Diffusing Background Dictionary for Hyperspectral Anomaly Detection

Yaochen Wu, Yu Meng, and Lei Sun*

School of Systems Science and Engineering, Sun Yat-sen University
wuych55@mail2.sysu.edu.cn
mengy25@mail2.sysu.edu.cn
sunlei8@mail.sysu.edu.cn

Abstract. The diffusion model (DM) has achieved remarkable results in image generation and has been used in hyperspectral image (HSI) processing. However, DM has not been directly applied in the HSI anomaly detection (HAD) task. In this paper, based on the characteristics of HSI and HAD tasks, we combine the advantages of model-driven and data-driven and propose the diffusion background dictionary method (DBD). DBD intrinsically combines the DM with the low-rank representation (LRR) model, using DM to get the crucial background dictionary tensor in the tensor LRR, so that it can accurately detect the anomalies. We also diffuse the multivariate normal distribution that approximates the HSI background based on the idea of the RX algorithm in the HAD, making it more suitable for suppressing the background. DBD combines the advantages of the three main groups of HAD methods, and the experimental results on real datasets prove its effectiveness. DBD can outperform several existing state-of-the-art methods in terms of detection accuracy, which proves the DM's potential in HAD.

Keywords: hyperspectral anomaly detection(HAD) · diffusion model (DM) · tensor low-rank representation(tensor LRR)

1 Introduction

Hyperspectral image (HSI) is captured by hyperspectral imaging sensors, distinguished from general RGB images by a spectral dimension with hundreds of contiguous narrow bands that encompass the visible, near-infrared, short-infrared, mid-infrared and even ultraviolet wavelengths of the electromagnetic spectrum [15]. As a result, HSI can characterize the subtle spectral differences between ground objects, which makes it widely used in fields such as environmental monitoring [47], smart agriculture [40], and mineral exploration [39]. Diffusion models (DMs) have risen to prominence in the field of deep generative modeling in recent years. Representing a cutting-edge paradigm, DMs [17,41,55] have disrupted the dominance of generative adversarial networks (GANs) in image generation and demonstrated vast potential across various applications, including computer vision [9], natural language processing [1,13], and multi-modal [23,31]

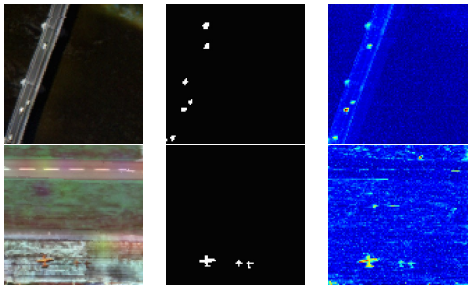


Fig. 1: HAD on two datasets. From left to right, input HSI, ground truth images, and Color detection maps from our method.

domains. As expected, the application of this promising method to HSI tasks also provides new opportunities for advancements in this field [5, 34, 51, 57].

Among the various HSI tasks, anomaly detection is an unsupervised binary classification problem that separates image pixels into background and anomaly, without prior knowledge of the anomaly targets, making it incredibly challenging. For hyperspectral anomaly detection (HAD), anomalies are generally atypical ground objects with irregular spectra [42]. Figure 1 shows the datasets in HAD, demonstrating that anomalies correspond to different ground objects in various scenarios. To date, HAD methods can be roughly divided into two categories, data-driven methods and model-driven methods.

Model-driven methods can be divided into two groups: statistical-based and low-rank-based. Most statistical methods are based on the Reed-Xiaoli (RX) algorithm, which assumes that the background is distributed by a multivariate Gaussian distribution. These methods estimate the probability density function of pixels globally or locally [3, 35], and obtain the pixel's anomalous level by calculating the degree of deviation from the background's distribution. With the advantages of simplicity and speed, it is commonly used in real-time anomaly detection [50]. The low-rank methods decompose the HSI into background and anomaly two parts, and characterize the spatial homogeneity and spectral correlation of the background by low-rank constraints, while sparse constraints are used to characterize the anomalies with a low probability of occurring and accounting for a small proportion of HSI [4, 46]. Wherein methods based on low-rank representation (LRR) further assume that the background lies in multiple low-rank subspaces represented by the dictionary, which is more in line with complex real-world scenarios than assuming that it exists in only one low-rank space [54]. Wang *et al*. proposed PCA-based tensor low-rank and sparse representation (PCA-TLRSR) [49] based on tensor nuclear norm (TNN) [6], which avoids unfolding the HSI tensor into a two-dimensional matrix and preserves the intrinsic structure of the HSI. Whether the model is matrix LRR or tensor LRR, the dictionary in the LRR model determines the detection effect, and the selection of the dictionary is a crucial problem in the LRR model [43].

Data-driven methods leverage deep learning for abstract feature representation and analysis of spectral data, achieving HAD through estimation and reconstruction. Supervised HAD methods [25,26] use labeled data for model training, guiding them to learn data distributions and understand HSI information, but obtaining such data is challenging. In contrast, HAD tasks inherently lack prior knowledge, making them more suitable for unsupervised learning. AutoEncoders [2, 32], GANs [20, 48], CNN-based [14] and Swin Transformer-based [27] methods have shown promise in unsupervised HAD. Background suppression diffusion model (BSDM) [34] was the first to introduce the DM into the HAD field, effectively suppressing the background in HSI and achieving excellent performance. However, the DM in BSDM needs to be used as a pre-processing module in combination with other HAD methods, and it cannot generate anomaly-only HSI. To overcome the issue of anomaly contamination in data-driven methods due to the lack of prior knowledge [52], numerous recent studies have coupled the LRR model into a deep neural network to explore interpretable deep network while enhancing anomaly detection capability [19, 24]. In addition, multiple AE combined with low-rank theory methods have been proposed [12, 28], showing excellent detection performance in experiments.

In order to solve the problem that DM cannot generate background or anomaly directly and the problem that it is difficult to select an appropriate dictionary in the LRR model, we propose the diffusion background dictionary method (DBD). DBD uses the DM to generate the dictionary which is crucial for detecting anomalies in the LRR model, and combines the DM with the tensor LRR model to solve the above problems. In addition, we improve the DM to make it more suitable to HAD based on the assumptions of the background in the RX algorithm, and adopt the TNN-based tensor LRR to ensure that the inherent multidimensional structure of HSI is not corrupted. Thus, DBD improves anomaly detection performance by combining the advantages of the DM to compress the background and the LRR model to characterize the background and explore the application of the DM to HAD. The main contributions of this work can be therefore summarized as follows:

(1) In this paper, the DM and tensor LRR models are combined and applied to HAD for the first time, and the diffusion background dictionary (DBD) method is proposed, which solves the problems of the DM and the LRR model and integrates their advantages.

(2) In this paper, based on the assumption that the background obeys a multivariate normal distribution by the statistical model-based methods in HAD, the DM used for HSI processing is redesigned to enhance its diffusion ability to the HSI background.

(3) Experimental comparisons of several HAD methods demonstrate that DBD achieves state-of-the-art performance across five real HSI datasets, showcasing outstanding anomaly detection capabilities and effectively suppressing background in a variety of complex scenarios.

2 Related Work

2.1 Dictionary Construction in LRR

To address the problem of selecting the dictionary in the LRR model, most of the methods directly use the original HSI as the dictionary [7, 45, 54] or use the features obtained by doing SVD decomposition on the original HSI [16, 44]. These methods do not prevent anomalous information from being included in the dictionary used to represent the background. Several studies [18, 37, 56] obtain the background dictionary based on the k-means clustering method. In addition, PCA-LRSR used the low-rank model to obtain dictionary [49], and Lin *et al*. proposed the spatial and spectral detector for the joint construction of dictionary [29]. But both of them still have the problems of needing to pre-set sensitive hyperparameters or being time-consuming. To avoid the above problems, some methods adaptively construct dictionaries in iterations based on the dictionary learning [8, 58], so that the dictionary can effectively contain the background information of the current HSI. On this basis, Lee *et al*. proposed LRR-Net [24] for learning background dictionary using DNN in iterations. LRR-Net is the first model-driven deep network in the field of HAD.

2.2 Diffusion Model on HSI Processing

Ho *et al*. introduced a novel paradigm of image generation model, denoising diffusion probabilistic model (DDPM) [17], which has brought DM to prominence in the field of deep generative. In recent years, several methods have also emerged in tasks related to HSI. In the field of classification, SpectralDiff [5] effectively mines the spectral-spatial distribution information of high dimensional data through iterative denoising and explicit construction of the data generation process. Zhou *et al*. proposed the MTMSD model [60] for multi-time-step and multi-stage diffusion feature learning in HSI classification. Hyperspectral image super-resolution (HISR) is a method to overcome the limitations of imaging mechanisms to obtain HSI with high spatial resolution. HSR-Diff [51] pioneers the use of DM in HISR task scenarios, showing their effectiveness. S2CycleDiff [36] achieves bidirectional spatial-spectral super-resolution with a conditional cyclic diffusion framework. Dong *et al*. proposed ISPDiff [11] to address long inference times and weak interpretability in DMs for HISR tasks. DiffUn [10] is the first work to apply DM to the semi-blind unmixing task of HSI, achieving sampling of underlying endmembers and corresponding abundances from the posterior distribution by combining spectral prior distributions from the spectral library with the inverse process of the DM. Zeng *et al*. proposed a self-supervised diffusion denoising model Diff-Unmix [57], which achieves HSI denoising through spectral unmixing and abundance generation.

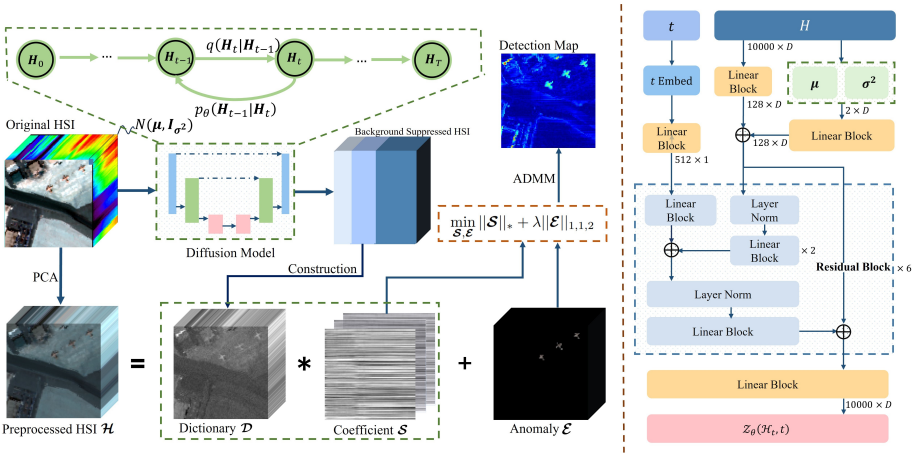


Fig. 2: The left part of the brown dashed line shows the flowchart of our DBD method, and the right part is a schematic of the network structure used for the diffusion model.

3 Method

3.1 Notations and Preliminaries

Some notations and preliminaries related to our method are introduced. Tensors, matrices, vectors, and scalars are denoted by boldface Euler script letters, boldface capital letters, boldface lowercase letters, and lowercase letters, respectively, *e.g.*, $\boldsymbol{\mathcal{A}}$, $\boldsymbol{A}$, $\boldsymbol{a}$ and a. For the vector where each element is the square of the corresponding element in $\boldsymbol{a}$ is denoted by $\boldsymbol{a}^2$. The diagonal matrix formed by the elements in the vector $\boldsymbol{a}$ is denoted $\boldsymbol{I}_{\boldsymbol{a}}$, *e.g.*, $\boldsymbol{I}_{\boldsymbol{a}} = \mathrm{diag}(\boldsymbol{a})$. A HSI data with H rows, W columns, and D spectral bands is denoted by $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$ and its (i,j,k)-th entry denoted as $\boldsymbol{\mathcal{A}}_{ijk}$. $\boldsymbol{\mathcal{A}}(i,:,:)$, $\boldsymbol{\mathcal{A}}(:,j,:)$ and $\boldsymbol{\mathcal{A}}(:,:,k)$ are used to denote the the i-th horizontal, j-th lateral and k-th frontal slice respectively. For subsequent convenience, the frontal slice $\boldsymbol{\mathcal{A}}(:,:,k)$ is also denoted as $\boldsymbol{A}^{(k)}$. The inner product between tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{R}^{H\times W\times D}$ defined as $\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i=1}^{D} \langle \boldsymbol{A}^{(i)}, \boldsymbol{B}^{(i)} \rangle$. The result of discrete Fourier transformation (DFT) of $\boldsymbol{\mathcal{A}}$ along the third dimension is denoted as $\widehat{\boldsymbol{\mathcal{A}}}$. The tensor Frobenius norm is denoted as $||\boldsymbol{\mathcal{A}}||_F = \sqrt{\sum_{i,j,k} \boldsymbol{\mathcal{A}}_{ijk}^2}$, and the tensor $l_{1,1,2}$ norm is used to characterize the anomaly tensor, which denoted as $||\boldsymbol{\mathcal{A}}||_{1,1,2} = \sum_{i,j} \sqrt{\sum_k \boldsymbol{\mathcal{A}}_{ijk}^2}$. The block circulant matrix $\texttt{bcirc}(\boldsymbol{\mathcal{A}}) \in \mathbb{R}^{HD\times WD}$ and the operator $\texttt{unfold}(\boldsymbol{\mathcal{A}}) \in \mathbb{R}^{HD\times W}$ of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$ are defined as follows

$$\texttt{bcirc}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{A}^{(D)} & \cdots & \boldsymbol{A}^{(2)} \\ \boldsymbol{A}^{(2)} & \boldsymbol{A}^{(1)} & \cdots & \boldsymbol{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(D)} & \boldsymbol{A}^{(D-1)} & \cdots & \boldsymbol{A}^{(1)} \end{bmatrix}, \quad \texttt{unfold}(\boldsymbol{\mathcal{A}}) = \begin{bmatrix} \boldsymbol{A}^{(1)} \\ \boldsymbol{A}^{(2)} \\ \vdots \\ \boldsymbol{A}^{(D)} \end{bmatrix}, \tag{1}$$

and fold is the inverse operator of unfold, $\texttt{fold}(\texttt{unfold}(\boldsymbol{\mathcal{A}})) = \boldsymbol{\mathcal{A}}$.

Definition 1 (Tensor-Tensor product (t-product) [33]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$, $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{W\times M\times D}$, the t-product $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{H\times M\times D}$ is

$$\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} = \texttt{fold}(\texttt{bcirc}(\boldsymbol{\mathcal{A}})\texttt{unfold}(\boldsymbol{\mathcal{B}})), \tag{2}$$

where the tensor product of $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ can be computed by bcirc, fold and unfold operators. By the property of t-product, $\boldsymbol{\mathcal{C}}$ can also be computed via DFT, that is, $\widehat{\boldsymbol{\mathcal{C}}} = \langle \widehat{\boldsymbol{\mathcal{A}}}, \widehat{\boldsymbol{\mathcal{B}}} \rangle$.

Definition 2. (Identity Tensor [22, 33]): For $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n_1\times n_1\times n_3}$, if $\boldsymbol{\mathcal{I}}$ satisfies that all the frontal slice of $\widehat{\boldsymbol{\mathcal{I}}}$ are $n_1 \times n_1$ identity matrices, i.e., $\widehat{\boldsymbol{I}}^{(i)} = I$, then $\boldsymbol{\mathcal{I}}$ is the identity tensor based on t-product.

Definition 3. (Tensor Transpose [22, 33]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$, if $\boldsymbol{\mathcal{A}}^T \in \mathbb{R}^{W\times H\times D}$ satisfies that each frontal slice of $\widehat{\boldsymbol{\mathcal{A}}^T}$ is the transpose matrix of corresponding slice of $\widehat{\boldsymbol{\mathcal{A}}}$, then $\boldsymbol{\mathcal{A}}^T$ is the tensor transpose of $\boldsymbol{\mathcal{A}}$ based on t-product.

Definition 4. (Orthogonal Tensor [22,33]): For $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{n_1\times n_1\times n_3}$, if it satisfies $\boldsymbol{\mathcal{Q}} * \boldsymbol{\mathcal{Q}}^T = \boldsymbol{\mathcal{I}}$, $\boldsymbol{\mathcal{Q}}$ is the orthogonal tensor based on t-product.

Definition 5. (F-diagonal Tensor [22, 33]): For $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{H\times W\times D}$, if all frontal slices of $\boldsymbol{\mathcal{S}}$ are diagonal matrices, then $\boldsymbol{\mathcal{S}}$ is the f-diagonal tensor based on t-product.

Definition 6. (Tensor SVD [22,33]): For $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$, it can be factorized as

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T, \tag{3}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{H\times H\times D}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{W\times W\times D}$ are orthogonal tensors, $\boldsymbol{\mathcal{S}} \in \mathbb{S}^{H\times W\times D}$ is an f-diagonal tensor. By the property of t-product, it can be easily deduced that $\widehat{\boldsymbol{A}}^{(i)} = \widehat{\boldsymbol{U}}^{(i)}\widehat{\boldsymbol{S}}^{(i)}\widehat{\boldsymbol{V}}^{(i)}$, where matrix SVD has been applied on all frontal slices of the tensor $\widehat{\boldsymbol{A}}$.

Definition 7. (Tensor Nuclear Norm [33, 59]): Let $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^T$ be the tensor SVD of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{H\times W\times D}$ and $d = min(H, W)$. The tensor nuclear norm of $\boldsymbol{\mathcal{A}}$ is defined as

$$\|\boldsymbol{\mathcal{A}}\|_* := \|\widehat{\boldsymbol{\mathcal{A}}}\|_* = \langle \widehat{\boldsymbol{\mathcal{S}}}, \widehat{\boldsymbol{\mathcal{I}}} \rangle = \sum_{i=1}^{D}\sum_{j=1}^{d} \sigma_j(\widehat{\boldsymbol{A}}^{(i)}), \tag{4}$$

where $\sigma_j(\widehat{\boldsymbol{A}}^{(i)})$ means the j-th singular value of $\widehat{\boldsymbol{A}}^{(i)}$.

3.2 Diffusing the Backgroud Dictionary

HSI data $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{H\times W\times D}$ typically contains some irregular spectral anomalies, but these anomalies constitute a very low proportion of the HSI. Therefore, in HAD tasks, HSI is usually considered to have a data distribution that is highly similar to the background. Consequently, in HAD, DM can be employed to suppress the background by treating it as noise and removing it from the HSI. The BSDM method, which first applied DM to HAD, assumes that the background

distribution is an univariate normal function. This assumption apparently ignores both that each ground object has different spectral characteristics and the spatial smoothness of the image. The added pseudo-background thus differs too much from the real background, making the neural network unable to learn to suppress the background effectively. RX algorithms detect anomalies by assuming that the background follows a multivariate Gaussian normal distribution. In order to comply with the mathematical principles of DM as well as to prevent overfitting, we assume that the vector of background ground objects obeys a multivariate Gaussian normal distribution with a diagonal matrix of covariances, *e.g.* $\mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_{\boldsymbol{\sigma}^2})$. After obtaining the mean and variance of each spectral dimension from the HSI data, similar to adding noise process in DDPM, start the forward process of adding the pseudo background in our method.

The forward phase q needs incrementally adding Gaussian noise $\boldsymbol{\mathcal{Z}} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_{\boldsymbol{\sigma}^2})$ to real sampled data $\boldsymbol{\mathcal{H}}_0 \sim q(\boldsymbol{\mathcal{H}})$ until it achieves a Gaussian distribution $\boldsymbol{\mathcal{H}}_T$ over T steps. The entire forward process $\boldsymbol{\mathcal{H}}_0 \to \boldsymbol{\mathcal{H}}_1 \to \cdots \to \boldsymbol{\mathcal{H}}_T$ with Gaussian transition kernel $q(\boldsymbol{\mathcal{H}}_t|\boldsymbol{\mathcal{H}}_{t-1})$ can be viewed as a Markov process:

$$q(\boldsymbol{\mathcal{H}}_{1:T}|\boldsymbol{\mathcal{H}}_0) = \prod_{t=1}^{T} q(\boldsymbol{\mathcal{H}}_t)|\boldsymbol{\mathcal{H}}_{t-1}), \tag{5}$$

$$q(\boldsymbol{\mathcal{H}}_t|\boldsymbol{\mathcal{H}}_{t-1}) = \mathcal{N}(\boldsymbol{\mathcal{H}}_t; \sqrt{1-\beta_t}\boldsymbol{\mathcal{H}}_{t-1} + \sqrt{\beta_t}\boldsymbol{\mu}, \beta_t \mathbf{I}_{\boldsymbol{\sigma}^2}) \quad s.t. \quad \beta_t = \frac{\lambda_\beta}{T}, \tag{6}$$

where $q(\boldsymbol{\mathcal{H}}_{1:T}|\boldsymbol{\mathcal{H}}_0)$ is the joint distribution of $\boldsymbol{\mathcal{H}}_{1:T}$ given $\boldsymbol{\mathcal{H}}_0$ derived from the Markov property and the probability multiplication rule. $\beta_t \in (0, 1)$ are hyperparameters pre-selected before model training, regulated by the hyperparameter λ_β. Similar to DDPM, we define $\alpha_t = 1 - \beta_t$ and $\bar{\alpha_t} = \prod_{i=1}^{t} \alpha_i$. The Gaussian transition kernel allows us to marginalize the joint distribution in Eq.5 to obtain the analytical form of $q(\boldsymbol{\mathcal{H}}_t|\boldsymbol{\mathcal{H}}_0)$ for all $t \in \{0, 1, \cdots, T\}$.

$$\boldsymbol{\mathcal{H}}_t \sim q(\boldsymbol{\mathcal{H}}_t|\boldsymbol{\mathcal{H}}_0) = \mathcal{N}(\boldsymbol{\mathcal{H}}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{\mathcal{H}}_0 + \gamma_t \boldsymbol{\mu}, \gamma_t^2 \mathbf{I}_{\boldsymbol{\sigma}^2}), \tag{7}$$

where

$$\gamma_t = \sum_{i=0}^{t-1} \sqrt{\frac{\bar{\alpha_t}\beta_{t-i}}{\bar{\alpha}_{t-i}}}. \tag{8}$$

By selecting a specific λ_β, γ_t can be approximated to 1, thus $\boldsymbol{\mathcal{H}}_T$ can be considered closely as a Gaussian distribution representing the background.

Starting from $\boldsymbol{\mathcal{H}}_T \sim p(\boldsymbol{\mathcal{H}}_T) \approx \mathcal{N}(\boldsymbol{\mu}, \mathbf{I}_{\boldsymbol{\sigma}^2})$, the reverse process iteratively suppressed HSI background by running a learnable Markov chain in the opposite direction of the forward process. The reverse Markov process $\boldsymbol{\mathcal{H}}_T \to \cdots \to \boldsymbol{\mathcal{H}}_1 \to \boldsymbol{\mathcal{H}}_0$ with Gaussian transition kernel parameterized by θ can be defined as:

$$p(\boldsymbol{\mathcal{H}}_{0:T}) = p(\boldsymbol{\mathcal{H}}_T) \prod_{t=1}^{T} p_\theta(\boldsymbol{\mathcal{H}}_{t-1}|\boldsymbol{\mathcal{H}}_t)), \tag{9}$$

$$p_\theta(\boldsymbol{\mathcal{H}}_{t-1}|\boldsymbol{\mathcal{H}}_t) = \mathcal{N}(\boldsymbol{\mathcal{H}}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{\mathcal{H}}_t, t), \sum\nolimits_\theta(\boldsymbol{\mathcal{H}}_t, t)), \tag{10}$$

where $\boldsymbol{\mu_\theta}$ and $\sum_\theta$ represent the learnable mean and variance. Specifically, similar to DDPM, we use the posterior distribution $q(\boldsymbol{\mathcal{H}}_{t-1}|\boldsymbol{\mathcal{H}}_t, \boldsymbol{\mathcal{H}}_0)$ to supervise the neural network $\mathcal{F}_\theta$ to predict the noise $\mathcal{Z}_\theta(\boldsymbol{\mathcal{H}}_t, t) = \mathcal{F}_\theta(\boldsymbol{\mathcal{H}}_t, t)$ in $\boldsymbol{\mathcal{H}}_t$ rather than directly predicting the mean and the covariance. Thus, each refinement step can be mathematically expressed as:

$$\boldsymbol{\mathcal{H}}_{t-1} = \frac{1}{\sqrt{\bar{\alpha}_t}}(\boldsymbol{\mathcal{H}}_t - \gamma_t \mathcal{Z}_\theta(\boldsymbol{\mathcal{H}}_t, t)), \tag{11}$$

and by iterating to obtain $\boldsymbol{\mathcal{H}}_{t-1}$ until $t = 1$, our method ultimately achieve the background-suppressed HSI $\boldsymbol{\mathcal{H}}_{BS}$.

We use the neural network structure on the right side of Fig. 2 to learn the background distribution. The input is an unfolded HSI $H \in \mathbb{R}^{HW \times D}$. All HSI are of size 100×100 with different spectral dimensions. The mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2$ of HSI are encoded to help learn the background distribution. We also use sine-cosine pairs embedding t, and Tanh as the activation function. The residual block is reused, and its input and output dimensions are varied from large to small and then to large, to construct a U-type network structure.

The forward process of the BSDM does not include the step of adding background from the anomaly-only HSI to the realistic HSI. Therefore, there is no guarantee that the anomaly-only HSI will be generated exactly. Thus we cannot preset the number of inference steps that generate the optimal results, and too many inference steps will result in negative background pixel values, but the anomalies are often preserved contrary to expectations. We therefore use the original HSI $\boldsymbol{\mathcal{H}}$ and the anomalous HSI generated by multiple inference steps to obtain the background dictionary $\boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{H}}_{BS}$. So $\boldsymbol{\mathcal{D}}$ will exclude the anomalies and the features of the background will be enhanced. This also solves the problem of predetermining the number of inference steps. In order to reduce the computation time of the subsequent tensor LRR model, the principal component analysis (PCA) method is used to reduce the dimensionality to obtain $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{H \times W \times d}$ and $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{H \times W \times d}$.

The tensor LRR model detects anomalies from HSI by modeling the HSI into two components: the background and the anomaly. The background has the property of low rank due to its spatial homogeneity and spectral correlation. Whereas anomalies are sparse due to their low probability of occurrence and a small proportion of the image. So the tensor LRR model is proposed as:

$$\min_{\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{E}}} ||\boldsymbol{\mathcal{S}}||_* + \lambda ||\boldsymbol{\mathcal{E}}||_{1,1,2} \qquad s.t. \quad \boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{E}}, \tag{12}$$

where $\boldsymbol{\mathcal{H}}$ is divided into the anomaly tensor $\boldsymbol{\mathcal{E}}$ and the background tensor, which is obtained through the tensor product of the background dictionary tensor $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{H \times W \times d}$ and the coefficient tensor $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{H \times W \times d}$. λ is hyperparameter to balance the three constraint terms. To portray the background as lying in

multiple low-rank subspaces, a low-rank constraint is imposed on the coefficient tensor. For anomalies, the $l_{1,1,2}$-norm is used to constrain $\boldsymbol{\mathcal{E}}$ to incorporate its joint sparsity. Then the overall flowchart of DBD is illustrated in Fig 2.

3.3 Solving the Optimized Model

ADMM is used to solve the optimization problem (12). The auxiliary variable $\boldsymbol{\mathcal{C}}$ is introduced to separate $||\boldsymbol{\mathcal{E}}||_{1,1,2}$ and $||\boldsymbol{\mathcal{S}}||_*$ from the constraint $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{E}}$. The optimization problem (12) is transformed into its equivalent as:

$$\min_{\boldsymbol{\mathcal{C}},\boldsymbol{\mathcal{E}}} ||\boldsymbol{\mathcal{C}}||_* + \lambda||\boldsymbol{\mathcal{E}}||_{1,1,2} \qquad s.t. \quad \boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} + \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{S}}. \tag{13}$$

The Augmented Lagrangian Method (ALM) is introduced to solve the problem (13), the following function needs to be minimized:

$$\begin{aligned} L(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{E}}, \boldsymbol{\mathcal{C}}, \boldsymbol{\mathcal{T}}_1, \boldsymbol{\mathcal{T}}_2, \mu) &= ||\boldsymbol{\mathcal{C}}||_* + \lambda||\boldsymbol{\mathcal{E}}||_{1,1,2} + \langle \boldsymbol{\mathcal{T}}_1, \boldsymbol{\mathcal{C}} - \boldsymbol{\mathcal{S}} \rangle \\ &+ \langle \boldsymbol{\mathcal{T}}_2, \boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}} \rangle + \frac{\mu}{2} \left(||\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}||_F + ||\boldsymbol{\mathcal{C}} - \boldsymbol{\mathcal{S}}||_F \right), \end{aligned} \tag{14}$$

where $\boldsymbol{\mathcal{T}}_1$ and $\boldsymbol{\mathcal{T}}_2$ denote the Lagrangian multipliers, μ denotes the penalty parameter. The ADMM, which updates one variable while fixing the others, can solve the previous function.

1) Update $\boldsymbol{\mathcal{C}}$:

$$\boldsymbol{\mathcal{C}}^{k+1} = \min_{\boldsymbol{\mathcal{C}}} ||\boldsymbol{\mathcal{C}}||_* + \frac{\mu^k}{2} \left|\left| \boldsymbol{\mathcal{C}} - \boldsymbol{\mathcal{S}}^k + \frac{\boldsymbol{\mathcal{T}}_1^k}{\mu^k} \right|\right|_F = TSVT\left(\boldsymbol{\mathcal{S}}^k - \frac{\boldsymbol{\mathcal{T}}_1^k}{\mu^k}, \frac{1}{\mu^k} \right). \tag{15}$$

This subproblem can be solved by tensor singular value thresholding (TSVT) [33].

2) Update $\boldsymbol{\mathcal{E}}$:

$$\boldsymbol{\mathcal{E}}^{k+1} = \min_{\boldsymbol{\mathcal{E}}} \frac{\mu^k}{2} \left|\left| \boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}}^k - \boldsymbol{\mathcal{E}} + \frac{\boldsymbol{\mathcal{T}}_2^k}{\mu^k} \right|\right|_F + \lambda||\boldsymbol{\mathcal{E}}||_{1,1,2}. \tag{16}$$

This subproblem can be solved by $l_{1,1,2}$ threshold operator [30] as follow

$$\boldsymbol{\mathcal{E}}^{k+1}(i,j,:) = \max\left(1 - \frac{\lambda}{\mu^k ||\boldsymbol{\mathcal{P}}(i,j,:)||_F}, 0 \right) * \boldsymbol{\mathcal{P}}(i,j,:), \tag{17}$$

where $i \in \{1, \ldots, H\}$, $j \in \{1, \ldots, W\}$, $\boldsymbol{\mathcal{P}} = \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{D}}^k * \boldsymbol{\mathcal{S}}^k + \boldsymbol{\mathcal{T}}_1^k / \mu^k$.

3) Update $\boldsymbol{\mathcal{S}}$:

$$\boldsymbol{\mathcal{S}}^{k+1} = \min_{\boldsymbol{\mathcal{S}}} \left|\left| \boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}} - \boldsymbol{\mathcal{E}}^{k+1} + \frac{\boldsymbol{\mathcal{T}}_2^k}{\mu^k} \right|\right|_F + \left|\left| \boldsymbol{\mathcal{C}}^{k+1} - \boldsymbol{\mathcal{S}} + \frac{\boldsymbol{\mathcal{T}}_1^k}{\mu^k} \right|\right|_F. \tag{18}$$

By setting the derivative of the equation with respect to $\boldsymbol{\mathcal{S}}$ to zero, the following expression is obtained:

$$\boldsymbol{\mathcal{S}}^{k+1} = \left(\boldsymbol{\mathcal{D}}^T * \boldsymbol{\mathcal{D}} + \boldsymbol{\mathcal{I}}\right)^{-1} * \left(\boldsymbol{\mathcal{D}}^T * \left(\boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{E}}^{k+1} + \frac{\boldsymbol{\mathcal{T}}_2^k}{\mu^k}\right) + \boldsymbol{\mathcal{C}}^{k+1} + \frac{\boldsymbol{\mathcal{T}}_1^k}{\mu^k}\right). \quad (19)$$

4) Update $\boldsymbol{\mathcal{T}}_1, \boldsymbol{\mathcal{T}}_2$ and μ : The Lagrange multipliers and the penalty parameter are updated as follows:

$$\begin{aligned} &\boldsymbol{\mathcal{T}}_1^{k+1} = \boldsymbol{\mathcal{T}}_1^k + \boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{C}}^{k+1}, \quad \boldsymbol{\mathcal{T}}_2^{k+1} = \boldsymbol{\mathcal{T}}_2^k + \boldsymbol{\mathcal{H}} - \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{E}}^{k+1}, \\ &\mu^{k+1} = \min(\rho\mu^k, \mu_{\max}), \end{aligned} \quad (20)$$

where the constant $\rho > 1$ is used to gradually increase the coefficient μ, and $\mu_{\max}$ is the upper bound of μ.

When the residual ϵ is less than ϵ_0, the iteration comes to an end. ϵ is calculated according to the following formula:

$$\epsilon = \max\left(\left\|\boldsymbol{\mathcal{E}}^{k+1} - \boldsymbol{\mathcal{E}}^k\right\|_F, \left\|\boldsymbol{\mathcal{D}}^{k+1} - \boldsymbol{\mathcal{D}}^k\right\|_F, \left\|\boldsymbol{\mathcal{S}}^{k+1} - \boldsymbol{\mathcal{S}}^k\right\|_F\right). \quad (21)$$

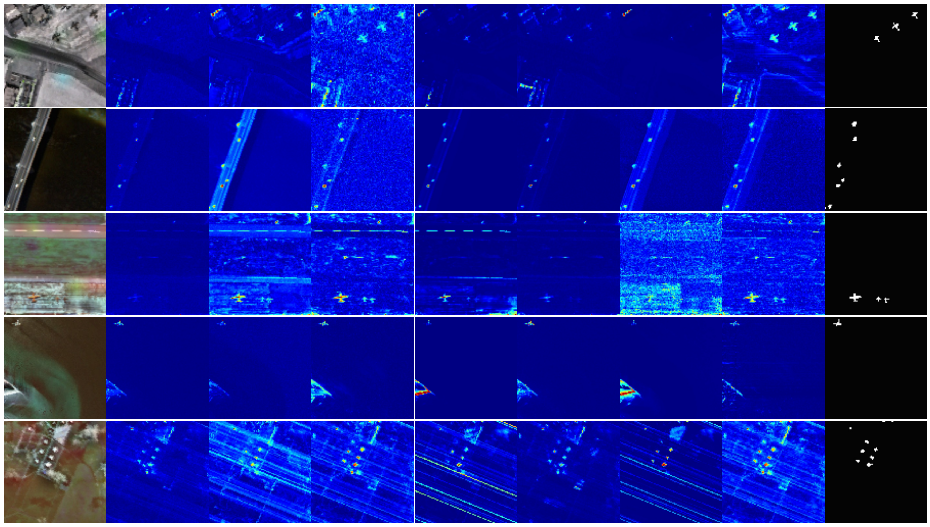
And the detection map $\boldsymbol{M}$ can be obtained as $\boldsymbol{M}(h, w) = \|\boldsymbol{\mathcal{E}}(h, w, :)\|_F$.

4 Experiments

4.1 Experimental Settings

Dataset Our method was evaluated on five distinct datasets. The first dataset, San Diego, encompasses the airport area in the city of USA San Diego [34]. It contains 58 anomalous pixels, corresponding to three airplanes. The second dataset, Pavia, captures the city Pavia in northern Italy [19], with 71 anomalous pixels represented by cars on a bridge. The third to fifth datasets are referred to as Airport, Beach, and Urban (ABU), respectively [21]. The Airport dataset includes 60 anomalous pixels, also corresponding to three airplanes. The Beach dataset comprises 19 anomalous pixels, representing an unidentified floating object on the water's surface. Finally, the Urban dataset contains 67 anomalous pixels consisting of the small buildings in the urban. All the test images are 100 $\times$ 100 pixels in size but have different spectral dimensions, with water absorption and low SNR bands removed.

Evaluation Measures The common quantitative analysis technique for assessing detection performance is the receiver operating characteristic (ROC) curve, which is calculated by comparing the anomaly detection map with the ground truth map. The area under the curve (AUC) represents the area of ROC. Specifically, AUC_{P_D,P_F}, $\mathrm{AUC}_{P_D,\tau}$ and $\mathrm{AUC}_{P_F,\tau}$ represents the area under the ROC curve of (P_D, P_F), (P_D, τ) and (P_F, τ), respectively. Additionally, $\mathrm{AUC}_{\mathrm{OD}}$ enable a more comprehensive and significant evaluation of the anomaly detection



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

Fig. 3: Color detection maps of the proposed DBD and the compared methods for five datasets: (top to bottom) San Diego, Pavia, ABU-airport, ABU-beach, and ABU-urban. (a) Pseudo-color image. (b) GRX. (c) LRASR. (d) PCA-TLRSR. (e) GAED. (f) GRX-BSDM. (g) LREN. (h) Our proposed DBD. (i) Ground truth.

method, defined as $\mathrm{AUC}_{\mathrm{OD}} = \mathrm{AUC}_{P_D,P_F} + \mathrm{AUC}_{P_D,\tau} - \mathrm{AUC}_{P_F,\tau}$. We used the $\mathrm{AUC}_{P_F,\tau}$ and $\mathrm{AUC}_{\mathrm{OD}}$ metrics to compare the detection performance of each method, where the $\mathrm{AUC}_{P_F,\tau}$ value tends to be better as it converges to 1 and the $\mathrm{AUC}_{\mathrm{OD}}$ value tends to be better as it converges to 2.

Implementation Details GRX [38] and other Five state-of-the-art methods are chosen to compare with our proposed DBD, which includes LRASR [54], PCA-TLRSR [49], GAED [53], GRX-BSDM [34] and LREN [19]. The GRX is the classical statistics-based model as the baseline method. LRASR transformed HSIs into matrices for HAD tasks based on matrix LRR, and PCA-TLRSR used the idea of tensor LRR. GAED is the advanced AE method for HAD. BSDM introduced DM to HAD as a pioneer work, GRX-BSDM is a complete HAD method using the BSDM. LREN combined DNN and low-rank theory for HAD and has high performance. The optimal parameters are selected for each method based on the corresponding literature. Our modified DM is trained with Adam optimizer, which has an initial learning rate of 1e-3 and decreases to 1e-5 with cosine scheduler, and λ_β is set to 0.015. A personal computer with an Intel Core i9-12900K, 32 GB of memory, NVIDIA 3090 Ti GPU, and 64-bit Windows 10 is used for all of the experiments.

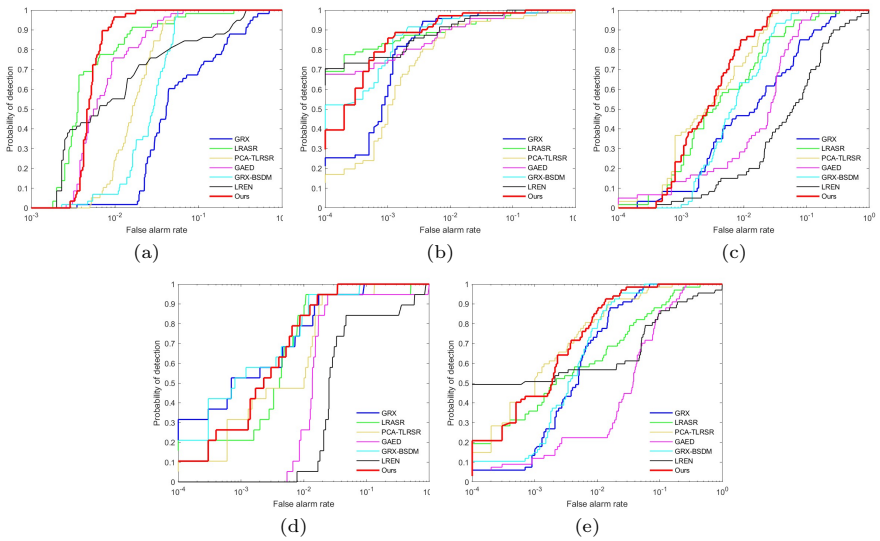


Fig. 4: ROC curves of of (P_D, P_F) for five datasets. (a) San Diego. (b) Pavia. (c) ABU-airport. (d) ABU-beach. (e) ABU-urban.

Table 1: AUC Values and Running Time (Unit:second) of Different Anomaly Detection Methods on the Seven Datasets

Datasets	AUC	GRX	LRASR	PCA-TLRSR	GAED	GRX-BSDM	LREN	Ours
San Diego	P_D, P_F	0.8884	0.9876	0.9802	0.9889	0.9707	0.9511	**0.9945**
	OD	0.9188	1.2094	1.1979	1.1151	1.0571	0.9723	**1.3385**
	Time	0.3148	7.9998	5.8821	21.3999	8.0257	65.0716	10.9280
Pavia	P_D, P_F	0.9933	0.9945	0.9793	0.9947	0.9951	0.9964	**0.9969**
	OD	1.1212	**1.4753**	1.2642	1.2117	1.1607	1.3197	1.3214
	Time	0.2819	69.2686	5.8464	11.1232	7.7713	59.3031	10.4237
Airport	P_D, P_F	0.9525	0.9797	0.9931	0.9681	0.9866	0.8730	**0.9945**
	OD	1.0005	1.2884	**1.4263**	1.2273	1.0956	1.0468	1.3508
	Time	0.3026	8.3175	5.8402	20.9851	7.7705	68.8871	19.8395
Beach	P_D, P_F	0.9912	0.9694	0.9859	0.9356	0.9927	0.8839	**0.9944**
	OD	1.3271	1.1480	1.3485	1.0396	**1.4230**	0.9754	1.2993
	Time	0.3079	45.8022	5.8991	20.6471	7.7526	71.3927	15.4711
Urban	P_D, P_F	0.9906	0.9672	0.9916	0.9446	0.9924	0.9188	**0.9946**
	OD	1.2463	1.0228	1.4227	1.2383	1.2974	1.2991	**1.4259**
	Time	0.3242	8.7505	5.7802	21.7135	7.7986	71.0485	20.3932

4.2 Results and Comparison with State of the Arts

The visual comparisons of various HAD methods across different datasets are shown in Figure 3. It is evident that GRX is susceptible to severe background contamination. Although GRX-BSDM somewhat improves the HAD results of GRX, it still fails to fully capture anomalies in HSI. LREN performs exceptionally well on Pavia and Urban but fails to accurately detect anomalies in other datasets, demonstrating weaker generalization capabilities. In the Airport dataset, compared to our method, LRASR, PCA-TLRSR, and GAED exhibit severe false alarm issues, and other methods fail to effectively detect anomalies. DBD clearly shows the shape of the airplane in its detection results, highlighting our advantage in identifying structured anomalies in HSI. On Beach, all methods except DBD exhibit notable false alarms, incorrectly giving high responses to objects in the lower right as anomalies. DBD also effectively detects anomalies at the edges of Urban HSI which most methods miss. However, unlike the LREN method, which shares this advantage, DBD erroneously identifies some backgrounds as anomalies. Overall, DBD outperforms other methods in background suppression, with anomalous pixels showing significantly stronger responses compared to the background.

As can be seen from the ROC curve of (P_D, P_F) in Fig. 4, the curve of our method is overall higher than the other methods on all datasets. In the San Diego, Pavia, and Urban datasets, the detection rate P_D of our method is significantly higher than other methods when the false alarm rate P_F is around 10^{-2}. And P_D reaches 0.9 in all datasets when P_F is around 10^{-2}. Therefore, it proves that the P_D of our method exceeds that of the other methods at extremely low P_F, and demonstrates that the DBD is able to discriminate between the background and the anomaly effectively. The ROC curves are interspersed with each other on the Airport and Beach datasets, which makes it difficult to visually evaluate the detection effect of each method, so the AUC values are further used to compare.

To more accurately express HAD efficiency, the AUC values of the ROC curves in Figure 4 have been calculated with greater precision. Table 1 presents the AUC values and running times achieved by various HAD methods across five datasets. Through comparison, it is evident that the DBD method achieves the highest AUC_{P_D,P_F} values across all datasets within an acceptable time range. In contrast, previous methods only achieve the optimal AUC_{P_D,P_F} on a single dataset. Specifically, the DBD method shows improvements in accuracy over the previous state-of-the-art by 0.56%, 0.05%, 0.14%, 0.17%, and 0.22% on the San Diego, Pavia, Airport, Beach, and Urban datasets, respectively. Meanwhile, our AUC_{OD} metric also achieves either the best or second-best performance across all datasets except Beach, indicating that the DBD model maintains a slightly low false alarm rate while achieving excellent anomaly detection performance. This effectively demonstrates that our method has a stronger detection capability compared to other HAD methods.

Table 2: AUC_{P_D,P_F} VALUE OF ABLATION EXPERIMENTS

Methods	San Diego	Pavia	Airport	Beach	Urban
tensor LRR	0.9635(-0.031)	0.9933(-0.003)	0.9242(-0.070)	0.9774(-0.017)	0.9908(-0.004)
DM	0.5701(-0.424)	0.9731(-0.023)	0.6717(-0.323)	0.9602(-0.034)	0.9814(-0.013)
BSDM-TLRR	0.9852(-0.009)	0.9937(-0.003)	0.9247(-0.070)	0.9839(-0.011)	0.9908(-0.004)
DBD	0.9945	0.9969	0.9945	0.9944	0.9946

4.3 Ablation Study

We further investigate the enhancement of the model's HAD capabilities introduced by the innovations presented in this paper. Table 2 demonstrates that DBD consistently outperforms tensor LRR and DM across all datasets. This indicates that the combination of DM with the LRR model, using DM to generate the background dictionary, significantly enhances the model's HAD capability. Furthermore, we process the results of BSDM in the same way as DBD for the background dictionary of the tensor LRR model, which is denoted as BSDM-TLRR. As shown in Table 2, the AUC_{P_D,P_F} of DBD surpasses that of BSDM-TLRR in all datasets. It verifies that the DM modified according to the nature of HSI data in this paper enhances the ability to detect anomalies. Moreover, BSDM-TLRR exhibits a lower AUC_{P_D,P_F} than PCA-TLRSR and GRX-BSDM on the ABU dataset, indicating that simply and directly combining BSDM with the tensor LRR model is ineffective for HAD.

5 Conclusion

Our DBD, as a new approach for HAD, obtains the dictionary tensor that is essential for the tensor LRR by processing the HSI generated by the modified DM. Introducing pseudo background noise that follows a mutually independent multivariate Gaussian normal distribution into the DM model can better suppress the background and prevent the background dictionary from containing anomalous information. Subsequent ADMM is used to solve the tensor LRR optimization problem. Experiments demonstrate the validity of the innovative points in this paper, and our method maintains excellent anomaly detection on all five datasets compared to the other state-of-the-art methods.

Acknowledgements This paper is supported by the Department of Education of Guangdong Province (No.2022ZDZX1006).

References

1. Austin, J., Johnson, D.D., Ho, J., Tarlow, D., Van Den Berg, R.: Structured denoising diffusion models in discrete state-spaces. Advances in Neural Information Processing Systems **34**, 17981–17993 (2021)
2. Bati, E., Çalışkan, A., Koz, A., Alatan, A.A.: Hyperspectral anomaly detection method based on auto-encoder. In: Image and Signal Processing for Remote Sensing XXI. vol. 9643, pp. 220–226. Spie (2015)
3. Borghys, D., Kåsen, I., Achard, V., Perneel, C.: Comparative evaluation of hyperspectral anomaly detectors in different types of background. In: Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII. vol. 8390, pp. 803–814 (2012)
4. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? Journal of the ACM (JACM) **58**(3), 1–37 (2011)
5. Chen, N., Yue, J., Fang, L., Xia, S.: Spectraldiff: A generative framework for hyperspectral image classification with diffusion models. IEEE Transactions on Geoscience and Remote Sensing (2023)
6. Chen, Z., Yang, B., Wang, B.: A preprocessing method for hyperspectral target detection based on tensor principal component analysis. Remote Sensing **10**(7), 1033 (2018)
7. Cheng, T., Wang, B.: Graph and total variation regularized low-rank representation for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **58**(1), 391–406 (2019)
8. Cheng, X., Mu, R., Lin, S., Zhang, M., Wang, H.: Hyperspectral anomaly detection via low-rank representation with dual graph regularizations and adaptive dictionary. Remote Sensing **16**(11) (2024). https://doi.org/10.3390/rs16111837, https://www.mdpi.com/2072-4292/16/11/1837
9. Croitoru, F.A., Hondru, V., Ionescu, R.T., Shah, M.: Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
10. Deng, K., Qian, Y., Nie, J., Zhou, J.: Diffusion model based hyperspectral unmixing using spectral prior distribution. IEEE Transactions on Geoscience and Remote Sensing (2024)
11. Dong, W., Liu, S., Xiao, S., Qu, J., Li, Y.: Ispdiff: Interpretable scale-propelled diffusion model for hyperspectral image super-resolution. IEEE Transactions on Geoscience and Remote Sensing (2024)
12. Feng, M., Zhu, Y., Yang, Y., Shu, Q.: Deep low-rank and piecewise-smooth constraint tensor model for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–14 (2023). https://doi.org/10.1109/TGRS.2023.3329639
13. Floto, G., Pour, M.M.A., Farinneya, P., Tang, Z., Pesaranghader, A., Bharadwaj, M., Sanner, S.: Diffudetox: A mixed diffusion model for text detoxification. arXiv preprint arXiv:2306.08505 (2023)
14. Fu, X., Jia, S., Zhuang, L., Xu, M., Zhou, J., Li, Q.: Hyperspectral anomaly detection via deep plug-and-play denoising cnn regularization. IEEE Transactions on Geoscience and Remote Sensing **59**(11), 9553–9568 (2021)
15. Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., Plaza, A.: Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. IEEE Geoscience and Remote Sensing Magazine **5**(4), 37–78 (2017). https://doi.org/10.1109/MGRS.2017.2762087

16. He, X., Wu, J., Ling, Q., Li, Z., Lin, Z., Zhou, S.: Anomaly detection for hyperspectral imagery via tensor low-rank approximation with multiple subspace learning. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–17 (2023). https://doi.org/10.1109/TGRS.2023.3270667
17. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
18. Huyan, N., Zhang, X., Zhou, H., Jiao, L.: Hyperspectral anomaly detection via background and potential anomaly dictionaries construction. IEEE Transactions on Geoscience and Remote Sensing **57**(4), 2263–2276 (2018)
19. Jiang, K., Xie, W., Lei, J., Jiang, T., Li, Y.: Lren: Low-rank embedded network for sample-free hyperspectral anomaly detection. Proceedings of the AAAI Conference on Artificial Intelligence **35**, 4139–4146 (05 2021). https://doi.org/10.1609/aaai.v35i5.16536
20. Jiang, T., Li, Y., Xie, W., Du, Q.: Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **58**(7), 4666–4679 (2020)
21. Kang, X., Zhang, X., Li, S., Li, K., Li, J., Benediktsson, J.A.: Hyperspectral anomaly detection with attribute and edge-preserving filters. IEEE Transactions on Geoscience and Remote Sensing **55**(10), 5600–5611 (2017). https://doi.org/10.1109/TGRS.2017.2710145
22. Kilmer, M.E., Braman, K., Hao, N., Hoover, R.C.: Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. SIAM Journal on Matrix Analysis and Applications **34**(1), 148–172 (2013)
23. Kim, G., Chun, S.Y.: Datid-3d: Diversity-preserved domain adaptation using text-to-image diffusion for 3d generative model. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14203–14213 (2023)
24. Li, C., Zhang, B., Hong, D., Yao, J., Chanussot, J.: Lrr-net: An interpretable deep unfolding network for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–12 (2023). https://doi.org/10.1109/TGRS.2023.3279834
25. Li, W., Wu, G., Du, Q.: Transferred deep learning for anomaly detection in hyperspectral imagery. IEEE Geoscience and Remote Sensing Letters **14**(5), 597–601 (2017)
26. Li, Y., Jiang, T., Xie, W., Lei, J., Du, Q.: Sparse coding-inspired gan for hyperspectral anomaly detection in weakly supervised learning. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–11 (2021)
27. Li, Z., Wang, Y., Xiao, C., Ling, Q., Lin, Z., An, W.: You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–18 (2023)
28. Lin, S., Zhang, M., Cheng, X., Shi, L., Gamba, P., Wang, H.: Dynamic low-rank and sparse priors constrained deep autoencoders for hyperspectral anomaly detection. IEEE Transactions on Instrumentation and Measurement **73**, 1–18 (2024). https://doi.org/10.1109/TIM.2023.3323997
29. Lin, S., Zhang, M., Cheng, X., Zhao, S., Shi, L., Wang, H.: Hyperspectral anomaly detection using spatial–spectral-based union dictionary and improved saliency weight. Remote Sensing **15**(14) (2023). https://doi.org/10.3390/rs15143609, https://www.mdpi.com/2072-4292/15/14/3609

30. Liu, G., Lin, Z., Yu, Y.: Robust subspace segmentation by low-rank representation. In: Proceedings of the 27th international conference on machine learning (ICML-10). pp. 663–670 (2010)
31. Liu, R., Wu, R., Van Hoorick, B., Tokmakov, P., Zakharov, S., Vondrick, C.: Zero-1-to-3: Zero-shot one image to 3d object. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9298–9309 (2023)
32. Liu, Y., Xie, W., Li, Y., Li, Z., Du, Q.: Dual-frequency autoencoder for anomaly detection in transformed hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–13 (2022)
33. Lu, C., Feng, J., Chen, Y., Liu, W., Lin, Z., Yan, S.: Tensor robust principal component analysis with a new tensor nuclear norm. IEEE Transactions on Pattern Analysis and Machine Intelligence **42**(4), 925–938 (2020). https://doi.org/10.1109/TPAMI.2019.2891760
34. Ma, J., Xie, W., Li, Y., Fang, L.: Bsdm: Background suppression diffusion model for hyperspectral anomaly detection. arXiv preprint arXiv:2307.09861 (2023)
35. Molero, J.M., Garzon, E.M., Garcia, I., Plaza, A.: Analysis and optimizations of global and local versions of the rx algorithm for anomaly detection in hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **6**(2), 801–814 (2013)
36. Qu, J., He, J., Dong, W., Zhao, J.: S2cyclediff: Spatial-spectral-bilateral cycle-diffusion framework for hyperspectral image super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4623–4631 (2024)
37. Qu, Y., Wang, W., Guo, R., Ayhan, B., Kwan, C., Vance, S., Qi, H.: Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition. IEEE Transactions on Geoscience and Remote Sensing **56**(8), 4391–4405 (2018)
38. Reed, I.S., Yu, X.: Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution. IEEE Transactions on Acoustics, Speech, and Signal Processing **38**(10), 1760–1770 (1990)
39. Ren, Z., Sun, L., Zhai, Q., Liu, X.: Mineral mapping with hyperspectral image based on an improved k-means clustering algorithm. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 2989–2992. IEEE (2019)
40. Sawant, S., Patole, Y., Sawant, C., Kumawat, D.: Food safety inspection and control using hyperspectral imaging: A review. SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology **12**(SUP 1), 204–206 (2020)
41. Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International conference on machine learning. pp. 2256–2265. PMLR (2015)
42. Su, H., Wu, Z., Zhang, H., Du, Q.: Hyperspectral anomaly detection: A survey. IEEE Geoscience and Remote Sensing Magazine **10**(1), 64–90 (2021)
43. Su, H., Wu, Z., Zhang, H., Du, Q.: Hyperspectral anomaly detection: A survey. IEEE Geoscience and Remote Sensing Magazine **10**(1), 64–90 (2022). https://doi.org/10.1109/MGRS.2021.3105440
44. Sun, S., Liu, J., Chen, X., Li, W., Li, H.: Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints. IEEE Transactions on Neural Networks and Learning Systems **34**(11), 8679–8692 (2023). https://doi.org/10.1109/TNNLS.2022.3152252
45. Sun, S., Liu, J., Li, W.: Spatial invariant tensor self-representation model for hyperspectral anomaly detection. IEEE Transactions on Cybernetics pp. 1–12 (2023). https://doi.org/10.1109/TCYB.2022.3233108

46. Sun, W., Liu, C., Li, J., Lai, Y.M., Li, W.: Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery. Journal of Applied Remote Sensing **8**(1), 083641–083641 (2014)
47. Wan, Y., Hu, X., Zhong, Y., Ma, A., Wei, L., Zhang, L.: Tailings reservoir disaster and environmental monitoring using the uav-ground hyperspectral joint observation and processing: A case of study in xinjiang, the belt and road. In: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium. pp. 9713–9716. IEEE (2019)
48. Wang, J., Guo, S., Hua, Z., Huang, R., Hu, J., Gong, M.: Cl-cagan: Capsule differential adversarial continual learning for cross-domain hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing (2024)
49. Wang, M., Wang, Q., Hong, D., Roy, S.K., Chanussot, J.: Learning tensor low-rank representation for hyperspectral anomaly detection. IEEE Transactions on Cybernetics **53**(1), 679–691 (2022)
50. Wang, Y., Cai, J., Zhou, J., Sun, J., Xu, Y., Zhang, Y., Wei, Z., Plaza, J., Plaza, A., Wu, Z.: Ce-rx: A collaborative cloud-edge anomaly detection approach for hyperspectral images. Remote Sensing **15**(17) (2023). https://doi.org/10.3390/rs15174242, https://www.mdpi.com/2072-4292/15/17/4242
51. Wu, C., Wang, D., Bai, Y., Mao, H., Li, Y., Shen, Q.: Hsr-diff: hyperspectral image super-resolution via conditional diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7083–7093 (2023)
52. Wu, Z., Tao, X., Paoletti, M.E., Haut, J.M., Pastor-Vargas, R., Plaza, A.: Deep unrolling network with active dictionary learning for hyperspectral anomaly detection. In: 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS). pp. 1–5 (2023). https://doi.org/10.1109/WHISPERS61460.2023.10431277
53. Xiang, P., Ali, S., Jung, S.K., Zhou, H.: Hyperspectral anomaly detection with guided autoencoder. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–18 (2022)
54. Xu, Y., Wu, Z., Li, J., Plaza, A., Wei, Z.: Anomaly detection in hyperspectral images based on low-rank and sparse representation. IEEE Transactions on Geoscience and Remote Sensing **54**(4), 1990–2000 (2015)
55. Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., Zhang, W., Cui, B., Yang, M.H.: Diffusion models: A comprehensive survey of methods and applications. ACM Computing Surveys **56**(4), 1–39 (2023)
56. Yang, Y., Zhang, J., Song, S., Liu, D.: Hyperspectral anomaly detection via dictionary construction-based low-rank representation and adaptive weighting. Remote Sensing **11**(2), 192 (2019)
57. Zeng, H., Cao, J., Zhang, K., Chen, Y., Luong, H., Philips, W.: Unmixing diffusion for self-supervised hyperspectral image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27820–27830 (2024)
58. Zhang, X., Ma, X., Huyan, N., Gu, J., Tang, X., Jiao, L.: Spectral-difference low-rank representation learning for hyperspectral anomaly detection. IEEE Transactions on Geoscience and Remote Sensing **59**(12), 10364–10377 (2021)
59. Zhang, Z., Ely, G., Aeron, S., Hao, N., Kilmer, M.: Novel methods for multilinear data completion and de-noising based on tensor-svd. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3842–3849 (June 2014)

60. Zhou, J., Sheng, J., Ye, P., Fan, J., He, T., Wang, B., Chen, T.: Exploring multi-timestep multi-stage diffusion features for hyperspectral image classification. IEEE Transactions on Geoscience and Remote Sensing (2024)