

MMR-HAD: Multiscale Mamba Reconstruction Network for Hyperspectral Anomaly Detection

Xiyu Fu[✉], *Member, IEEE*, Ting Zhang, Juan Cheng, and Sen Jia[✉], *Senior Member, IEEE*

Abstract—Hyperspectral anomaly detection (HAD) aims to identify anomalous objects from hyperspectral images (HSIs) whose spectral features significantly deviate from their surroundings. Existing HAD methods still reconstruct some anomalies during the background reconstruction process, which can seriously affect the detection accuracy. Consequently, inspired by Mamba’s ability to effectively model long sequences, we propose a multiscale Mamba reconstruction network for HAD (MMR-HAD) by enhancing the representation of the background and inhibit anomalies from being reconstructed. MMR-HAD first removes most of the anomalous pixels in HSIs using the random mask (RM) strategy, which reduces the interference of anomalous pixels on the background reconstruction and makes the background features more prominent. To further filter out the small amount of residual anomalous pixels, we propose the multiscale dilated attention background enhancement (MDABE) mechanism, which enhances the background representation. Finally, we apply the multiscale dynamic feature fusion (MDFF) strategy to reconstruct the background image, further extracting and strengthening the background information, thus obtaining a pure background image. MMR-HAD, which focuses on generating a pure background, has been experimentally validated on seven real hyperspectral datasets. The results demonstrate that it excels in background enhancement and anomaly suppression, significantly improving detection accuracy. Introducing Mamba offers a promising solution for HAD, with substantial potential for practical application.

Index Terms—Deep learning (DL), hyperspectral anomaly detection (HAD), hyperspectral image (HSI), image reconstruction, Mamba, random mask (RM).

Received 8 September 2024; revised 7 January 2025; accepted 5 February 2025. Date of publication 13 February 2025; date of current version 25 June 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 42301375 and Grant 62271327; in part by the Project of Department of Education of Guangdong Province under Grant 2023KCXTD029; in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2022A1515110076 and Grant 2022A1515011290; in part by Shenzhen Science and Technology Program under Grant RCJC20221008092731042, Grant JCYJ20220818100206015, Grant JCYJ20240813141635047, Grant KQTD20200909113951005, and Grant KCXFZ20240903093100001; in part by the Research Team Cultivation Program of Shenzhen University under Grant 2023JCT002; and in part by the Open Research Fund of Shaanxi Key Laboratory of Optical Remote Sensing and Intelligent Information Processing under Grant KF20230303. (Corresponding author: Sen Jia.)

Xiyu Fu is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China, and also with Shaanxi Key Laboratory of Optical Remote Sensing and Intelligent Information Processing, Xi’an 710119, China (e-mail: fuxy0623@szu.edu.cn).

Ting Zhang and Sen Jia are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: 2300271069@email.szu.edu.cn; senjia@szu.edu.cn).

Juan Cheng is with Shaanxi Key Laboratory of Optical Remote Sensing and Intelligent Information Processing, Xi’an 710119, China (e-mail: chengjuan@opt.ac.cn).

Digital Object Identifier 10.1109/TGRS.2025.3541353

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) are 3-D cubes composed of hundreds of spectral images, where each pixel contains rich spatial and spectral information [1], [2], [3], which can be used to distinguish objects on the ground. Therefore, HSIs have been shown to have unique advantages in many applications, such as land-cover classification [4], [5], [6], dynamic monitoring of the environment [7], [8], and anomaly detection [9], [10].

Hyperspectral anomaly detection (HAD) is an unsupervised binary classification problem for identifying anomalous objects from HSIs whose spectral features significantly deviate from their surroundings [9]. Anomalies are objects that occupy a tiny region in the image and have spectral features that are different from their surroundings, and these anomalies are usually irregular and have a low probability of occurrence. Up to now, reconstructing the background first and then separating the anomalies is a common approach used to recognize anomalies. Existing hyperspectral anomaly detectors can be mainly categorized into three main groups: statistical-based models, representation-based models, and DL-based models.

A. Statistical-Based Models

The Reed–Xiaoli (RX) [11] detector is the most classical statistically based HAD method, which assumes that the background data conforms to a multivariate Gaussian distribution, and detects anomalous targets by calculating the Mahalanobis distance of each pixel from the background. Based on this, a series of variants are derived, such as global RX (GRX) [12], which uses the data of the whole image to model the background, thus obtaining the overall statistical information, and is suitable for homogeneous scenes. Localized RX (LRX) [12] models the background by using the data of the local region around the test pixel only and can better handle complex and nonuniform scenes and reduce the interference of anomalies on the background estimation. However, these RX methods have certain drawbacks: first, a single normal distribution model cannot adequately describe the complex background in real hyperspectral scenes. Second, the background statistics are easily affected by noisy pixels and anomalous pixels, affecting the detection results. Many improved methods have been proposed to address the above shortcomings. The kernel RX detector [13] effectively handles the problem of spectral indivisibility in linear space by mapping the data to a high-dimensional feature space and kernelizing the RX model.

Weighted RX (W-RXD) [14] improves the estimation of background information by employing a method that reduces the weight of anomalous pixels and assigns a higher weight to background pixels. Clustering-based RX detector (CBAD) [15] identifies anomalies by clustering the distribution of pixel values in HSIs and recognizing anomalies in each cluster. In [16], an algorithm called subspace RX (SSRX) was proposed to reduce the interference of anomalous pixels in background modeling by eliminating background dimensions with high variance.

B. Representation-Based Models

Representation-based models usually achieve anomaly detection by constructing an efficient representation of hyperspectral data. Chen et al. [17] detected anomalies by representing pixels in HSIs as sparse linear combinations of training samples. However, the limitations of this method are the high computational complexity required to construct the dictionary and the lack of robustness to noise, so robust principal component analysis (RPCA) [18] was proposed, which is a typical method based on low-rank sparse matrix decomposition, and enhances the robustness to noise by decomposing the hyperspectral data into a low-rank matrix and a sparse matrix. Similarly, Guo et al. [19] introduced background endmember representation and proposed a novel HAD method termed learnable background endmember with subspace representation (LEBSR). The collaborative representation-based detector (CRD) proposed by Li and Du [20] uses the strategy of representing background pixels as approximate representations of their spatial neighbors. In contrast, anomalies cannot be represented by adjacent pixels. This method has a high dependence on the distance-weighted regularization matrix and parameter adjustment. In [21], the problem of abnormal pixel contamination in hyperspectral data was solved by introducing spatial principal component analysis (PCA) to extract main pixel information, combined with the CRD method. Xu et al. [22] proposed a low-rank and sparse representation (LRASR), which separates background and anomalies by representing background pixels as low-rank matrices and adding sparse regularization terms to the representation coefficients. Guo et al. [23] proposed a method that utilizes structural incoherence constraints, first-order statistics, and background decentralization to enhance anomaly separation, while a mixed noise model and hierarchical optimization improve robustness and accuracy. Considering the special characteristics of HSIs, which contain not only spectral information but also rich spatial information, many methods based on tensor decomposition have been designed. Li et al. [24] proposed a tensor-based detector (PTA) by decomposing the HSI into a low-rank background tensor and a sparse anomaly tensor. Representation-based models often require parameter settings, which hinders the performance and applications in practical scenes.

C. DL-Based Models

In recent years, DL has shown great potential in the field of computer vision and has been successfully applied to HAD.

In the existing research on HAD, the commonly used algorithms are convolutional neural network (CNN), autoencoder (AE), generative adversarial network (GAN), and Transformer.

CNN-based anomaly detectors utilize multilayered convolution to capture deep features in the HSI. Li et al. [25] proposed a transfer-based deep CNN, where the multilayered CNN is responsible for capturing the differences between the center pixel and its surrounding pixels and classifying each pixel based on a similarity measure, but the disadvantage is that it requires a large number of data samples. Fu et al. [9] proposed a CNN-based HAD method using the plug-and-play technique. Zhang et al. [26] proposed a lightweight CNN method based on residual learning and background estimation, focusing on directly learning anomalous features to simplify HAD. In [27] and [28], a novel strategy based on CNN is used to realize the task of background reconstruction, and finally the anomalies are separated to realize the detection of anomalous targets, and its unique strategy provides a new research idea for HAD.

AE-based networks tend to identify anomalies through the large amount of residuals generated during the background reconstruction process. Lei et al. [29] proposed a HAD method (SLDR) based on spectral learning, which uses AE to reconstruct the spectrum and detect anomalies through spectral error map (SEM) and spectral angular distance (SAD). They also apply some constraints on the encoder to produce more accurate background reconstructions. Chang et al. [30] proposed a hypersection anomaly detection algorithm based on sparse AE, which is trained by using pixels in the entire window and the outer window, respectively, and uses the difference in reconstruction errors to identify anomalies. Fan et al. [31] proposed a robust anomaly detector (RGAE) based on AE by introducing the graph regularization term of $\ell_{2,1}$ -norm and superpixel segmentation to enhance the robustness of AE in HSI anomaly detection. To effectively fuse the spectral and spatial information, Lu et al. [32] employed manifold learning to extract the local spatial structure, then combined it with an AE model. However, when processing complex scenes, AE will also reconstruct some anomalies, resulting in higher reconstruction errors.

Considering the limited performance of AE in complex scenes, GANs have subsequently emerged. GANs utilize adversarial training of generators and discriminators to reconstruct the background image of the original HSI [33] and others generated images as close to the background as possible by training the GAN and employing detectors to detect anomalies. Unsupervised hyperspectral anomaly detectors have also been proposed. Jiang et al. [34] proposed an unsupervised discriminative reconstruction HAD method based on GAN and AE. The background reconstruction is learned by the AE network and the anomaly samples are highlighted using the GAN. Li et al. [35] proposed weakly supervised detectors by incorporating a regularization network in the GAN and using the GAN to learn discriminative potential reconstruction, which improves the robustness to anomalies and noise. Jiang et al. [36] proposed a weakly supervised discriminative learning method for HAD using a spectrally constrained GAN, which focuses mainly on learning a discriminative end-to-end

reconstruction network. Luo et al. [37] introduced an adversarial sample generation-based multiscale Siamese network that generates high-quality training samples through adversarial learning and utilizes background suppression loss functions to enhance target-background discrimination.

Although GANs perform well in generating data, due to the instability of training in GANs, Transformer-based methods appear and provide new research ideas in the field of HAD. The unique self-attention mechanism in Transformer can better capture long-distance dependencies in hyperspectral data [38]. Li et al. [39] combined Swin Transformer with UNet to reconstruct the background HSI for anomaly detection. Lian et al. [40] proposed a new gating network based on Transformer, which uses two specialized branches to enhance background reconstruction and suppress abnormal reconstruction, respectively. They also introduce an adaptive gating unit to adjust the working status of the branches. Xiao et al. [41] extracted features from global and local features through a dual-window mask converter to fully utilize the spatial-spectral characteristics of HSIs. An adaptive weighted loss function is also proposed to accurately suppress the reconstruction of abnormal targets. Wu and Wang [42] addressed the problem of insufficient representation of pixel spatial correlation in HSIs by introducing a spectral Transformer network. He et al. [43] proposed a transformer network that gradually captures spatial-spectral features by using global and local transformer blocks layer by layer, while introducing hidden diagonal masks to suppress the problem of excessive abnormal reconstruction, further improving detection accuracy. Wu and Wang [44] proposed a Transformer-based autoencoder framework that successfully solves the problems of nonlinear mixing and insufficient spatial information in HAD by combining local and global spatial information. The successful application of Transformer to HAD further motivates us to explore more powerful modeling methods of long-distance dependency for the background reconstruction in HAD task.

D. Motivation and Contribution

In summary, methods for HAD still suffer from problems such as a limited number of samples and the possibility of reconstructing some anomalies during the background reconstruction process, which seriously affect the detection accuracy. To address these problems, in this article, we propose a self-supervised multiscale Mamba reconstruction network for HAD (MMR-HAD) to enhance the representation of the background and suppress the reconstruction of anomalies, thus gaining good anomaly detection performance. Our method is divided into three main steps: the first step removes most of the anomalous pixels in the HSI by random mask (RM) strategy. The second step further filters out a small portion of the anomalies by the multiscale dilated attention background enhancement (MDABE) mechanism, which in turn strengthens the background representation. Finally, we enhance the background representation by the multiscale dynamic feature fusion (MDFF) strategy to reconstruct the background image and finally obtain the pure background. According to our investigation, the proposed MMR-HAD is the first to introduce

Mamba into the field of HAD. The main contributions of our proposed MMR-HAD are summarized as follows.

- 1) We propose a random mask strategy that aims to effectively exclude anomalous interference in HSIs by replacing the pixels with the mean value of randomly selected neighboring pixels. The strategy flexibly adapts to anomalous targets with various morphology and distribution, avoiding the problem that a fixed mask cannot cover all anomalous pixels. Reconstruction using these random neighborhood pixels can minimize the influence of anomalous pixels on the background reconstruction, and ultimately obtain a purer background image.
- 2) A multiscale dilated attention background enhancement module is designed to enhance the reconstruction of the background of HSIs by synthesizing feature information at different scales through the combination of multiscale dilated convolution and attention mechanisms. MSABE employs convolutional branches with varying dilation rates to capture fine features and global context. It leverages a multihead attention mechanism to highlight critical background regions, effectively mitigating the influence of anomalous pixels.
- 3) We constructed a multiscale Mamba reconstruction network, in which an MDFF strategy is proposed to fully utilize the flexible and efficient feature expression capabilities of the Mamba network. By dynamically selecting and fusing features at different scales, MDFF organically combines global and local information and enhances the expression ability of background information, thus generating a pure background image without the interference of anomalous targets.

The rest of this article is organized as follows. Section II describes the details of the model of MMR-HAD. Section III presents a large number of experimental results and analyses, and Section IV concludes the full paper.

II. PROPOSED METHOD

The general flow of our proposed MMR-HAD is shown in Fig. 1. First, the original HSI is subjected to pixel-shuffle downsampling operation to reduce the spatial correlation of neighboring pixels in the HSI [28], and then the RM strategy is applied to the downsampling image. Next, the HSI after RM is fed into MMR-HAD for the reconstruction task of the background. Specifically, given $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ an HSI after performing RM, where H , W , and C denote the height, width, and the number of spectral channels of the HSI, respectively. $\mathbf{X}$ first undergoes shallow feature extraction by the shallow feature extraction module, which is composed of three 3×3 convolutions. Then, the presence of anomalies is further weakened by the proposed MDABE module, achieving the enhancement of the background. After that, multiscale background information is extracted by Mamba-in-Mamba (MiM) hierarchical encoder [45] and fused using the proposed MDGG strategy. Finally, the background is constructed by the image reconstruction module. After the processing of the MMR-HAD, pixel-shuffle upsampling is applied to the reconstructed background to obtain the HSI

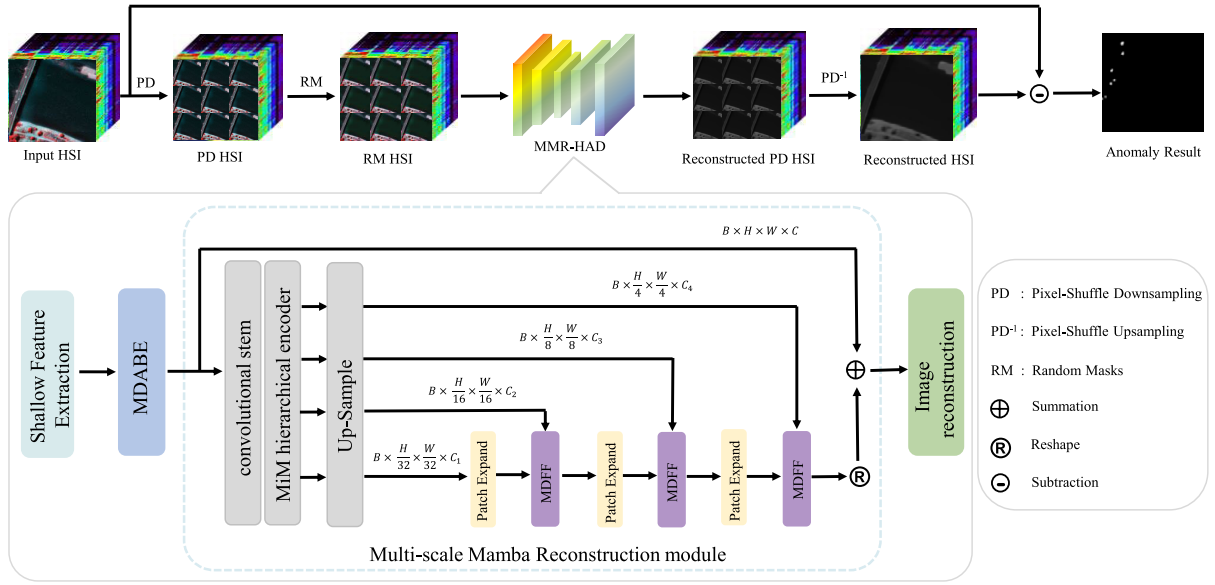


Fig. 1. Flowchart of the proposed MMR-HAD for HAD.

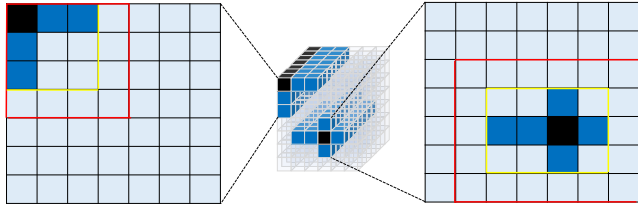


Fig. 2. Diagram of RM. The left side represents a random pixel value of 4, and the right side represents a random pixel value of 5.

with pure backgrounds. In the anomaly detection stage, the reconstruction errors of the original HSI and the reconstructed image with pure background are calculated to get the final anomaly detection result.

A. Random Mask

In order to prevent the interference of anomalies during the background reconstruction process, we propose an RM strategy, as shown in Fig. 2. We aim to remove as many anomalous pixels as possible from the HSI. The strategy traverses each pixel in the HSI in turn (the black pixels in Fig. 2) and generates a set of random neighboring pixels (blue pixels) around each pixel point. The neighboring pixels are randomly generated in four directions (top, right, bottom, and left). Then, the pixels in the yellow rectangle enclosing these neighboring pixels are considered to be anomalous pixels or highly likely to be anomalous pixels, while the pixels between the yellow rectangle and the red rectangle serve as the neighboring background pixels. Finally, the value of the black pixel is replaced by the average of the neighboring background pixels. As a result, we reduce the influence of anomalous pixels by removing the vast majority of anomalies and reconstructing the values of the black pixels from the surrounding background pixels. In addition, the RM are set up to cope with the irregularities in the shape and size of the anomalous targets in HSIs. If a fixed mask is used, the pixels

of anomalous targets may not be completely covered and the anomalous interference cannot be effectively removed. The advantage of the RM strategy is that it can flexibly cope with a variety of complex anomalous target morphologies in HSIs, and by randomly selecting neighboring pixels, it effectively reduces the influence of the anomalous pixels on the background reconstruction, resulting in a purer background image for the final reconstruction.

B. Multiscale Dilated Attention Background Enhancement

The HSI obtained after RM may still retain a small portion of anomalies, we need to suppress the expression of anomalies as much as possible to obtain a purer background. To achieve this goal, we propose a multiscale dilated attention background enhancement module, as shown in Fig. 3. Since HSIs are different from ordinary RGB images in that each pixel contains a large amount of spectral information and the background region has complex and extensive contextual information, how to effectively utilize the information contained in the background is crucial for background reconstruction. To fully utilize the background information in HSIs, in MSABE, we use three dilation convolutions with different dilation rates to increase the extent of the receptive field. Specifically, the MSABE module consists of three parallel branches, each of which possesses a 1×1 convolution with shared weights and a 3×3 convolution with different dilation rates. The dilation rates of the three branches are set to 1, 2, and 3, respectively. The multiscale dilated attention will facilitate the acquisition of subtle features and global information in the reconstructed background. The dilated attention will also help filter out some abnormal pixels in the features. Mathematically, assuming that the input feature map is $\mathbf{M}$, the output of each branch is denoted as

$$\mathbf{M}_i = \text{Conv}_{3 \times 3, d=i}(\text{Conv}_{1 \times 1}(\mathbf{M})), \quad i \in \{1, 2, 3\} \quad (1)$$

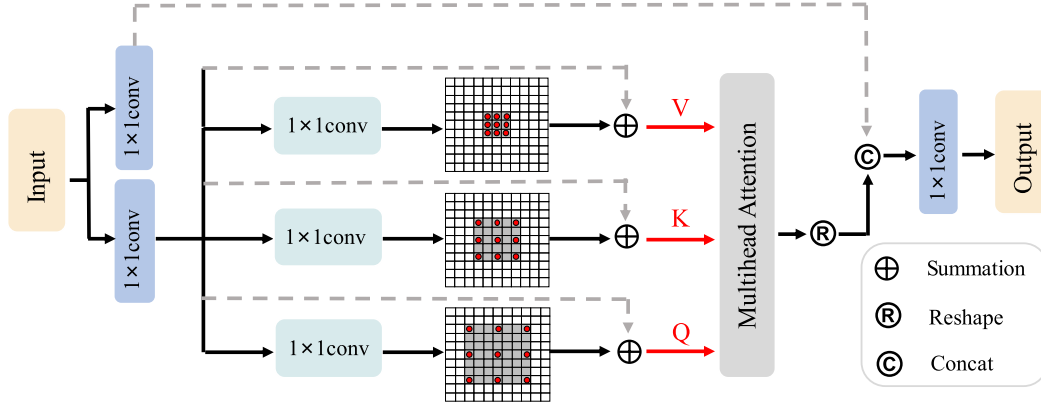


Fig. 3. Flowchart of the MDABE module.

where $\text{Conv}_{1 \times 1}$ represents a 1×1 convolution layer and $\text{Conv}_{3 \times 3, d=i}$ represents a 3×3 convolution layer with a dilation rate of i .

The outputs of each of the three branches M_i are residually summed to fully preserve the shallow features, and the results obtained are denoted as

$$\mathbf{Y}_i = \mathbf{M}_i \oplus \mathbf{M}, \quad i \in \{1, 2, 3\} \quad (2)$$

where $\oplus$ represents element-wise summation.

To further enhance the feature representation, we use feature maps at different scales as inputs to the query ($\mathbf{Q}$), key ($\mathbf{K}$), and value ($\mathbf{V}$) in the attention mechanism, i.e., $\mathbf{Y}_3$ as $\mathbf{Q}$, $\mathbf{Y}_2$ as $\mathbf{K}$, and $\mathbf{Y}_1$ as $\mathbf{V}$. With this setup, the multihead attention mechanism can synthesize the feature information at different scales to capture the multilevel contextual relationships contained in the input feature maps, which are denoted, respectively, as

$$\mathbf{Q} = \mathbf{Y}_3, \quad \mathbf{K} = \mathbf{Y}_2, \quad \mathbf{V} = \mathbf{Y}_1. \quad (3)$$

Finally, the multihead attentional mechanism is introduced to pay more attention to the expression of context and suppress the expression of anomalies. The attention mechanism is represented as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{D}}\right)\mathbf{V} \quad (4)$$

where $\text{Softmax}(\cdot)$ represents the Softmax function to normalize the attention weights into a probability distribution and D is the column dimension of $\mathbf{Q}$ and $\mathbf{K}$.

Through the combination of multiscale dilation convolution and multihead attention mechanism, MSABE can synthesize feature information of different scales, improve the ability to capture complex background features in HSIs, and enhance the performance of background reconstruction. MSABE is not only able to capture the features in a wider range and reduce the influence of anomalies but also focus on the important background features through the attention mechanism, which can improve the accuracy of the background reconstruction and the Robustness.

C. Multiscale Mamba Reconstruction

HSI contains both spatial information and numerous spectral bands. However, this richness of information also brings higher computational complexity and data redundancy, placing higher demands for background generation models. Recently, Mamba has demonstrated its strong potential for modeling complex data, and it has achieved remarkable success in several areas [45], [46], especially in handling long sequence data. Therefore, we propose a multiscale Mamba reconstruction module, which aims to utilize the advantages of Mamba to improve the performance of HAD. Specifically, visual sentences and visual words are first generated by the convolutional stem, i.e., the local patch of the HSI is considered as visual sentences and they are further decomposed into subpatches as visual words to further explore the locality. Then, the subpatches are fed into the MiM hierarchical encoder module to extract background features at different scales. In the decoder stage, to fuse the features extracted by Mamba, we propose an MDFF module as shown in Fig. 4. MDFF achieves the fusion by dynamically selecting important features based on their global information during fusion. The whole process is done in two separate phases after the mixed convolutions are used to extract multiscale features from the feature map. The first stage is to dynamically adjust the features on the channel dimension (indicated by blue arrows in Fig. 4). The second stage obtains global spatial information by simple convolution (indicated by orange arrows in Fig. 4). Finally, the features obtained in the two stages are combined to obtain the final features.

First, three parallel branches with different sizes of convolution kernels are first used to extract the background feature information under different receptive fields, the sizes of the convolution kernels are set to 3×3 , 5×5 , and 7×7 , which are capable of capturing features at different scales. Next, the outputs of the above different scale convolutions are spliced in the channel dimension to preserve the multiscale features. In this way, we can fuse the features at each scale to obtain a rich feature map. A 1×1 convolutional layer is then applied on the spliced feature map, which not only preserves the important feature information but also reduces the computational cost by linearly combining the information from different channels. Specifically, the input feature maps

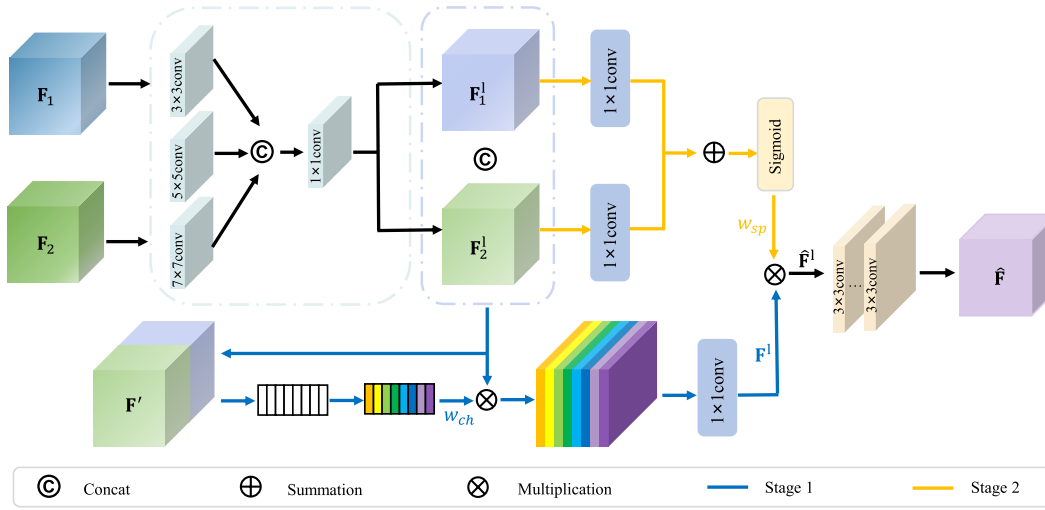


Fig. 4. Flowchart of the MDFF module.

are F_1 and F_2 , respectively. F_1 represents the multiscale semantic features extracted from the MiM layered encoder, which includes four feature maps of different resolutions and can capture multiscale information from high resolution to low resolution. F_2 is the high-resolution feature map obtained by upsampling in the PatchExpand module. This module restores the spatial size of low-resolution features layer by layer and adjusts the number of channels to further enhance the spatial details and contextual information of the feature map. The generated F_2 gradually restores high-resolution spatial details and performs MDFF with the corresponding scale F_1 feature map through the MDFF module, thereby fully integrating multiscale spatial and semantic information. Taking the F_1 as an example, the process of F_1 can be expressed as

$$F_{i \times i} = \text{ReLU}(\text{BN}(\text{Conv}_{i \times i}(F_1))), \quad i \in \{3, 5, 7\} \quad (5)$$

$$F_{\text{concat}} = [F_{3 \times 3}, F_{5 \times 5}, F_{7 \times 7}] \quad (6)$$

$$F_1^l = \text{Conv}_{1 \times 1} * F_{\text{concat}} \quad (7)$$

where $\text{BN}(\cdot)$ represents batch normalization, $\text{ReLU}(\cdot)$ represents the rectified linear unit activation function, and $[]$ represents concatenating feature maps along the channel dimension. Similarly, F_2 is processed through the same operations to obtain F_2^l . Feature maps F_1^l and F_2^l are then concatenated to obtain F'

$$F' = [F_1^l, F_2^l]. \quad (8)$$

In the first stage, to enable the subsequent modules to effectively utilize the fused features, the number of channels of the feature maps has to be reduced back to the original value by a channel compression mechanism. The channel compression of MDFF is not achieved by a simple 1×1 convolution but is guided and optimized by the global channel information w_{ch} . Specifically, the global average pool is carried out in the spatial dimension to obtain the compressed feature map. The compressed feature map is then passed sequentially through a fully connected, ReLU activation function, a fully connected layer, and a Sigmoid function to generate the channel weight. Finally, w_{ch} is multiplied with F' channel by channel to obtain

the channel-weighted feature map, which is further refined and fused by using the 1×1 convolution layer to obtain F^l . By filtering the channels according to the feature significance, the network is guided to retain key features and discard irrelevant features. The whole process can be expressed as

$$w_{\text{ch}} = \text{Sigmoid}(\text{Linear}(\text{ReLU}(\text{Linear}(\text{AvgP}(F'))))) \quad (9)$$

$$F^l = \text{Conv}_{1 \times 1}(w_{\text{ch}} \otimes F') \quad (10)$$

where $\text{AvgP}(\cdot)$ represents the global average pool, $\text{Linear}(\cdot)$ represents a fully connected layer, $\text{Sigmoid}(\cdot)$ represents the Sigmoid activation function, and $\otimes$ means element-wise multiplication.

In the second stage, to efficiently model the spatial dependencies between the local feature maps, the global spatial information w_{sp} is obtained by a 1×1 convolutional layer combined with the feature maps F_1^l and F_2^l , followed by a Sigmoid activation. The feature maps are then adjusted using weights generated by the Sigmoid activation function, highlighting the significant spatial regions in the image while suppressing unimportant areas, which can be expressed as

$$w_{\text{sp}} = \text{Sigmoid}(\text{Conv}_{1 \times 1}(F_1^l \oplus F_2^l)) \quad (11)$$

$$\hat{F}^l = w_{\text{sp}} \otimes F^l. \quad (12)$$

Finally, the feature map $\hat{F}^l$ is reconstructed using three 3×3 convolutions to obtain $\hat{F}$.

III. EXPERIMENTAL RESULTS

In this section, we perform extensive experiments on seven real hyperspectral datasets to validate the effectiveness of our method.

A. Dataset Introduction

1) *Cat Island*: This dataset was acquired by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor near Cat Island on September 12, 2010 [47], which has a spatial size of 150×150 , spatial resolution of 17.2 m, and a total of 188 spectral band counts. The anomaly in the image is an

airplane, which consists of 19 anomalous pixels, or 0.08% of the total pixels.

2) *Bay Champagne*: This dataset, also acquired by the AVIRIS sensor [47], contains 100×100 pixels, of which 11 pixels are considered abnormal targets, accounting for 0.11% of the total pixels. The spatial resolution is 4.4 m and consists of 188 spectral bands after the removal of unsuitable bands.

3) *Texas Coast*: This dataset was acquired at Texas Coast, captured by the AVIRIS sensor [47], with an image having a spatial extent of 100×100 and a spatial resolution of 17.2 m. A total of 204 spectral bands were retained after excluding the noise bands. The main anomalies are a few houses, including 67 anomalous pixels, representing 0.67% of the total number of pixels.

4) *Los Angeles*: The acquisition scene of this dataset [47] was in Loss Angeles, captured by the AVIRIS sensor on November 9, 2011, which consisted of 100×100 pixels with a spatial resolution of 7.1 m, a total of 205 spectral bands, and a total of up to 272 anomalous pixels, accounting for 2.72% of the total pixels.

5) *HAD-100*: The HAD-100 dataset was developed by Li et al. [39], containing rich and realistic scenes. From this dataset, we chose one of the images (ang20191004t185054_27) and referred it to HAD-100. This dataset was obtained by the AVIRIS-NG sensor, with a resolution of 8.4 m and a size of $75 \times 75 \times 425$. There are a total of 251 abnormal pixels, accounting for 4.46% of the total number of pixels.

6) *Gulfport*: This dataset was obtained by the AVIRIS sensor at Gulfport [47], where the anomalous objects were three different sized and shaped aircraft with a total of 60 anomalous pixels, accounting for 0.6% of the total number of pixels. The spatial range of the image is 100×100 , with a spatial resolution of 3.4 m and a total of 191 spectral bands.

7) *Szu-Data*: This dataset was captured by the Specim FX10 hyperspectral camera installed on the drone platform of Yuehai Campus of Shenzhen University. It is a hyperspectral dataset with realistic and complex scenes. The size of the dataset is 120×120 , with a total of 25 abnormal pixels, accounting for 0.17% of the total number of pixels. This data consists of 112 bands with a spectral resolution of 5.5 nm, and a spatial resolution of 0.1 m.

B. Experimental Setup

1) *Comparison Methods*: We introduced nine state-of-the-art hyperspectral anomaly methods for comparison, including traditional methods and DL-based methods. They are DCAE [42], RGAE [31], GAED [48], GTVLRR [49], PCA-TLRSR [50], Auto-AD [51], BS³LNet [27], DirectNet [52], and PDBSNet [28]. Among them, GTVLRR and PCA-TLRSR are representative representation-based methods, and the rest are state-of-the-art DL-based methods proposed in the last three years.

2) *Experimental Settings*: All experiments covered in this article are based on the PyTorch framework. The DL-based methods are implemented on a cluster of NVIDIA GPUs (GeForce RTX 4060) and Intel Core i7 processors. The

software stack runs on Linux system Ubuntu 22.04 and the traditional methods run on MATLAB 2021b. We used the Adam optimizer [53] with a learning rate set to $1e^{-4}$ and weight decay set to $1e^{-5}$ to optimize the performance of the MMR-HAD model efficiently.

3) *Evaluation Metrics*: We used evaluation metrics recognized in the field of HAD to assess the performance of MMR-HAD and other comparative methods. The evaluation metrics include anomaly detection map, box plot [54], 2-D receiver operating characteristic (ROC) curve, and area under the ROC curve (AUC) [55]. Among them, the anomaly detection map can visually represent the background suppression effect and anomaly detection. Through the color change, we can quickly identify the location and intensity of the abnormal pixels to evaluate the effectiveness of MMR-HAD. The box plot, also called statistical separability analysis, is used to evaluate the separability between background and anomaly. ROC presents the probability of detection (PD) versus the false alarm rate (FAR) for all possible thresholds, and curves close to the upper left corner indicate better detection performance. AUC is further calculated by integrating the area under the ROC curve, and a larger AUC indicates a higher accuracy. Ideally, an AUC of 1 means that all anomalous targets are detected.

C. Comparative Experimental Analysis

In this article, we show the superiority of our proposed model MMR-HAD through a large number of experiments, and the results of seven real datasets are analyzed in detail in the following, in which the AUC values of the datasets are shown in Table I, the color anomaly maps are shown in Figs. 5–11, the ROC curve chart is shown in Fig. 12, and the box plot is shown in Fig. 13. These seven datasets can be simply categorized into three groups according to the size of the anomaly targets. Cat Island, Szu-data, and Bay Champagne contain small targets; Texas Coast and Gulfport contain medium targets; Los Angeles and HAD-100 belong to large targets.

1) *Cat Island*: The AUC value of the proposed method on this dataset is 0.9949, which is higher than the rest of the state-of-the-art methods. As can be seen from the anomaly maps, our method recognizes the anomaly more accurately. While RGAE, GAED, and Auto-AD identify the backgrounds in the lower left corner as anomalies, leading to lower AUC values. Their ROC curve and box plot are shown in Figs. 12(a) and 13(a), respectively. In Fig. 12(a), the red curve of MMR-HAD is above the other curves and closest to the upper left corner, which proves that MMR-HAD achieves effective background reconstruction and successfully suppresses the reconstruction of anomalies. In Fig. 13(a), the blue boxes of DCAE and MMR-HAD are very low, which indicates better background reconstruction.

2) *Bay Champagne*: This dataset has smaller anomaly targets and simpler backgrounds. It can be seen from the AUC values and the anomaly maps that our method is in the lead and has some superiority, while the performance of PCA-TLRSR, DCAE, DirectNet, and PDBSNet is also very good. From the

TABLE I
AUC SCORES OF TEN HAD METHODS ON SEVEN DATASETS

Datasets	Methods									
	GTVLRR	PCA-TLRSR	DCAE	RGAE	GAED	Auto-AD	BS ³ LNet	DirectNet	PDBSNet	Ours
Cat Island	0.9787	0.9823	0.9863	0.9394	0.9222	0.9749	0.9788	0.9733	0.9775	0.9983
Bay Champagne	0.9917	0.9964	0.9954	0.8664	0.9892	0.9919	0.9820	0.9950	0.9934	0.9972
Texas Coast	0.9512	0.9926	0.9884	0.9822	0.9417	0.9833	0.9679	0.9883	0.9927	0.9958
Los Angeles	0.9422	0.9859	0.9960	0.9948	0.9925	0.9959	0.9174	0.9648	0.9946	0.9965
HAD-100	0.8642	0.9760	0.9841	0.9817	0.9506	0.9737	0.9388	0.9883	0.9851	0.9932
Gulfport	0.9881	0.9930	0.9405	0.7615	0.9678	0.9585	0.8820	0.9673	0.9845	0.9964
Szu-data	0.9733	0.9707	0.9811	0.9753	0.9806	0.9269	0.9400	0.8225	0.7993	0.9858
Average	0.9556	0.9853	0.9817	0.9288	0.9635	0.9722	0.9438	0.9571	0.9610	0.9947

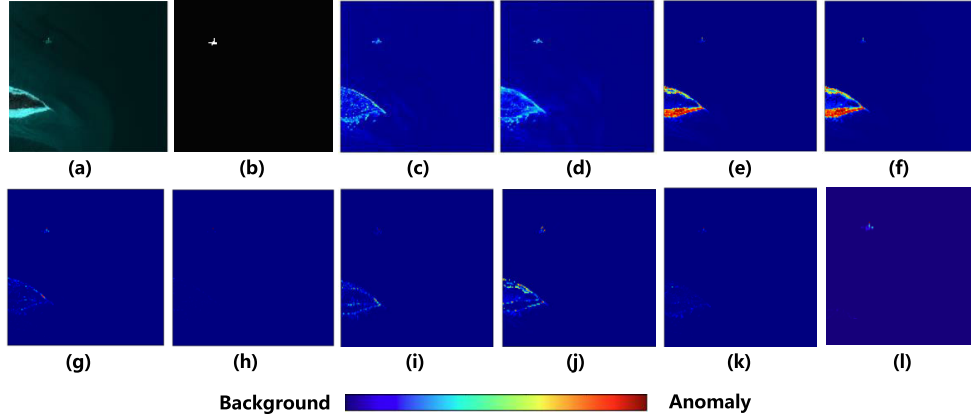


Fig. 5. Color anomaly maps obtained by different algorithms for the Cat Island dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

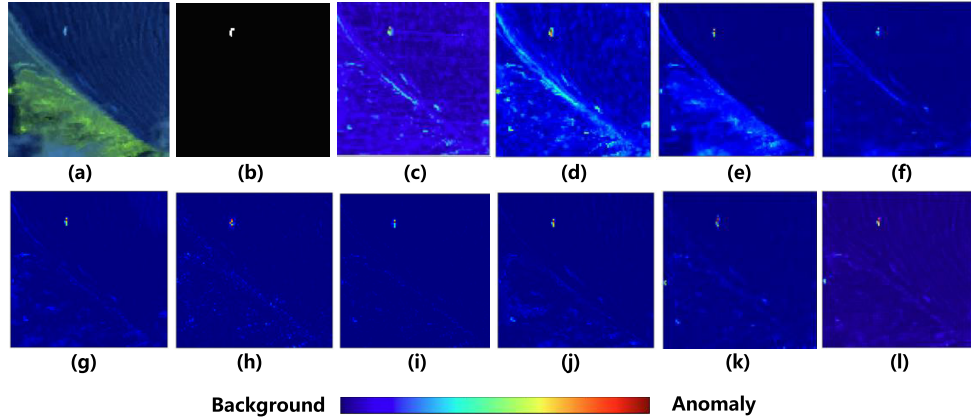


Fig. 6. Color anomaly maps obtained by different algorithms for the Bay Champagne dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

overall ROC curve in Fig. 12(b), MMR-HAD exhibits higher PD and is closer to the upper left corner. Fig. 13(b) shows the box plot of this dataset, and it can be seen that our method MMR-HAD still has some superiority in terms of separability between anomalies and background.

3) *Texas Coast*: This dataset has a large number of interfering factors in the background, so the task of anomaly detection will be relatively difficult. As can be seen from the AUC values and the anomaly maps, the performance of our method is the best. From the anomaly maps of GTVLRR, GAED, and Auto-AD, it can be seen that these three methods are highly affected by the background. From Figs. 12(c) and 13(c), it can be seen

that the five detectors, DCAE, RGAE, DirectNet, PDBSNet, and MMR-HAD, can effectively separate the background from the anomalies. It is proved that the proposed MMR-HAD can effectively suppress the background and separate the anomalous targets.

4) *Los Angeles*: The anomaly of this dataset is more obvious, as can be seen from Table I, the AUC value of our method is as high as 0.9966, which is the highest AUC value of all methods, but the AUC values of DCAE, RGAE, Auto-AD, and PDBSNet are all very high, just slightly lower than our method, which indicates that these methods for Los Angeles dataset all show good performance. However,

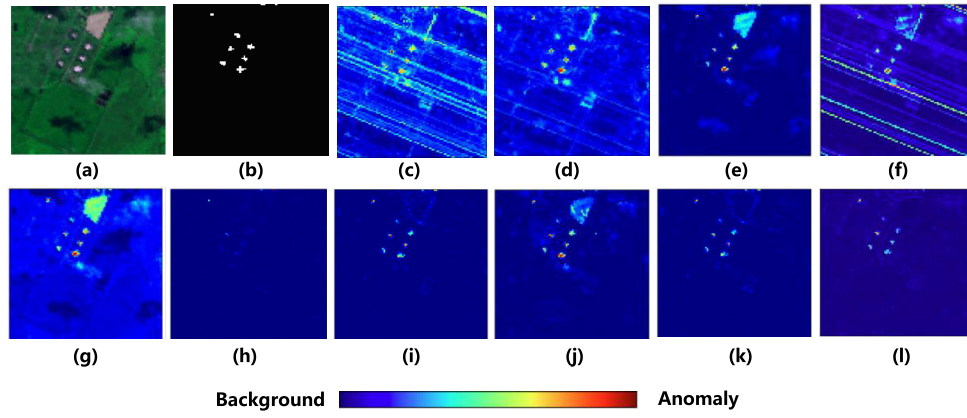


Fig. 7. Color anomaly maps obtained by different algorithms for the Texas Coast dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

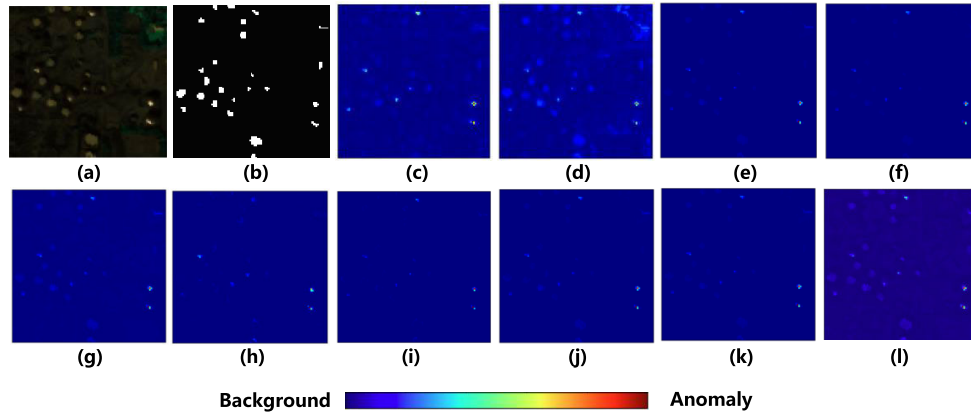


Fig. 8. Color anomaly maps obtained by different algorithms for the Los Angeles dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

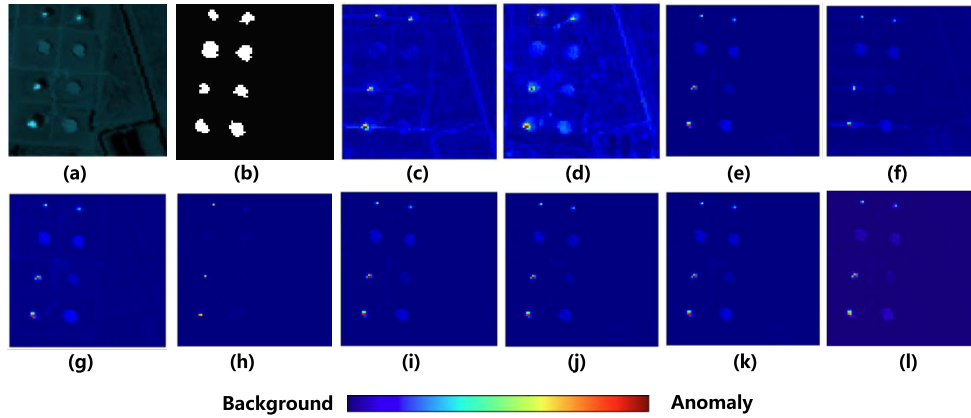


Fig. 9. Color anomaly maps obtained by different algorithms for the HAD-100 dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

their anomaly result plots are not visually obvious and it is difficult to distinguish the distribution of detected anomalies. Its ROC curve and box plot are shown in Figs. 12(d) and 13(d), respectively. As can be seen in Fig. 13(d), the blue box of MMR-HAD is narrow enough. When the false alarm rate value is larger than 0.01, the red curves are all located above the other curves, proving that the proposed MMR-HAD has a higher detection rate than other methods.

5) *HAD-100*: This dataset has larger anomaly targets than all the other four datasets, and as can be seen from the resultant graphs, except for BS³LNet which detects only a small number of anomalies, the rest of the methods detect most of the anomalies and show good performance. However, from the AUC value, it can be seen that our method has absolute superiority over the competitors. As can be seen from the box plot of Fig. 13(e), MMR-HAD performs well in terms of

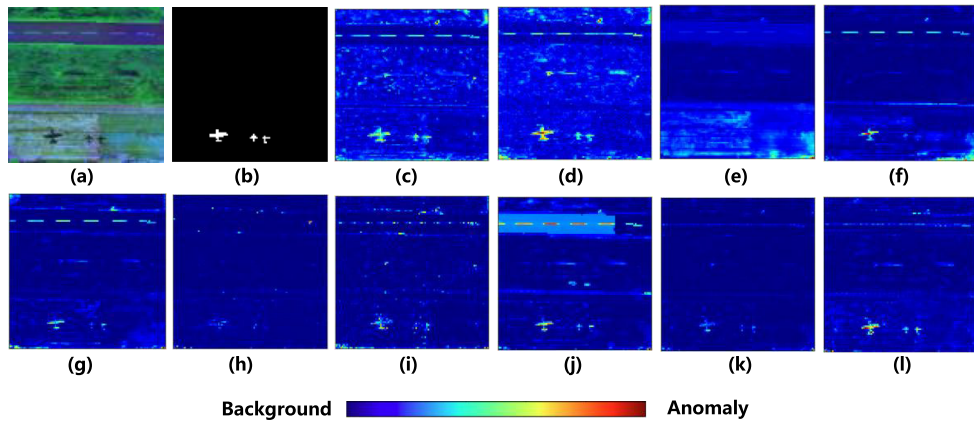


Fig. 10. Color anomaly maps obtained by different algorithms for the Gulfport dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

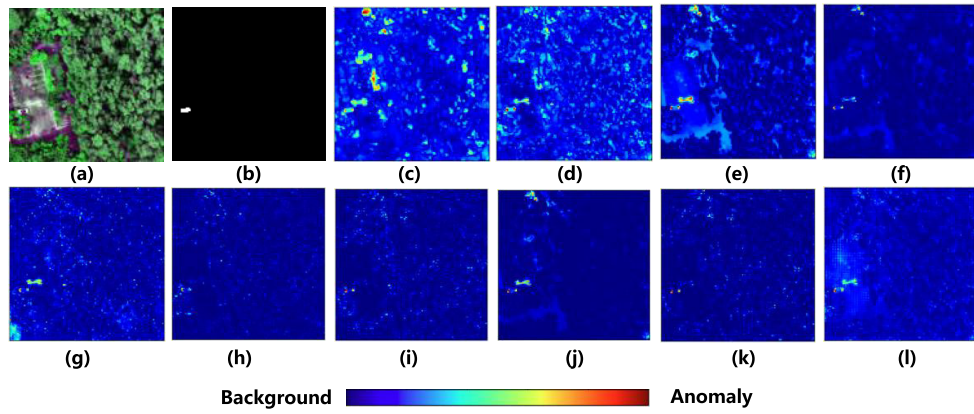


Fig. 11. Color anomaly maps obtained by different algorithms for the Szu-data dataset. (a) False color. (b) Ground truth. (c) GTVLRR. (d) PCA-TLRSR. (e) RGAE. (f) GAED. (g) Auto-AD. (h) BS³LNet. (i) DirectNet. (j) DCAE. (k) PDBSNet. (l) Ours.

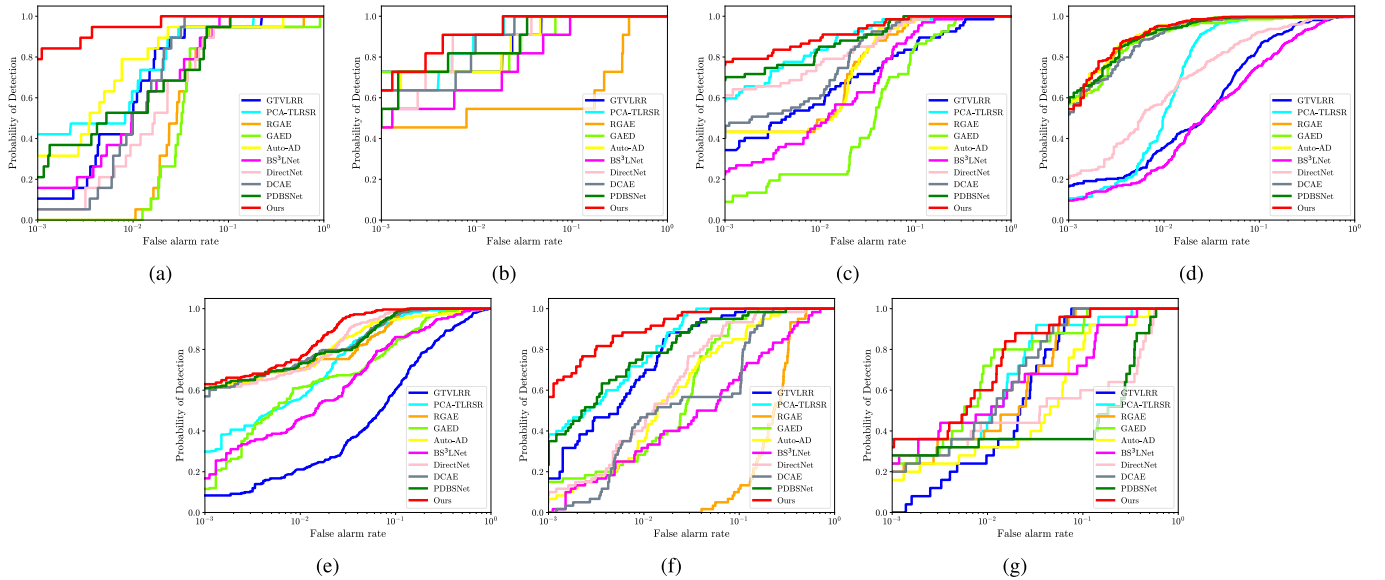


Fig. 12. ROC curves obtained by different methods for seven datasets. (a) Cat Island. (b) Bay Champagne. (c) Texas Coast. (d) Los Angeles. (e) HAD-100. (f) Gulfport. (g) Szu-data.

background anomaly separability and the degree of background suppression compared to the other methods. Meanwhile, from the ROC curve plot in Fig. 12(e), we can see that the ROC curve of MMR-HAD is closer to the upper left corner.

6) *Gulfport*: The anomalous targets in the Gulfport dataset consist of three aircraft with distinct shapes and sizes. From Fig. 10, it can be seen that the performance of RGAE and BS³LNet is not satisfactory, especially when identifying

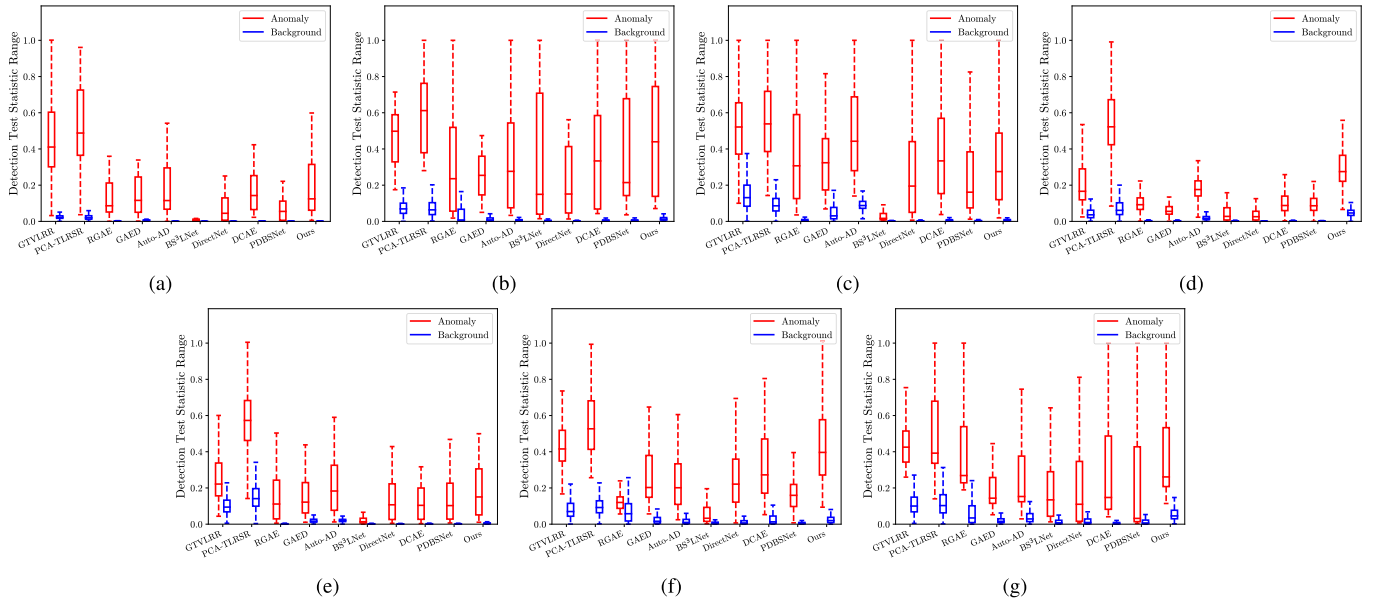


Fig. 13. Separability maps obtained by different methods for seven datasets. (a) Cat Island. (b) Bay Champagne. (c) Texas Coast. (d) Los Angeles. (e) HAD-100. (f) Gulfport. (g) Szu-data.

the shapes of two small aircraft. This is mainly because these methods fail to fully capture the key information of abnormal targets when dealing with targets with significant morphological changes, resulting in poor detection performance. Although GTVLRR, PCA-TLRSR, GAED, Auto-AD, DCAE, and DirectNet perform well in highlighting anomalous targets, they fail to effectively remove the interference of background structures while retaining background information, which affects the accurate recognition of anomalous targets. On the contrary, our proposed MMR-HAD method demonstrates superior performance in handling these anomalous targets, accurately identifying three aircraft of different sizes and effectively suppressing background interference while highlighting anomalous targets. From the ROC curve in Fig. 12(f), it can be seen that the ROC curve of MMR-HAD always outperforms other methods and has significant advantages. From the box plot in Fig. 13(f), it can be seen that the interval between the anomaly box and the background box of our model is the largest, showing the strong separability of anomalous targets and background.

7) *Szu-Data*: The background of the Szu-data dataset is relatively complex, with a large number of trees as interference factors, making the task of detecting abnormal targets difficult. Fig. 11 shows that GTVLRR, PCA-TLRSR, and RGAE exhibit significant performance fluctuations, failing to effectively suppress background interference and resulting in a weak ability to highlight and separate anomalous targets. Although other methods such as GTVLRR, PCA-TLRSR, and Auto-AD can constrain the background to some extent, the recognition performance of anomalous targets is reduced. In contrast, MMR-HAD can detect anomalous vehicles well under complex background conditions, demonstrating significant advantages. The ROC curve in Fig. 12(g) shows our model has higher detection performance than other methods. From the box plot in Fig. 13(g), it can also be seen that our method has strong separability between anomalous targets and

backgrounds, further verifying the advantages of MMR-HAD in handling complex backgrounds.

D. Ablation Study

This section conducts ablation studies on seven datasets to evaluate the effectiveness of RM, MDABE, and MDFF, as shown in Table II.

- 1) After removing the RM module, the AUC scores of all datasets showed a significant decrease, especially on the Texas Coast, Los Angeles, and HAD-100 datasets, where the performance degradation was particularly evident. This indicates that the RM module plays a crucial role in removing abnormal pixels from data, especially in complex backgrounds with large and clustered abnormal targets. The RM module can effectively suppress the interference of abnormal pixels on background reconstruction. By introducing the RM module, we can significantly reduce the impact of abnormal pixels, thereby improving the detection performance in subsequent processing stages.
- 2) After removing the MDABE module, although the model performance has decreased, it can still maintain a relatively stable effect. This module utilizes convolutions of different scales to capture details and global features, and focuses on important background areas through attention mechanisms, reducing the influence of anomalous pixels. After removing this module, the accuracy of background reconstruction decreased, further demonstrating the positive role of the MDABE module in improving background feature extraction and enhancing robustness.
- 3) In the experiment of removing the MDFF module, the performance of the model slightly decreased on all the datasets. The removal of this module further validates the importance of MDFF in feature extraction and background information enhancement. Specifically,

TABLE II
AUC SCORES OF ABLATION STUDY ON SEVEN DATASETS

Modules			Datasets and AUC values						
RM	MDABE	MDFF	Cat Island	Bay Champagne	Texas Coast	Los Angeles	HAD-100	Gulfport	Szu-data
✗	✓	✓	0.9390	0.9466	0.8348	0.7931	0.7264	0.8945	0.8854
✓	✗	✓	0.9850	0.9879	0.9873	0.9943	0.9892	0.9686	0.9793
✓	✓	✗	0.9898	0.9918	0.9934	0.9945	0.9842	0.9904	0.9841
✓	✓	✓	0.9983	0.9972	0.9958	0.9965	0.9932	0.9964	0.9858

TABLE III
RUNNING TIMES (IN SECONDS) OF THE TEN CONSIDERED DETECTORS

Datasets	Methods									
	GTVLRR	PCA-TLRSR	DCAE	RGAE	GAED	Auto-AD	BS ³ LNet	DirectNet	PDBSNet	Ours
Cat Island	183.6259	15.1027	0.0451	0.0408	0.0738	0.1649	0.7307	0.7092	0.7191	1.0476
Bay Champagne	77.1548	6.8643	0.0359	0.0354	0.0604	0.1762	0.7028	0.6981	0.7008	1.0298
Texas Coast	89.7765	6.5868	0.0419	0.0355	0.0613	0.1744	0.6998	0.6973	0.7059	1.0257
Los Angeles	53.6805	6.6594	0.0413	0.0357	0.0607	0.1541	0.6952	0.6978	0.7048	1.0408
HAD-100	46.8024	3.8124	0.0310	0.0247	0.0468	0.1754	0.6960	0.6955	0.7074	0.9501
Gulfport	78.1602	6.8132	0.0304	0.0264	0.0477	0.1686	0.6966	0.6942	0.7001	1.0446
Szu-data	108.0734	9.6560	0.0513	0.0255	0.0502	0.1786	0.6920	0.6909	0.6996	1.1693

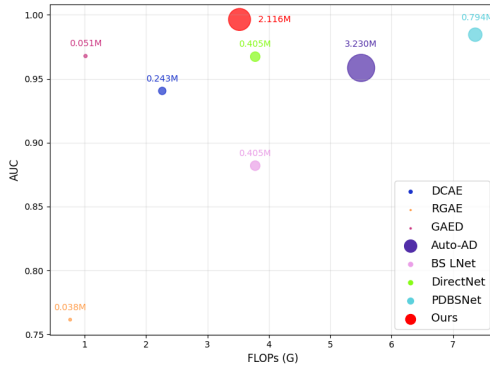


Fig. 14. Model complexity of eight DL-based methods.

the MDFF module effectively combines global and local information by dynamically selecting and fusing features of different scales, thereby enhancing the expressive power of background information. In this way, the MDFF module can generate pure background images that are not affected by abnormal target interference.

Experimental results have shown that the MMR-HAD model performs well on various datasets, demonstrating the powerful ability of our method to remove outlier pixels and enhance background representation. The RM strategy and multiscale extended attention module each undertake important tasks in our method, and their synergistic effect significantly improves the performance of the model. The RM module first simplifies data input by removing most of the abnormal pixels. On this basis, the MDABE module further optimizes the reconstruction of background features. By conducting ablation experiments, we verified the synergistic effect between these modules and further demonstrated that their organic combination in the overall framework plays an important role in improving the model's ability and robustness.

E. Model Complexity

In this section, we mainly evaluate the computational complexity of MMR-HAD with the competitors. Fig. 14 shows

the floating-point operations (FLOPs, unit: G), model parameters (Params, unit: M), and AUC values. The horizontal axis represents the computational complexity of the model, specifically, the FLOPs value, which represents the amount of floating-point operations required for the model to perform detections. The vertical axis represents the detection performance of the model, measured by AUC values. The size of the dots in the figure reflects the number of parameters in the model. All experimental data were evaluated based on the Gulfport dataset, ensuring the objectivity and reliability of the experiment.

From Fig. 14, it can be seen that the AUC value of MMR-HAD ranks above all methods. This indicates that our model is significantly superior to other comparison methods in terms of accuracy and has stronger detection capabilities. Although RGAE is the lightest in terms of FLOPs and parameter count, with minimal computational complexity and parameter count, its AUC value is significantly lower than MMR-HAD. Compared to PDBSNet and Auto-AD, our model significantly reduces computational complexity and demonstrates higher computational efficiency. This advantage enables MMR-HAD to reduce the consumption of computing resources while ensuring detection accuracy. In terms of parameter count, our model has significantly fewer circular dots than Auto-AD, indicating that our model has fewer parameters. Compared with DirectNet and BS³LNet, our model not only significantly improves AUC performance, but also maintains a moderate level of computational complexity. Overall, MMR-HAD achieves a good balance between computational complexity and parameter count, effectively reducing the consumption of computing resources while maintaining high performance, which will facilitate the anomalous target detection of large-scale hyperspectral datasets.

Table III shows the inference time of different anomaly detection methods on seven hyperspectral datasets. From the results in Table III, it can be seen that our model significantly reduces the running time compared to traditional methods. Although lightweight methods such as DCAE, RGAE, and

GAED perform outstandingly in inference speed and have shorter running times, their performance is limited. Compared to other more complex DL methods, our model is comparable in terms of runtime but demonstrates significant advantages in detection accuracy. Overall, our method achieves a balance between accuracy and efficiency when compared with all the DL-based competitors.

IV. CONCLUSION

In this article, we propose a multiscale Mamba reconstruction network for HAD. MMR-HAD removes most of the anomalous pixels in HSIs by utilizing the RM strategy to reduce the interference of these anomalous pixels on the background reconstruction. We introduce the MDABE module to deal with the remaining anomalous pixels to further enhance the background representation. Finally, the background image is reconstructed by the multiscale dynamic feature fusion module to further refine and enhance the background information. As a result, a purer background image is obtained. Numerous experimental results show that MMR-HAD exhibits superior detection performance and outperforms many existing HAD methods. Our study introduces a new perspective on integrating Mamba into HAD tasks, which can serve as a cornerstone for further research in this direction.

REFERENCES

- [1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [2] X. Fu, H. Liang, and S. Jia, "Mixed noise-oriented hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526916.
- [3] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10438–10454, Dec. 2021.
- [4] X. Fu, X. Zhou, Y. Fu, P. Liu, and S. Jia, "Progressive semantic enhancement network for hyperspectral and LiDAR classification," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Dec. 20, 2024, doi: 10.1109/TNNLS.2024.3513979.
- [5] H. Yang, H. Yu, K. Zheng, J. Hu, T. Tao, and Q. Zhang, "Hyperspectral image classification based on interactive transformer and CNN with multilevel feature fusion network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
- [6] L. Zhuang, M. K. Ng, X. Fu, and J. M. Bioucas-Dias, "Hy-demosaicing: Hyperspectral blind reconstruction from spectral subsampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515815.
- [7] W. He, X. Fu, N. Li, Q. Ren, and S. Jia, "LGCT: Local-global collaborative transformer for fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5537114.
- [8] X. Fu, Y. Guo, M. Xu, and S. Jia, "Hyperspectral image denoising via robust subspace estimation and group sparsity constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5512716.
- [9] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.
- [10] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500117.
- [11] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Aug. 1990.
- [12] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [13] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [14] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.
- [15] M. J. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.
- [16] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, vol. 3, Mar. 2004, pp. 1818–1824.
- [17] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.
- [18] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [19] T. Guo, L. He, F. Luo, X. Gong, L. Zhang, and X. Gao, "Learnable background endmember with subspace representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5501513.
- [20] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [21] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.
- [22] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [23] T. Guo, L. He, F. Luo, X. Gong, Y. Li, and L. Zhang, "Anomaly detection of hyperspectral image with hierarchical antinoise mutual-incoherence-induced low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510213.
- [24] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.
- [25] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.
- [26] J. Zhang et al., "A light CNN based on residual learning and background estimation for hyperspectral anomaly detection," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 132, Aug. 2024, Art. no. 104069.
- [27] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS³LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504218.
- [28] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, "PDB-SNet: Pixel-shuffle downsampling blind-spot reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5511914.
- [29] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7406–7417, Oct. 2020.
- [30] S. Chang, B. Du, and L. Zhang, "A sparse autoencoder based hyperspectral anomaly detection algorithm using residual of reconstruction error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5488–5491.
- [31] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511314.
- [32] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1527–1537, Mar. 2020.
- [33] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "GAN-based hyperspectral anomaly detection," in *Proc. 28th Eur. Signal Process. Conf. (EUSIPCO)*, Jan. 2021, pp. 1891–1895.

- [34] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, Jul. 2020.
- [35] Y. Li, T. Jiang, W. Xie, J. Lei, and Q. Du, "Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512811.
- [36] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, "Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022.
- [37] F. Luo, S. Shi, T. Guo, Y. Dong, L. Zhang, and B. Du, "AGMS: Adversarial sample generation-based multiscale Siamese network for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5536713.
- [38] S. Jia, Z. Min, and X. Fu, "Multiscale spatial-spectral transformer network for hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 96, pp. 117–129, Aug. 2023.
- [39] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, and W. An, "You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5506718.
- [40] J. Lian, L. Wang, H. Sun, and H. Huang, "GT-HAD: Gated transformer for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 2, pp. 3631–3645, Feb. 2025.
- [41] S. Xiao, T. Zhang, Z. Xu, J. Qu, S. Hou, and W. Dong, "Anomaly detection of hyperspectral images based on transformer with spatial-spectral dual-window mask," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1414–1426, 2023.
- [42] Z. Wu and B. Wang, "Background reconstruction via 3D-transformer network for hyperspectral anomaly detection," *Remote Sens.*, vol. 15, no. 18, p. 4592, Sep. 2023.
- [43] X. He et al., "Global-to-local spatial-spectral awareness transformer network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5530422.
- [44] Z. Wu and B. Wang, "Transformer-based autoencoder framework for nonlinear hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5508015.
- [45] T. Chen et al., "MiM-ISTD: Mamba-in-mamba for efficient infrared small target detection," 2024, *arXiv:2403.02148*.
- [46] J. Yao, D. Hong, C. Li, and J. Chanussot, "SpectralMamba: Efficient mamba for hyperspectral image classification," 2024, *arXiv:2404.08489*.
- [47] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [48] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, "Hyperspectral anomaly detection with guided autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538818.
- [49] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [50] M. Wang, Q. Wang, D. Hong, S. K. Roy, and J. Chanussot, "Learning tensor low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 679–691, Jan. 2023.
- [51] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.
- [52] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, and A. Plaza, "Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5504115.
- [53] M. Reyad, A. M. Sarhan, and M. Arafa, "A modified Adam algorithm for deep neural network optimization," *Neural Comput. Appl.*, vol. 35, no. 23, pp. 17095–17112, Aug. 2023.
- [54] D. F. Williamson, R. A. Parker, and J. S. Kendrick, "The box plot: A simple visual method to interpret data," *Ann. Internal Med.*, vol. 110, no. 11, pp. 916–921, 1989.
- [55] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2020.



Xiyu Fu (Member, IEEE) received the B.S. degree from Wuhan University, Wuhan, China, in 2012, and the M.S. and Ph.D. degrees from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2015 and 2019, respectively.

He is currently an Assistant Professor at Shenzhen University, Shenzhen, China. His research interests include hyperspectral image restoration, anomaly detection, and super-resolution.



Ting Zhang received the B.E. degree from Northwest Minzu University, Lanzhou, China, in 2023. She is pursuing the master's degree with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Her research interests include hyperspectral anomaly detection and object tracking.



Juan Cheng received the B.S. degree from Chang'an University, Xi'an, China, in 2016, the M.S. degree from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2019, and the Ph.D. degree from the Aerospace Information Research Institute, Chinese Academy of Sciences, in 2022.

She is currently a Special Research Assistant at Shaanxi Key Laboratory of Optical Remote Sensing and Intelligent Information Processing, Xi'an, China. Her research interests include remote sensing image analysis and application.



Sen Jia (Senior Member, IEEE) received the B.S. and Ph.D. degrees from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

Since 2008, he has been with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, where he is currently a Full Professor. His research interests include hyperspectral image processing, signal and image processing, and machine learning.