presents the background and preliminaries. Section 4 describes the proposed REDNMF method. Section 5 reports the experimental results and their analysis. Finally, Section 6 concludes the paper.

2. Related work

This section reviews hyperspectral anomaly detection methods, from traditional statistical detectors (e.g., RX and kernel-based variants) to reconstruction-driven approaches such as dictionary learning and matrix factorization (especially NMF and robust NMF). We then summarize recent deep learning based HAD methods that leverage spatial–spectral representation learning, attention, transformers, and generative priors for improved robustness in complex backgrounds.

2.1. Traditional methods

Traditional methods rely heavily on statistical differences between background and anomaly components [19]. The primary challenge arises when background information is lacking or varies spatially among images [22]. Developed by Reed and Xiaoli in 1990, the RX detector is a widely recognized method for anomaly detection, which works by measuring the statistical deviation of each pixel from the background in hyperspectral images [20]. A second-order matched filtering approach is used to compute the Mahalanobis distance between a test pixel and the background. When the entire image is utilized for background modeling, the method is referred to as Global RX (GRX); in contrast, if local statistical features are employed, it is termed Local RX (LRX) [21]. In practice, the background in hyperspectral images is often highly complex and cannot be accurately modeled by a simple multivariate Gaussian distribution. Accordingly, kernel-based detection methods, such as the nonlinear kernel RX algorithm, have shown superior performance compared to conventional approaches [23,24].

Recent approaches, such as dictionary learning and reconstruction error analysis [25], aim to detect anomalies with reduced dependence on background information. Dictionary learning constructs a set of basis vectors that sparsely represent the data, emphasizing salient features while suppressing noise. Reconstruction error analysis then identifies anomalies by measuring significant discrepancies between the original data and its reconstructed counterpart [26]. A key challenge in applying these methods is their strong sensitivity to the quality of the learned basis vectors. If the basis is not sufficiently representative, data reconstruction deteriorates, leading to inaccurate anomaly detection. Furthermore, when the data exhibit high noise levels or complex spatial–spectral patterns, distinguishing anomalies from noise becomes difficult, reducing detection reliability. In addition, diverse image backgrounds and varying noise conditions can further compromise model robustness in real-world scenarios. Finally, the high spectral dimensionality of hyperspectral data increases computational complexity, limiting efficiency [27].

2.2. Matrix factorization-based methods

Matrix factorization (MF) is widely used to extract low-dimensional representations from high-dimensional datasets by decomposing the original matrix into simpler components [50–55]. This reduces data complexity while preserving essential structure. Traditional MF techniques, such as principal component analysis (PCA) and singular value decomposition (SVD), generate low-rank matrices containing both positive and negative values, which can cancel each other out during reconstruction and reduce interpretability. Nonnegative matrix factorization (NMF) improves upon this by enforcing a non-negativity constraint, ensuring that only additive combinations of components are used [28]. This property enables part-based, interpretable representations [29]. NMF aims to find two nonnegative matrices, $\boldsymbol{W}$ and $\boldsymbol{H}$, that minimize the reconstruction error, typically measured by squared Euclidean distance or generalized Kullback–Leibler divergence between $\boldsymbol{X}$ and $\boldsymbol{WH}$ [30]. In hyperspectral analysis, NMF is particularly suitable because reflectance values are inherently nonnegative, allowing decomposition into physically meaningful basis spectra and coefficients. It has proven effective for data representation, clustering, and capturing similar data patterns while highlighting deviations [31]. Despite its advantages, NMF can struggle to produce robust low-dimensional representations when data contain substantial noise or outliers [32].

Basic NMF, which uses the least squares error function, performs best only when there is minimal noise and few outliers [33]. The L_2 loss function used in basic NMF is sensitive to extreme values, which can lead to inaccurate pattern detection in the presence of moderate to severe outliers. To address this issue, several robust M-estimator-based loss functions have been introduced to replace the squared Euclidean distance. Hamza and Brady [34] proposed hypersurface cost-based NMF (HCNMF), which minimizes a hypersurface cost function but suffers from slow optimization due to a complex line search method. Kong et al. [33] developed $L_{2,1}$-NMF, which replaces the Frobenius norm with the $L_{2,1}$ norm to measure reconstruction errors, making it less sensitive to outliers and effective for data with Laplacian noise. However, the non-smooth loss function of $L_{2,1}$-NMF complicates the factorization process. Manhattan NMF (MahNMF [35]), based on L_1-norm, improves upon this by approximating the non-smooth loss function with a smooth one and using Nesterov's method [36] for optimization. Despite these advancements, L_1-norm and $L_{2,1}$-norm-based methods, as well as their combination, Huber's function-based NMF methods [37], can still struggle with heavy outliers. To overcome these challenges, Liutkus et al. [38] introduced Cauchy NMF, which employs an isotropic Cauchy distribution to evaluate reconstruction error and optimizes it using maximum likelihood. Additional robust NMF variants, including CIM-NMF and rCIM-NMF, utilize Correntropy, a nonlinear similarity measure associated with the Welsch M-estimator in information-theoretic learning (ITL) [56], to effectively capture the probability of similarity between random variables in their joint space [39]. In this context, Zhang et al. [40] propose Robust Nonnegative Matrix Factorization (RNMF), an advanced framework designed to address the limitations of traditional NMF when dealing with corrupted data. By representing the data matrix as $\boldsymbol{X} \approx \boldsymbol{WH} + \boldsymbol{S}$, where $\boldsymbol{S}$ is a sparse error matrix capturing inconsistencies, RNMF enhances robustness by effectively separating noise while preserving meaningful features. This robust formulation significantly improves performance in tasks such as clustering and face recognition, demonstrating its capability in handling complex and noisy datasets.

Allan et al. [17] utilize NMF to cluster Aviation Safety Reporting System (ASRS) documents by generating feature vectors that provide a sum-of-parts term representation. These feature and coefficient matrices are then used to group documents and map known anomalies, while dominant features in test documents generate relevance scores to identify potential issues. Similarly, Tosyali et al. [41] apply NMF to cluster citation networks represented as directed graphs, calculating an outlier score for each node based on these clusters to rank nodes by their degree of outlierness. This method detects graph-based anomalies by considering both direct and indirect citation links, and it has been validated on both a small artificial dataset and real-world U.S. patent data. Furthermore, Ahmed et al. [31] introduce Neighborhood Structure

Assisted Nonnegative Matrix Factorization (NS-NMF) to improve dimensionality reduction in anomaly detection. By incorporating a minimum spanning tree model, NS-NMF more effectively captures neighborhood structures, enhancing anomaly differentiation. Moreover, Aiba et al. [42] present a system designed to improve anomaly detection in low Signal-to-Noise Ratio (SNR) environments by utilizing early-stage sound source extraction. Using NMF with Generalized Gaussian Distribution (GGD-NMF) for preprocessing, this system significantly enhances detection performance under challenging conditions, with experimental results confirming its effectiveness in low SNR environments. Collectively, these studies address critical challenges in anomaly detection, such as enhancing accuracy, feature representation, and performance in noisy environments. However, they still highlight ongoing difficulties in ensuring robust and reliable anomaly detection across diverse datasets and varying levels of outliers and noise.

NMF has emerged as a powerful tool for anomaly detection in hyperspectral imagery, with several NMF-based techniques developed to effectively extract the distinct spectral signatures of anomalies from the background. Zhang et al. [43] present the Dual-View Anomaly Detection (DVAD) algorithm, which employs Spectral Angular Distance (SAD) and NMF-based sparse spectral unmixing to improve detection at both pixel and subpixel levels. The DVAD algorithm leverages spatial consistency and spectral unmixing techniques to enhance anomaly detection performance. The approach effectively combines dual-view data to address challenges in hyperspectral imaging and achieve more accurate results in complex environments. Furthermore, the similarity constrained convex NMF (SC-CNMF) [44] method improves anomaly detection by integrating similarity constraints into CNMF, which refines the background matrix and improves detection accuracy, especially in cases involving mixed pixels. The Robust NMF-Based Background Reconstruction[45] method further advances anomaly detection by introducing a robust objective function to better manage noise and outliers, leading to more accurate background reconstruction.

2.3. Deep learning-based methods

Deep learning has become a dominant paradigm for hyperspectral image (HSI) processing by learning rich spatial–spectral representations for restoration and detection. Starting from deep learning for hyperspectral images and denoising, recent works mitigate inherent HSI noise through systematic paradigm evaluations [57], unsupervised cooperation of low-rank and deep spatial priors [58], and 3D spatial–spectral attention transformers [59]. Building on these advances, deep learning based hyperspectral anomaly detection (HAD) increasingly adopts transformer architectures (e.g., gated transformers and transformer-guided reconstruction) to capture long-range dependencies and enhance anomaly–background separability [60,61], self-supervised multiscale learning to reduce label dependence [62], and autoencoder-style feature aggregation with tailored attention mechanisms to strengthen discriminative cues [63,64]. More recently, generative modeling (diffusion-based background suppression) and alternative backbones (e.g., deep-prior Mamba networks) have been explored to better suppress complex backgrounds and improve robustness [65,66], while graph attention-based models explicitly leverage pixel relationships for noisy and nonlinear scenes [67].

3. Background

In this section, we introduce the related models and key notation essential to our research. We start by defining the primary notation used throughout the paper, then provide an overview of Basic NMF. We proceed to introduce the encoder and decoder concepts, explaining their roles within our framework. Finally, we explore Robust NMF, emphasizing its significance and applications.

3.1. Notation

In this paper, the matrices are represented by bold uppercase letters, such as $\boldsymbol{D}$. For a matrix $\boldsymbol{D} \in \mathbb{R}^{m\times n}$, $\boldsymbol{d}_i$ denotes the ith column and $\boldsymbol{d}^{(j)}$ denotes the jth row. Scalars are indicated by italicized lowercase letters like b, while vectors are shown as bold lowercase letters, such as $\boldsymbol{b}$. The transpose and trace of $\boldsymbol{D}$ are denoted by $\boldsymbol{D}^\top$ and $\mathrm{Tr}(\boldsymbol{D})$, respectively. The Frobenius norm of $\boldsymbol{D}$ is given by $\|\boldsymbol{D}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} D_{ij}^2}$, and the $L_{2,1}$-norm is given by $\|\boldsymbol{D}\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m} D_{ij}^2}$, where D_{ij} represents the (i,j)th entry of $\boldsymbol{D}$.

3.2. Basic nonnegative matrix factorization

Given a nonnegative data matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$, NMF aims to find two nonnegative matrices $\boldsymbol{W} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{H} \in \mathbb{R}^{r\times n}$ such that $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$. The nonnegative constraints lead to a parts-based representation because they allow only additive, not subtractive, combinations [28]. To achieve approximate factorization, we must define cost functions to measure the quality of the approximation. Popular cost functions include the squared Euclidean distance between $\boldsymbol{X}$ and $\boldsymbol{W}\boldsymbol{H}$, which is defined as:

$$\min_{\boldsymbol{W},\boldsymbol{H}\geq 0} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2, \tag{1}$$

where the Frobenius norm $\|\cdot\|_F$ quantifies the reconstruction error. The update rules for $\boldsymbol{W}$ and $\boldsymbol{H}$ are as follows:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \circ \frac{\boldsymbol{X}\boldsymbol{H}^\top}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top}, \qquad \boldsymbol{H} \leftarrow \boldsymbol{H} \circ \frac{\boldsymbol{W}^\top\boldsymbol{X}}{\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{H}},$$

where $\circ$ indicates the pointwise (Hadamard) product.

3.3. Symmetric encoder-decoder factorization

Traditional NMF functions as a decoder, focusing on reconstructing the input matrix from a lower-dimensional representation. However, it lacks an encoder to explicitly map the input data to a latent space. By integrating an encoder into the basic matrix decomposition, the process is refined, allowing for a direct mapping of the input to its latent representation. This combination not only enhances the interpretability and significance of the resulting representations but also improves the model's resilience to noise and outliers in the data. The encoder–decoder architecture minimizes both the reconstruction error of the input data and the reconstruction of the latent factors simultaneously. This dual reconstruction enables the model to capture fundamental patterns within the data more effectively [68]. Formally, the encoder representation can be defined as:

$$\boldsymbol{H} \approx \boldsymbol{W}^\top\boldsymbol{X}, \tag{2}$$

where this step serves as the encoding process, mapping the input data $\boldsymbol{X}$ to the latent representation $\boldsymbol{H}$. The composite objective function is formulated as:

$$\min_{\boldsymbol{W},\boldsymbol{H}\geq 0} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F^2 + \|\boldsymbol{H} - \boldsymbol{W}^\top\boldsymbol{X}\|_F^2. \tag{3}$$

The rules for updating $\boldsymbol{W}$ and $\boldsymbol{H}$ are presented as follows:

$$\boldsymbol{W} \leftarrow \boldsymbol{W} \circ \frac{2\boldsymbol{X}\boldsymbol{H}^\top}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top + \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{W}}, \quad \boldsymbol{H} \leftarrow \boldsymbol{H} \circ \frac{2\boldsymbol{W}^\top\boldsymbol{X}}{\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{H} + \boldsymbol{H}}.$$

The objective function is designed to minimize reconstruction errors from both the decoder and encoder, quantified using the Frobenius norm. The constraints $\boldsymbol{W}, \boldsymbol{H} \geq 0$ ensure non-negativity, maintaining interpretability in hyperspectral data analysis. By incorporating the encoder–decoder mechanism, the model reduces reconstruction errors in both the input data and latent variables, enhancing its ability to derive interpretable representations and increasing robustness against noise and anomalies [69].

3.4. Robust nonnegative matrix factorization

RNMF effectively addresses the limitations of traditional NMF methods in handling significant data errors. Although conventional NMF techniques, which are based on squared Euclidean distance or modified Kullback-Leibler divergence, are appropriate for managing small-scale noise, they are often insufficient in practical applications such as image processing and remote sensing due to substantial data inaccuracies [40]. In contrast, RNMF employs robust loss functions that are better suited to this challenge. Drawing on principal component analysis, RNMF improves the algorithm's ability to accurately represent and extract features from noisy hyperspectral data. By accounting for arbitrary corruption within the data matrix, RNMF provides a more reliable approximation of the nonnegative matrix $\boldsymbol{X}$.

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H} + \boldsymbol{S}. \tag{4}$$

In RNMF, decomposition involves nonnegative matrices $\boldsymbol{W} \in \mathbb{R}^{m\times r}$ and $\boldsymbol{H} \in \mathbb{R}^{r\times n}$, along with a sparse matrix $\boldsymbol{S} \in \mathbb{R}^{m\times n}$. The sparse matrix $\boldsymbol{S}$ helps to safeguard $\boldsymbol{W}$ and $\boldsymbol{H}$ from potential corruption, thereby improving the robustness of the decomposition compared to traditional NMF methods. Similar to classic NMF, RNMF utilizes the squared Euclidean distance between the original matrix $\boldsymbol{X}$ and its approximation $\boldsymbol{W}\boldsymbol{H}+\boldsymbol{S}$. Consequently, the objective function for RNMF is expressed as:

$$\min_{\boldsymbol{W},\boldsymbol{H},\boldsymbol{S}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H} - \boldsymbol{S}\|_F^2 + \lambda\|\boldsymbol{S}\|_1, \quad \text{s.t.} \quad \boldsymbol{W},\boldsymbol{H} \geq 0, \tag{5}$$

where the $\|\boldsymbol{S}\|_1$ norm is used to enforce sparsity in $\boldsymbol{S}$, ensuring that the solution promotes sparsity by driving some of the elements of $\boldsymbol{S}$ to zero. According to soft-thresholding, the update rule for matrix $\boldsymbol{S}$ is as follows:

$$S_{ij} = \begin{cases} E_{ij} - \frac{\lambda}{2}, & \text{if } E_{ij} \geq \frac{\lambda}{2}, \\ E_{ij} + \frac{\lambda}{2}, & \text{if } E_{ij} < -\frac{\lambda}{2}, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

where $\boldsymbol{E} = \boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}$ denotes the reconstruction error between the data matrix and its low-rank approximation.

4. Proposed model

In this section, we present Robust Asymmetric Encoder–Decoder Nonnegative Matrix Factorization (REDNMF), which is an innovative framework for hyperspectral anomaly detection. By combining Robust NMF (RNMF) with an autoencoder architecture (Fig. 1), it effectively captures the pixel-wise structure of hyperspectral data, preserving inter-sample relationships and extracting detailed spectral features. A key innovation is the incorporation of a robust error matrix $\boldsymbol{S}$, which isolates anomalies, improves noise resistance, and improves the accuracy of anomaly detection. Furthermore, group sparsity regularization using the $L_{2,1}$ norm is tailored to the unique spectral redundancy among hyperspectral bands, enabling more precise and reliable identification of anomalies. The optimization method leverages an iterative update scheme to efficiently solve the non-convex objective, ensuring convergence while balancing reconstruction fidelity, robustness, and sparsity.

4.1. Basic model: robust asymmetric encoder-decoder NMF

The primary objective of this work is to present a robust encoder–decoder matrix factorization method that not only enhances the approximation of NMF but also improves its capability to distinguish between background data and anomalies. Specifically, the encoder component refines the data representation by using the anomaly-free input ($\boldsymbol{X} - \boldsymbol{S}$), where $\boldsymbol{S}$ is the anomaly matrix. The matrix $\boldsymbol{S}$ is iteratively estimated during the optimization process, effectively isolating anomalies from the normal data. This separation allows the model to focus on learning the underlying structure of the background while capturing anomalous components in $\boldsymbol{S}$.

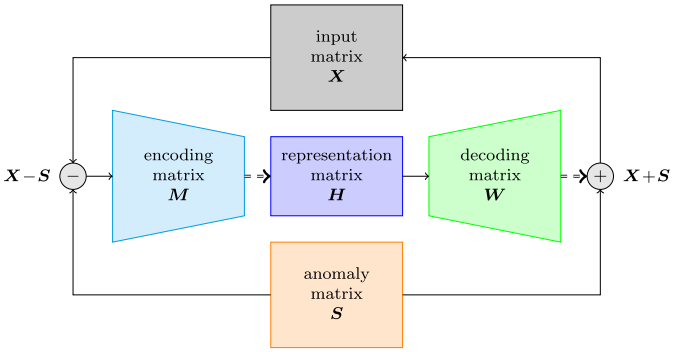

Fig. 1. Diagram of the proposed robust asymmetric Encoder–Decoder NMF (REDNMF) model for hyperspectral anomaly detection.

The model is designed to learn a robust representation of the input data by encoding the matrix $\boldsymbol{X}$ into two nonnegative matrices: the basis matrix $\boldsymbol{W}$ and the representation matrix $\boldsymbol{H}$. This encoding process enables the model to estimate a clean background component $\boldsymbol{X} - \boldsymbol{S} \approx \boldsymbol{W}\boldsymbol{H}$, which captures the underlying structure of the normal data. By subtracting this estimated anomaly $\boldsymbol{S}$ from the original input $\boldsymbol{X}$, the model effectively isolates anomalies, resulting in an anomaly-free representation. This encoder-based framework ensures that the learned components are resilient to the influence of anomalies during factorization.

4.1.1. Symmetric encoder-decoder NMF and projective NMF

In traditional symmetric encoder–decoder frameworks, the model approximates the input data $\boldsymbol{X}$ using a product of two nonnegative matrices $\boldsymbol{W}$ and $\boldsymbol{H}$, and assumes that the encoder can be approximated by the transpose of the basis matrix. The factorization is given by:

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}, \tag{7}$$

$$\boldsymbol{H} \approx \boldsymbol{W}^\top\boldsymbol{X}. \tag{8}$$

The symmetric encoder–decoder constraint $\boldsymbol{H} \approx \boldsymbol{W}^\top\boldsymbol{X}$ yields the projective NMF [49] $\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{W}^\top\boldsymbol{X}$, which restricts the reconstruction operator to be symmetric positive semidefinite and (approximately) projective onto span($\boldsymbol{W}$). Equivalently, symmetry ties the analysis step to the synthesis dictionary, which is appropriate only under restrictive conditions (e.g., approximately orthonormal bases). In hyperspectral imagery, however, background spectra are highly correlated, and endmember signatures are typically non-orthogonal due to spectral similarity and linear mixing, so the transpose $\boldsymbol{W}^\top$ is generally not an optimal inverse mapping from spectra to latent coefficients or abundances. From a model-class viewpoint, the symmetric/projective model restricts the effective background operator to $\boldsymbol{W}\boldsymbol{W}^\top$, whereas the asymmetric formulation induces the more general low-rank operator $\boldsymbol{W}\boldsymbol{M}$ (still under nonnegativity), increasing expressiveness while preserving the interpretability of $\boldsymbol{W}$ as a background/endmember basis and $\boldsymbol{H}$ as nonnegative activations.

4.1.2. Asymmetric EDNMF

To overcome the limitations of the symmetric formulation, we introduce an asymmetric encoder–decoder framework in which the encoding matrix $\boldsymbol{M}$ is independently learned, instead of being tied to $\boldsymbol{W}^\top$. In this model, the input data $\boldsymbol{X}$ is factorized as:

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}, \quad \text{and} \quad \boldsymbol{H} \approx \boldsymbol{M}\boldsymbol{X}.$$

We decouple synthesis and analysis by learning an independent encoder $\boldsymbol{M}$. For fixed $\boldsymbol{H}$, estimating $\boldsymbol{M}$ corresponds to a (nonnegative) multivariate regression problem:

$$\min_{\boldsymbol{M}\geq 0} \|\boldsymbol{H} - \boldsymbol{M}\boldsymbol{X}\|_F^2.$$

This modification allows the encoder $\boldsymbol{M}$ to adapt more flexibly to the data, improving its capacity to model diverse data distributions. By decoupling the encoder from the basis matrix, the model gains the ability to learn more expressive and informative latent features. This flexibility is particularly beneficial when the underlying data structure is complex. The asymmetric encoder-decoder optimization is expressed as:

$$\min_{\boldsymbol{W},\boldsymbol{H},\boldsymbol{M}\geq 0} \|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}\|_F^2+\|\boldsymbol{H}-\boldsymbol{M}\boldsymbol{X}\|_F^2.$$

To expose the underlying structure, consider the unconstrained least-squares solution (ignoring the nonnegativity constraint for interpretation). $\boldsymbol{M}^\star$ is a data-dependent map and, in general, differs from $\boldsymbol{W}^\top$. Therefore, enforcing symmetry via $\boldsymbol{M}=\boldsymbol{W}^\top$ can introduce systematic bias in the inferred latent representation when $\boldsymbol{W}$ is not orthonormal.

4.1.3. Robust asymmetric EDNMF

To further enhance robustness against anomalies, we extend the robust asymmetric EDNMF formulation by introducing an anomaly matrix $\boldsymbol{S}$. The model is designed to separate the input matrix $\boldsymbol{X}$ into a clean background component and an anomaly component. The factorization is defined as:

$$\boldsymbol{X}\approx\boldsymbol{W}\boldsymbol{H}+\boldsymbol{S},\quad \text{and}\quad \boldsymbol{H}\approx\boldsymbol{M}(\boldsymbol{X}-\boldsymbol{S}).$$

In this setting, $\boldsymbol{W}\boldsymbol{H}$ captures the normal background structure, while $\boldsymbol{S}$ isolates the anomalous parts of the data. The model aims to estimate a robust representation of the input data by learning the matrices $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{M}$, and $\boldsymbol{S}$. The asymmetric EDNMF optimization problem is expressed as:

$$\min_{\boldsymbol{W},\boldsymbol{H},\boldsymbol{M},\boldsymbol{S}} \|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}-\boldsymbol{S}\|_F^2+\beta\|\boldsymbol{H}-\boldsymbol{M}(\boldsymbol{X}-\boldsymbol{S})\|_F^2,\quad \text{s.t.}\quad \boldsymbol{W},\boldsymbol{H},\boldsymbol{M}\geq 0, \tag{9}$$

where parameter β balances encoder and decoder reconstruction, with the subtraction of the anomaly matrix $\boldsymbol{S}$ from input $\boldsymbol{X}$ during encoding, ensuring that the latent features $\boldsymbol{H}$ are learned from clean data. This yields more accurate and interpretable representations. Incorporating $\boldsymbol{S}$ into the optimization enables the model to iteratively isolate anomalies, enhancing precision in anomaly detection and improving the decoder's reconstruction of normal data. The integration of RNMF with encoder–decoder matrix factorization boosts robustness, while the matrix $\boldsymbol{M}$ adapts feature extraction to the data structure. The sparse anomaly matrix $\boldsymbol{S}$ further sharpens the model's sensitivity to subtle anomalies. This synergy of robust matrix factorization and anomaly modeling results in a powerful system capable of learning meaningful representations and accurately identifying outliers in high-dimensional data, with applications in hyperspectral imaging and other anomaly-sensitive domains.

4.2. Enhancing sparsity with the $L_{2,1}$ norm

The L_1 norm is frequently employed as a convex approximation of the L_0 norm, which is seldom used due to its inherent optimization difficulties. However, when applied to input matrices with large and varying values, the L_1 norm may induce uneven sparsity across different parts of the data, heavily sparsifying some regions while leaving others relatively dense. This localized sparsity can fail to capture the global structure of anomalies, especially in high-dimensional data like hyperspectral images, ultimately resulting in suboptimal detection performance. To address this issue, we propose utilizing the $L_{2,1}$ norm, which first computes the L_2 norm across each sample (e.g., each pixel spectrum) and then promotes sparsity at the sample level. By enforcing a stronger and more consistent sparsity constraint, the $L_{2,1}$ norm yields solutions with fewer non-zero coefficients and better alignment with the sample-wise structure inherent in hyperspectral data, thereby enhancing the accuracy and robustness of anomaly detection.

The $L_{2,1}$ norm operates on a pixel-wise operation, allowing each pixel to either exhibit predominantly small (near-zero) magnitudes or relatively significant values. This facilitates the identification of anomalous pixels through the enforcement of sparsity. By applying sparsity at the pixel level, we are able to independently analyze each pixel, marking non-anomalous pixels with zero values in the matrix $\boldsymbol{S}$, while distinguishing the small subset of pixels identified as anomalies. This approach enhances the precision of anomaly detection, leading to more reliable anomaly identification in hyperspectral images.

This refined sparsity improves the separation between background data and anomalies, leading to more accurate anomaly detection. By integrating the $L_{2,1}$-norm regularization into the basic objective function, the base model is formulated as follows:

$$\min_{\boldsymbol{W},\boldsymbol{H},\boldsymbol{M},\boldsymbol{S}} \|\boldsymbol{X}-\boldsymbol{W}\boldsymbol{H}-\boldsymbol{S}\|_F^2+\beta\|\boldsymbol{H}-\boldsymbol{M}(\boldsymbol{X}-\boldsymbol{S})\|_F^2+\lambda\|\boldsymbol{S}\|_{2,1},\ \text{s.t. } \boldsymbol{W},\boldsymbol{M},\boldsymbol{H}\geq 0, \tag{10}$$

This formula represents the main objective function of our REDNMF model.

4.3. Optimization

The proposed non-convex model aims to decompose the data matrix $\boldsymbol{X}$ by introducing the matrices $\boldsymbol{W}$, $\boldsymbol{M}$, $\boldsymbol{H}$, and $\boldsymbol{S}$, formulating the problem as an optimization task. To solve this optimization problem, we adopt an alternating minimization strategy, iteratively updating each variable while keeping the others fixed. This approach facilitates convergence toward a good local minimum of the objective function. The update rules are derived using multiplicative updates, ensuring that the nonnegativity constraints of the model are inherently preserved. The objective function (10) is further expanded as follows, incorporating the trace terms and regularization:

$$\begin{aligned} J =&\mathrm{Tr}(\boldsymbol{X}^\top\boldsymbol{X}+\boldsymbol{H}^\top\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{H}+\boldsymbol{S}^\top\boldsymbol{S}-2\boldsymbol{X}^\top\boldsymbol{W}\boldsymbol{H}-2\boldsymbol{X}^\top\boldsymbol{S}+2\boldsymbol{H}^\top\boldsymbol{W}^\top\boldsymbol{S})\\ &+\beta\boldsymbol{H}^\top\boldsymbol{H}+\beta\boldsymbol{X}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{X}+\beta\boldsymbol{S}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{S}-2\beta\boldsymbol{H}^\top\boldsymbol{M}\boldsymbol{X}+2\beta\boldsymbol{H}^\top\boldsymbol{M}\boldsymbol{S}\\ &-2\beta\boldsymbol{X}^\top\boldsymbol{M}^\top\boldsymbol{M}\boldsymbol{S})+\lambda\|\boldsymbol{S}\|_{2,1}. \end{aligned}$$

The derivative of the objective function J concerning the matrix $\boldsymbol{W}$ is derived to ensure optimization and is expressed as follows:

$$\frac{\partial J}{\partial \boldsymbol{W}}=2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top-2\boldsymbol{X}\boldsymbol{H}^\top+2\boldsymbol{S}\boldsymbol{H}^\top \tag{11}$$

This formulation captures the gradient necessary for refining $\boldsymbol{W}$ during the optimization process. The derivative of the objective function J concerning the matrix $\boldsymbol{M}$ is computed to guide the optimization process. The resulting expression is as follows:

$$\begin{aligned} \frac{\partial J}{\partial \boldsymbol{M}} =&2\beta\boldsymbol{M}\boldsymbol{X}\boldsymbol{X}^\top+2\beta\boldsymbol{M}\boldsymbol{S}\boldsymbol{S}^\top-2\beta\boldsymbol{H}\boldsymbol{X}^\top+2\beta\boldsymbol{H}\boldsymbol{S}^\top\\ &-2\beta\boldsymbol{M}\boldsymbol{X}\boldsymbol{S}^\top-2\beta\boldsymbol{M}\boldsymbol{S}\boldsymbol{X}^\top \end{aligned} \tag{12}$$

This gradient formulation is crucial for updating $\boldsymbol{M}$ to minimize the objective function effectively. The gradient of the objective function J concerning $\boldsymbol{H}$ is derived to enable efficient optimization updates for $\boldsymbol{H}$. The gradient is expressed as follows:

$$\frac{\partial J}{\partial \boldsymbol{H}}=2\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{H}-2\boldsymbol{W}^\top\boldsymbol{X}+2\boldsymbol{W}^\top\boldsymbol{S}+2\beta\,\boldsymbol{H}-2\beta\,\boldsymbol{M}\boldsymbol{X}+2\beta\,\boldsymbol{M}\boldsymbol{S}. \tag{13}$$

The update rule for $\boldsymbol{W}$ is given by Eq. (14) to ensure non-negativity and optimality.

$$\boldsymbol{W}\leftarrow\boldsymbol{W}\circ\frac{\boldsymbol{X}\boldsymbol{H}^\top+\boldsymbol{S}^-\boldsymbol{H}^\top}{\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top+\boldsymbol{S}^+\boldsymbol{H}^\top}, \tag{14}$$

where we split the matrix $\boldsymbol{A}$ into its positive and negative components as

$$A_{ij}^+=\frac{|A_{ij}|+A_{ij}}{2},\qquad A_{ij}^-=\frac{|A_{ij}|-A_{ij}}{2}.$$

The update rule for $\boldsymbol{M}$, as expressed in Eq. (15), is designed to iteratively refine the encoder matrix for better representation.

$$\boldsymbol{M}\leftarrow\boldsymbol{M}\circ\frac{\boldsymbol{H}(\boldsymbol{X}+\boldsymbol{S}^-)^\top+\boldsymbol{M}(\boldsymbol{X}\boldsymbol{S}^{+\top}+\boldsymbol{S}^+\boldsymbol{X}^\top+[\boldsymbol{S}\boldsymbol{S}^\top]^-)}{\boldsymbol{M}(\boldsymbol{X}\boldsymbol{X}^\top+\boldsymbol{X}\boldsymbol{S}^{-\top}+\boldsymbol{S}^-\boldsymbol{X}^\top+[\boldsymbol{S}\boldsymbol{S}^\top]^+)+\boldsymbol{H}\boldsymbol{S}^{+\top}}. \tag{15}$$

Algorithm 1 Robust asymmetric Encoder-Decoder NMF (REDNMF).

Input: data matrix $\boldsymbol{X} \in \mathbb{R}^{d\times n}$, number of latent factors r, balance parameter β, sparseness parameter λ, and numbers of iterations *iter*;
Output: decoding matrix $\boldsymbol{W}$, encoding matrix $\boldsymbol{M}$, coefficient matrix $\boldsymbol{H}$, and anomaly matrix $\boldsymbol{S}$;

```
Initialize W, H, M randomly;
for t ← 1 to iter do
    Compute the residual matrix E ← X − WH;
    Update anomaly matrix S according to (17);
    Update decoding matrix W according to (14);
    Update encoding matrix M according to (15);
    Update coefficient matrix H according to (16);
end for
Return pixel anomaly scores a = BRAD(S);
```

The update rule for $\boldsymbol{H}$, provided in Eq. (16), adjusts the coefficient matrix to minimize reconstruction error while maintaining consistency with the encoded structure.

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \circ \frac{\boldsymbol{W}^\top \boldsymbol{X} + \boldsymbol{W}^\top \boldsymbol{S}^- + \beta \boldsymbol{M}\boldsymbol{X} + \beta \boldsymbol{M}\boldsymbol{S}^-}{\boldsymbol{W}^\top \boldsymbol{W}\boldsymbol{H} + \boldsymbol{W}^\top \boldsymbol{S}^+ + \beta \boldsymbol{H} + \beta \boldsymbol{M}\boldsymbol{S}^+}. \tag{16}$$

The update rule for the noise matrix $\boldsymbol{S}$, presented in Eq. (17), adjusts each element based on the residual matrix $\boldsymbol{E}$ and the regularization parameter λ. This process ensures that the sparsity constraint is applied effectively, setting elements to zero when their norms fall below the threshold λ and scaling others proportionally to maintain consistency.

$$S_{ij} = \begin{cases} E_{ij} - \lambda \frac{E_{ij}}{\|E_{i,:}\|}, & \|E_{i,:}\| > \lambda \\ 0, & \|E_{i,:}\| \le \lambda. \end{cases} \tag{17}$$

The detailed steps and procedural flow of the proposed method are thoroughly described and illustrated in Algorithm 1. This algorithm summarizes the main components and update rules, providing a clear and concise overview of the entire approach. The BRAD [45] is a background reconstruction–based anomaly detector that produces pixel-level anomaly scores by modeling local spatial context through the mean and covariance of neighboring pixels in the anomaly matrix $\boldsymbol{S}$. To support reproducibility and further exploration, the implementation code is publicly available at https://github.com/shirin-moradi/REDNMF.

4.4. Convergence analysis

In this subsection, we show that the multiplicative update in (14) yields a *monotonic non-increase* of the objective function with respect to $\boldsymbol{W}$ when $\boldsymbol{H}$, $\boldsymbol{M}$ and $\boldsymbol{S}$ are fixed, following the auxiliary-function strategy of Lee and Seung [30].

Subproblem in $\boldsymbol{W}$. When $\boldsymbol{H}$ and $\boldsymbol{S}$ are fixed, all terms in J that do not depend on $\boldsymbol{W}$ are constants and can be ignored. The $\boldsymbol{W}$-dependent part of the objective can be written as

$$J(\boldsymbol{W}) = \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H} - \boldsymbol{S}\|_F^2 = \|(\boldsymbol{X} - \boldsymbol{S}) - \boldsymbol{W}\boldsymbol{H}\|_F^2, \tag{18}$$

where we define $\boldsymbol{A} = \boldsymbol{X} - \boldsymbol{S}$ (note that $\boldsymbol{A}$ can be sign-indefinite). Expanding (18) gives

$$J(\boldsymbol{W}) = \mathrm{Tr}\big(\boldsymbol{A}^\top \boldsymbol{A}\big) - 2\,\mathrm{Tr}\big(\boldsymbol{A}^\top \boldsymbol{W}\boldsymbol{H}\big) + \mathrm{Tr}\big(\boldsymbol{H}^\top \boldsymbol{W}^\top \boldsymbol{W}\boldsymbol{H}\big). \tag{19}$$

The gradient with respect to $\boldsymbol{W}$ is

$$\nabla_{\boldsymbol{W}} J = 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top - 2(\boldsymbol{X} - \boldsymbol{S})\boldsymbol{H}^\top = 2\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top - 2\boldsymbol{X}\boldsymbol{H}^\top + 2\boldsymbol{S}\boldsymbol{H}^\top. \tag{20}$$

Positive/negative splitting. Since $\boldsymbol{S}$ (and thus $\boldsymbol{A}$) may contain negative entries, we use the standard decomposition

$$\boldsymbol{S} = \boldsymbol{S}^+ - \boldsymbol{S}^-, \qquad S_{ij}^+ = \frac{|S_{ij}| + S_{ij}}{2}, \quad S_{ij}^- = \frac{|S_{ij}| - S_{ij}}{2}, \tag{21}$$

so that $\boldsymbol{S}^+, \boldsymbol{S}^- \ge 0$ and $\boldsymbol{X} - \boldsymbol{S} = \boldsymbol{X} - \boldsymbol{S}^+ + \boldsymbol{S}^-$. With this, we can write the (element-wise) stationarity/KKT condition for nonnegativity ($W_{ik} \ge 0$) in the Lee and Seung [30] form by separating the gradient into positive and negative parts:

$$\big[\nabla_{\boldsymbol{W}} J\big]_{ik} = 2\Big(\underbrace{\big[\boldsymbol{W}\boldsymbol{H}\boldsymbol{H}^\top + \boldsymbol{S}^+\boldsymbol{H}^\top\big]_{ik}}_{\text{positive part}} - \underbrace{\big[\boldsymbol{X}\boldsymbol{H}^\top + \boldsymbol{S}^-\boldsymbol{H}^\top\big]_{ik}}_{\text{negative part}} \Big). \tag{22}$$

Auxiliary function and monotonicity. Define an auxiliary function $G(\boldsymbol{W}, \boldsymbol{W}^{(t)})$ for $J(\boldsymbol{W})$ such that

$$G(\boldsymbol{W}, \boldsymbol{W}) = J(\boldsymbol{W}), \qquad G(\boldsymbol{W}, \boldsymbol{W}^{(t)}) \ge J(\boldsymbol{W}) \quad \forall\, \boldsymbol{W} \ge 0. \tag{23}$$

Following [30], one can construct G by majorizing the quadratic term $\mathrm{Tr}(\boldsymbol{H}^\top \boldsymbol{W}^\top \boldsymbol{W}\boldsymbol{H})$ using Jensen-type inequalities, yielding a separable upper bound over the entries W_{ik}. Minimizing this auxiliary function with respect to $\boldsymbol{W}$ gives the multiplicative update

$$\boldsymbol{W}^{(t+1)} \leftarrow \boldsymbol{W}^{(t)} \circ \frac{\boldsymbol{X}\boldsymbol{H}^\top + \boldsymbol{S}^-\boldsymbol{H}^\top}{\boldsymbol{W}^{(t)}\boldsymbol{H}\boldsymbol{H}^\top + \boldsymbol{S}^+\boldsymbol{H}^\top}, \tag{24}$$

which is exactly (14). By the auxiliary-function property (23), we have

$$J(\boldsymbol{W}^{(t+1)}) \le G(\boldsymbol{W}^{(t+1)}, \boldsymbol{W}^{(t)}) \le G(\boldsymbol{W}^{(t)}, \boldsymbol{W}^{(t)}) = J(\boldsymbol{W}^{(t)}), \tag{25}$$

thus the objective is non-increasing after each $\boldsymbol{W}$-update when $\boldsymbol{H}$ and $\boldsymbol{S}$ are fixed. In addition, because the update is multiplicative, nonnegativity is preserved automatically: if $\boldsymbol{W}^{(t)} \ge 0$, then $\boldsymbol{W}^{(t+1)} \ge 0$. The above analysis shows that the $\boldsymbol{W}$-update monotonically decreases the objective function when $\boldsymbol{H}$ and $\boldsymbol{S}$ are fixed. Since the algorithm alternately updates $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{M}$, and $\boldsymbol{S}$, and each subproblem is solved in a descent manner, the overall objective value is non-increasing and the algorithm converges to a stationary point. The update of $\boldsymbol{S}$ is obtained via a row-wise soft-thresholding operator, which exactly solves the $L_{2,1}$-regularized subproblem and guarantees objective decrease [70].

4.5. Complexity analysis

This section provides a detailed analysis of the time and space complexity of the proposed optimization algorithm. Let $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ denote the hyperspectral data matrix, where m is the number of spectral bands and n is the number of pixels, with $m \ll n$. At each iteration, the proposed algorithm updates the variables $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{M}$, and $\boldsymbol{S}$ using multiplicative update rules. The updates of $\boldsymbol{W}$, $\boldsymbol{H}$, and $\boldsymbol{S}$ are dominated by matrix multiplications involving $\boldsymbol{X}$, $\boldsymbol{W}$, and $\boldsymbol{H}$, resulting in a computational complexity of $O(mnr)$, where r denotes the latent dimension and $r \ll m, n$. The update of the encoder matrix $\boldsymbol{M}$ involves computing matrix products such as $\boldsymbol{X}\boldsymbol{X}^\top$, $\boldsymbol{S}\boldsymbol{S}^\top$, and $\boldsymbol{X}\boldsymbol{S}^\top$, each requiring $O(m^2 n)$ operations. Since the number of spectral bands m is typically much smaller than the number of pixels n in hyperspectral imagery, this term dominates the overall computational cost. Consequently, the per-iteration time complexity of the proposed method is $O(m^2 n)$, and the total computational complexity over T iterations is $O(Tm^2 n)$. In terms of memory consumption, the dominant storage requirements arise from the data matrix and the factor matrices, including $\boldsymbol{X}$, $\boldsymbol{W}$, $\boldsymbol{H}$, $\boldsymbol{M}$, and $\boldsymbol{S}$. The overall space complexity is therefore $O(mn + mr + rn)$, which is linear with respect to the data size and suitable for large-scale hyperspectral anomaly detection tasks.

5. Experiments

To evaluate the performance of the proposed Robust Asymmetric Encoder–Decoder NMF (REDNMF) model for hyperspectral anomaly

detection, we conducted experiments on a variety of real and synthetic hyperspectral datasets. These datasets differ in spatial resolution, spectral dimensionality, and anomaly characteristics, providing realistic and diverse testing conditions. The REDNMF model is benchmarked against several representative methods, including matrix factorization-based approaches (NMF [28], RNMF [40], L1NMF [45], and EWNMF [71]), advanced low-rank and collaborative representation-based detectors (LSUNRSORAD and LSAD-CR-IDW [24]), and deep learning-based methods, including the multi-scale residual convolutional autoencoder (MSNet) [62] and the diffusion-based background suppression model (BSDM) [65]. This comparison enables a comprehensive evaluation across different algorithmic paradigms. Standard evaluation metrics are used to assess detection accuracy and false alarm rate. Additionally, parameter sensitivity analysis is conducted to examine the effects of regularization coefficients β and λ, and the convergence behavior of the optimization process is analyzed to ensure stability and efficiency. Experimental results demonstrate the accuracy, robustness, and adaptability of REDNMF across diverse hyperspectral anomaly detection scenarios.

5.1. Evaluation metrics

To quantitatively assess the anomaly detection performance of the proposed REDNMF model, two widely recognized metrics are employed: the Receiver Operating Characteristic (ROC) curve and the Area Under the ROC Curve (AUC). The ROC curve depicts the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) over a range of decision thresholds, providing a comprehensive view of the model's discriminative ability. The AUC is a scalar value summarizing the ROC curve. A higher AUC value indicates a greater ability to distinguish anomalous pixels from the background, with a maximum of 1.0 representing perfect separation and a value of 0.5 corresponding to random performance. These metrics are particularly suitable for hyperspectral anomaly detection, where robust target–background discrimination is critical.

5.2. Results

To ensure a comprehensive evaluation of the proposed method, experiments were conducted on two distinct categories of hyperspectral datasets: (1) synthetic datasets with well-defined and controlled anomalies, and (2) real-world datasets containing diverse anomaly types embedded within complex backgrounds. This dual evaluation strategy provides a rigorous test of the model's robustness, as it assesses performance under both idealized and realistic conditions. The chosen datasets encompass a broad range of application scenarios, from urban environments and coastal monitoring to aerial surveillance and simulated anomaly-rich settings. For each dataset, key attributes, including spatial resolution, number of spectral bands, anomaly type, and contextual complexity, are summarized in Table 1. This selection aims to capture varying data distributions and anomaly characteristics, enabling a systematic demonstration of the REDNMF model's accuracy, generalization capacity, and resilience in detecting anomalies across diverse spectral-spatial contexts.

5.2.1. Synthetic dataset evaluation

The Salinas dataset, a widely recognized benchmark for hyperspectral anomaly detection, was employed for the controlled experiments. It consists of hyperspectral imagery with 120×120 spatial pixels and 204 spectral bands. Anomalies are artificially injected to serve as a precisely defined ground truth, enabling an accurate and quantitative assessment of detection performance. Such controlled conditions allow direct, fair comparisons between competing algorithms while minimizing confounding factors from environmental variability. Fig. 2 reports the 3D ROC curves of REDNMF and the competing methods on the synthetic dataset, where the x-axis denotes the FPR, the y-axis represents the decision threshold, and the z-axis denotes the TPR. As shown, REDNMF achieves an AUC of 0.9910, which is substantially higher than the conventional matrix-factorization baselines, including NMF (0.9582), RNMF (0.9487), and EWNMF (0.9447), and also improves upon the sparsity-regularized variant L1NMF (0.9823). In particular, REDNMF yields consistent gains of $+0.0328$, $+0.0423$, and $+0.0463$ AUC over NMF, RNMF, and EWNMF, respectively, indicating stronger discrimination between anomalies and background under controlled conditions. Beyond the overall AUC, the ROC curve of REDNMF is visibly closer to the top-left corner, demonstrating a more favorable sensitivity-specificity trade-off. Specifically, REDNMF attains a high true positive rate at very low false positive rates, which is critical for anomaly detection scenarios where even a small increase in false alarms can be costly. Compared with deep learning–based methods such as MSNet (AUC = 0.9952) and BSDM (AUC = 0.9930), REDNMF (AUC = 0.9910)

Table 1
Detailed specifications of the hyperspectral anomaly detection (HAD) datasets used in this study, including spatial dimensions, spectral band counts, and the types of anomalies present. These datasets span both synthetic and real-world scenarios, enabling a comprehensive evaluation of the proposed REDNMF method across controlled environments and complex natural scenes.

Datasets	Size	**Anomaly type**
Salinas	$120 \times 120 \times 204$	Simulated Targets
ABU-Airport-1	$100 \times 100 \times 205$	airplanes
ABU-Airport-4	$100 \times 100 \times 191$	airplanes
ABU-Beach-1	$150 \times 150 \times 188$	Ship
ABU-Beach-2	$100 \times 100 \times 193$	Ship
ABU-Beach-3	$100 \times 100 \times 188$	Ship
ABU-Beach-4	$150 \times 150 \times 102$	Ship
ABU-Urban-1	$100 \times 100 \times 207$	houses
ABU-Urban-2	$100 \times 100 \times 207$	houses
ABU-Urban-3	$100 \times 100 \times 191$	houses
ABU-Urban-4	$100 \times 100 \times 191$	houses
ABU-Urban-5	$100 \times 100 \times 205$	houses
AVIRIS	$100 \times 100 \times 189$	airplanes
AVIRIS-sunlight	$100 \times 100 \times 189$	airplanes
Pavia	$150 \times 150 \times 102$	vehicles
HyMap	$280 \times 800 \times 126$	vehicles & fabric

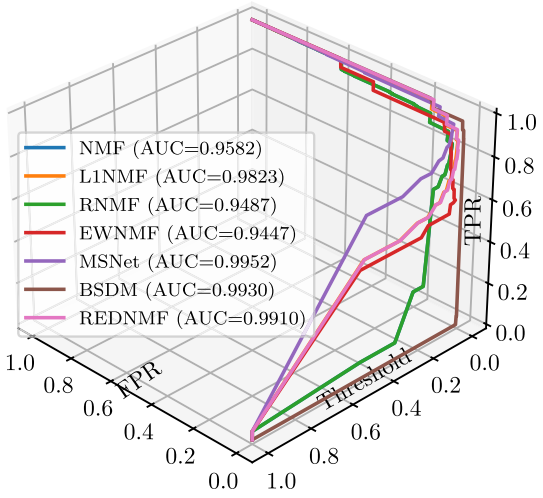

Fig. 2. 3D ROC curves of the REDNMF and the compared methods on the synthetic dataset.

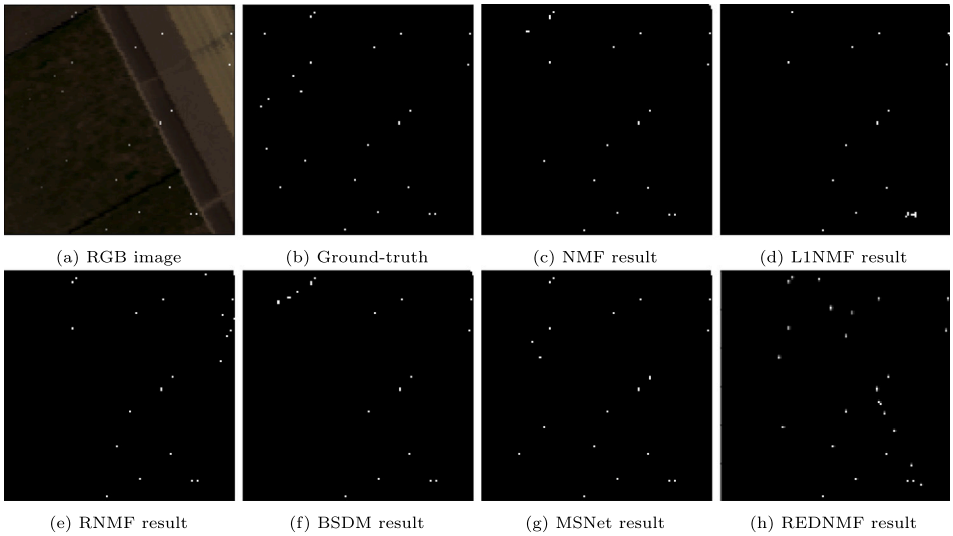

Fig. 3. Qualitative comparison of anomaly detection results on the Salinas-Synthetic dataset. (a) RGB composite image, (b) ground-truth anomaly map, and detection results obtained by (c) NMF, (d) L1NMF, (e) RNMF, (f) BSDM, (g) MSNet, and (h) the proposed REDNMF method.

remains highly competitive, exhibiting only a marginal performance gap while maintaining a similarly steep rise in the low-FPR regime. These results demonstrate that REDNMF consistently outperforms classical NMF-family methods and achieves performance comparable to deep learning models on the synthetic benchmark.

Qualitative Detection Results: Fig. 3 presents the qualitative anomaly detection results of different methods on the synthetic Salinas dataset. The ground-truth map shows sparsely distributed point-like anomalies embedded within a relatively homogeneous background, including several weak targets that are difficult to detect. Among the classical NMF-based methods (NMF, L1NMF, and RNMF), the detection maps exhibit either incomplete target recovery or noticeable background interference. In particular, basic NMF and RNMF tend to produce residual background responses, while L1NMF improves sparsity but still misses some subtle anomalies. BSDM and MSNet generate cleaner detection maps with stronger suppression of background noise; however, they may fail to capture certain low-contrast or isolated targets. In comparison, the proposed REDNMF produces detection results that more closely match the ground-truth distribution. It successfully recovers most of the sparse anomaly points while maintaining clear background suppression. The detected anomalies exhibit better spatial consistency and sharper localization, reflecting the effectiveness of explicit anomaly modeling and pixel-wise structured sparsity. Although a few isolated responses remain, the overall detection pattern of REDNMF demonstrates a more balanced trade-off between sensitivity and false alarm suppression. These qualitative results further validate the robustness and discriminative capability of the proposed framework on the synthetic benchmark.

5.2.2. Real-world datasets

To thoroughly evaluate the robustness, adaptability, and real-world applicability of the proposed REDNMF framework, we performed extensive experiments on a series of real-world hyperspectral datasets from the well-established benchmark collection. This benchmark is widely recognized for its high complexity and diversity, encompassing varied environmental conditions and imaging scenarios that challenge anomaly detection algorithms. From the public repositories, we selected multiple representative scenes. Additionally, we included the HyMap dataset, a large-scale hyperspectral image with a sparse distribution of anomaly pixels, posing a challenging scenario for anomaly detection. These cover a wide range of real-world environments, including dense urban areas, complex airport infrastructures, and dynamic coastal zones. Each scene possesses unique spectral signatures, spatial structures, and anomalous patterns, such as small aircraft, ships, industrial facilities, and other man-made objects, providing an ideal testing ground for assessing generalization and detection accuracy under challenging operational conditions.

ROC Curve Analysis: Fig. 4 presents the 3D ROC curves of REDNMF and the competing methods across multiple real-world hyperspectral datasets. The three-dimensional visualization illustrates the relationship among the false positive rate (FPR), decision threshold, and true positive rate (TPR), providing a comprehensive view of detection behavior across varying operating points. Overall, REDNMF exhibits consistently favorable ROC characteristics, with surfaces positioned closer to the ideal region corresponding to high TPR and low FPR. In most datasets, REDNMF achieves superior or highly competitive AUC values compared with classical NMF-based methods (NMF, L1NMF, RNMF, and EWNMF), demonstrating its robustness under complex real-world conditions. In particular, on challenging scenes such as Airport-1 and Urban-3, REDNMF achieves noticeable performance improvements, reflecting its enhanced ability to suppress background clutter while preserving anomaly responses. Compared with deep learning–based approaches, including MSNet and BSDM, REDNMF remains highly competitive. Although deep models may attain slightly higher AUC values on certain datasets, the performance gap is generally marginal. Moreover, REDNMF maintains a steep rise in the low-FPR region, which is especially important in practical anomaly detection scenarios where controlling false alarms is critical. The consistent ROC behavior across diverse environments, including airport, urban, coastal, and airborne scenes, demonstrates that REDNMF generalizes well to varying spectral-spatial complexities. These results confirm that the proposed framework not only outperforms traditional factorization-based detectors but also achieves performance comparable to advanced deep learning methods in real-world hyperspectral anomaly detection tasks.

Quantitative Performance: Table 2 reports the AUC values of REDNMF and competing methods across ten real-world hyperspectral datasets. Overall, REDNMF achieves the most consistent and competitive performance, outperforming classical detectors (LSUNRSORAD, LSAD-CR-IDW) and conventional NMF-based approaches (NMF, L1NMF, RNMF, EWNMF), while remaining highly competitive with deep learning models such as MSNet and BSDM. On challenging airport scenes, REDNMF attains 0.9357 (Airport-1) and 0.9826 (Airport-4), demonstrating strong robustness to structural complexity. For beach datasets, it achieves

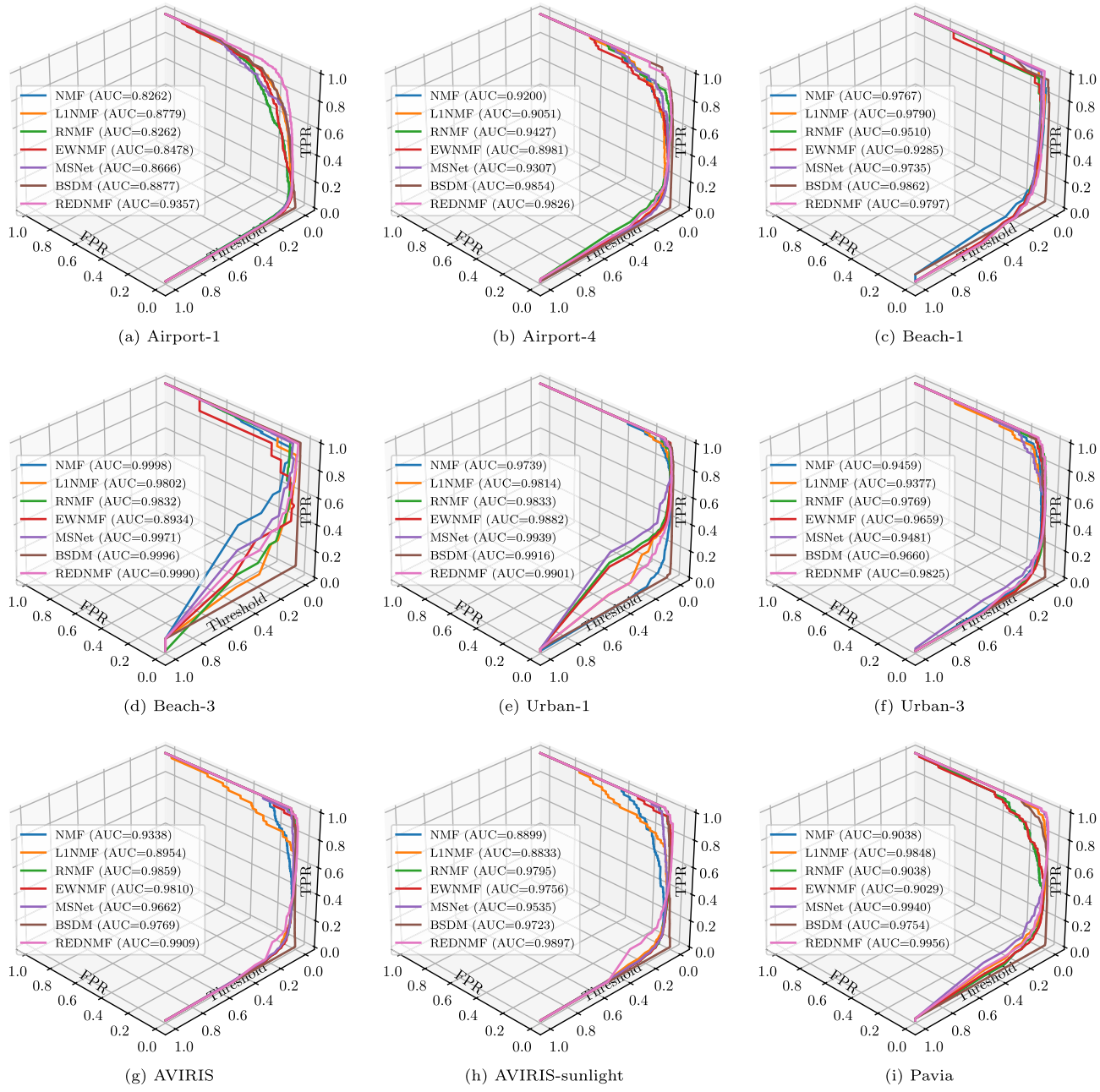

Fig. 4. 3D ROC curves of REDNMF and the comparison methods on nine real-world datasets.

Table 2
Evaluation of hyperspectral anomaly detection algorithms using real-world datasets.

Method	Airport-1	Airport-4	Beach-1	Beach-3	Beach-4	Urban-1	Urban-3	AVIRIS	AVIRIS-sun	Pavia	HyMap	Avg
LSUNRSORAD	0.6136	0.5313	0.7961	0.8834	0.8143	0.6153	0.8877	0.8815	0.8658	0.8693	0.4891	0.7498
LSAD-CR-IDW	0.5060	0.4994	0.7777	0.8845	0.5605	0.5626	0.6435	0.6373	0.6216	0.6251	0.4372	0.6141
NMF	0.8451	0.9200	0.9767	**0.9998**	0.9074	0.9739	0.9458	0.9338	0.8899	0.9038	0.6615	0.9052
L1NMF	0.8823	0.9051	0.9790	0.9802	0.9842	0.9814	0.9377	0.8954	0.8832	0.9855	0.7043	0.9198
RNMF	0.8338	0.9427	0.9510	0.9832	0.9845	0.9833	0.9769	0.9859	0.9794	0.8990	0.7289	0.9317
EWNMF	0.8478	0.8981	0.9285	0.8934	0.9082	0.9881	0.9658	0.9810	0.9755	0.9035	0.6437	0.9031
MSNet	0.8666	0.9307	0.9735	0.9973	0.9952	**0.9953**	0.9481	0.9662	0.9535	0.9940	0.8160	0.9488
BSDM	0.8877	0.9822	**0.9862**	0.9995	0.9743	0.9897	0.9615	0.9728	0.9724	0.9719	0.7608	0.9508
REDNMF	**0.9357**	**0.9826**	0.9797	0.9990	**0.9955**	0.9901	**0.9825**	**0.9909**	**0.9897**	**0.9958**	**0.8243**	**0.9696**

0.9797 (Beach-1), 0.9955 (Beach-4), and near-saturated performance on Beach-3 (0.9990). In urban environments, REDNMF maintains high accuracy with 0.9901 (Urban-1) and 0.9825 (Urban-3). It also performs strongly on benchmark datasets, reaching 0.9909 (AVIRIS), 0.9897 (AVIRIS-sunlight), and 0.9958 (Pavia). On the HyMap dataset, REDNMF achieves an AUC of 0.8243. Notably, all methods exhibit relatively lower performance on this dataset, reflecting the increased difficulty posed by its large spatial scale and highly sparse anomaly distribution. Despite

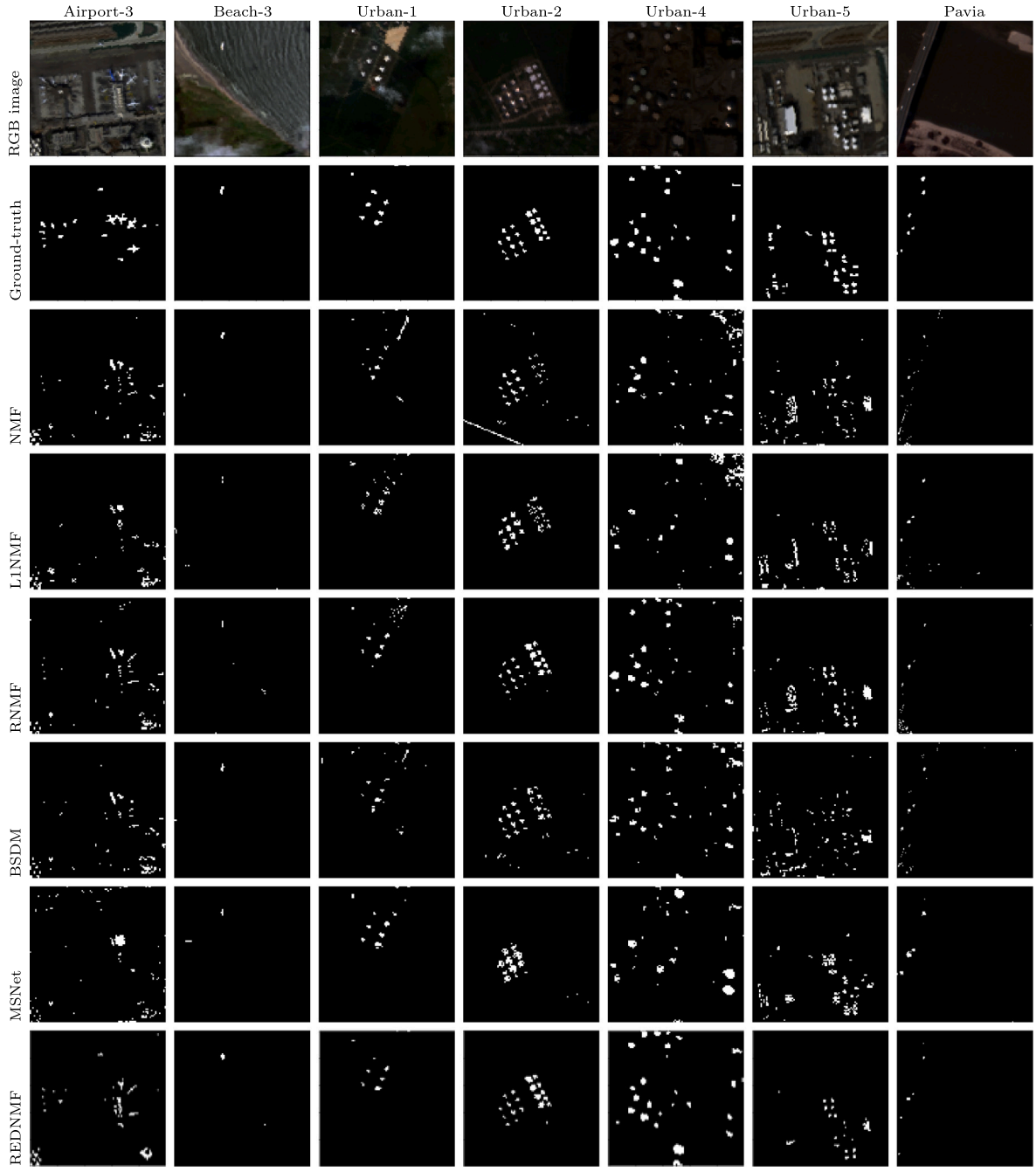

Fig. 5. Qualitative comparison of anomaly detection results on Airport-3, Beach-3, Urban-1, Urban-2, Urban-4, Urban-5, and Pavia datasets. From top to bottom: RGB images, ground truth maps, and detection results obtained by NMF, L1NMF, RNMF, BSDM, MSNet, and the proposed REDNMF method.

these challenges, REDNMF remains competitive, highlighting its robustness in handling complex hyperspectral scenes with limited anomaly presence. Averaged across all datasets, REDNMF obtains the highest mean AUC, confirming its robustness and generalization ability. These results demonstrate that the proposed method consistently surpasses traditional factorization-based detectors and achieves performance comparable to advanced deep learning approaches in diverse real-world scenarios.

Qualitative Detection Results: Fig. 5 presents qualitative comparisons of anomaly detection results across several representative real-world scenes, including airport, beach, urban, and Pavia datasets. These scenes encompass diverse background conditions, ranging from relatively smooth coastal regions to highly structured and cluttered urban environments, providing a comprehensive evaluation of detection robustness. Classical NMF-based methods often exhibit either incomplete anomaly recovery or residual background responses. In complex scenes such as airport and urban areas, these methods tend to confuse structural edges and textured regions with anomalies, leading to scattered false activations or blurred anomaly shapes. Although sparsity-regularized variants improve background suppression to some extent, subtle or low-contrast targets are occasionally missed. Deep learning approaches, including MSNet and BSDM, generally produce cleaner anomaly maps with stronger background suppression. However, in certain cases they may oversmooth fine details or partially suppress weak and spatially small targets, especially when anomalies exhibit limited spectral contrast relative to their surroundings. In comparison, REDNMF

Table 3
Robustness analysis under additive Gaussian noise. AUC values are reported for three noise intensities.

$\sigma = 0.01$ (Low Noise)											
Method	Airport-1	Airport-4	Beach-1	Beach-3	Beach-4	Urban-1	Urban-3	AVIRIS	AVIRIS-sun	Pavia	HyMap
NMF	0.8302	0.9066	0.9661	0.9880	0.8951	0.9590	0.9320	0.9185	0.8794	0.8670	0.6417
L1NMF	0.8659	0.8957	0.9688	0.9727	0.9730	0.9721	0.9234	0.8840	0.8684	0.9734	0.6839
RNMF	0.8179	0.9321	0.9436	0.9758	0.5766	0.9709	0.9643	0.9749	0.9662	0.8889	0.7068
EWNMF	0.8329	0.8882	0.9198	0.8870	0.8991	0.9804	0.9536	0.9698	0.9619	0.8909	0.6234
MSNet	0.8589	0.9230	0.9681	0.9911	**0.9900**	**0.9832**	0.9359	0.9543	0.9424	0.9846	0.7916
BSDM	0.8792	0.9717	0.8900	**0.9944**	0.9688	0.9800	0.9556	0.9669	0.9610	0.9666	0.7362
REDNMF	**0.9255**	**0.9722**	**0.9723**	0.9943	0.9856	0.9814	**0.9721**	**0.9845**	**0.9838**	**0.9882**	**0.8069**
$\sigma = 0.05$ (Moderate Noise)											
Method	Airport-1	Airport-4	Beach-1	Beach-3	Beach-4	Urban-1	Urban-3	AVIRIS	AVIRIS-sun	Pavia	HyMap
NMF	0.7916	0.8683	0.9378	0.9681	0.8633	0.9304	0.9030	0.8929	0.8423	0.8383	0.6083
L1NMF	0.8396	0.8639	0.9436	0.9528	0.9519	0.9417	0.8997	0.8599	0.8418	0.9526	0.6615
RNMF	0.7932	0.9030	0.9166	0.9554	0.5544	0.9462	0.9435	0.9510	0.9410	0.8703	0.6732
EWNMF	0.7999	0.8587	0.8960	0.8631	0.8740	0.9494	0.9279	0.9454	0.9337	0.8737	0.5968
MSNet	0.8272	0.8970	0.9420	0.9756	0.9646	**0.9653**	0.9135	0.9351	0.9238	0.9675	0.7537
BSDM	0.8580	0.9479	0.8742	**0.9795**	0.9485	0.9575	0.9332	0.9457	0.9439	0.9488	0.7019
REDNMF	**0.9020**	**0.9527**	**0.9528**	0.9773	**0.9694**	0.9622	**0.9520**	**0.9664**	**0.9614**	**0.9708**	**0.7714**
$\sigma = 0.10$ (High Noise)											
Method	Airport-1	Airport-4	Beach-1	Beach-3	Beach-4	Urban-1	Urban-3	AVIRIS	AVIRIS-sun	Pavia	HyMap
NMF	0.7336	0.8224	0.8939	0.9319	0.8270	0.8830	0.8544	0.8454	0.7930	0.7996	0.5712
L1NMF	0.7923	0.8234	0.9067	0.9281	0.9109	0.9060	0.8627	0.8194	0.8018	0.9216	0.6187
RNMF	0.7480	0.8628	0.8840	0.9333	0.5174	0.9128	0.9004	0.9154	0.9025	0.8392	0.6369
EWNMF	0.7557	0.8145	0.8596	0.8361	0.8353	0.9103	0.8905	0.9097	0.8963	0.8346	0.5531
MSNet	0.7944	0.8634	0.9178	0.9505	0.9392	0.9296	0.8877	0.9099	0.8905	0.9373	0.7124
BSDM	0.8191	0.9194	0.8497	**0.9603**	0.9187	0.9276	0.9038	0.9148	0.9112	0.9246	0.6586
REDNMF	**0.8671**	**0.9241**	**0.9272**	0.9569	**0.9430**	**0.9320**	**0.9302**	**0.9368**	**0.9306**	**0.9509**	**0.7328**

generates detection maps that more closely resemble the ground-truth distributions in terms of sparsity, localization accuracy, and structural consistency. The explicit anomaly modeling and pixel-wise structured sparsity enable effective separation of background and anomalous components, even under heterogeneous textures and illumination variations. Anomalies are sharply localized with limited background interference, while maintaining sensitivity to subtle spectral deviations. Overall, the qualitative results demonstrate that REDNMF achieves a favorable balance between sensitivity and false-alarm control, producing coherent and well-localized anomaly maps across diverse and challenging real-world scenarios.

5.3. Robustness analysis

To evaluate the stability of the proposed REDNMF under degraded sensing conditions, we conducted robustness experiments by injecting additive Gaussian noise into the normalized hyperspectral data, i.e., $\boldsymbol{X}_{\text{noisy}} = \boldsymbol{X} + \mathcal{N}(0, \sigma^2)$, with noise intensities $\sigma \in \{0.01, 0.05, 0.10\}$. These levels correspond to mild, moderate, and strong corruption, respectively. Table 3 reports the AUC results for all compared methods across the real-world datasets. As expected, detection performance gradually decreases as noise intensity increases; however, the degradation behavior is nonlinear, with relatively small performance drops at low noise and more noticeable declines at higher intensities. Compared with classical matrix factorization methods (NMF, L1NMF, RNMF, and EWNMF), REDNMF exhibits consistently smaller AUC reductions, particularly under moderate and strong noise. This improved stability can be attributed to three key design elements: (i) the explicit anomaly matrix, which prevents corruption from contaminating the low-rank background representation; (ii) the asymmetric encoder–decoder structure that learns latent features from the cleaned input $\boldsymbol{X} - \boldsymbol{S}$; and (iii) the pixel-wise structured sparsity enforced via the $L_{2,1}$ norm, which aligns with the spectral structure of hyperspectral anomalies. Deep learning methods such as MSNet and BSDM also demonstrate competitive robustness; nevertheless, REDNMF maintains slightly more consistent performance across heterogeneous scenes, indicating stronger algorithmic resilience to noise. Overall, these results confirm that the proposed framework achieves stable and reliable anomaly detection under varying noise conditions.

5.4. Parameter analysis

To better understand the robustness and stability of the proposed REDNMF method, we conducted a comprehensive sensitivity analysis with respect to its two key regularization parameters, λ and β. These parameters control the balance between reconstruction fidelity and the imposed regularization constraints, and thus play a crucial role in the trade-off between model flexibility and generalization capability. The evaluation was carried out by systematically varying λ over the set $\{0.001, 0.01, 0.1, 1, 10, 50, 100, 200, 500\}$ and β over the set $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$. For each (λ, β) pair, we computed the detection accuracy of REDNMF on six representative hyperspectral datasets: Salinas, Airport-1, Airport-4, Beach-1, Urban-1, and Urban-3. These datasets were selected to cover both synthetic and real-world acquisition scenarios, ensuring that the conclusions drawn from this analysis are not dataset-specific but instead reflect the general behavior of the method.

The results, presented in Fig. 6, reveal distinct patterns across datasets. For most datasets, the model achieves high detection accuracy within moderately sized ranges of λ and β, particularly when $\lambda \in [0.1, 100]$ and $\beta \in [1, 1000]$. This region corresponds to a balance between enforcing the desired constraints and allowing sufficient flexibility for data adaptation. Conversely, extreme parameter settings, either too small or too large, tend to cause noticeable performance degradation. Extremely small values may under-regularize the model, leading to overfitting and noisy detection maps, while excessively large values may over-constrain the model, hindering its ability to adapt to the spectral and spatial variability of the target scenes. Interestingly, while the optimal ranges are generally consistent across datasets, subtle variations can be observed. For example, Airport-1 and Airport-4 exhibit slightly narrower optimal regions compared to Beach-1, which appears more tolerant to a wider range of β values. Such differences may be attributed to the inherent spectral complexity and background variability present

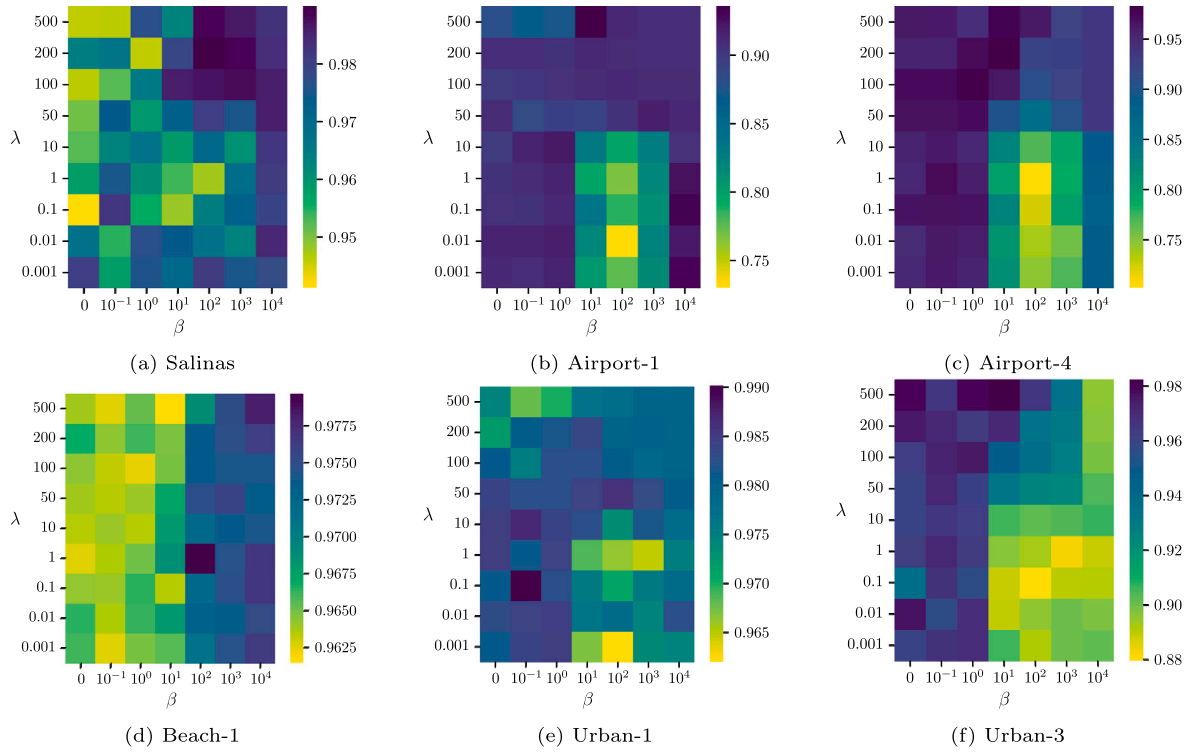

Fig. 6. Effect of different β and λ values on the AUC performance of REDNMF.

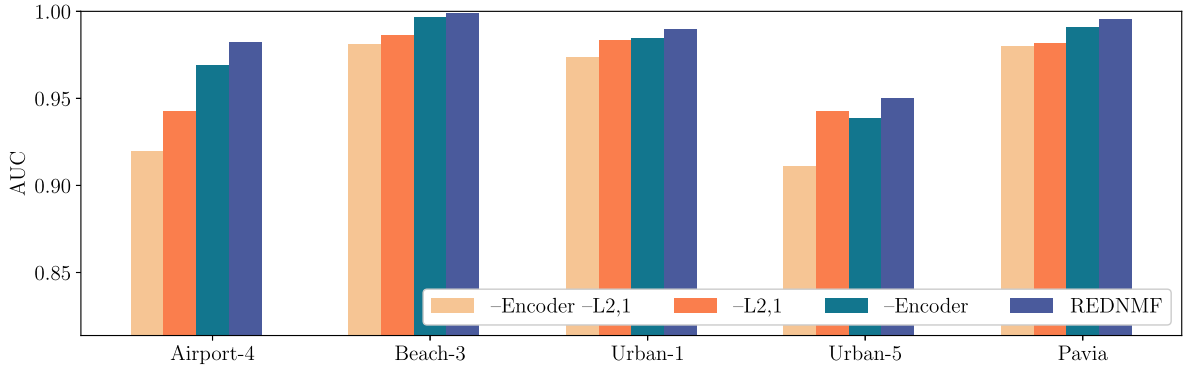

Fig. 7. Ablation study of REDNMF evaluating the contributions of the encoder and the $L_{2,1}$ sparsity regularization by progressively removing each component from the full model.

in each dataset. Fig. 6 shows that REDNMF is robust to λ and β within a practical range: high AUC is consistently achieved for moderate λ (about 10–100) and small-to-moderate β (about 10^{-1}–10^{1}), while extreme values degrade performance. Thus, exhaustive tuning is unnecessary; we use $(\lambda, \beta) = (50, 1)$ as a stable default.

5.5. Ablation study

We evaluate the contribution of each component in REDNMF through four ablation variants that selectively remove the encoder module and the $L_{2,1}$ sparsity constraint. AUC results of six real-world datasets are shown in Fig. 7. The –Encoder –L2,1 variant removes both components, leaving only a basic decoder-side NMF with anomaly learning. This configuration produces the poorest results across all metrics. Without a learned encoder or structured sparsity, the model fails to capture discriminative representations or effectively isolate anomalous components. The –L2,1 variant keeps the encoder–decoder architecture but removes $L_{2,1}$ regularization. Performance improves over the first variant, yet remains suboptimal. Without structured sparsity to enforce row-wise sparse anomaly patterns, the model separates anomalies less effectively. Architectural capacity alone proves insufficient for robust anomaly learning. The –Encoder variant removes the encoder while retaining $L_{2,1}$ sparsity regularization. Here, latent representations come solely from decoder-side factorization, with anomalies constrained through structured sparsity. Performance surpasses both variants lacking $L_{2,1}$ regularization. Sparsity plays a critical role in producing compact and interpretable anomaly representations, even without an encoder. The full REDNMF model combines the encoder–decoder structure with $L_{2,1}$ sparsity regularization. This configuration achieves the best results across all metrics. The encoder strengthens representation learning while $L2,1$ regularization isolates anomalous components. The two components work synergistically; both are necessary for optimal anomaly detection performance.

In addition to quantitative comparisons, Fig. 8 presents the corresponding anomaly maps for different ablated variants of REDNMF.

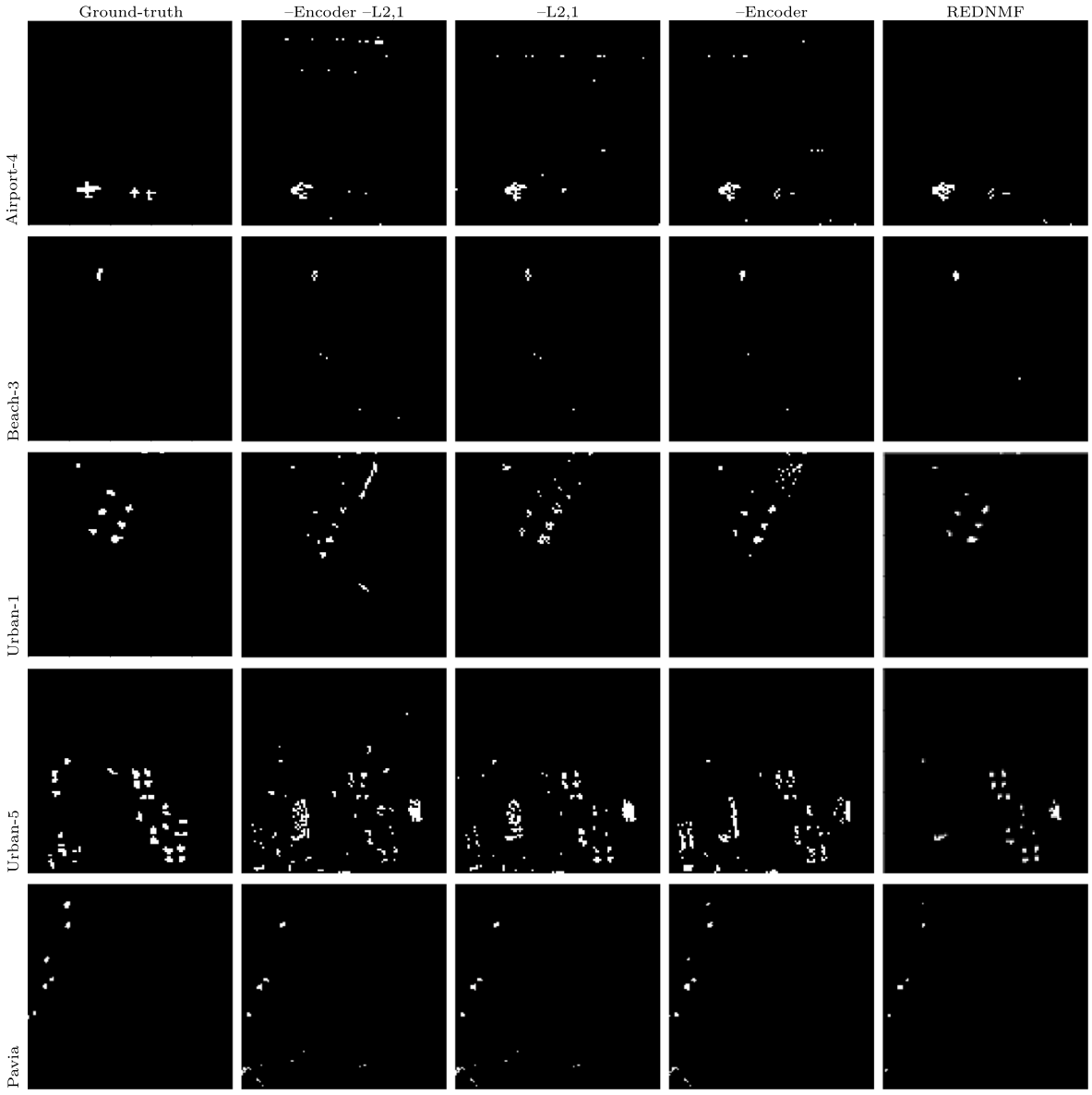

Fig. 8. Ablation study of REDNMF evaluating the contributions of the encoder module and the $L_{2,1}$ sparsity regularization by progressively removing each component from the full model, with qualitative visualization results illustrating their impact on anomaly localization and background suppression.

Visually, removing the $L_{2,1}$-norm regularization results in increased background responses and reduced sparsity, leading to scattered false activations. When the asymmetric encoder–decoder structure is omitted, anomaly localization becomes less consistent, and certain weak targets are partially suppressed. In contrast, the full REDNMF model produces cleaner anomaly maps with sharper localization and improved separation between background and anomalous regions. These visual differences further demonstrate the complementary roles of structured sparsity and asymmetric self-representation in enhancing detection robustness.

5.6. Convergence analysis

The purpose of this experiment is to analyze and compare the convergence behaviors of NMF, L1NMF, RNMF, and the proposed REDNMF in terms of relative reconstruction error over 500 iterations. As shown in Fig. 9, all methods except L1NMF exhibit a rapid decrease in error during the initial iterations, indicating efficient extraction of the dominant low-rank background structure at the early stage of optimization. Standard NMF achieves the lowest final reconstruction error, demonstrating strong data-fitting capability; however, such aggressive minimization may absorb anomalous structures into the learned subspace, potentially

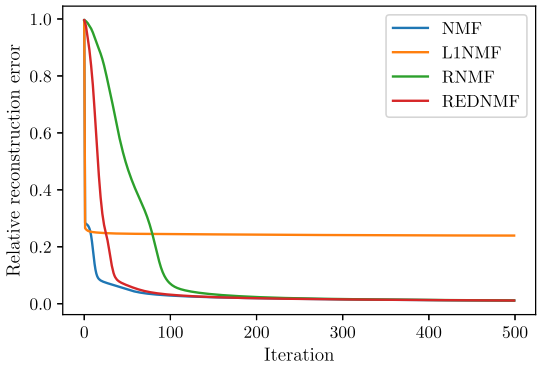

Fig. 9. Convergence curves of NMF, L1NMF, RNMF, and the proposed REDNMF in terms of relative reconstruction error versus iteration number.

Table 4
Runtime comparison (in seconds) of different methods across multiple datasets. The reported values represent the average execution time for each method on each dataset, highlighting computational efficiency differences among competing approaches.

Method	Salians	Airport-1	Airport-4	Beach-1	Beach-3	Beach-4	Urban-1	Urban-3	AVIRIS	AVIRIS-sunlight	Pavia
NMF	1.6304	1.2236	1.1767	2.2090	1.1282	7.4270	1.1825	1.1771	1.1466	1.1486	10.5725
L1NMF	6.0128	3.1755	2.9556	8.6416	2.9513	7.8503	3.1455	2.9306	2.8777	2.8912	7.2455
EWNMF	17.3117	19.7998	17.1234	39.3543	16.2058	17.0707	16.8219	15.5718	13.2288	12.8169	16.3034
RNMF	8.1156	13.6865	11.0474	23.4434	10.2993	14.1117	12.2929	11.4137	11.2655	11.1179	13.4937
BSDM	15.8649	12.2502	11.5762	21.5896	11.2358	21.1036	11.4658	11.2909	11.0501	11.0657	20.1128
MSNet	12.5159	9.1319	8.69986	18.9539	9.3532	17.14288	9.6451	10.1915	11.0594	9.8710	16.9189
REDNMF	14.1642	8.8662	10.5426	19.1988	17.0960	11.8130	8.9149	12.7363	8.8317	7.8626	10.3542

reducing anomaly separability. In contrast, L1NMF quickly stagnates at a relatively high error level, suggesting insufficient representation capacity to accurately model the background structure. The robust formulations, RNMF and REDNMF, display fast, smooth, and stable convergence. Although their final background reconstruction error is slightly higher than that of standard NMF, this controlled fitting behavior reflects a more balanced decomposition that preserves anomalous components instead of forcing them into the low-rank background. Overall, REDNMF achieves rapid and stable convergence while maintaining robustness, providing a favorable trade-off between accurate background modeling and reliable anomaly separation.

5.7. Computational efficiency analysis

Besides detection accuracy, computational efficiency is an important consideration for practical hyperspectral anomaly detection. To evaluate the runtime performance of the proposed REDNMF method, we compare the average execution time of all competing algorithms, including classical matrix factorization–based approaches and deep learning–based models. All experiments were implemented in Python and carried out on a workstation equipped with an RTX 4060 Laptop GPU. All methods were evaluated under the same hardware and software environment to guarantee fairness in computational efficiency comparisons. Each method was executed multiple times, and the reported results correspond to the average running time. Table 4 presents the execution time (in seconds) of the compared methods across multiple datasets. As expected, traditional NMF-based algorithms generally exhibit the lowest computational cost due to their relatively simple multiplicative update rules. In contrast, deep learning–based methods, such as MSNet and BSDM, typically require higher computation time because of multi-layer architectures, convolutional operations, and large-scale parameter optimization. Although REDNMF introduces an asymmetric encoder and structured L2,1 sparsity regularization, its runtime remains competitive. In most cases, REDNMF is moderately slower than basic NMF variants but comparable to or slightly faster than deep learning–based approaches. Importantly, when considering the substantial improvements in detection performance demonstrated in Section 5, REDNMF achieves a favorable balance between computational efficiency and detection accuracy. Overall, REDNMF provides competitive runtime performance while delivering significantly improved robustness and detection capability, making it suitable for real-world hyperspectral anomaly detection applications.

5.8. Discussion

The experimental results indicate that the performance gains of REDNMF are not merely due to the inclusion of additional regularization terms, but arise from the proposed joint coupling mechanism between anomaly estimation and representation learning. Unlike conventional NMF and RNMF methods, where latent representations are learned directly from corrupted observations, REDNMF enforces that the latent representation is derived from anomaly-suppressed data through $\boldsymbol{H} \approx \boldsymbol{M}(\boldsymbol{X} - \boldsymbol{S})$. This design ensures that anomalous components are explicitly removed prior to encoding, leading to a cleaner, more stable, and more discriminative latent space. As a result, the learned representation is less affected by outliers and better reflects the underlying background structure, particularly in scenarios with strong, dense, or spatially structured anomalies. In contrast to encoder-decoder NMF variants that lack explicit anomaly modeling, and robust NMF approaches that treat anomaly estimation and representation learning as loosely coupled or sequential processes, the proposed formulation introduces a bidirectional dependency between $\boldsymbol{S}$ and $\boldsymbol{H}$ within a unified optimization framework. Specifically, the estimation of the anomaly matrix influences the encoding process, while the learned representation simultaneously guides anomaly separation and reconstruction. This mutual interaction enables the model to progressively refine both anomaly isolation and background reconstruction throughout the optimization process, rather than relying on fixed or decoupled steps. As observed in the experimental results, this leads to more consistent and stable improvements across diverse datasets, especially in complex scenes where anomaly contamination significantly degrades the performance of traditional methods.

6. Conclusion

In conclusion, this work proposes REDNMF, a novel and robust framework for hyperspectral anomaly detection that directly addresses the inherent limitations of conventional low-rank matrix approximation techniques. By jointly learning a sparse anomaly matrix, employing an asymmetric encoder-decoder architecture for self-representation, and enforcing structured pixel-wise sparsity through the L2,1 norm, REDNMF achieves superior anomaly localization and robustness to noise. Unlike existing methods that rely solely on residual-based anomaly cues, REDNMF explicitly disentangles anomalous components from the background, leading to cleaner separation and more interpretable results. The integration of an encoder pathway ensures consistent and informative latent representations, enabling better generalization across complex and noisy scenarios. Experimental evaluations on multiple real-world hyperspectral datasets demonstrate the effectiveness of the proposed approach, achieving consistent improvements over state-of-the-art baselines.

While REDNMF is effective in capturing global spectral and spatial structures, it lacks explicit mechanisms to model local spatial patterns, which are often key to detecting fine-grained or small-scale anomalies. For future work, patch-based extensions of REDNMF could be explored to incorporate localized context to capture spatially coherent structures. Additionally, integrating deep architectures may improve the representation learning capacity and enable more expressive, nonlinear feature modeling. Finally, extending the framework to semi-supervised or unsupervised settings and enabling real-time anomaly detection in streaming data would significantly broaden its practical applicability.

CRediT authorship contribution statement

Shirin Moradi: Writing – original draft, Visualization, Software, Investigation. **Amjad Seyedi:** Writing – review & editing, Visualization, Software, Methodology, Investigation, Conceptualization. **Wafa Barkhoda:** Writing – review & editing, Validation, Investigation.

Fardin Akhlaghian Tab: Writing – review & editing, Validation, Supervision, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Amjad Seyedi reports that financial support was provided by the European Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Amjad Seyedi acknowledges the support by the European Union (ERC Consolidator, eLinoR, no 101085607).

Data availability

Data will be made available on request.

References

[1] X. Qian, Y. Huo, G. Cheng, C. Gao, X. Yao, W. Wang, Mining high-quality pseudoinstance soft labels for weakly supervised object detection in remote sensing images, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–15, https://doi.org/10.1109/TGRS.2023.3266838

[2] M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: a meta-review, Remote Sens. Environ. 236 (2020) 111402.

[3] H. Zha, Y. Guo, Y. Liu, X. Zhang, S. Xiao, G. Gao, G. Wu, The characteristic analysis of esdd and nsdd detection of composite insulators based on hyperspectral technology, IEEE Trans. Instrum. Meas. 72 (2023) 1–8.

[4] J.-J. Ying, S.-Q. Yin, W.-Y. Yang, H. Liu, X. Li, Hyperspectral image target recognition based on YOLO model, in: Sixth Conference on Frontiers in Optical Imaging and Technology: Imaging Detection and Target Recognition, vol. 13156, SPIE, 2024, pp. 221–227.

[5] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, A. Plaza, Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15, https://doi.org/10.1109/TGRS.2024.3351179

[6] J. Qu, W. Dong, Y. Yang, T. Zhang, Y. Li, Q. Du, Cycle-refined multidecision joint alignment network for unsupervised domain adaptive hyperspectral change detection, IEEE Trans. Neural Netw. Learn. Syst. (2024) 1–14, https://doi.org/10.1109/TNNLS.2023.3347301

[7] Y. Gao, B. Pan, X. Xu, X. Song, Z. Shi, A reversible generative network for hyperspectral unmixing with spectral variability, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15, https://doi.org/10.1109/TGRS.2024.3403926

[8] C. He, Y. Xu, Z. Wu, Z. Wei, Connecting low-level and high-level visions: a joint optimization for hyperspectral image super-resolution and target detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–16.

[9] J. Lu, Z. Wu, Y. Lan, X. Deng, J. Huang, Study on the prediction model of litchi downy blight damage based on IOT and hyperspectral data fusion, IEEE Internet Things J. 11 (16) (2024) 27184–27200.

[10] Y. Sun, J. Hu, D. Yuan, Y. Chen, Y. Liu, Q. Zhang, W. Chen, Hyperspectral classification of hazardous materials based on deep learning, Sustainability 15 (9) (2023) 7653.

[11] L. Xiong, X. Chen, J. Schneider, Direct robust matrix factorizatoin for anomaly detection, in: 2011 IEEE 11th International Conference on Data Mining, IEEE, 2011, pp. 844–853.

[12] V.C. Gogineni, K. Müller, M. Orlandić, S. Werner, Lightweight autonomous autoencoders for timely hyperspectral anomaly detection, IEEE Geosci. Remote Sens. Lett. 21 (2024) 1–5.

[13] Y. Yan, Y. Wang, Y. Hu, Y. Li, Federated learning-based anomaly detection for environment monitoring sensor networks, IEEE Sens. Lett. 8 (8) (2024) 1–4.

[14] A. Diallo, H.M. Ray, N.R. Ahmed, Falco: flight and ai-enabled coordination and object-detection for search and rescue missions, in: AIAA SCITECH 2024 Forum, 2024, pp. 2769.

[15] N.K. Saini, P. Singh, S. Gnanavel, Invasive flower species detection using CNN and alert system, in: AIP Conference Proceedings, vol. 3075, AIP Publishing, 2024.

[16] Z. Li, Y. Zhu, M. Van Leeuwen, A survey on explainable anomaly detection, ACM Trans. Knowl. Discov. Data 18 (1) (2023) 1–54.

[17] E.G. Allan, M.R. Horvath, C.V. Kopek, B.T. Lamb, T.S. Whaples, M.W. Berry, Anomaly detection using nonnegative matrix factorization, in: Survey of Text Mining II: Clustering, Classification, and Retrieval, 2008, pp. 203–217.

[18] F.D. Van der Meer, H.M.A. Van der Werff, F.J.A. Van Ruitenbeek, C.A. Hecker, W.H. Bakker, M.F. Noomen, M. Van Der Meijde, E.J.M. Carranza, J.B. De Smeth, T. Woldai, Multi-and hyperspectral geologic remote sensing: a review, Int. J. Appl. Earth Obs. Geoinf. 14 (1) (2012) 112–128.

[19] H. Zhang, Y. Li, Y. Zhang, Q. Shen, Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network, Remote Sens. Lett. 8 (5) (2017) 438–447.

[20] I.S. Reed, X. Yu, Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution, IEEE Trans. Acoust. Speech Signal Process. 38 (10) (1990) 1760–1770.

[21] S. Matteoli, M. Diani, G. Corsini, A kurtosis-based test to efficiently detect targets placed in close proximity by means of local covariance-based hyperspectral anomaly detectors, in: 2011 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), IEEE, 2011, pp. 1–4.

[22] Y. Chen, N.M. Nasrabadi, T.D. Tran, Hyperspectral image classification via kernel sparse representation, IEEE Trans. Geosci. Remote Sens. 51 (1) (2012) 217–231.

[23] H. Kwon, N.M. Nasrabadi, Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 43 (2) (2005) 388–397.

[24] K. Tan, Z. Hou, F. Wu, Q. Du, Y. Chen, Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation, Remote Sens. 11 (11) (2019) 1318.

[25] H. Su, Z. Wu, H. Zhang, Q. Du, Hyperspectral anomaly detection: a survey, IEEE Geosci. Remote Sens. Mag. 10 (1) (2022) 64–90, https://doi.org/10.1109/MGRS.2021.3105440

[26] S. Matteoli, M. Diani, G. Corsini, A tutorial overview of anomaly detection in hyperspectral images, IEEE Aerosp. Electron. Syst. Mag. 25 (7) (2010) 5–28.

[27] H. Zhang, W. He, L. Zhang, H. Shen, Q. Yuan, Hyperspectral image restoration using low-rank matrix recovery, IEEE Trans. Geosci. Remote Sens. 52 (8) (2013) 4729–4743.

[28] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[29] W. He, H. Zhang, L. Zhang, Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 9 (9) (2016) 4267–4279.

[30] D. Lee, H.S. Seung, Algorithms for non-negative matrix factorization, Adv. Neural Inf. Process. Syst. 13 (2000).

[31] I. Ahmed, X.B. Hu, M.P. Acharya, Y. Ding, Neighborhood structure assisted non-negative matrix factorization and its application in unsupervised point-wise anomaly detection, J. Mach. Learn. Res. 22 (34) (2021) 1–32.

[32] N. Wang, B. Du, L. Zhang, An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (2) (2013) 554–569.

[33] D. Kong, C. Ding, H. Huang, Robust nonnegative matrix factorization using l21-norm, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, 2011, pp. 673–682.

[34] A.B. Hamza, D.J. Brady, Reconstruction of reflectance spectra using robust nonnegative matrix factorization, IEEE Trans. Signal Process. 54 (9) (2006) 3637–3642.

[35] N. Guan, D. Tao, Z. Luo, J. Shawe-Taylor, Mahnmf: Manhattan non-negative matrix factorization, arXiv preprint arXiv:1207.3438, 2012.

[36] Y. Nesterov, Smooth minimization of non-smooth functions, Math. Program. 103 (2005) 127–152.

[37] C. Ding, D. Kong, Nonnegative matrix factorization using a robust error function, in: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2012, pp. 2033–2036.

[38] A. Liutkus, D. Fitzgerald, R. Badeau, Cauchy nonnegative matrix factorization, in: 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE, 2015, pp. 1–5.

[39] S. Peng, B. Chen, L. Sun, W. Ser, Z. Lin, Constrained maximum correntropy adaptive filtering, Signal Process. 140 (2017) 116–126.

[40] L. Zhang, Z. Chen, M. Zheng, X. He, Robust non-negative matrix factorization, Front. Electr. Electron. Eng. China 6 (2011) 192–200.

[41] A. Tosyali, J. Kim, J. Choi, Y. Kang, M.K. Jeong, New node anomaly detection algorithm based on nonnegative matrix factorization for directed citation networks, Ann. Oper. Res. 288 (1) (2020) 457–474.

[42] A. Aiba, M. Yoshida, D. Kitamura, S. Takamichi, H. Saruwatari, Noise robust acoustic anomaly detection system with nonnegative matrix factorization based on generalized Gaussian distribution, IEICE Trans. Inf. Syst. 104 (3) (2021) 441–449.

[43] J. Zhang, X. Zhang, L. Jiao, Dual-view hyperspectral anomaly detection via spatial consistency and spectral unmixing, Remote Sens. 15 (13) (2023) 3330.

[44] W. Zhang, X. Lu, X. Li, Similarity constrained convex nonnegative matrix factorization for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 57 (7) (2019) 4810–4822.

[45] S. Xiaorui, C. Lingyan, L. Caiping, D. Tao, Z. Haojun, W. Heng, Robust nonnegative matrix factorization based background reconstruction for hyperspectral image anomaly detection, in: 2022 7th International Conference on Signal and Image Processing (ICSIP), IEEE, 2022, pp. 316–321.

[46] S. Soleymanbaigi, A. Seyedi, F. Akhlaghian Tab, F. Daneshfar, Encoder-decoder nonnegative matrix factorization with β-divergence for data clustering, Pattern Recognit. 171 (2026) 112211.

[47] S. Soleymanbaigi, A. Seyedi, F. Daneshfar, F. Akhlaghian Tab, Semantic encoder-decoder nonnegative matrix factorization with kullback-leibler divergence, Int. J. Mach. Learn. Cybern. 17 (4) (2026) 157.

[48] A. Seyedi, N. Gillis, Encoder-decoder symmetric nonnegative matrix tri-factorization for graph clustering, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2026.

[49] Z. Yang, E. Oja, Linear and nonlinear projective nonnegative matrix factorization, IEEE Trans. Neural Netw. 21 (5) (2010) 734–749.

[50] W. Barkhoda, A. Seyedi, N. Gillis, F. Akhlaghian Tab, Instance-wise distributionally robust nonnegative matrix factorization, Pattern Recognit. 169 (2026) 111732.

[51] W. Barkhoda, A. Seyedi, N. Gillis, F. Akhlaghian Tab, Distributionally robust nonnegative matrix factorization with self-paced adaptive multi-loss fusion, Inf. Sci. 728 (2026) 122823.

[52] A. Mohammadi, S.A. Seyedi, F. Akhlaghian Tab, R. Pir Mohammadiani, Diverse joint nonnegative matrix tri-factorization for attributed graph clustering, Appl. Soft Comput. 164 (2024) 112012.
[53] N. Jabari, A. Seyedi, R. Mahmoodi, F. akhlaghian Tab, Contrastive calibration on consensus and complementary multi-view representations, Pattern Recognit. 176 (2026) 113291.
[54] M. Faraji, A. Seyedi, F. Akhlaghian Tab, Robust log-based multi-label feature selection with dynamic label correlation and relevance-redundancy optimization, Knowl.-based Syst. 342 (2026) 115825.
[55] S. Ghodsi, A. Seyedi, T. Le Quy, F. Karimi, E. Ntoutsi, A deep latent factor graph clustering with fairness-utility trade-off perspective, in: 2025 IEEE International Conference on Big Data (BigData), IEEE, 2025, pp. 1705–1714.
[56] J.C. Principe, Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives, Springer Science & Business Media, 2010.
[57] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, Y. Xiao, Hyperspectral image denoising: from model-driven, data-driven, to model-data-driven, IEEE Trans. Neural Netw. Learn. Syst. 35 (10) (2024) 13143–13163.
[58] Q. Zhang, Q. Yuan, M. Song, H. Yu, L. Zhang, Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising, IEEE Trans. Image Process. 31 (2022) 6356–6368, https://doi.org/10.1109/TIP.2022.3211471
[59] Q. Zhang, Y. Dong, Y. Zheng, H. Yu, M. Song, L. Zhang, Q. Yuan, Three-dimension spatial–spectral attention transformer for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–13.
[60] J. Lian, L. Wang, H. Sun, H. Huang, Gt-had: gated transformer for hyperspectral anomaly detection, IEEE Trans. Neural Netw. Learn. Syst. 36 (2) (2025) 3631–3645.
[61] S.-S. Young, C.-H. Lin, Z.-C. Leng, Unsupervised abundance matrix reconstruction transformer-guided fractional attention mechanism for hyperspectral anomaly detection, IEEE Trans. Neural Netw. Learn. Syst. 36 (5) (2025) 9150–9164.
[62] H. Liu, X. Su, X. Shen, X. Zhou, MSNet: self-supervised multiscale network with enhanced separation training for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–13.
[63] X. Cheng, Y. Huo, S. Lin, Y. Dong, S. Zhao, M. Zhang, H. Wang, Deep feature aggregation network for hyperspectral anomaly detection, IEEE Trans. Instrum. Meas. 73 (2024) 1–16.
[64] R. Wang, J. Hu, Gaussian-inspired attention mechanism for hyperspectral anomaly detection, IEEE Geosci. Remote Sens. Lett. 22 (2025) 1–5.
[65] J. Ma, W. Xie, Y. Shi, X. Xiang, Y. Li, L. Fang, BSDM: background suppression diffusion model for hyperspectral anomaly detection, IEEE Trans. Circuits Syst. Video Technol. (2025) 1.
[66] L. Li, B. Wang, DPMN: deep prior mamba network for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 63 (2025) 1–16.
[67] A. Ruhan, Q. Gao, X. Zhang, W. Feng, S.K. Ali, Hyperspectral anomaly detection leveraging spatial attention and right-shifted spectral energy, Plos One 20 (9) (2025) 1–35.
[68] B.-J. Sun, H. Shen, J. Gao, W. Ouyang, X. Cheng, A non-negative symmetric encoder-decoder approach for community detection, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017, pp. 597–606.
[69] M. Ranzato, Y.-L. Boureau, Y. Cun, et al., Sparse feature learning for deep belief networks, Adv. Neural Inf. Process. Syst. 20 (2007).
[70] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 171–184.
[71] J. Wei, C. Tong, B. Wu, Q. He, S. Qi, Y. Yao, Y. Teng, An entropy weighted nonnegative matrix factorization algorithm for feature representation, IEEE Trans. Neural Netw. Learn. Syst. 34 (9) (2022) 5381–5391.

Author biography

Shirin Moradi has completed her master's degree in Artificial Intelligence and Robotics in the Department of Computer Engineering, University of Kurdistan. Her research interests include machine learning, representation learning, anomaly detection, and remote sensing. She received her bachelor's degree in Software Engineering.

Amjad Seyedi is a PhD student in Matrix Theory and Optimization in the Department of Mathematics and Operational Research, Faculté Polytechnique, University of Mons, Mons, Belgium. He received his Master's in Artificial Intelligence from the Department of Computer Engineering at the University of Kurdistan in 2018. His work mainly focused on matrix factorization and low-rank approximation.

Wafa Barkhoda is a faculty member in the Department of Information Technology at Kermanshah University of Technology, Iran. He is currently a PhD student in Artificial Intelligence and Robotics in the Department of Computer Engineering, University of Kurdistan, Iran. Barkhoda received his M.S. degree in Computer Engineering from Kurdistan University in 2010. His research interests include machine learning, representation learning, unsupervised learning, robustness, and data clustering.

Fardin Akhlaghian Tab is currently an Associate Professor of Computer Engineering at the University of Kurdistan, Iran, and his research focuses on machine learning pattern recognition, and computer vision. He did his Ph.D. in Computer Vision at the University of Wollongong in 2005. He holds a master's degree from Tarbiat Modares University, Tehran, in 1992.