Pattern Recognition 169 (2026) 111945

Contents lists available at ScienceDirect

Pattern Recognition

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

Leveraging multi-class background description and token dictionary representation for hyperspectral anomaly detection

Zhiwei Wang [a], Kun Tan [a,b,*], Xue Wang [a], Wen Zhang [c]

[a] Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai 200241, China
[b] School of Geospatial Artificial Intelligence, East China Normal University, Shanghai 200241, China
[c] Shanghai Municipal Institute of Surveying and Mapping, Shanghai 200063, China

ARTICLE INFO

Keywords:
Hyperspectral anomaly detection
Data augmentation
Multi-class background description
Token dictionary representation

ABSTRACT

Hyperspectral anomaly detection is aimed at distinguishing between background and anomalous regions in hyperspectral images, and plays a crucial role iSn various applications. However, the existing deep learning methods face challenges when dealing with complex background distributions and insufficient training data. In this article, we propose a novel multi-class background description transformer network (MBDTNet) to address the problems of imprecise background distribution learning and poor anomaly detection. Firstly, we propose an image-level end-to-end data augmentation method based on self-supervised training, which enhances the diversity and quantity of the training samples through adaptive clustering and spatial masking strategies. Secondly, based on the principles of low-rank representation, a sparse self-attention mechanism based on token dictionary representation is designed to help the model focus on key background features and guide the model in recognizing anomalies. Finally, a token dictionary learning mechanism for multi-class background description is established by combining Gaussian discriminant analysis with a conditional distance function, and intra-class and inter-class losses are designed to enhance the model's ability to separate background and anomalies. Experiments on five benchmark datasets demonstrate the superiority and applicability of the proposed MBDTNet method, showing that it outperforms the current state-of-the-art hyperspectral anomaly detection methods.

1. Introduction

Hyperspectral images (HSIs) acquire scene representations through hundreds of contiguous spectral bands, encoding rich material-specific signatures in the spectral-spatial domain [1,2]. Unsupervised hyperspectral anomaly detection identifies spectrally distinctive pixels deviating from background statistics without prior target knowledge [3]. These anomalies typically exhibit stochastic spatial distribution patterns, ultra-low occurrence probabilities, and subpixel-scale manifestations that challenge conventional detection paradigms [4]. For example, a tank in a forest background, a vehicle in a city, and artificially placed panels can all be referred to as anomalies.

The current hyperspectral anomaly detection models can be categorized into statistics-based models, representation-based models, and deep learning based models. The statistics-based models formulate anomaly detection as a hypothesis testing problem, where background pixels are assumed to follow a parameterized probability distribution (e.g., multivariate Gaussian), with anomalies exhibiting statistically significant deviations from the learned distribution [5]. Notable statistical model algorithms include the Reed-Xiaoli (RX) detector [6] and support vector data description (SVDD) [7]. The RX algorithm implements a Gaussian-based anomaly detector by constructing the Mahalanobis distance metric as the detection statistic [6]. Building on the classical RX algorithm, several extensions have been proposed, such as local RX [8], weighted RX [9], and the random selection based anomaly detector (RSAD) [10]. Unlike RX's parametric density estimation, SVDD [7] adopts nonparametric boundary learning that is particularly effective for multi-modal distributions. Various integrated methods based on kernel transformation analysis have also been proposed to enhance the robustness of the background model [11]. However, the statistics-based methods rely on distribution assumptions that may not hold in complex real-world backgrounds, leading to a decline in anomaly detection performance.

The representation-based models leverage the inherent properties of hyperspectral imagery to detect anomalies, aiming to design more effective detection models from the perspectives of collaborative

* Corresponding author.
E-mail address: tankuncu@gmail.com (K. Tan).

https://doi.org/10.1016/j.patcog.2025.111945
Received 22 February 2025; Received in revised form 5 May 2025; Accepted 31 May 2025
Available online 3 June 2025

representation [12], low-rank and sparse representation [13,14], and topological structure [15–17]. Improvements to collaborative representation based models have focused on enhancing the background adaptability, primarily through optimization of the background dictionary [18], spatial optimization [19] joint sparsity regularization [20]. For instance, Zhang et al [21] adopted a self-paced learning strategy to optimize background atoms, which improved the model's generalization ability. The low-rank and sparse representation models take advantage of the low-rank property of the background and the sparse characteristics of the anomaly to locate anomalies [22]. These models are iteratively updated through strategies such as dictionary learning [23], tensor decomposition [24,25], and orthogonal subspace projection [26]. For example, Qin et al [25] proposed generalized non-convex low-order tensor representation, which efficiently detects targets by establishing a unified solver for fast and effective detection. Ren et al [27] proposed a unified nonconvex anomaly detection framework, HADGSMs, which introduces generalized shrinkage mappings to more precisely approximate the penalty terms in LRR-based models. Alternatively, topological structure based models, such as the structure tensor and guided filter (STGF) [15], chessboard topology-based anomaly detection (CTAD) [16], and attribute and edge-preserving filters [17], enhance detection performance by leveraging spatial topology, without requiring complex modeling processes. However, the aforementioned representation-based methods rely on constructing a background dictionary specific to each HSI, which limits their transferability across different scenarios and results in high computational complexity.

Recently, deep learning has advanced rapidly in the field of hyperspectral imagery, bringing more novel ideas to anomaly detection [28–30]. Deep learning networks, such as autoencoders (AE) [31], generative adversarial networks (GANs) [32,33], and transformer networks [34,35] have been successfully applied to anomaly detection. The AE-based approach generally follows a two-step detection paradigm, where the network learns the background distribution and detects anomalies using reconstruction errors [36–38]. For example, Wang et al [39] proposed Auto-AD based on the U-Net architecture, which reconstructs the original image using an encoder-decoder structure and detects anomalies by leveraging reconstruction errors. Wang et al [40] proposed a dual-window-inspired reconstruction network (DirectNet) to predict the center pixel using the outer window information, which improved the accuracy of background reconstruction. Wang et al [35] utilized the transformer network and incorporated finite spatial wise attention to precisely reconstruct the image. Moreover, GANs are capable of learning the probabilistic distribution of multivariate data, thereby enhancing the stability of the model [32]. Weakly supervised [41,42] and semi-supervised [43,44] strategies have also been applied to GAN frameworks to improve networks stability. For instance, Li et al [45] proposed a background search strategy to extract training samples and trained a sparse coding GAN in a weakly supervised manner, detecting targets in the latent space. To fully utilize both global and local spatial features, Li et al [34] proposed an approach for unsupervised learning of spatial contextual features between background and anomalies and on anomaly-free images with random masking. Lian et al [46] proposed a gated transformer anomaly detection network by introducing a content matching mechanism and adaptive gating units, effectively distinguishing between background and anomalies using spatial-spectral similarity. Subsequently, the anomaly enhancement transformation network (AETNet) [34] and transferred direct detection (TDD) [47] establish a one-step detection paradigm by unsupervised learning of spatial contextual features between background and anomalies on anomaly-free HSIs with random masking, addressing the limitations of the traditional two-step or multi-step anomaly detection networks. In addition, scholars have developed model-driven deep networks for anomaly detection by coupling physical models with deep learning [3,48–51]. Specifically, the model-driven deep mixture network (MDMN) [48] and the low-rank representation network (LRR-Net) [49] transform regularization parameters into trainable parameters of networks while emphasizing the interpretability of the model. Shen et al [51] proposed the deep denoising dictionary tensor, which integrates low-rank and deep denoising priors into the dictionary and coefficient tensors, enhancing the separation of background and anomaly.

Deep learning methods have shown promising results in hyperspectral anomaly detection, but upon review, several challenges remain that require further investigation and resolution:

1) The one-step anomaly detection approach directly learn the feature differences between background and anomalies, quickly outputting detection results [34,47]. However, when the training samples contain contamination from anomalous pixels, the model struggles to accurately learn the boundaries between background and anomalies. Therefore, constructing an effective clean sample extraction strategy remains requires further exploration.
2) Most deep learning methods assume that background samples belong to a single class. However, as the scenes in HSIs become larger and the background more complex, the assumption of a single background class often fails to accurately reflect the true diversity of the background. Although some research has explored multi-class background separation [37,45] and endmember extraction [26], addressing the differences between the background classes in complex hyperspectral scenes remains an open problem that requires further investigation.
3) Transformer models have been widely applied in hyperspectral anomaly detection, with a focus primarily on capturing spatial context features or spectral features [46,47]. However, these models have not fully explored the potential of low-rank and dictionary learning, which limits their performance in anomaly detection. Therefore, integrating low-rank representation with the transformer architecture and optimizing the learning of background low-rank structures remain unresolved issues.

In this article, to tackle the above challenges, we propose a multi-class background description transformer network (MBDTNet) for hyperspectral anomaly detection. Firstly, to provide stable and pure training samples, an image patch based anomaly data augmentation strategy is introduced, which increases the diversity of the training samples through adaptive clustering and spatial random masking processes. Simultaneously, MBDTNet is trained in a self-supervised manner, utilizing unsupervised data to generate background pseudo-labels, thereby supporting multi-class background description. In addition, based on the concepts of low-rank representation and dictionary learning, a self-attention mechanism based on token dictionary sparse representation is incorporated. By using the token dictionary, each sample is represented as a sparse combination of dictionary elements, helping the model accurately identify the key features of the background classes, thereby reducing attention to non-key features. To group the background classes together and effectively separate anomalies, Gaussian discriminant analysis (GDA) is integrated into a deep neural network framework. This integration allows the model to learn spherical decision boundaries for each background class, ensuring a compact representation of the background. The network can then distinguish anomalous regions that lie outside the decision boundaries, through adaptive spectral-spatial feature integration that enhances operational resilience across diverse hyperspectral scenarios. Finally, a composite guided loss function is proposed, which consists of binary cross-entropy (BCE) loss, intra-class loss, and inter-class loss. This loss function enables the model to focus more on enhancing the distinction between background classes and improving the detection capability for anomalous samples. The main contributions of the MBDTNet method are summarized as follows:

1) MBDTNet is an image-level end-to-end network that directly outputs an anomaly detection map, and is capable of handling large-scale

HSIs with complex backgrounds. A patch-based anomaly data augmentation strategy is established that trains the model in a self-supervised manner, enhancing its ability to detect anomalies accurately in practical applications.

2) A self-attention mechanism based on token dictionary sparse representation is proposed, where the attention weights are used to represent the degree of background feature reconstruction. This mechanism enhances the model's focus on key background features, helping to highlight the differences between background and anomalies.
3) A multi-class background description method is proposed that constructs spherical decision boundaries for each class and dynamically generates token dictionaries, effectively capturing the complex distribution of hyperspectral backgrounds. In addition, intra-class and inter-class losses are designed to enhance the distinction between background and anomalies.

2. Related work

2.1. Low-rank representation and dictionary learning

The LRR [49] assumes that the background exhibits low-rank properties, while anomalies are characterized by their sparsity. HSI data can be defined as $\mathbf{X} = [x_1, x_2, \ldots, x_N] \in \Re^{N\times b}$, where x_i represents the i-th spectral vector with b dimensions, and N denotes the total number of pixels. Thus, the $\mathbf{X}$ can be decomposed into the background component and the anomaly component, as follows:

$$\mathbf{X} = \mathbf{DA} + \mathbf{S} \tag{1}$$

where $\mathbf{D}$ is the background dictionary, and $\mathbf{A}$ refers to the representation coefficients. The optimization problem for the LRR model can be denoted as:

$$\min_{L,S} \| \mathbf{A} \|_* + \beta \| \mathbf{S} \|_{2,1}, \text{s.t.} \mathbf{X} = \mathbf{DA} + \mathbf{S} \tag{2}$$

where $\| \mathbf{A} \|_*$ represents the nuclear norm of the $\mathbf{A}$, and $\| \mathbf{S} \|_{2,1}$ denotes the $l_{2,1}$-norm of the $\mathbf{S}$.β is a regularization parameter. To better isolate the anomaly information, the desired learned dictionary should, as much as possible, contain only the spectra of the background [14]. Typically, a dictionary learning algorithm based on gradient iteration is employed. Firstly, an initial dictionary $\mathbf{D}$ is randomly generated, and is then iteratively updated using a gradient algorithm, as follows:

$$\mathbf{D}^{(n+1)} = \mathbf{D}^{(n)} - \mu \sum_{i=1}^{M} (\mathbf{D}^{(n)}\mathbf{A}_i - x_i)\mathbf{A}_i^T \tag{3}$$

where μ is the step size in each iteration, and M is the number of samples selected in each iteration.

2.2. Deep support vector data description

Anomaly detection models can also be considered as learning a model that accurately describes normal data, with deviations from the model considered to be anomalies. This approach is commonly referred to as one-class classification. SVDD [7] aims to find a closed spherical boundary, known as a hypersphere, that encloses the normal data, while anomalies are located outside this hypersphere. The learning objective of SVDD is to learn the center c, radius R, and the feature mapping Φ associated with its kernel, with the observation equation as follows:

$$\min_{c,R,\xi} R^2 + \frac{1}{\nu n} \sum_{i=1}^{N} \xi_i \, s.t \forall_i, \| \Phi(x_i) - c \|^2 \le R^2 + \xi_i, \xi_i \ge 0 \tag{4}$$

where ξ_i is the soft boundary, and ν is the penalty coefficient, which is

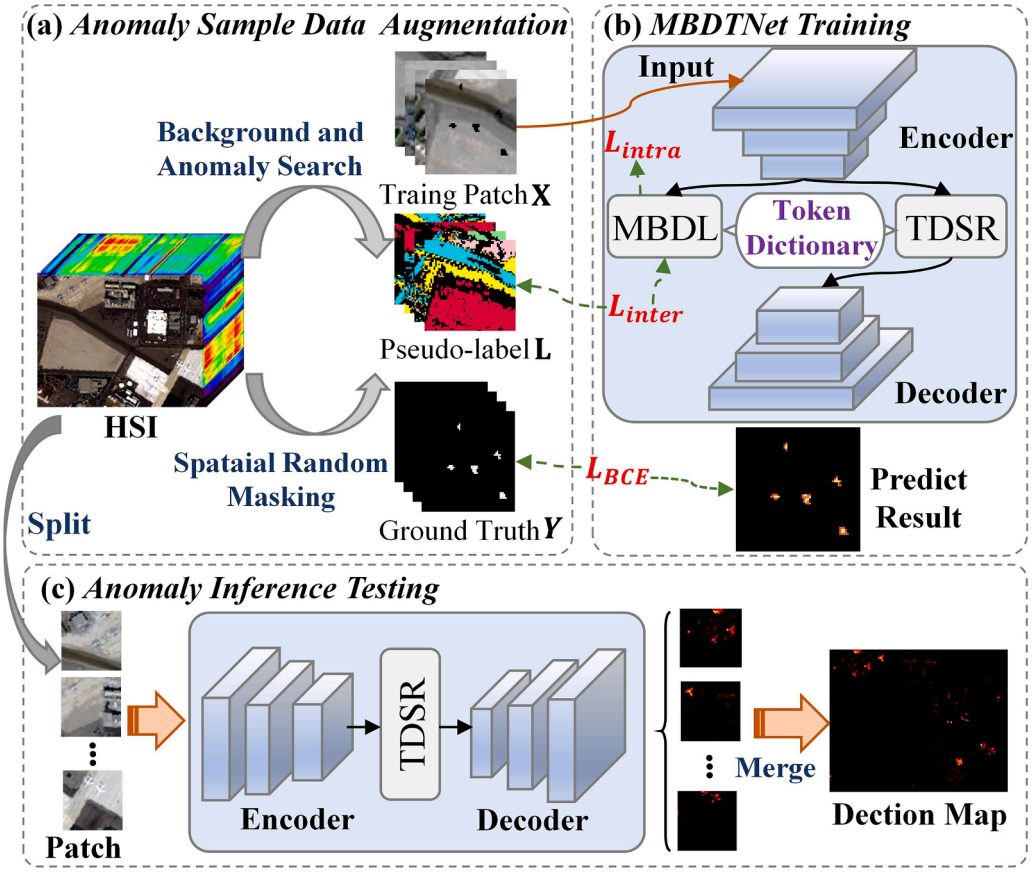

Fig. 1. Flowchart of the proposed MBDTNet architecture.

used to control the volume of the hypersphere.

Deep SVDD trains a deep neural network (DNN) to fit the network outputs into a hypersphere of minimal volume. However, Deep SVDD [52] is limited to single-class data description, as it assumes that all the samples in the dataset belong to the same class. Deep multi-class data description (MCDD) [53] introduces the concept of multi-class data description, employing multiple-sphere modeling instead of single-sphere modeling, where each sphere represents a background class. Deep-MCDD optimizes the DNN to map training samples into a latent space close to the center c_k, ultimately defining a hypersphere of minimal volume centered at c_k. Given N training samples (x_i, y_i) from K different classes, the objective of Deep-MCDD can be described as follows:

$$\min_{\mathscr{W},c,r} \sum_{k=1}^{K} R_k^2 + \frac{1}{\nu N} \sum_{n=1}^{N} \max\left\{0, \zeta_{ik}\left(\| f(x_i; \mathscr{W}) - c_k \|^2 - R_k^2\right)\right\} \tag{5}$$

where $\mathscr{W}$ is the network parameter of the trained model, ζ_{ik} represents the class assignment indicator, and ν determines the strength of each hypersphere in enclosing the corresponding training samples. Subsequently, for a given test data point x_t, the anomaly score in Deep-MCDD can be defined based on the distance from this point to the center of the hypersphere, expressed as $\| f(x_t; \mathscr{W}) - c_k \|^2$. Therefore, in complex HSI scenes, it is crucial to develop an unsupervised MCDD method that groups the various background categories into the same attribute category while helps to detect anomalies.

3. Proposed method

The proposed MBDTNet method consists of three main components: the anomaly sample data augmentation strategy, MBDTNet network training, and anomaly inference testing. An overview of the architecture of MBDTNet is provided in Fig. 1. Firstly, an image-level data augmentation method is introduced, which includes adaptive clustering and spatial random masking. Next, the MBDTNet architecture includes two key modules: token dictionary sparse representation (TDSR) and multi-class background description learning (MBDL), which are used to train the model by feeding the training samples into the network. TDSR module is capable to describe the background features by introducing token dictionary, in which each pixel is represented as a sparse linear combination of dictionary atoms. Simultaneously, MBDL module constructs spherical decision boundaries for different classes separately and utilizes tokens to represent the feature information of each class. The token dictionary provided by MBDL enhances the anomaly detection capability of TDSR, while the sparse modeling of TDSR further promotes the discriminative feature extraction. Finally, during the testing phase, the test HSI is divided into patches, and each patch is passed through the test network. The patches are then merged to obtain the final detection map.

3.1. Training sample augmentation

Due to the coexistence of diverse background and anomalies in large-scale HSIs, the traditional methods are often limited by the imbalance between background and anomaly samples. To address this issue, an anomaly data augmentation strategy is proposed that combines adaptive clustering and random masking to dynamically identify the background and anomaly categories while ensuring the availability of sufficient training samples. Specifically, a background and anomaly search strategy is established, aimed at extracting high-confidence background and anomaly classes. Given that HSIs typically contain multiple background classes with relatively abundant pixels, while anomaly pixels are sparse density-based spatial clustering of applications with noise (DBSCAN) [37] is employed to adaptively determine the number of ground object classes based on the spectral characteristics of the HSI.

As illustrated in Fig. 2, the spectral data of the HSI are input into DBSCAN to generate a clustering label map. Each class label's pixel count is then sorted by size, and a reasonable pixel threshold ζ is set to distinguish the pseudo background and anomaly classes. The background and anomaly search strategy is defined as follows:

$$\mathscr{L}(k_i) = \begin{cases} b_i \in \mathscr{B}_D, sum(k_i)/N > \zeta \\ a_i \in \mathscr{A}_D, sum(k_i)/N \le \zeta \end{cases} \tag{6}$$

where k_i represents the pixel set of class i. $sum(\cdot)$ is used to counting the number of pixels. N denotes the total number of pixels. ζ is a threshold parameter used to determine whether the current class belongs to anomaly or background. $\mathscr{B}_D$ represents the pseudo background set, and $\mathscr{A}_D$ represents the pseudo anomaly set.

Next, an $n \times n$ window slides over the image, starting from the top-left corner and traversing the entire image with a fixed step size. For each patch within the sliding window, the pixel labels of the patch are checked. The patch will be excluded if contains potential anomaly pixels $x_i \in \mathscr{A}_D$. Simultaneously, the Euclidean distance from the pixel to its cluster center is calculated to serve as a probability estimate, which helps in selecting pseudo-label samples based on their reliability. Pixels with a distance greater than 0.7 are classified as belonging to that class, and the remaining pixels are matched based on their location indices to generate the pseudo-label map for the current window.

Furthermore, a random masking strategy [34] is used to simulate

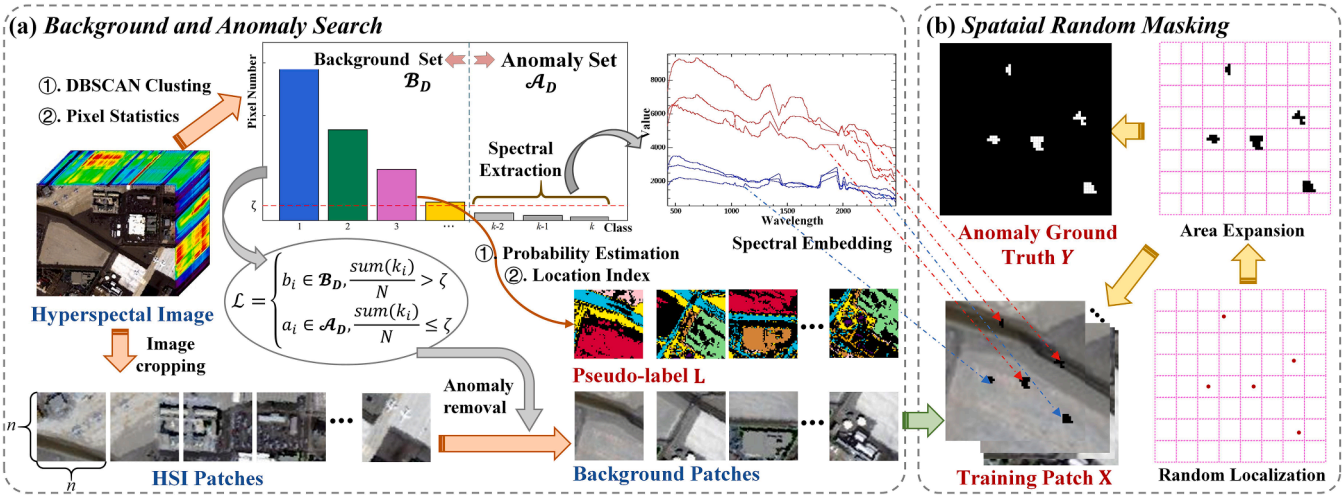

Fig. 2. Flowchart of training sample augmentation.

anomalous targets. As shown in Fig. 2, the random masking strategy simulates target shapes with irregular and random sizes by dividing the mask map **M** into k patches. The anomaly spectra, which are extracted from the pseudo anomaly set $\mathscr{A}_D$ using the DBSCAN algorithm, are embedded into the random mask regions. It is important to emphasize that the pseudo anomaly labels only serve as auxiliary supervision to indicate potential target regions and are not intended to represent true semantic-level anomaly annotations. The target areas are randomly selected and set to 1, while the background is set to 0, with the final mask map used to generate the ground-truth map for the targets. In addition, the regions set to 1 in the mask map **M** are not assigned background pseudo-labels and are defined as belonging to the "other" class. Finally, the training data consist of three components: $\mathbf{X} \in \Re^{n\times n\times b}$, representing the HSI patch; the background pseudo-label $\mathbf{L} \in \Re^{n\times n}$ corresponding to each training sample; and the anomaly ground-truth map $\mathbf{Y} \in \Re^{n\times n}$. Therefore, the primary task of MBDTNet from the background pseudo-label guides the model to distinguish features among different background categories. In contrast, the ground truth of anomaly indicates the anomaly regions and serves as a reference for the model's output during training.

3.2. MBDTNet architecture

Fig. 3 illustrates the proposed MBDTNet, which is based on a U-Net framework. The encoder of MBDTNet is built based on a vision transformer (ViT) [54] to effectively capture multi-scale features during the encoding process of HSIs. The encoder consists of two feature extraction stages, each composed of a ViT block. The first stage ends with a downsampling layer for dimensionality reduction. The structure of the ViT module, shown in Fig. 4(a), divides the image into fixed-size patches and linearly maps them into vectors. The self-attention mechanism inherently models patch relationships and global anomaly-background interactions through its long-range dependency learning capability. The downsampling layer is a 2 × 2 strided convolutional layer that reduces feature map's spatial resolution. At the end of the encoding process, the feature maps $\boldsymbol{F}_1 \in \Re^{c1\times n\times n}$ and $\boldsymbol{F}_2 \in \Re^{c2\times n\times n}$ are fused to feature map $\boldsymbol{F}_r$ using feature fusion module. As shown in Fig. 4(b), a bilinear interpolation layer upsamples the feature map $\boldsymbol{F}_2$, followed by convolution, batch normalization, and ReLU to obtain the fine-scale feature map. The $\boldsymbol{F}_1$ branch employ a channel attention (CA) [55] mechanism to selectively enhance discriminative feature channels.

After the feature extraction by the encoder, the TDSR model is

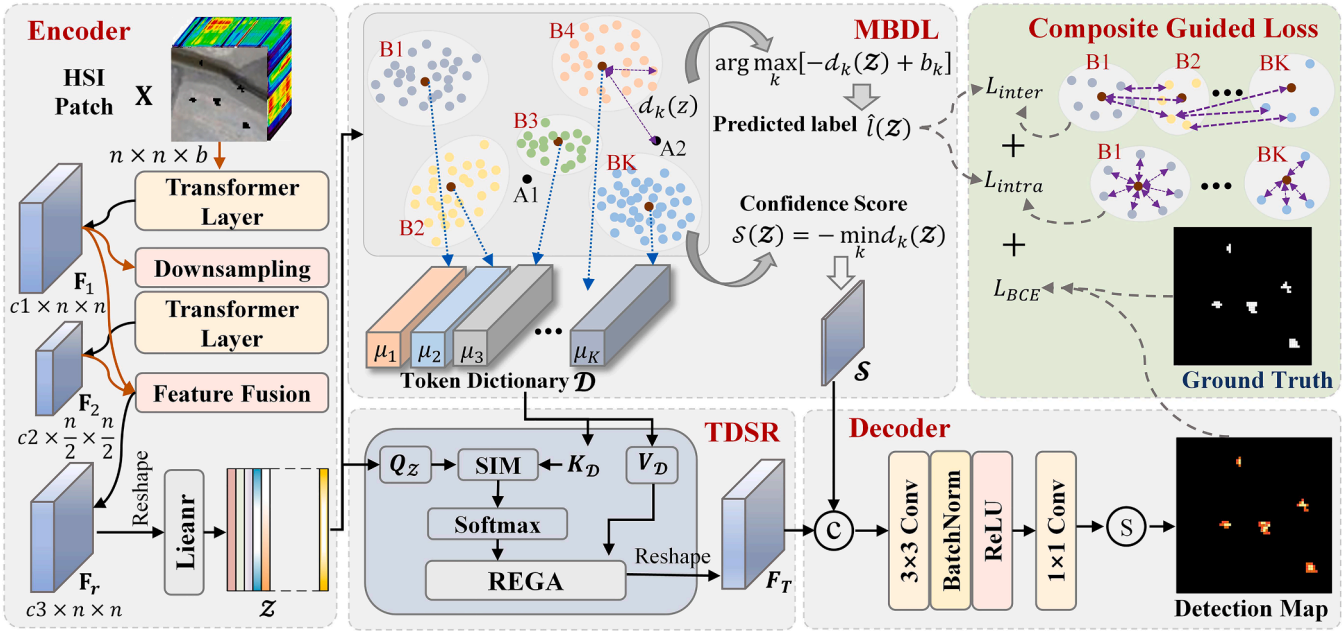

Fig. 3. Overview illustration of the proposed MBDTNet network.

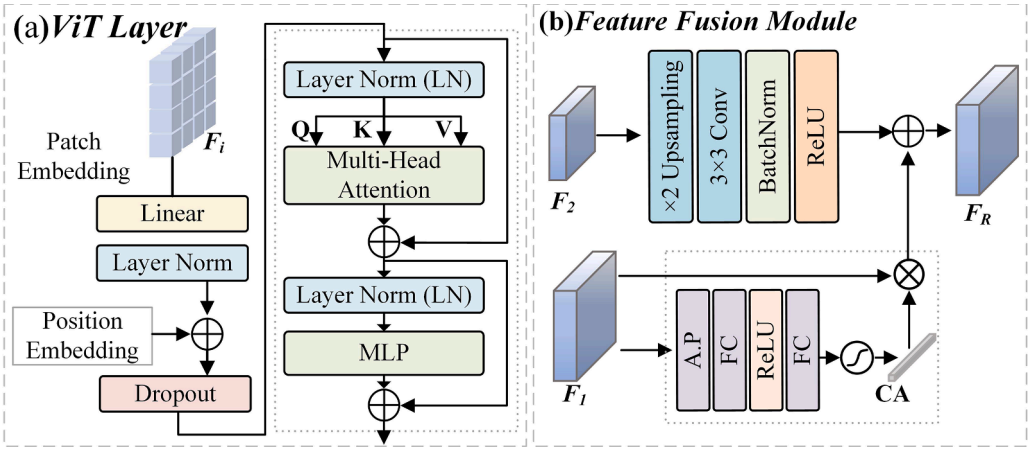

Fig. 4. Structure diagram of the ViT layer and feature fusion module of MBDTNet.

further utilized to generate the enhanced feature $\boldsymbol{F_T} \in \Re^{c3\times n\times n}$, with the goal of emphasizing the anomalous regions in the HSI while suppressing the influence of background regions, as described in Section 3.3. The features generated by the encoder are reshaped and then passed through a DNN network for multi-class background description learning (MBDL), which generates a token dictionary while computing the anomaly confidence score $\mathscr{S} \in \Re^{1\times n\times n}$, as detailed in Section 3.4. Finally, by concatenating the features $\boldsymbol{F_T}$ and $\mathscr{S}$, the decoder module generates an anomaly prediction map $\boldsymbol{M} \in \Re^{n\times n}$ with a 0–1 distribution. The decoder progressively reduces feature channels via convolutional layers, ultimately producing a single-channel output that is transformed into the final target probability map through sigmoid activation. Finally, in the testing phase, the test HSI is divided into several $n \times n$ patches, which are input into the MBDTNet network. Each patch generates an anomaly probability map, and these maps are then merged to obtain the anomaly detection map.

3.3. Token dictionary sparse representation

In LRR and dictionary learning, the representation coefficients $\boldsymbol{A}$ of the background dictionary represent how the original data are expressed as a weighted combination of dictionary elements, which helps improve the compactness and interpretability of the feature representation. In self-attention based methods, the attention weights between different elements are determined by calculating the normalized inner product. Specifically, the inner product between the query $\boldsymbol{Q}$ and the key $\boldsymbol{K}$ is computed to measure their similarity, and the attention weights $\boldsymbol{A}$ are then obtained by normalizing through Softmax:

$$\boldsymbol{A} = Softmax\left(\boldsymbol{Q}\boldsymbol{K}^T \Big/ \sqrt{d}\right) \tag{7}$$

The LRR and self-attention mechanisms both focus on effectively capturing the key features of data by optimizing the model parameters. These methods share a commonality in their approach to similarity calculation and weight allocation, as both aim to emphasize the important components by evaluating the relationships between elements within the data. Therefore, drawing from the concepts of LRR and token dictionary learning [56], we propose an attention mechanism based on TDSR, as shown in Fig. 5. Specifically, an additional token dictionary $\mathscr{D} \in \Re^{K\times d}$ is introduced and dynamically updated using multi-class background dictionary learning (Section 3.3), resulting in a compact set of background feature basis vectors. In other words, the token dictionary $\mathscr{D}$ serves as a parameterized memory collection that stores background features. The learned token dictionary $\mathscr{D}$ is then used to generate the key dictionary $\boldsymbol{K}_{\mathscr{D}}$ and the value dictionary $\boldsymbol{V}_{\mathscr{D}}$, with the input features $\mathscr{Z} \in \Re^{N\times d}$ being employed to generate query tokens:

$$Q_z = ZW^Q, K_z = \mathscr{D}W^K, V_{\mathscr{D}} = \mathscr{D}W^V \tag{8}$$

where W^Q, W^K, and W^V are the linear transformations. Furthermore, the attention map $\boldsymbol{A} \in \Re^{N\times K}$ is calculated using the cosine similarity between the query token $\boldsymbol{Q}_{\mathscr{Z}}$ and the key dictionary token $\boldsymbol{K}_{\mathscr{D}}$, as expressed by the following formula:

$$\boldsymbol{A} = Softmax(SIM_{cos}(\boldsymbol{Q}_{\mathscr{Z}}, \boldsymbol{K}_{\mathscr{D}}) / \tau) \tag{9}$$

where τ is a learnable parameter used to scale the similarity value. $SIM_{cos}(\boldsymbol{Q}_{\mathscr{Z}}, \boldsymbol{K}_{\mathscr{D}})$ denotes the cosine similarity between $\boldsymbol{Q}_{\mathscr{Z}}$ and $\boldsymbol{K}_{\mathscr{D}}$, and is calculated as follows:

$$SIM_{cos}(\boldsymbol{Q}_{\mathscr{Z}}, \boldsymbol{K}_{\mathscr{D}}) = \frac{\boldsymbol{Q}_{\mathscr{Z}} \cdot \boldsymbol{K}_{\mathscr{D}}}{\| \boldsymbol{Q}_{\mathscr{Z}} \| \, \| \boldsymbol{K}_{\mathscr{D}} \|} \tag{10}$$

where $\| \boldsymbol{Q}_{\mathscr{Z}} \|$ and $\| \boldsymbol{K}_{\mathscr{D}} \|$ represent the Euclidean norms of $\boldsymbol{Q}_{\mathscr{Z}}$ and $\boldsymbol{K}_{\mathscr{D}}$, respectively. Next, the Softmax function is applied to convert the similarity values into an attention map $\boldsymbol{A}$, which represents the strength of the association between each query token $\boldsymbol{Q}_{\mathscr{Z}}$ and the elements in the dictionary $\mathscr{D}$. Normal background samples can be precisely represented by a sparse linear combination of dictionary atoms. In contrast, anomaly samples which deviated from the background distribution require more dictionary atoms for reconstruction which leads to higher reconstruction errors. As shown in Fig. 4, a reconstruction error-guided attention (REGA) mechanism is proposed based on the reconstruction error, which improves the detection performance by integrating the reconstruction error into the calculation of attention weights. The mathematical formulation for the REGA mechanism is given as follows:

$$\mathscr{Z}^o = Softmax(\| \mathscr{Z} - \boldsymbol{A}\boldsymbol{V}_{\mathscr{D}} \|)\mathscr{Z} \tag{11}$$

where $\| \mathscr{Z} - \boldsymbol{A}\boldsymbol{V}_{\mathscr{D}} \|$ represents the sparse component, i.e., the anomaly component, which captures how well the feature $\mathscr{Z}$ can be reconstructed using the dictionary $\boldsymbol{V}_{\mathscr{D}}$ and attention weights $\boldsymbol{A}$.

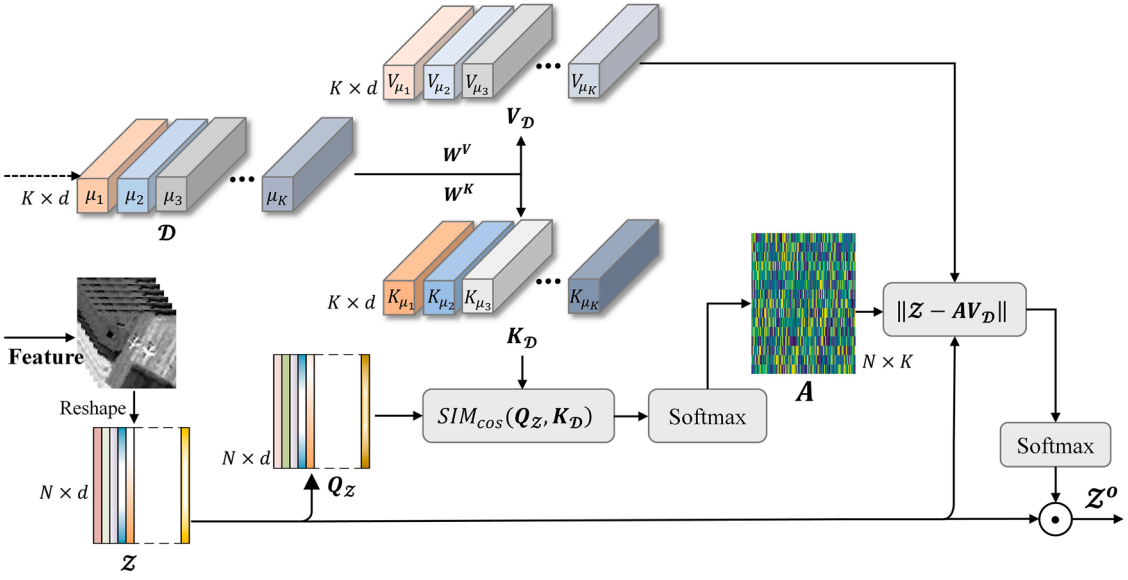

Fig. 5. Structure diagram of the reconstruction error-guided attention mechanism.

3.4. Multi-class background description learning

Based on the concept of Deep SVDD, the main objective of MBDTNet is to optimize the feature extractor so that the latent representations of the background samples from the same class are clustered together, forming an independent hypersphere with minimal volume. However, when dealing with multiple background classes in hyperspectral data, Deep SVDD faces challenges in defining clear decision boundaries. MCDD [53] determines each class by learning multiple Gaussian distributions through the network, which helps classify a test sample into that class. As a result, K hypersphere classifiers are integrated into the network using Deep-MCDD, where each hypersphere represents a distinct background class. The centers of these K hyperspheres collectively form a background dictionary, providing a stable reference for the TDSR module.

In the background distribution of HSI, different background classes typically exhibit distinct statistical characteristics. Modeling these background classes using a Gaussian distribution can help identify potential differences between them and anomalies. GDA assumes that each class follows a Gaussian distribution, making it particularly effective for handling the background distribution in HSI. Given that each class-conditional distribution follows a multivariate Gaussian distribution $P(\mathscr{Z}|y=k) = \mathscr{N}(f(x)|\mu_k,\Sigma_k)$, it is assumed that the class priors follow a Bernoulli distribution $P(y=k) = \beta_k/(\Sigma_{k'}\beta_{k'})$. Furthermore, it is assumed that the covariance and standard deviation of each class are isotropic (i.e., $\Sigma_k = \sigma_k^2 I$). Under these assumptions, the posterior probability of a sample x_t belonging to class k is described as:

$$\begin{aligned} P(y=k|x) &= \frac{P(y=k)P(x|y=k)}{\Sigma_{k'}P(y=k')P(x|y=k')} \\ &= \frac{\exp\left(-\frac{\|f(x;\mathscr{W})-\mu_k\|^2}{2\sigma_k^2} - \log\sigma_k^d + \log\beta_k\right)}{\Sigma_{k'}\exp\left(-\frac{\|f(x;\mathscr{W})-\mu_{k'}\|^2}{2\sigma_{k'}^2} - \log\sigma_{k'}^d + \log\beta_{k'}\right)} \end{aligned} \tag{12}$$

where $f(x;\mathscr{W})$ represents the feature representation of sample x.

However, during the learning process of deep network models, there is no guarantee that the class-conditional distributions will adhere to a Gaussian distribution, as no explicit constraint ensures the alignment with the Gaussian assumption or the true class means. To reinforce the GDA assumption, we minimize the reverse Kullback-Leibler (KL) divergence $KL(\mathscr{P}_k\|\mathscr{N}(\mu_k,\sigma_k^2 I))$ for each class, where the empirical class-conditional distribution $\mathscr{P}_k = \frac{1}{N_k}\sum_{y_i=k}\vartheta(x-f(x_i;\mathscr{W}))$ is constructed through deep feature averaging. The KL divergence is then formulated as:

$$\begin{aligned} KL(\mathscr{P}_k\|\mathscr{N}(\mu_k,\sigma_k^2 I)) &= -\int \frac{1}{N_k}\sum_{y_i=k}\vartheta(x-f(x_i;\mathscr{W}))\log\left[\frac{1}{(2\pi\sigma_k^2)^{\frac{d}{2}}}\exp\left(-\frac{\|x-\mu_k\|^2}{2\sigma_k^2}\right)\right]dx \\ &+ \int \frac{1}{N_k}\sum_{y_i=k}\vartheta(x-f(x_i;\mathscr{W}))\log\left[\frac{1}{N_k}\sum_{y_i=k}\vartheta(x-f(x_i;\mathscr{W}))\right]dx \\ &= -\frac{1}{N_k}\sum_{y_i=k}\log\left[\frac{1}{(2\pi\sigma_k^2)^{\frac{d}{2}}}\exp\left(-\frac{\|f(x_i;\mathscr{W})-\mu_k\|^2}{2\sigma_k^2}\right)\right] + \log\frac{1}{N_k} \\ &= \frac{1}{N_k}\sum_{y_i=k}\left(\frac{\|f(x_i;\mathscr{W})-\mu_k\|^2}{2\sigma_k^2} + \log\sigma_k^d\right) + c \end{aligned} \tag{13}$$

where the constant terms are merged and defined as c, and N_k is the number of training samples for class k. By minimizing the KL divergence across all the classes, the model enhances data structural representation, enforcing $P(\mathscr{Z}|y=k)$ in the latent space to conform to Gaussian priors. To facilitate model optimization and accurately represent the multi-class background spheres, the class-conditional probabilities are used as a confidence measure of how likely it is that a sample belongs to a particular class in the feature space. Therefore, the distance function d_k is expressed as follows:

$$\begin{aligned} d_k(\mathscr{Z}) &= -\log P(\mathscr{Z}|y=k) = -\log\mathscr{N}(f(x;\mathscr{W})|\mu_k,\sigma_k^2 I) \\ &\approx \frac{\|f(x_i;\mathscr{W})-\mu_k\|^2}{2\sigma_k^2} + \log\sigma_k^d \end{aligned} \tag{14}$$

The learning objective of MBDL is to align the priori distribution of the training data with the class-conditional distribution in the feature space, while ensuring that the classes are distinguishable from one another in a multi-class classification setting. To achieve this, the MBDL objective is shown in the following equation:

$$\min_{W,\mu,\sigma,b}\frac{1}{N}\sum_{i=1}^{N}\left[d_{y_i}(x_i) - \frac{1}{\nu}\log\frac{\exp(-d_{y_i}(x_i)+b_{y_i})}{\sum_{k=1}^{K}\exp(-d_k(x_i)+b_k)}\right] \tag{15}$$

where ν is the regularization factor. Therefore, the trainable parameters of MBDL module include the weights $\mathscr{W}$ of the MBDTNet encoder, the class mean μ_k, the standard deviation σ_k, and the bias b_k. In the MBDTNet training phase, the class centers need to be initialized first, i.e., $\mu_k = \frac{1}{N_k}\sum_{y_i=k}f(x;\mathscr{W})$. Finally, the class means μ_k can represent the multi-class background token dictionary $\mathscr{D}$ in the feature space.

Based on the established multiple background hyperspheres and the distance function d_k, the possibility and confidence of a sample belonging to each class can be calculated. Notably, if the test sample lies outside all of the background hyperspheres, it is likely to be classified as an anomaly. A distance function d_k is used to define the confidence score $\mathscr{S}(\mathscr{Z})$:

$$\mathscr{S}(\mathscr{Z}) = -\min_k d_k(\mathscr{Z}) \tag{16}$$

where a higher $\mathscr{S}(\mathscr{Z})$ indicates that the sample is closer to a background class, suggesting that it is more likely to be a normal sample. Conversely, a lower $\mathscr{S}(\mathscr{Z})$ (i.e., a larger distance) suggests that the sample may be anomalous. Therefore, the confidence score $\mathscr{S}(\mathscr{Z})$ helps distinguish between background and anomaly samples. As shown in Fig. 3, the confidence score is integrated with the other features and input into the anomaly detector.

3.5. MBDTNet loss function

1) *Classification Loss*: During the training sample construction process, the random masking strategy generates labels for both background and anomalies. Consequently, the probability of the anomaly is computed using the decoder module. To supervise the network training, the BCE loss L_{BCE} is introduced. The L_{BCE} is defined as follows:

$$L_{BCE} = -\frac{1}{hw}\sum_{i=1}^{hw}[y_i \log(p_i) + (1-y_i)\log(1-p_i)] \tag{17}$$

where y_i denotes the ground-truth label, p_i represents the predicted probability for the target, and hw is the number of samples.

2) *Inter-Class Loss*: The inter-class loss L_{inter} enhances feature separability by optimizing inter-class differences, guiding the model to learn clearer decision boundaries for the background class and thereby causing the features of anomalies to significantly deviate from the normal background distribution. The distance function d_k reflects the relationship between a sample and its class center, ensuring that the distinction between classes is preserved. Since the MBDL aims to maximize the posterior probability of the background class for each sample $P(l = l_i|\mathscr{Z})$, the distance function d_k computes the predicted label through maximum a posteriori estimation. The expression for this is as follows:

$$\widehat{l}(\mathscr{Z}) = \arg\max_k P(l=k|\mathscr{Z}) = \arg\max_k[-d_k(\mathscr{Z}) + b_k] \tag{18}$$

where b_k is the bias term associated with class k, providing flexibility to the decision process by allowing an adjustment parameter for the decision boundary of each class. Therefore, cross-entropy loss is utilized to quantify the disparity between the predicted label and the pseudo-label $\boldsymbol{L}$. The L_{inter} loss is as follows:

$$L_{inter} = -\sum_{k=1}^{K} l_k \log(p_k) \tag{19}$$

3) *Intra-Class Loss*: The intra-class loss L_{intra} constrains the distance between the features of background samples and their corresponding class centers, forcing the features of background samples from the same class to be more compactly distributed in the feature space. At the same time, L_{intra} allows for reasonable diversity in the background by utilizing multiple subclass centers, maintaining compactness within each subclass. To quantify the similarity between a sample and the background class centers, the spectral angle distance (SAD) is employed. The expression for L_{intra} is as follows:

$$L_{intra} = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{i=1}^{m}\arccos\left(\frac{\mathscr{Z}_i \cdot \mu_k}{\| \mathscr{Z}_i \| \| \mu_k \|}\right) \tag{20}$$

where m is the number of samples in each class k. $\mathscr{Z}_i$ is the feature vector of the i-th sample in class k. By optimizing the L_{intra} loss, the model enhances the cohesion of samples within each class, thereby improving the recognition of background classes, which in turn helps to boost the anomaly detection performance.

4) *Total Loss*: The total loss of MBDTNet can be expressed as follows:

$$L = L_{BCE} + L_{inter} + L_{intra} \tag{21}$$

By jointly optimizing these three losses, MBDTNet effectively integrates information from the anomaly target classification, inter-class loss, and intra-class loss. This comprehensive loss function enables MBDTNet to adapt more flexibly to the complex relationships between background and anomalies, thereby enhancing its ability to accurately identify anomalous samples.

4. Experimental results and analysis

4.1. Datasets

1) *San Diego Dataset*: This dataset is publicly available hyperspectral data, acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the San Diego airport area in California, USA. The San Diego dataset covers a wavelength range of 370–2510 nm and contains 189 bands. By cropping the original image, the scene area used in the experiment was 200 × 250 pixels, with a spatial resolution of 3.5 m. As illustrated in Fig. 6(a), three airplanes are considered the targets to be detected.
2) *Viareggio Dataset*: This dataset was acquired by the push broom-style hyperspectral Sistema Iperspettrale Modulare Galileo Avionica sensor in the suburbs of Viareggio, Italy. The spectral range is from 400 nm to 1000 nm and contains 511 bands. As shown in Fig. 5(b), the size of the dataset is 375 × 450 pixels, with a spatial resolution of 0.6 m. The scene contains three cars, four panels, and two reference calibration tarps.
3) *Avon Dataset*: This dataset was acquired the ProSpecTIR-VS sensor system, with the acquisition area located in the southern part of Avon, Rochester, New York, USA. After geographic and mosaicking preprocessing, a region of size 400 × 400 pixels was selected for this study, with a spatial resolution of 1 m, as shown in Fig. 6(c). The wavelength range spans from 400 nm to 2450 nm. The scene containing 25 grid-patterned tarps and three red or blue felt pads are considered anomalies.
4) *Qingpu Dataset:* This dataset was acquired over the Shanghai region of China using the Airborne Multi-Modular Imaging Spectrometer (AMMIS). The wavelength range is 400–1000 nm and contains 250 bands. The spatial resolution of the images is 0.75 m. As shown in

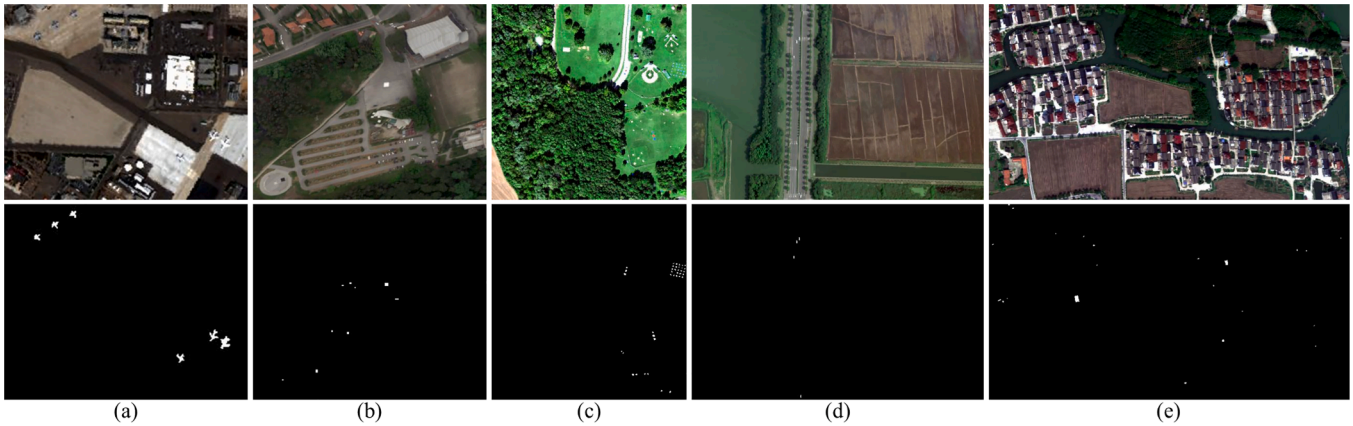

Fig. 6. The five HSI datasets are displayed as follows: the first row shows the pseudo-color images of the datasets, and the second row shows the corresponding ground-truth maps. (a) San Diego. (b) Viareggio. (c) Avon. (d) Qingpu-I. (e) Qingpu-II.

Fig. 6(d), after cropping the original strip data, the Qingpu-I dataset has a size of 600 × 400 pixels, with four vehicles considered as the anomalies. As shown in Fig. 6(e), the Qingpu-II dataset has a size of 400 × 740 pixels, with the blue rooftops considered the targets to be detected.

4.2. Experimental setup

1) Evaluation Metrics: The baseline was made up of nine popular anomaly algorithms: the global RX (GRX) detector [6], CRD [12], LSMAD [57], RGAE [31], DFAN [37], Auto-AD [39], DeepLR [50], PDBSNet [36] BockNet [38], and GT-HAD [46]. The GRX, CRD and LSMAD are traditional algorithms. RGAE, DFAN, Auto-AD, DeepLR, PDBSNet, BockNet, and GT-HAD are all deep learning models that utilize reconstruction error to detect anomalies.
2) Parameter Settings: The hardware device used in the experiments was a computer with an Intel Core i7-10700 K CPU, an NVIDIA GeForce RTX 3070 GPU. For the software environment, all the algorithms were computed in Python 3.9, implemented using Torch-GPU 2.0.1 and CUDA 11.3. For MBDTNet, we employed the Adam optimizer with a learning rate of 0.001. The MBDTNet took 64 × 64 patches as input and was optimized over 200 epochs with a batch size of 32. In this study, the threshold parameter ζ was set to 10‰ × N. After applying data augmentation strategy with rotation and flipping, the number of pure background training sample for the five datasets was expanded to 82, 84, 168, 93, and 157 samples, respectively. For the comparison algorithms, the optimal parameters were set for traditional algorithms. For RGAE, Auto-AD, DeepLR, DFAN, PDBSNet, and BockNet, the parameters were dynamically adjusted based on the specific characteristics of each dataset to achieve an optimal performance.
3) Evaluation Metrics: The performance of the target detection algorithms is evaluated here using 3D receiver operating characteristic (ROC) curves [58], the area under the curve (AUC) [58]. The 3D ROC curves serve as a fundamental evaluation metric, plotted based on the detection probability P_d, false alarm rate P_f, and threshold τ. By projection, this generates 2D ROC curves, namely, the ROC curve of (P_d, P_f), the ROC curve of (P_d, τ), and the ROC curve of (P_f, τ). Using the aforementioned three ROC curves, the following AUCs can be derived: $AUC_{(P_d,P_f)}$, $AUC_{(P_d,\tau)}$, and $AUC_{(P_f,\tau)}$. In addition, there are several derivative versions of AUC, namely, AUC_{TD}, AUC_{BS}, AUC_{ODP}, AUC_{TDBS}, and AUC_{SNPR}. AUC_{TD} quantifies the performance of the anomaly detection, while AUC_{BS} evaluates the background suppression performance. AUC_{ODP}, AUC_{TDBS}, and AUC_{SNPR} provide a comprehensive assessment of the algorithm's target detection capability.

4.3. Detection results

Fig. 7 presents the anomaly maps of the various methods on the San Diego dataset. It can be observed that DFAN, BockNet, and MBDTNet successfully detect all six airplane targets, whereas the other methods fail to identify the targets completely. However, LSMAD, RGAE, DFAN and GT-HAD incorrectly identify building rooftops as anomalies due to the complex background features. Notably, the proposed MBDTNet method not only identifies aircrafts but also accurately delineates anomaly boundaries. Fig. 8 illustrates the detection maps for the Viareggio dataset. GRX, CRD, and DFAN exhibit poor background suppression, making the target detection challenging. PDBSNet and BockNet

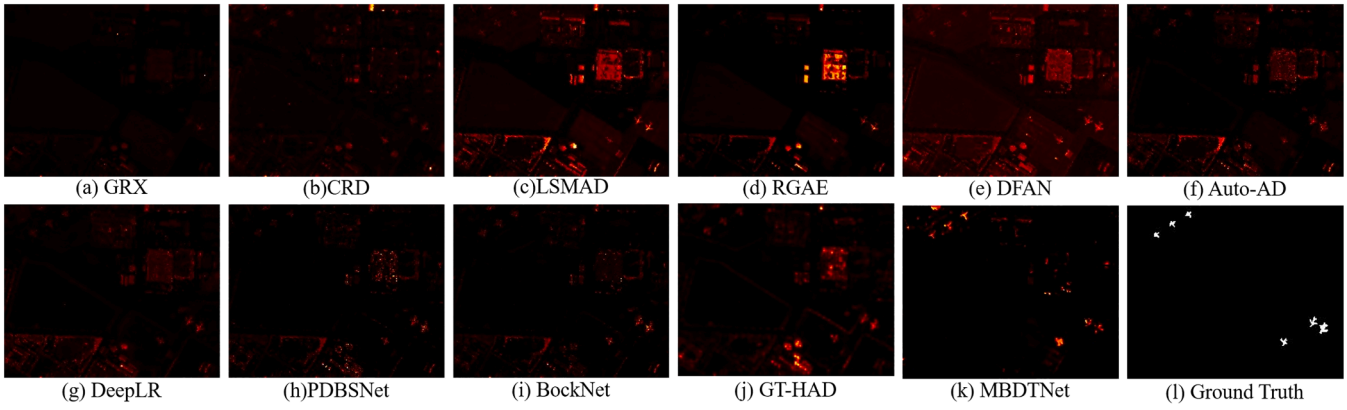

Fig. 7. Anomaly detection maps of the different methods on the San Diego dataset.

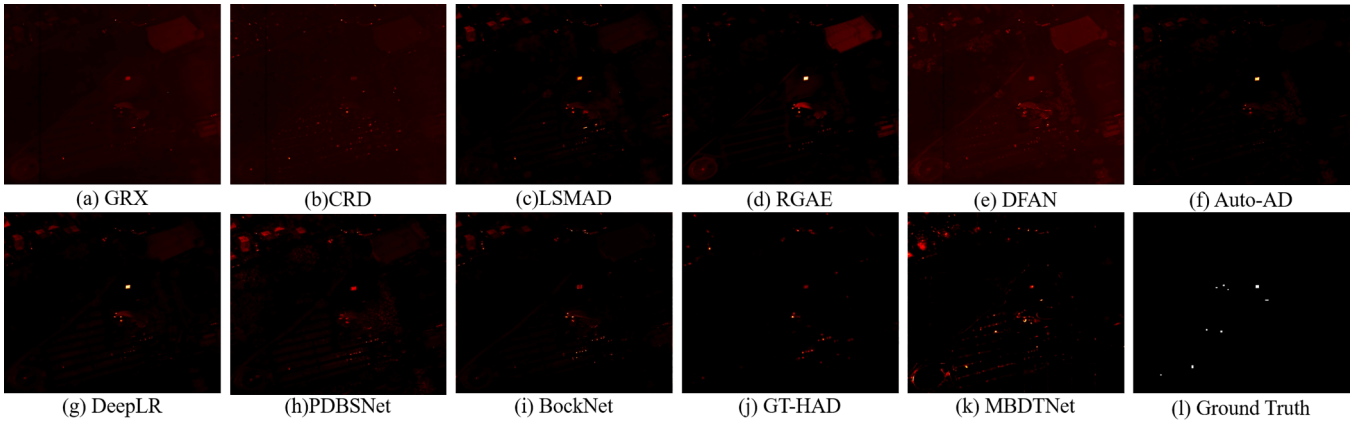

Fig. 8. Anomaly detection maps of the different methods on the Viareggio dataset.

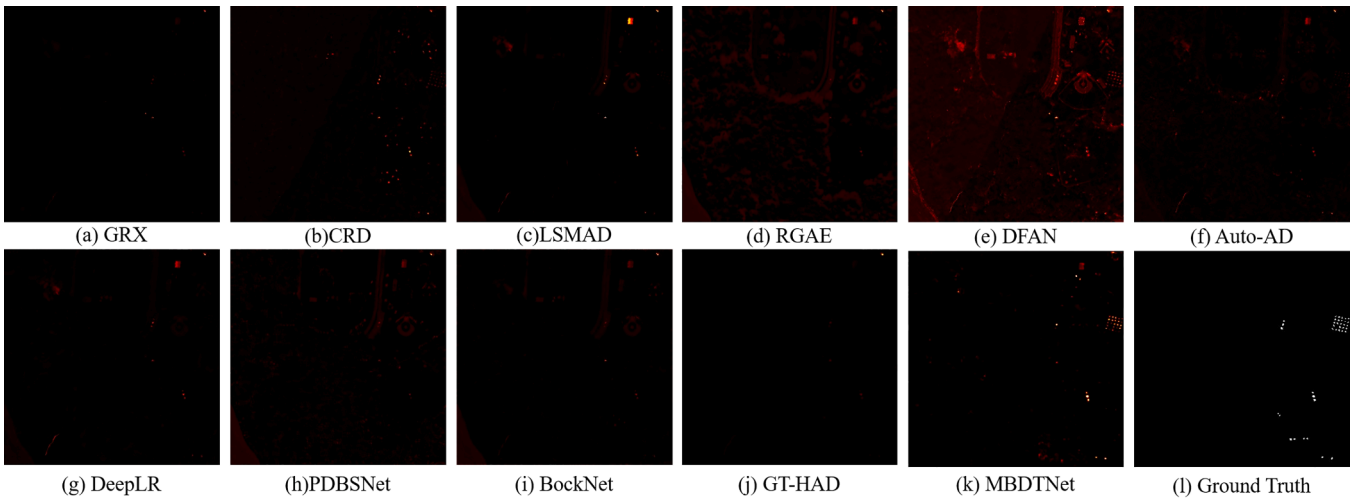

Fig. 9. Anomaly detection maps of the different methods on the Avon dataset.

misclassify a small portion of the vehicles in the center as anomalies. In contrast, LSMAD, GT-HAD, and MBDTNet obtain superior performances, and MBDTNet achieves the best target detection results. As shown in Fig. 9, GRX, Auto-AD, DeepLR, PDBSNet, BockNet, and GT-HAD fail to detect the anomalous targets and exhibit large false alarms. RGAE and DFAN display slightly weaker background suppression capabilities. Although the detection performance of DFAN and MBDTNet appears similar, MBDTNet significantly outperforms DFAN in background suppression. For the Qingpu-I dataset, the detection results are shown in Fig. 10. RGAE, DFAN, PDBSNet, and GT-HAD produce false alarms targets. However, BockNet and MBDTNet accurately identify the targets, with high precision. As shown in Fig. 11, RGAE, DFAN, and BockNet exhibit poor background suppression, whereas the traditional methods perform comparatively better on the Qingpu-II dataset. Among the deep learning based methods, MBDTNet accurately detects targets of varying scales. Overall, the proposed MBDTNet method not only successfully detects the targets of different scales but also maintains a high level of background suppression, demonstrating its robustness and effectiveness.

Fig. 12 shows the ROC curves for the different methods across the five datasets. The ROC curve of (P_d, P_f) being close to the top-left corner and the ROC curve of (P_d, τ) being close to the top-right corner indicates superior detection performances. Meanwhile, the ROC curve of (P_f, τ) appearing near the bottom-left corner suggests stronger background suppression capabilities. From the ROC curves of (P_d, P_f) on the five datasets, MBDTNet consistently appears near the top-left corner, demonstrating its strong detection capability. Similarly, in the ROC curves of (P_d, τ), MBDTNet is predominantly located in the top-right corner, significantly outperforming the other detection algorithms and maintaining a high detection probability even at high thresholds. For the

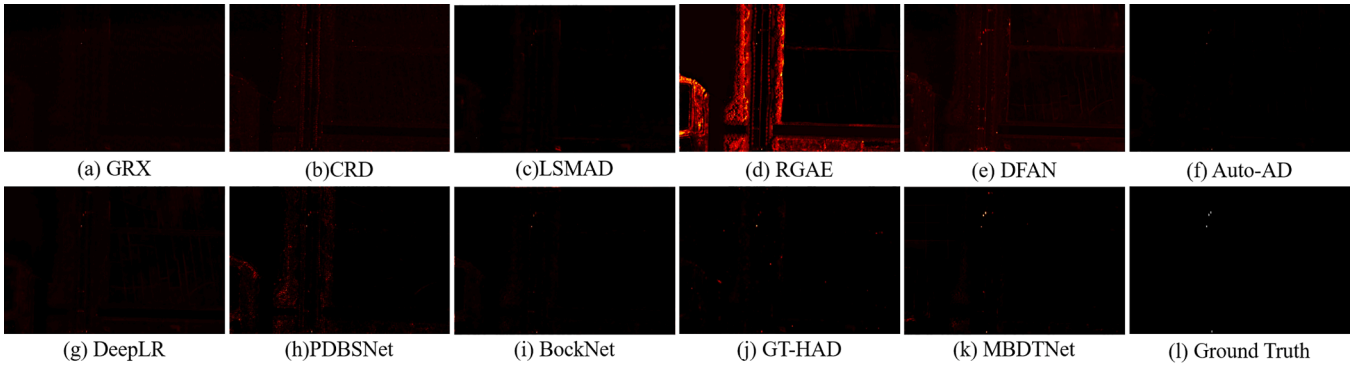

Fig. 10. Anomaly detection maps of the different methods on the Qingpu-I dataset.

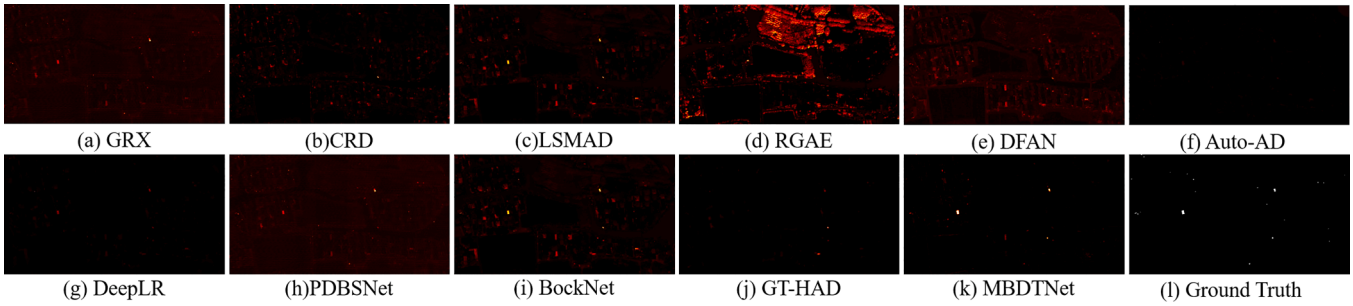

Fig. 11. Anomaly detection maps of the different methods on the Qingpu-II dataset.

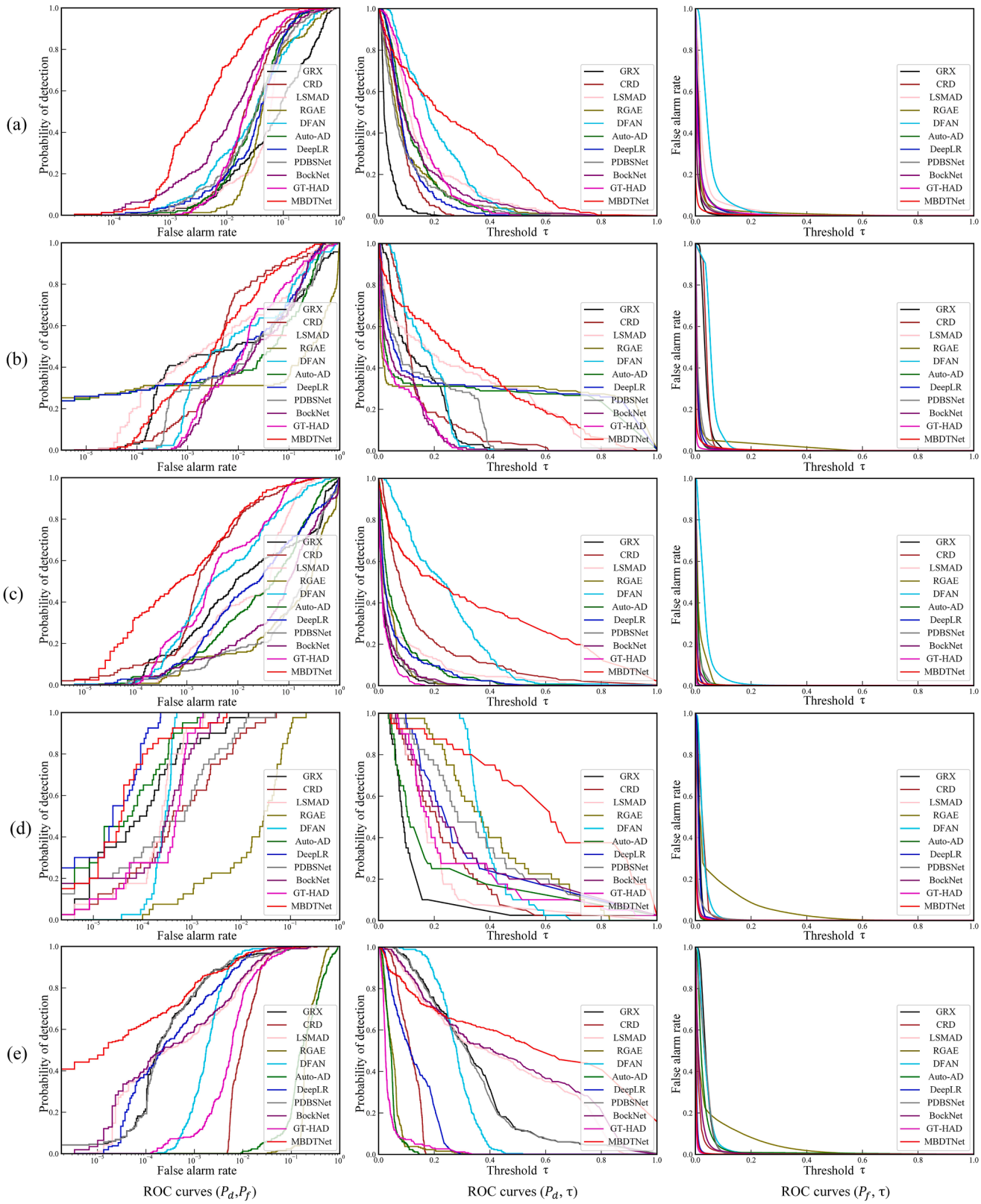

Fig. 12. ROC curves for each detection method on the four datasets. (a) San Diego. (b) Viareggio. (c)Avon. (d) Qingpu-I. (e) Qingpu-II.

ROC curves (P_f, τ) of the five datasets, GT-HAD and MBDTNet are optimally positioned in the bottom-left corner, indicating its effectiveness in suppressing background interference. In addition, the eight deep learning based anomaly detection algorithms exhibit similar performances for the ROC curve (P_f, τ), generally outperforming the traditional algorithms. However, the performance of Auto-AD, DeepLR, PDBSNet, and BockNet fluctuates across the five datasets, exhibiting inconsistency and suggesting a lack of robustness in adapting to diverse background complexities.

Table 1 present the AUC scores of each detection method on the four

Table 1
AUC scores of the different methods on the five datasets.

Dataset	AUC	GRX	CRD	LSMAD	RGAE	DFAN	Auto-AD	DeepLR	PDBSNet	BockNet	GT-HAD	MBDTNet
San Diego	$AUC_{(P_d,P_f)}\uparrow$	0.8349	0.9563	0.9223	0.9077	0.9244	0.9458	0.9438	0.9361	0.9549	0.9623	**0.9865**
	$AUC_{(P_d,\tau)}\uparrow$	0.0332	0.0855	0.1625	0.1082	0.2015	0.1222	0.0997	0.0997	0.1350	0.1496	**0.2792**
	$AUC_{(P_f,\tau)}\downarrow$	0.0124	0.0191	0.0341	0.0193	0.0529	0.0183	0.0177	0.0094	0.0091	0.0183	**0.0043**
	$AUC_{TD}\uparrow$	0.8681	1.0418	1.0847	1.0159	1.1259	1.0680	1.0435	1.0358	1.0900	1.1119	**1.2657**
	$AUC_{BS}\uparrow$	0.8224	0.9371	0.8881	0.8884	0.8714	0.9275	0.9261	0.9267	0.9459	0.9440	**0.9821**
	$AUC_{ODP}\uparrow$	0.8557	1.0226	1.0505	0.9966	1.0729	1.0497	1.0257	1.0264	1.0809	1.0936	**1.2614**
	$AUC_{TDBS}\uparrow$	0.0207	0.663	0.1282	0.0889	0.1485	0.1039	0.0819	0.0903	0.1259	0.1313	**0.2748**
	$AUC_{SNPR}\uparrow$	2.66	4.45	4.75	5.60	3.80	6.68	5.61	10.62	14.88	8.16	**64.22**
Viareggio	$AUC_{(P_d,P_f)}\uparrow$	0.6884	0.9414	0.9345	0.6202	0.9082	0.8846	0.9199	0.8870	0.9140	0.9302	**0.9716**
	$AUC_{(P_d,\tau)}\uparrow$	0.0446	0.0785	0.2774	0.2904	0.1671	0.2722	0.3038	0.1524	0.0797	0.0681	0.3152
	$AUC_{(P_f,\tau)}\downarrow$	0.0558	0.0356	0.0121	0.0235	0.0583	0.0073	0.0105	0.0138	0.0082	**0.0026**	0.0056
	$AUC_{TD}\uparrow$	0.7329	1.0199	1.2120	0.9105	1.0753	1.1568	1.2238	1.0394	0.9938	0.9982	**1.2869**
	$AUC_{BS}\uparrow$	0.6325	0.9058	0.9224	0.5966	0.8498	0.8773	0.9094	0.8732	0.9058	0.9276	**0.9659**
	$AUC_{ODP}\uparrow$	0.6771	0.9843	1.1998	0.8869	1.0170	1.1495	1.2133	1.0256	0.9856	0.9956	**1.2812**
	$AUC_{TDBS}\uparrow$	0.0112	0.0429	0.2652	0.2668	0.1088	0.2649	0.2933	0.1385	0.0715	0.0655	0.3096
	$AUC_{SNPR}\uparrow$	0.79	2.12	22.80	12.31	2.86	37.19	28.91	11.03	9.71	26.17	**55.76**
Avon	$AUC_{(P_d,P_f)}\uparrow$	0.8359	0.9849	0.9463	0.6790	0.9560	0.8754	0.9009	0.8439	0.7807	0.9781	**0.9867**
	$AUC_{(P_d,\tau)}\uparrow$	0.0276	0.1482	0.0787	0.0398	0.2355	0.0712	0.0633	0.0351	0.0336	0.0221	**0.3365**
	$AUC_{(P_f,\tau)}\downarrow$	0.0033	0.0095	0.0045	0.0154	0.0351	0.0098	0.0036	0.0056	0.0052	**0.0008**	0.0023
	$AUC_{TD}\uparrow$	0.8635	1.1331	1.0251	0.7188	1.1915	0.9466	0.9643	0.8791	0.8143	1.0002	**1.3232**
	$AUC_{BS}\uparrow$	0.8326	0.9753	0.9418	0.6636	0.9208	0.8656	0.8973	0.8383	0.7755	0.9772	**0.9843**
	$AUC_{ODP}\uparrow$	0.8603	1.1236	1.0206	0.7035	1.1563	0.9368	0.9607	0.8735	0.8092	0.9993	**1.3210**
	$AUC_{TDBS}\uparrow$	0.0243	0.1387	0.0742	0.0244	0.2003	0.0614	0.0597	0.0296	0.0284	0.0212	**0.3342**
	$AUC_{SNPR}\uparrow$	8.42	15.52	17.51	2.58	6.69	7.25	17.58	6.31	6.48	25.87	**146.43**
Qingpu-I	$AUC_{(P_d,P_f)}\uparrow$	0.9990	0.9956	0.9996	0.9597	0.9996	**0.9998**	**0.9998**	0.9968	0.9993	0.9949	0.9997
	$AUC_{(P_d,\tau)}\uparrow$	0.1276	0.2395	0.2058	0.4108	0.3908	0.2250	0.3322	0.3741	0.3249	0.2712	**0.6120**
	$AUC_{(P_f,\tau)}\downarrow$	0.0157	0.0262	0.0083	0.0562	0.0294	0.0043	0.0133	0.0085	0.0043	0.0044	**0.0031**
	$AUC_{TD}\uparrow$	1.1266	1.2352	1.2054	1.3705	1.3905	1.2248	1.3321	1.3710	1.3242	1.2707	**1.6117**
	$AUC_{BS}\uparrow$	0.9833	0.9694	0.9912	0.9034	0.9703	0.9954	0.9866	0.9883	0.9950	0.9952	**0.9965**
	$AUC_{ODP}\uparrow$	1.1109	1.2090	1.1971	1.3143	1.3611	1.2205	1.3188	1.3624	1.3199	1.2664	**1.6086**
	$AUC_{TDBS}\uparrow$	0.1118	0.2133	0.1974	0.3546	0.3614	0.2207	0.3188	0.3655	0.3205	0.2669	**0.6089**
	$AUC_{SNPR}\uparrow$	8.09	9.13	24.64	7.30	13.28	51.63	24.90	43.76	74.55	62.83	**197.11**
Qingpu-II	$AUC_{(P_d,P_f)}\uparrow$	0.9958	0.9854	0.9918	0.7523	0.9969	0.9803	0.9973	0.9825	0.9929	0.9891	**0.9974**
	$AUC_{(P_d,\tau)}\uparrow$	0.3383	0.1124	0.4308	0.0531	0.2802	0.0415	0.1252	0.0502	0.4554	0.0398	**0.5489**
	$AUC_{(P_f,\tau)}\downarrow$	0.0402	0.0150	0.0176	0.0490	0.0394	0.0042	0.0049	0.0042	0.0175	**0.0027**	0.0036
	$AUC_{TD}\uparrow$	1.3341	1.0978	1.4227	0.8053	1.2771	1.0218	1.1224	1.0327	1.4483	1.0288	**1.5462**
	$AUC_{BS}\uparrow$	0.9556	0.9704	0.9743	0.7032	0.9575	0.9762	0.9923	0.9783	0.9753	0.9863	**0.9938**
	$AUC_{ODP}\uparrow$	1.2939	1.0827	1.4051	0.7563	1.2377	1.0177	1.1175	1.0285	1.4308	1.0262	**1.6528**
	$AUC_{TDBS}\uparrow$	0.2981	0.0973	0.4132	0.0040	0.2407	0.0373	0.1202	0.0459	0.4378	0.0371	**0.5453**
	$AUC_{SNPR}\uparrow$	8.42	7.45	24.49	1.08	7.11	9.96	25.40	11.82	25.94	14.91	**153.77**

Table 2
Inference time and network parameter count of the different methods on the five datasets.

Dataset	San Diego		Viareggio		Avon		Qingpu-I		Qingpu-II	
Method	Time/s	Params(M)	Test/s	Params(M)	Test/s	Params(M)	Test/s	Params(M)	Test/s	Params(M)
GRX	3.97	-	18.09	-	14.77	-	18.97	-	20.16	-
CRD	58.96	-	337.39	-	1827.89	-	165.78	-	1537.75	-
LSMAD	33.13	-	340.53	-	478.84	-	266.51	-	297.43	-
RGAE	0.26	0.04	1.68	0.10	2.23	0.72	1.51	0.50	2.27	0.50
DFAN	4.54	0.08	20.10	0.51	16.52	0.26	22.54	0.13	26.78	0.13
Auto-AD	0.06	0.32	1.01	0.37	0.25	0.35	0.65	0.33	0.40	0.33
DeepLR	0.10	0.10	0.30	0.15	0.36	0.13	0.38	0.11	0.86	0.11
PDBSNet	0.19	0.68	1.51	0.72	0.90	0.70	1.04	0.69	1.64	0.69
BockNet	1.18	0.13	1.62	0.22	1.48	0.18	1.02	0.15	1.32	0.15
GT-HAD	0.59	0.61	0.41	0.43	0.47	0.43	0.65	0.31	0.80	0.31
MBDTNet	1.67	1.48	1.84	1.51	2.07	1.50	2.16	1.49	2.20	1.49

datasets. For the San Diego dataset, MBDTNet achieves the highest results across all the AUC measures, with the most representative $AUC_{(P_d,P_f)}$ and AUC_{ODP} improving by 0.0302 and 0.1805, respectively, compared to the second-best scores. The other comparison algorithms, such as CRD, DFAN, and BockNet, demonstrate competitive performances, ranking second across the different AUC metrics. For the Viareggio dataset, the proposed MBDTNet algorithm achieves $AUC_{(P_d,P_f)}$ and AUC_{ODP} scores of 0.9716 and 1.2812, respectively, along with the best

Table 3
$AUC_{(P_d,P_f)}$ and $AUC_{(P_f,\tau)}$ scores for the different modules on the four datasets.

TDSR	MBDL	San Diego		Viareggio		Avon		Qingpu-I		Qingpu-II	
×	×	0.9668	0.0075	0.9577	0.0055	0.9690	0.0032	0.9990	0.0063	0.9910	0.0084
×	√	0.9740	0.0077	0.9672	**0.0035**	0.9822	0.0039	**0.9998**	0.0036	0.9924	0.0058
√	×	0.9745	0.0046	0.9709	0.0059	0.9643	0.0023	0.9996	0.0047	0.9961	0.0049
√	√	**0.9865**	**0.0043**	**0.9716**	0.0056	**0.9867**	**0.0021**	0.9997	**0.0031**	**0.9974**	**0.0036**

Table 4
$AUC_{(P_d,P_f)}$ and $AUC_{(P_f,\tau)}$ scores for the different loss functions on the four datasets.

Methods	San Diego		Viareggio		Avon		Qingpu-I		Qingpu-II	
L_{BCE}	0.9799	0.0090	0.9692	0.0047	0.9842	0.0075	0.9995	0.0042	0.9962	0.0034
$L_{BCE} + L_{intra}$	**0.9868**	0.0053	0.9694	0.0050	0.9865	0.0035	0.9993	0.0037	0.9970	0.0038
$L_{BCE} + L_{inter}$	0.9848	0.0046	0.9707	**0.0034**	0.9859	0.0049	**0.9997**	0.0034	0.9916	**0.0031**
$L_{BCE} + L_{inter} + L_{intra}$	0.9865	**0.0043**	**0.9716**	0.0056	**0.9867**	**0.0023**	**0.9997**	**0.0031**	**0.9974**	0.0036

results across the other AUC metrics. On the Avon dataset, MBDTNet obtains the best detection performance with the lowest $AUC_{(P_f,\tau)}$. Although DFAN achieves higher $AUC_{(P_d,\tau)}$ and AUC_{TD}, it AUC_{BS} is the lowest, indicating weaker ability of background suppression. For the Qingpu-I dataset, despite its $AUC_{(P_d,P_f)}$ score not being the highest, the proposed method still achieves the best scores in AUC_{TD} and AUC_{BS}, reinforcing its strong overall performance. On the Qingpu-II dataset, MBDTNet outperforms all the comparison algorithms, achieving the highest AUC_{TD}, AUC_{BS}, and AUC_{ODP} scores across all the metrics. GT-HAD achieves the best performance on the $AUC_{(P_f,\tau)}$, but its overall detection performance is lower. In addition, it can be observed that, although Auto-AD secures the second-best average $AUC_{(P_d,\tau)}$ score, its anomaly detection performance remains suboptimal. Meanwhile, DeepLR and BockNet achieve near-optimal results across several datasets but exhibit a relatively poor performance on certain datasets, highlighting their limited adaptability and stability. In summary, MBDTNet demonstrates superior anomaly detection capabilities while effectively suppressing background interference. Its adaptability to complex background variations and targets of different scales underscores its robustness and reliability in diverse scenarios.

Moreover, Table 2 presents the inference time and number of parameters of different detection algorithms on five datasets. Among these traditional methods, CRD and LSMAD have relative long computation times, but the detection results show instability in complex background scenarios. The RGAE, Auto-AD, and DeepLR algorithms rely on pixel-level inference by vanilla autoencoders or convolutional neural networks, and thus the inference time is faster. By contract, MBDTNet adopts a pyramidal global spectral-spatial modeling strategy and combines a 64 × 64 sliding window mechanism for inference, which significantly increases the computational complexity. In terms of parameter count, MBDTNet is also higher than other networks, mainly due to the incorporation of the pyramid structure and the ViT module. Compared to the DFAN algorithm, MBDTNet maintains a faster inference speed with a larger number of parameters. Overall, MBDTNet achieves a significant improvement in detection performance through controlled computational and parameter overhead, making it particularly suitable for anomaly detection.

4.4. Ablation study

1) *Model Structure Analysis*: To validate the effectiveness of the TDSR and MBDL modules within the MBDTNet network, we evaluated the model performance with four different network structures. As shown in Table 3, the $AUC_{(P_d,P_f)}$ and $AUC_{(P_f,\tau)}$ scores are used to assess the impact of the different models. The base model for the ablation experiments was based on the encoding module and detection module. When only the TDSR module was used, the token dictionary update method followed the approach in [56]. It can be seen that the introduction of the TDSR module improves the $AUC_{(P_d,P_f)}$ score and reduces the $AUC_{(P_f,\tau)}$ score. Specifically, when using MBDL alone, the model achieves significant improvements on most datasets, particularly on the Qingpu-I dataset, indicating that the MBDL module helps enhance the model's ability to model the relationship between the background and anomalies. Furthermore, the token dictionary learned through the MBDL module, combined with the TDSR algorithm, the model's performance is significantly improved, achieving the best accuracy on the San Diego and Qingpu-I datasets. This demonstrates that the joint use of the TDSR and MBDL modules can effectively enhance the model's performance. On the Viareggio and Avon datasets, the improvement in accuracy is relatively smaller, but still outperforms the results obtained by using TDSR or MBDL alone, validating the complementary nature of these two approaches. In summary, the combination of the TDSR and MBDL modules enables MBDTNet to achieve the best performance, demonstrating its superiority in anomaly detection.

2) *Loss Function Analysis*: To validate the performance of the

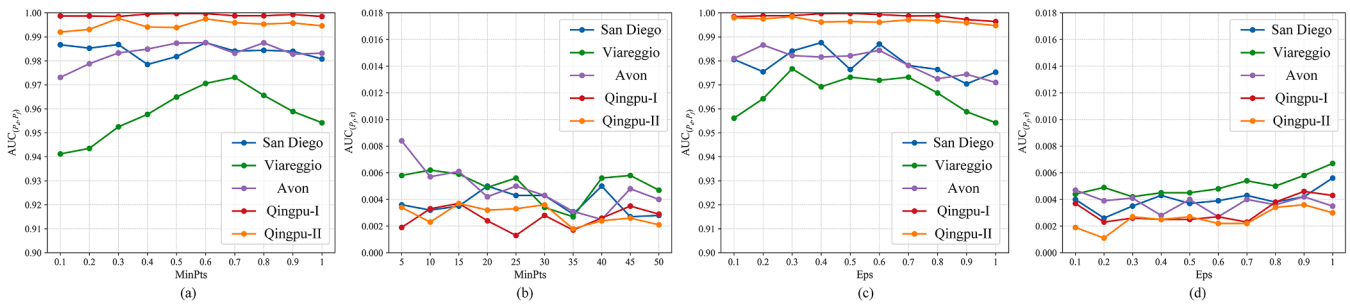

Fig. 13. Parameter analysis of DBSCAN. (a) $AUC_{(P_d,P_f)}$ of MinPts. (b) $AUC_{(P_f,\tau)}$ of MinPts. (c) $AUC_{(P_d,P_f)}$ of Eps. (d) $AUC_{(P_f,\tau)}$ of Eps.

MBDTNet composite guided loss function, ablation experiments were conducted using BCE loss as a baseline, with variations incorporating the L_{intra} loss and the L_{inter} loss. Table 4 lists the AUC scores on the four datasets under the different loss combinations. The results indicate that the combined effect of the L_{intra} and L_{inter} losses effectively suppresses background interference while enhancing the $AUC_{(P_d,P_f)}$ score. Specifically, the combination of L_{BCE} with both L_{intra} and L_{inter} achieves the highest accuracy on the San Diego, Qingpu-I, and Qingpu-II datasets, demonstrating that the dual loss functions help to better separate anomalies from background features. Whereas on the Viareggio and Avon datasets, the different loss combinations improve slightly, the accuracy remains largely stable across configurations. A detailed analysis of the different loss functions reveals that L_{intra} and L_{inter} optimize the MBDL model by emphasizing the inter-class separability and intra-class consistency.

3) *DBSCAN Parameter Analysis*: To validate the sensitivity of the DBSCAN parameters, additional two control experiments were performed in this study, as shown in Fig. 13. In the first experiment, the neighborhood radius *Eps* was fixed as 0.3 and the minimum number of neighbor points *MinPts* was varied from 5 to 50. In another experiment, *MinPts* was fixed as 40 and Eps was varied from 0.1 to 1. The first experiment showed that the $AUC_{(P_d,\ P_f)}$ generally improves as the number of *MinPts* increased, especially on the Qingpu-I and Qingpu-II datasets. Despite fluctuations in $AUC_{(P_f,\tau)}$, the performance remained low, indicating effective background suppression and anomaly detection. In the second experiment, although each dataset exhibits fluctuations under different *Eps* values, their performance remains relatively stable. The San Diego, Viareggio, and Avon datasets are slightly sensitive to both *MinPts* and *Eps*, whereas the Qingpu-I and Qingpu-II datasets shows lower sensitivity. In summary, DBSCAN parameters should be adjusted according to the characteristics of each dataset to achieve optimal detection performance.

5. Conclusion

In this article, we have proposed a multi-class background description transformer network (MBDTNet) for hyperspectral anomaly detection. Firstly, considering the limitations of the training data and the complexity of the background variations, pseudo-label generation and spatial random masking are employed to produce sufficient and diverse training samples. Next, a self-attention mechanism based on TDSR is introduced, improving the model's prioritize features of the different background categories. By integrating GDA with a deep network, MBDTNet leverages MBDL learning to learn the conditional distribution of each class and uses a distance function to infer the background token dictionary and localize anomalies. In addition, a composite guided loss function that combines BCE loss, intra-class loss, and inter-class loss is introduced to improve the stability of the model training. The experimental results demonstrated that MBDTNet achieved a superior performance and faster inference on five large-scale hyperspectral datasets.

CRediT authorship contribution statement

Zhiwei Wang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Kun Tan:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Xue Wang:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Wen Zhang:** Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by Yangtze River Delta Science and Technology Innovation Community Joint Research (Basic Research) Project (No. 2024CSJZN1300), Shanghai Municipal Education Commission Science and Technology Project(2024AI02002), National Natural Science Foundation of China (No. 42171335) and National Civil Aerospace Project of China (No. D040102).

Data availability

Data will be made available on request.

References

[1] L. Yang, F. Zhang, P.S.-P. Wang, X. Li, Z. Meng, Multi-scale spatial-spectral fusion based on multi-input fusion calculation and coordinate attention for hyperspectral image classification, Pattern Recognit. 122 (2022) 108348.
[2] C. Shi, Y. Liu, M. Zhao, C.-M. Pun, Q. Miao, Attack-invariant attention feature for adversarial defense in hyperspectral image classification, Pattern Recognit. 145 (2024) 109955.
[3] C. Li, B. Zhang, D. Hong, X. Jia, A. Plaza, J. Chanussot, Learning disentangled priors for hyperspectral anomaly detection: A coupling model-driven and data-driven paradigm, IEEE Trans. Neural Netw. Learn. Syst. (2024) 1–14.
[4] J. Zhang, Z. Yang, Y. Song, DC-AD: A divide-and-conquer method for few-shot anomaly detection, Pattern Recognit. 162 (2025) 111360.
[5] J.A. Malpica, J.G. Rejas, M.C. Alonso, A projection pursuit algorithm for anomaly detection in hyperspectral imagery, Pattern Recognit. 41 (11) (2008) 3313–3327.
[6] 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.
[7] P. Gurram, H. Kwon, Support-vector-based hyperspectral anomaly detection using optimized kernel parameters, IEEE Geosci. Remote Sens. Lett. 8 (6) (2011) 1060–1064.
[8] J.M. Molero, E.M. Garzon, I. Garcia, Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (2) (2013) 801–814.
[9] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, A. Plaza, Weighted-RXD and linear filter-based RXD: improving background statistics estimation for anomaly detection in hyperspectral imagery, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7 (6) (2014) 2351–2366.
[10] L.Zhang Du, Random-selection-based anomaly detector for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens. 49 (5) (2010) 1578–1589.
[11] K. Wu, G. Xu, Y. Zhang, B. Du, Hyperspectral image target detection via integrated background suppression with adaptive weight selection, Neurocomputing 315 (2018) 59–67.
[12] W. Li. Q. Du, Collaborative representation for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 53 (3) (2014) 1463–1474.
[13] Q. Ling, Y. Guo, Z. Lin, W. An, A constrained sparse representation model for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 57 (4) (2018) 2358–2371.
[14] Y. Niu, B. Wang, Hyperspectral anomaly detection based on low-rank representation and learned dictionary, Remote Sens. 8 (4) (2016) 289.
[15] W. Li, Q. Du, B. Zhang, Combined sparse and collaborative representation for hyperspectral target detection, Pattern Recognit. 48 (12) (2015) 3904–3916.
[16] W. Xie, T. Jiang, Y. Li, X. Jia, Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 57 (7) (2019) 4218–4230.
[17] S. Li, K. Zhang, Q. Hao, P. Duan, X. Kang, Hyperspectral anomaly detection with multiscale attribute and edge-preserving filters, IEEE Geosci. Remote Sens. Lett. 15 (10) (2018) 1605–1609.
[18] Y. Yang, H. Su, Z. Wu, Q. Du, Saliency-guided collaborative-competitive representation for hyperspectral anomaly detection, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16 (2023) 6843–6859.
[19] Y. Lu, X. Zheng, H. Xin, H. Tang, R. Wang, F. Nie, Ensemble and random collaborative representation-based anomaly detector for hyperspectral imagery, Signal Process. 204 (2023) 108835.
[20] C. Li, D. Zhu, C. Wu, B. Du, L. Zhang, Global overcomplete dictionary-based sparse and nonnegative collaborative representation for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–14.
[21] C. Zhang, H. Su, X. Lei, Z. Wu, Y. Yang, Z. Xue, Q. Du, Self-paced probabilistic collaborative representation for anomaly detection of hyperspectral images, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–10.
[22] X. Hu, Z. Li, L. Luo, H.R. Karimi, D. Zhang, Dictionary trained attention constrained low rank and sparse autoencoder for hyperspectral anomaly detection, Neural Netw. 181 (2025) 106797.
[23] X. Cheng, R. Mu, S. Lin, M. Zhang, H. Wang, Hyperspectral anomaly detection via low-rank representation with dual graph regularizations and adaptive dictionary, Remote Sens. 16 (11) (2024) 1837.
[24] M. Feng, Y. Zhu, Y. Yang, Q. Shu, Deep low-rank and piecewise-smooth constraint tensor model for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–14.

[25] H. Qin, Q. Shen, H. Zeng, Y. Chen, G. Lu, Generalized nonconvex low-rank tensor representation for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–12.
[26] C.-I. Chang, Effective anomaly space for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–24, s.
[27] L. Ren, L. Gao, M. Wang, X. Sun, J. Chanussot, HADGSM: A unified nonconvex framework for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15.
[28] X. Wang, F. Zhang, K. Zhang, W. Wang, X. Dun, J. Sun, Learning spatial-spectral dual adaptive graph embedding for multispectral and hyperspectral image fusion, Pattern Recognit. 151 (2024) 110365.
[29] S. Huang, Z. Liu, W. Jin, Y. Mu, Superpixel-based multi-scale multi-instance learning for hyperspectral image classification, Pattern Recognit. 149 (2024) 110257.
[30] Z. Li, F. Xiong, J. Lu, J. Wang, D. Chen, J. Zhou, Y. Qian, Multi-domain universal representation learning for hyperspectral object tracking, Pattern Recognit. 162 (2025) 111389.
[31] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, J. Ma, Hyperspectral anomaly detection with robust graph autoencoders, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–14.
[32] W. Xie, J. Lei, B. Liu, Y. Li, X. Jia, Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection, Neural Netw. 119 (2019) 222–234.
[33] Z. Wang, X. Wang, K. Tan, B. Han, J. Ding, Z. Liu, Hyperspectral anomaly detection based on variational background inference and generative adversarial network, Pattern Recognit 143 (2023) 109795.
[34] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, W. An, You only train once: learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–18.
[35] Z. Wang, D. Ma, G. Yue, B. Li, R. Cong, Z. Wu, Self-supervised hyperspectral anomaly detection based on finite spatial-wise attention, IEEE Trans. Geosci. Remote Sens. 62 (2023) 1–18.
[36] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, A.J. Plaza, PDBSNet: pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–14.
[37] X. Cheng, Y. Huo, S. Lin, Y. Dong, S. Zhao, M. Zhang, Deep feature aggregation network for hyperspectral anomaly detection, IEEE Trans. Instrum. Meas. 73 (2024) 5033016.
[38] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, A. Plaza, BockNet: blind-block reconstruction network with a guard window for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–16.
[39] S. Wang, X. Wang, L. Zhang, Y. Zhong, Auto-AD: autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–14.
[40] 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.
[41] T. Jiang, W. Xie, Y. Li, J. Lei, Q. Du, Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection, IEEE Trans. Neural Netw. Learn. Syst. 33 (11) (2021) 6504–6517.
[42] H. Qin, W. Xie, Y. Li, K. Jiang, J. Lei, Q. Du, Weakly supervised adversarial learning via latent space for hyperspectral target detection, Pattern Recognit 135 (2023) 109125.
[43] K. Jiang, W. Xie, Y. Li, J. Lei, G. He, Q. Du, Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 58 (7) (2020) 5224–5236.
[44] X. Chen, Y. Zhang, Y. Dong, B. Du, Generative self supervised learning with spectral spatial masking for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–13.
[45] YY. Li, T. Jiang, W. Xie, J. Lei, Q. Du, Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–11.
[46] J. Lian, L. Wang, H. Sun, H. Huang, GT-HAD: gated transformer for hyperspectral anomaly detection, IEEE Trans. Neural Netw. Learn. Syst. 36 (2) (2024) 1–15.
[47] J. Li, X. Wang, S. Wang, H. Zhao, Y. Zhong, One step detection paradigm for hyperspectral anomaly detection via spectral deviation relationship learning, IEEE Trans. Geosci. Remote Sens. 62 (2024) 1–15.
[48] Y. Li, K. Jiang, W. Xie, J. Lei, X. Zhang, Q. Du, A model-driven deep mixture network for robust hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–16.
[49] C. Li, B. Zhang, D. Hong, J. Yao, J. Chanussot, LRR-net: an interpretable deep unfolding network for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–12.
[50] S. Wang, X. Wang, L. Zhang, Y. Zhong, Deep low-rank prior for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–17.
[51] Q. Shen, Z. Liu, H. Wang, Y. Xu, Y. Chen, Y. Liang, D^3T: deep denoising dictionary tensor for hyperspectral anomaly detection, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 18 (2024) 3713–3727.
[52] F. Sohrab, J. Raitoharju, A. Iosifidis, M. Gabbouj, Multimodal subspace support vector data description, Pattern Recognit. 110 (2021) 107648.
[53] D. Lee, S. Yu, H. Yu, Multi-class data description for out-of-distribution detection, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 1362–1370.
[54] A. Diko, D. Avola, M. Cascio, L. Cinque, ReViT: enhancing vision transformers feature diversity with attention residual connections, Pattern Recognit. 156 (2024) 110853.
[55] H.A. Amirkolaee, M. Shi, M. Mulligan, TreeFormer: a semi-supervised transformer-based framework for tree counting from a single high resolution image, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–15.
[56] L. Zhang, Y. Li, X. Zhou, X. Zhao, S. Gu, Transcending the limit of local window: advanced super-resolution transformer with adaptive token dictionary, in: Proc. IEEE Conf. Comput. Vis. Pattern Recog., CVPR, 2024, pp. 2856–2865.
[57] Y. Zhang, B. Du, L. Zhang, S. Wang, A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 54 (3) (2015) 1376–1389.
[58] C.-I. Chang, An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis, IEEE Trans. Geosci. Remote Sens. 59 (6) (2020) 5131–5153.