GT-HAD: Gated Transformer for Hyperspectral Anomaly Detection

Jie Lian, Lizhi Wang, *Member, IEEE*, He Sun, and Hua Huang, *Senior Member, IEEE*

***Abstract*— Hyperspectral anomaly detection (HAD) aims to distinguish between the background and anomalies in a scene, which has been widely adopted in various applications. Deep neural network (DNN)-based methods have emerged as the predominant solution, wherein the standard paradigm is to discern the background and anomalies based on the error of self-supervised hyperspectral image (HSI) reconstruction. However, current DNN-based methods cannot guarantee correspondence between the background, anomalies, and reconstruction error, which limits the performance of HAD. In this article, we propose a novel gated transformer network for HAD (GT-HAD). Our key observation is that the spatial–spectral similarity in HSI can effectively distinguish between the background and anomalies, which aligns with the fundamental definition of HAD. Consequently, we develop GT-HAD to exploit the spatial–spectral similarity during HSI reconstruction. GT-HAD consists of two distinct branches that model the features of the background and anomalies, respectively, with content similarity as constraints. Furthermore, we introduce an adaptive gating unit to regulate the activation states of these two branches based on a content-matching method (CMM). Extensive experimental results demonstrate the superior performance of GT-HAD. The original code is publicly available at https://github.com/jeline0110/GT-HAD, along with a comprehensive benchmark of state-of-the-art HAD methods.**

***Index Terms*— Content similarity, gating unit, hyperspectral anomaly detection (HAD), transformer.**

I. Introduction

THE hyperspectral image (HSI) captures the power distribution of a scene as 3-D data, which delineates the spectral intensity for each wavelength at every pixel location [1], [2], [3]. The rich spectral and spatial information in HSI has proven beneficial for diverse scene analysis and understanding tasks. One such task is hyperspectral anomaly detection (HAD), which focuses on distinguishing between the background and anomalies in a scene. Recently, HAD has found applications in various domains, including remote sensing, military surveillance [4], and mineral exploration [5].

Manuscript received 1 June 2023; revised 14 December 2023; accepted 9 January 2024. Date of publication 12 February 2024; date of current version 6 February 2025. This work was supported by the National Natural Science Foundation of China under Grant 62322204, Grant 62131003, Grant 62072038, and Grant 62301534. *(Corresponding author: Lizhi Wang.)*

Jie Lian and Lizhi Wang are with the School of Computer Science, Beijing Institute of Technology, Beijing 100081, China (e-mail: lianjie@bit.edu.cn; wanglizhi@bit.edu.cn).

He Sun is with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: sunhe@aircas.ac.cn).

Hua Huang is with the School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China (e-mail: huahuang@bnu.edu.cn).

Digital Object Identifier 10.1109/TNNLS.2024.3355166

According to the strategies employed to distinguish between the background and anomalies, existing HAD methods can be categorized into three groups: 1) statistics-based methods [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], 2) representation-based methods [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], and 3) deep neural network (DNN)-based methods [30], [31], [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42]. Statistics-based methods assume that the background follows a specific data distribution, and the pixels that do not conform to this data distribution are regarded as anomalies. However, the background in different scenarios may follow distinct data distributions, and thus restricting the background to a single data distribution would limit the generalization ability of the model. Representation-based methods can be classified into two categories. The first category assumes that the background can be approximately represented by several basis signals extracted from the HSI, while anomalies cannot. The second category utilizes matrices or tensors to represent the background and anomaly components, which are then solved within a mathematical optimization framework. However, representation-based methods typically introduce a large number of parameters that require manual tweaking, subject to individual experience, which results in uncertainty regarding method accuracy.

Owning to the powerful capability of feature modeling, DNN-based methods have emerged as the mainstream solutions for HAD. Unlike other methods, DNN-based methods do not rely on prior distribution assumptions or manual parameter tweaking. Instead, DNN-based methods employ the self-supervised HSI reconstruction as a proxy task of HAD, where the background can be accurately reconstructed by the network while anomalies cannot. Thus, the reconstruction errors, i.e., the differences between the original HSI and the reconstructed one, are treated as indicators of the anomalies.

However, it is not always guaranteed that the correspondence between the background, anomalies, and reconstruction error will hold in current DNN-based methods. The fundamental reason is that DNNs tend to prioritize the reconstruction of HSI contents with simple data distributions, which results in reconstruction difficulties of the background with complex data distribution and reconstruction over-fitting of the anomalies with simple data distribution. Such a scenario deviates from the basic assumption of the DNN-based methods that anomalies should have significantly higher reconstruction errors than the background.

To address such issues, feature constraints should be imposed on the background and anomalies to guide the HSI reconstruction. Specifically, the ability of DNN to reconstruct the background can be enhanced by enhancing the background

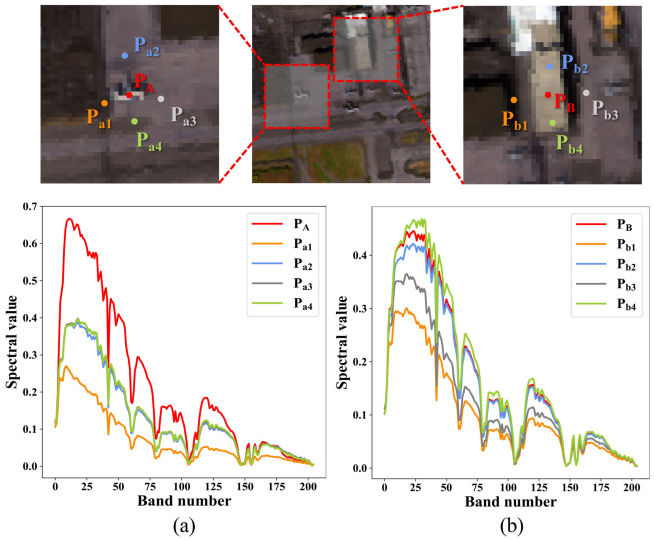

Fig. 1. Example intuitively shows the content similarity in HSI. (a) Spectral curves of P_A–P_{a4}. (b) Spectral curves of P_B–P_{b4}. The background (P_B) exhibits high similarity with its surrounding contents (P_{b1}–P_{b4}), while the anomaly (P_A) displays different characteristics and exhibits low similarity with the surrounding contents (P_{a1}–P_{a4}).

features, while preventing DNN from over-fitting the anomalies can be achieved by weakening the anomaly features.

Recently, several DNN-based methods have been proposed to incorporate specific constraints to guide the HSI reconstruction. The autonomous HAD network (Auto-AD) [31] assumes that the network can reconstruct the background while anomalies appear as reconstruction errors, and the errors are used as adaptive weights to further suppress the anomalies reconstruction. However, Auto-AD overlooks the fact that it is difficult to reconstruct the background with a complex data distribution, which results in reconstruction errors that may mistakenly suppress the background reconstruction. The deep low-rank (LR) prior-based method (DeepLR) [32] assumes that the reconstructed HSI exhibits an LR characteristic, while the anomalies deviate from this distribution and appear as reconstruction errors. Yet the high-frequency information in the reconstructed background would be lost when enforcing an LR distribution, which increases false alarm rates (FARs) in HAD. Existing constraints primarily concentrate on suppressing the anomalies reconstruction, while neglecting the reconstruction quality of the background. This limits their applicability in certain scenarios. Hence, further investigation is necessary to develop more robust and accurate DNN-based HAD methods.

Back to the basics of HAD, the anomalies are defined as such pixels that possess different spectral characteristics from their surrounding contents. This definition implies that the background exhibits high similarity with its surrounding contents, whereas anomalies exhibit low similarity, as depicted in Fig. 1. This principle motivates us to exploit the content similarity as the feature constraint to enhance the background features and weaken the anomaly features, thus guiding the reconstruction of the background and anomalies.

In this article, we propose a novel gated transformer network for HAD (GT-HAD), which leverages the exceptional capabilities of the transformer in capturing content similarity [43], [44], [45], [46], [47], [48], aligning with our motivation in exploiting content similarity in HSI reconstruction. The proposed GT-HAD comprises two branches dedicated to enhancing background features and weakening anomaly features by mining the content similarity in HSI. In addition, we develop a gating unit to regulate the activation states of each branch based on a content-matching method (CMM). Comprehensive experiments demonstrate superior performance compared with state-of-the-art HAD methods. We believe our study would propel further research on integrating transformers into HAD tasks and encourage the exploration of additional possibilities for HAD methods.

The main contributions of our work are summarized as follows.

1) We introduce a new GT-HAD, which leverages content similarity to guide the reconstruction of both background and anomalies.
2) We develop a gating unit to regulate the activation states of different branches in GT-HAD, based on a CMM.
3) We perform comprehensive comparative experiments and ablation studies on six HSI datasets to demonstrate the superiority and effectiveness of our proposed GT-HAD.

The remaining sections of this article are structured as follows. Section II provides an overview of related works on HAD, Section III presents the details of GT-HAD, Section IV presents the experimental results and analysis, and finally, Section V concludes the article.

II. Related Work

In this section, we briefly review the related work of HAD in statistics-based, representation-based, and DNN-based methods, respectively.

A. Statistics-Based

Statistics-based methods are the earliest proposed and widely used methods for the HAD task. The Reed–Xiaoli (RX) algorithm [6], based on the principle of the generalized likelihood ratio test (GLRT), is widely recognized as the benchmark method in this field. The RX assumes that the background follows a multivariate Gaussian distribution, and then the anomalies are detected by estimating the Mahalanobis distance between each test pixel and the background. Inspired by the classical RX, various improved versions are subsequently proposed, such as the local RX (LRX) [7], the weighted RX (WRX) [8], and the subspace RX (SRX) [9]. However, as abnormal targets often involve multiple pixels, RX-based methods relying on test point vector calculation may exhibit suboptimal detection performance. To address this issue, a two-step GLRT (2S-GLRT) method [10] is proposed, which considers the background information around each test pixel and aggregates neighboring pixels to detect multipixel anomalies. Besides the GLRT-based methods, many kernel-based HAD methods are proposed in the literature [11], [12], [13], [14]. The distributed online one-class support vector machine (doCSVM) method [13] maps the data into the kernel space and then separates the anomalies from the background. The kernel isolation forest detection (KIFD) method [14] assumes that the anomalies are more easily isolated than the background in kernel space and constructs an isolation forest to detect isolated pixels in the image. Moreover, different from the above methods, the recent MsRFQFT method [17] analyzes the discriminative properties of background and anomalies in the frequency domain.

Yet most statistics-based methods heavily rely on data assumptions, which limit their generalization capabilities in

different application scenarios. Furthermore, in the presence of complex scenes and noisy data, statistics-based methods may fail to capture subtle features or changes, potentially resulting in missed detection or false positives.

B. Representation-Based

In addition to statistics-based methods, representation theory has recently been applied to develop HAD algorithms. The collaborative representation detection (CRD) algorithm [18] is the pioneer of the representation-based methods, which assumes that background pixels can be approximated by a linear combination of their neighboring spatial pixels, while abnormal pixels cannot. Following the original CRD, a series of improved CRD-based methods have also been proposed, e.g., the sparse CRD (SCRD) [19], the weighted CRD (WCRD) [20], and the recursive CRD (RCRD) [21]. Unlike pixel-wise detection in CR-based methods, LR and sparse representation (LRASR) methods [22], [23], [24], [25] focus on exploiting the pixel correlations in HSI and become more attractive. The LR and CRD (LRCRD) method [22] utilizes a nuclear norm and a weighted $l2$-norm to regularize the representation coefficient to combine the global structure and local attributes of HSI. Similarly, the graph and total variation regularized LR representation (GTVLRR) method [25] exploits the graph Laplacian to regularize the representation coefficient to preserve the local geometrical structure and spatial relationships in HSI. Yet most LRASR methods convert 3-D HSI into a 2-D matrix, which destroys the inherent 3-D structure property of HSI. To tackle this problem, several tensor representation-based methods [26], [27], [28], [29] are proposed. The prior-based tensor approximation (PTA) method [26] uses a third-order tensor to preserve the data structure for integrated consideration of all the dimensions. The principal component analysis (PCA)-based tensor LRASR (PCA-TLRSR) method [29] utilizes a 3-D tensor LR model to separate the LR background part from HSI, then detects the anomalies using the rest sparse tensor.

However, most representation-based methods require manual parameter tweaking, which increases the complexity of the algorithm. In addition, the detection results are sensitive to parameter selection and adjustment, leading to uncertainty in the accuracy of the methods.

C. DNN-Based

Different from the preceding two kinds of HAD methods, DNN-based methods [30], [31], [32], [33], [34], [35], [36] aim to distinguish between the background and anomalies according to the error of the self-supervised HSI reconstruction, which is free of prior distribution assumption or manual parameter tweaking. Auto-AD [31] leverages a fully convolutional autoencoder to reconstruct the original HSI from noise input, with the anomalies appearing as the reconstruction errors. WeaklyAD [35] employs a spectral-constrained generative adversarial network (GAN) for HSI reconstruction, where abnormal pixels demonstrate higher reconstruction errors compared with background pixels. Besides the conventional reconstruction error-based methods, some variant methods [37], [38], [39] have been proposed. The deep support vector data description (DSVDD) classifier transforms anomaly detection into a classification problem, which uses the deep features to train a compact hypersphere. At the test phase, once the feature of a sample is outside the hypersphere, the sample is classified as an anomaly. In addition, some methods [40], [41], [42] use the reconstruction networks as feature extractors and then detect anomalies in the deep feature. The LR embedded network (LREN) [42] learns spectral features through the joint training of an auto-encoder and a Gaussian mixture model, followed by an LR representation for generating detection results based on the learned deep feature.

Although the methods above have achieved some positive results, they neglect to utilize the inherent data characteristics of HSI to distinguish between the background and anomalies, which limits their effectiveness in certain instances. In contrast, our proposed GT-HAD focuses on exploring the content similarity in HSI and utilizes the content similarity to process the background and anomalies, thus increasing the discrimination between the background and anomalies.

III. Method

Drawing inspiration from the content similarity in HSI and the exceptional performance of the transformer in capturing such similarity, we present a novel GT-HAD. This section introduces the GT-HAD and illustrates the implementation details.

A. Overall Framework

The anomaly detection process of GT-HAD is divided into two stages, as depicted in Fig. 2. In the training stage, the network is optimized through the mean square error (MSE) loss function, where the HSI cubes are used as training data. Our proposed network is highly lightweight and consists of a newly developed gated transformer block (GTB) and two 3×3 convolution layers. Two feature-modeling branches exist in the GTB, namely, the anomaly-focused branch (AFB) and background-focused branch (BFB), which are used to reconstruct the background and anomalies, respectively. The activation states of AFB and BFB are regulated by a gating unit, with the gating states determined by a CMM. In the detection stage, the reconstruction errors of all HSI cubes are used to generate an anomaly map as the final detection result. The overall process is described as follows.

Given a 3-D HSI data, $X \in \mathbb{R}^{H \times W \times D}$, where H, W, and D represent the height, width, and number of spectral bands of HSI, respectively. We first apply a sliding window to partition X into N overlapped HSI cubes $\mathcal{X} = \{x_i\}_{i=1}^{N}$, with the cube size being $\mathcal{H} \times \mathcal{W} \times D$. The default values for the cube height $\mathcal{H}$ and cube width $\mathcal{W}$ are set to 9, and the sliding stride of the window is set to 3. Subsequently, these HSI cubes are fed into the proposed network to be reconstructed.

First, a 3×3 convolution layer is applied to map HSI cube x_i into feature map $f_i \in \mathbb{R}^{\mathcal{H} \times \mathcal{W} \times C}$, where C is the number of channels and is set to 64 by default. Then, the feature map f_i is fed into GTB to generate an updated feature map $\hat{f}_i \in \mathbb{R}^{\mathcal{H} \times \mathcal{W} \times C}$. The gating unit determines whether the AFB or BFB processes the feature map f_i. Finally, the reconstructed HSI cube $\hat{x}_i \in \mathbb{R}^{\mathcal{H} \times \mathcal{W} \times D}$ can be obtained by applying another 3×3 convolution layer to model the updated feature map $\hat{f}_i$. To optimize the network, we use MSE as the loss function

$$\mathcal{L}(\theta) = \frac{1}{B} \sum_{i=1}^{B} \left\| x_i - \hat{x}_i \right\|_2^2 \tag{1}$$

where θ refers to the learnable parameters in the proposed net and B represents the batch size per training iteration.

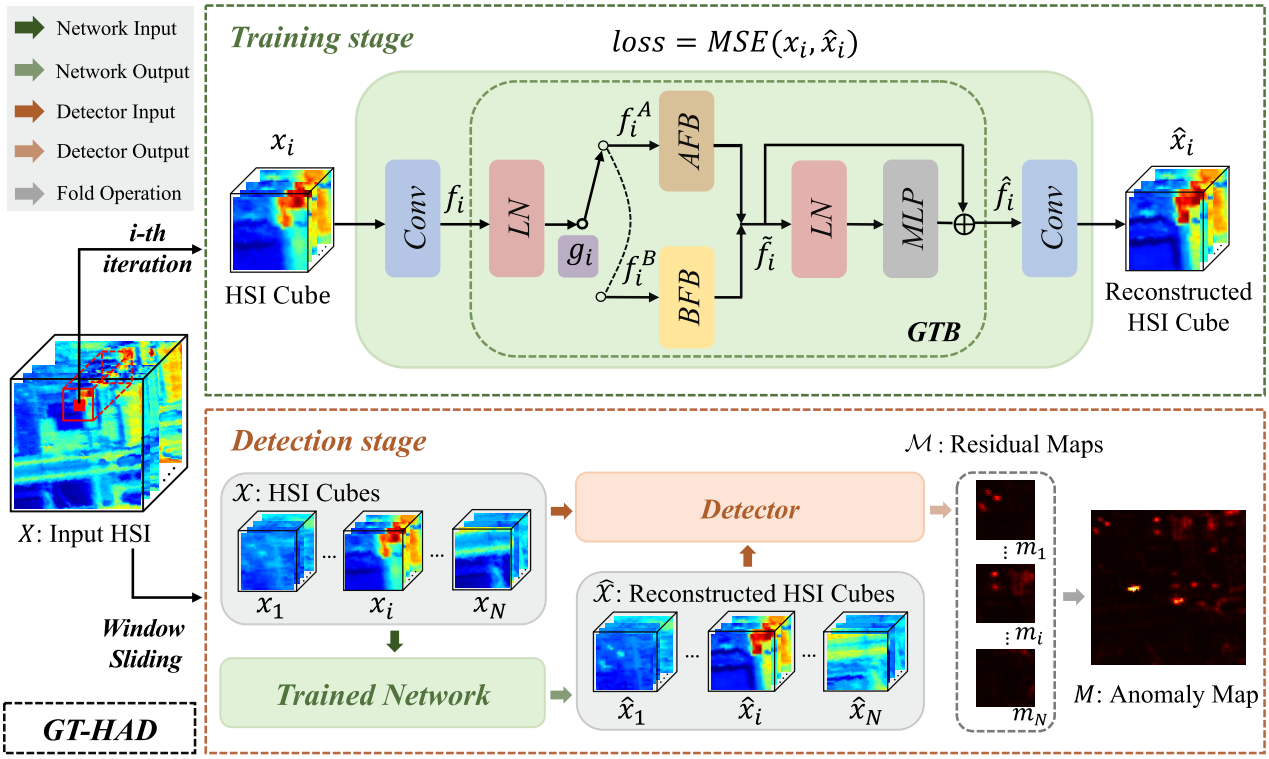

Fig. 2. Framework of GT-HAD. The proposed network is designed based on a transformer-like architecture and trained using numerous HSI cubes. In the detection stage, the reconstruction errors of the HSI cubes are utilized for detecting anomalies.

Once the training is completed, anomaly detection can be performed. We first obtain the residual map $m_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}}$ by measuring the difference between the HSI cube x_i and its corresponding reconstructed one $\hat{x}_i$. Then, all residual maps $\mathcal{M} = \{m_i\}_{i=1}^{N}$ are aggregated and folded to generate the final anomaly map $M \in \mathbb{R}^{H\times W}$.

B. Gated Transformer Block

As shown in Fig. 2, GTB is composed of two components: a gated dual-branch network (GDBN) and a feed-forward network (FFN). GDBN comprises a single layer normalization (LN) and two distinct feature-modeling branches activated/deactivated by a gating unit. For each input feature map f_i of GDBN, we use g_i to symbolize the corresponding gating state. Specifically, if feature map f_i contains anomaly features, g_i equals 0, and AFB is activated. Conversely, if feature map f_i solely comprises background features, g_i equals 1, and BFB is activated. The processing flow of GDBN is represented as

$$\tilde{f}_i = \begin{cases} \mathrm{AFB}(\mathrm{LN}(f_i)), & \text{if } g_i = 0 \\ \mathrm{BFB}(\mathrm{LN}(f_i)), & \text{if } g_i = 1 \end{cases} \tag{2}$$

where feature map $\tilde{f}_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times C}$ is the output of GDBN.

FFN comprises an LN and a multilayer perceptron (MLP), and the skip connection is exploited for bridging the information flow between the head and the tail of FFN. MLP consists of two linear layers with a GELU activation in between. The first linear layer expands the channel dimension by a factor of 2, and the second linear layer restores the channel dimension to the original one. The computation process of FNN is formulated as

$$\hat{f}_i = \mathrm{MLP}\big(\mathrm{LN}\big(\tilde{f}_i\big)\big) + \tilde{f}_i \tag{3}$$

where feature map $\hat{f}_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times C}$ is the final output of GTB.

Next, we introduce the details of AFB and BFB, respectively.

1) Anomaly-Focused Branch: We denote the input feature map of AFB as $f_i^A \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times C}$. The chief goal of AFB is to remove the anomaly features from the feature map f_i^A, thereby suppressing the anomalies' reconstruction.

To achieve this goal, we propose a feature redefinition approach where we divide the feature map f_i^A into several parts and redefine each part by taking a weighted sum of its surrounding parts, as illustrated in Fig. 3(a). Anomalies typically exhibit low similarity with surrounding contents. Consequently, when a specific part contains anomaly features, this part cannot be accurately defined by its surrounding parts, which results in the removal of anomaly features. To better capture the content similarity in HSI, we perform the feature redefinition process at the patch level instead of the pixel level, as the patch contains both spectral and spatial information, which ensures the integrity of the HSI content.

Specifically, we partition the feature map f_i^A into J_A nonoverlapped feature patches $\mathcal{P}_A = \{p_a\}_{a=1}^{J_A}$ with a patch size of $L_A \times C$, where $L_A = h_A w_A$ represents the flattened spatial dimension of the patch. The default values for the patch height h_A and patch width w_A are set to 3. Then, the similarity weight between a feature patch p_a and its adjacent feature patch p_j is calculated as

$$\mathcal{S}(p_a, p_j) = \langle p_a W_A, p_j W_A \rangle \tag{4}$$

where $W_A \in \mathbb{R}^{C\times(C/4)}$ represents a learnable weight matrix and $\langle\cdot,\cdot\rangle$ means the dot-product. Subsequently, the similarity weight $\mathcal{S}(p_a, p_j) \in \mathbb{R}^1$ is collected across the search region $\mathcal{Z}$ and fed into a soft-max function to obtain the correlation

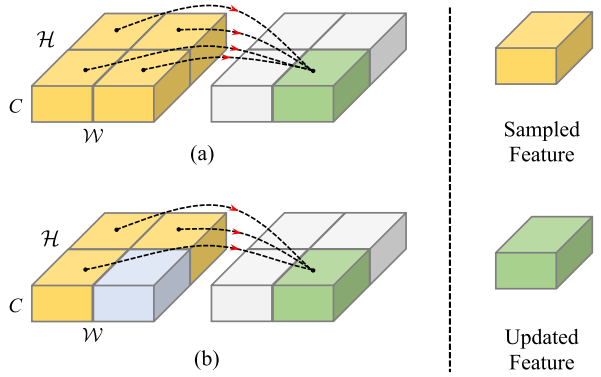

Fig. 3. Methods of modeling the features of AFB and BFB, respectively. (a) For each part in feature map f_i^A, generate a brand new feature for it by taking a weighted sum of its surrounding parts. (b) For each part in feature map f_i^B, generate an enhanced feature for it by aggregating its surrounding parts.

weight as

$$\omega_{aj} = \frac{\exp\left(\mathcal{S}\left(p_a, p_j\right)\right)}{\sum_{z\in\mathcal{Z}} \exp\left(\mathcal{S}\left(p_a, p_z\right)\right)} \tag{5}$$

where $\mathcal{Z} = \{z | 1 \le z \le J_A \text{ and } z \ne a\}$ is the set of indexes of neighboring feature patches of p_a. Then, patch-wise features are aggregated according to the correlation weights

$$\hat{p}_a = \sum_{j\in\mathcal{Z}} \omega_{aj} p_j \tag{6}$$

where $\hat{p}_a \in \mathbb{R}^{L_A\times C}$ represents the redefined feature patch. Finally, the output of AFB is obtained as

$$\tilde{f}_i = \mathcal{I}\left(\left\{\hat{p}_a\right\}_{a=1}^{J_A}\right) \tag{7}$$

where $\mathcal{I}(\cdot)$ represents the reshape and fold operation to combine this set of redefined feature patches into a new feature map $\tilde{f}_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times C}$.

2) Background-Focused Branch: We denote the input feature map of BFB as $f_i^B \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times C}$. The primary objective of BFB is to enhance the background features in the feature map f_i^B, thus strengthening the background reconstruction.

Here, we develop a feature enhancement approach to achieve this objective. Unlike anomalies, the background exhibits high similarity with surrounding contents. Hence, each part of the feature map f_i^B is enhanced by aggregating its surrounding parts, as illustrated in Fig. 3(b). Previous works [49], [50], [51] have shown that leveraging image structure prior is effective for image reconstruction, and we anticipate that it would also be advantageous for background reconstruction. Taking this into consideration, we perform the feature enhancement process at the patch level, as the patch retains structural information.

The entire procedure of BFB is regarded as a patch-level self-attention calculation process. To be specific, we first partition the feature map f_i^B into J_B nonoverlapped feature patches $\mathcal{P}_B = \{p_b\}_{b=1}^{J_B}$ with a patch size of $L_B \times C$, where the size of L_B is equal to L_A and $J_B = J_A$. For the convenience of calculation, we reshape the feature patches $\mathcal{P}_B$ into a tensor format $T_B \in \mathbb{R}^{J_B\times L_B\times C}$. Then, the self-attention computation is restricted to each nonoverlapped feature patch.

The feature patches tensor T_B is first projected into query $Q \in \mathbb{R}^{J_B\times L_B\times(C/4)}$, key $K \in \mathbb{R}^{J_B\times L_B\times(C/4)}$, and value $V \in \mathbb{R}^{J_B\times L_B\times C}$

$$Q = T_B W_B, \quad K = T_B W_B, \quad V = T_B \tag{8}$$

where $W_B \in \mathbb{R}^{C\times(C/4)}$ represents a learnable weight matrix and shares the parameters with W_A. Next, for the dimension matching requirement, Q, K, and V are, respectively, reshaped to $\hat{Q} \in \mathbb{R}^{J_B\times H}$, $\hat{K} \in \mathbb{R}^{J_B\times H}$, and $\hat{V} \in \mathbb{R}^{J_B\times H}$, where $H = (L_B C/4)$. Finally, attention is calculated by the self-attention mechanism

$$\tilde{f}_i = \mathcal{R}\left(\text{Softmax}\left(\hat{Q}\hat{K}^T\right)\hat{V}\right) \tag{9}$$

where $\hat{K}^T$ represents the transpose of matrix $\hat{K}$ and $\mathcal{R}(\cdot)$ represents the dimension reshape operation.

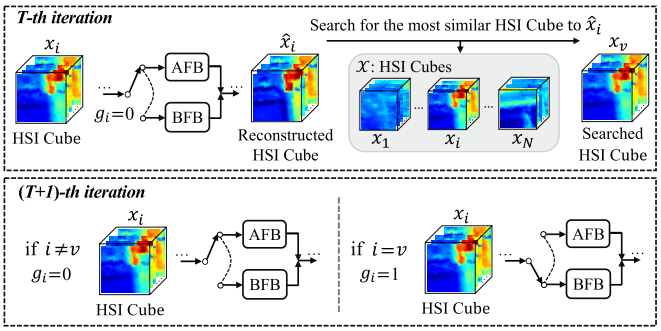

Fig. 4. Illustration of CMM procedure. During training, the activation states of AFB and BFB are controlled by the gating unit, whose operational states are determined by CMM.

C. Content-Matching Method

In the process of training, the activation states of AFB and BFB are regulated by the gating unit, whose control states are determined by a CMM. This section provides a detailed explanation of CMM.

To record the gating states during feature maps $\{f_i\}_{i=1}^N$ processing in GTB, we utilize the notation $\mathcal{G} = [g_1, \ldots, g_i, \ldots, g_N]$. Initially, all elements in gating states list $\mathcal{G}$ are initialized to zero, which indicates that all feature maps $\{f_i\}_{i=1}^N$ are fed into AFB at the start of the training. When feature map f_i solely comprises background features and passes through the AFB, its primary contents can be preserved since each feature patch in the feature map f_i can be approximately defined by its adjacent feature patches. As a result, the corresponding reconstructed HSI cube $\hat{x}_i$ exhibits good reconstruction quality. Conversely, if feature map f_i contains anomaly features, such features will be removed when the former passes through the AFB, which results in poor reconstruction quality of the reconstructed HSI cube $\hat{x}_i$, due to the loss of contents. The CMM determines the presence of anomaly features in the feature map f_i by evaluating the content matching degree between the HSI cube x_i and the reconstructed one $\hat{x}_i$ and then sets the gating states.

The specific CMM procedure is illustrated in Fig. 4. The CMM searches for the most similar one to the reconstructed HSI cube $\hat{x}_i$ within the HSI cubes set $\{x_i\}_{i=1}^N$. If the searched HSI cube x_v and the reconstructed one $\hat{x}_i$ have identical indexes, the reconstructed one $\hat{x}_i$ is deemed to have no contents loss and has consistent contents with the original HSI cube x_i. It implies that the feature map f_i solely comprises background features and should be fed into the BFB in the next training iteration, with the gating state g_i transitioning from 0 to 1.

Fig. 5. Compare the residual responses before and after using RD. (a) Airport scene. (b) Abnormal target. (c) Residual response before using RD. (d) Residual response after using RD.

Conversely, if the indexes are distinct, the reconstructed HSI cube $\hat{x_i}$ is deemed to have a content loss and exhibit a low content matching degree with the original one x_i. It implies that the feature map f_i contains anomaly features and should continue to pass through the AFB in the next training iteration, with the gating state g_i remaining at 0. Notably, the Euclidean distance is used to estimate the similarities between the reconstructed HSI cube $\hat{x_i}$ and the HSI cubes $\{x_i\}_{i=1}^N$, when searching for the most similar HSI cube pair.

As the training progresses, CMM gradually updates the values of the elements in gating states list $\mathcal{G}$. Then, an increasing number of feature maps solely comprising background features are directed to BFB and separated from AFB to focus on the background reconstruction.

D. Anomaly Detection

After completing the training, the reconstruction errors of all image pixels are used to generate the anomaly map. The residual cube is denoted as $r_i = (x_i - \hat{x_i})^2 \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times D}$, and then we sum the residual cube r_i along the band dimension to obtain the residual map $m_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}}$. Finally, the anomaly map is generated as

$$M = \mathcal{F}\left(\{m_i\}_{i=1}^N\right) \tag{10}$$

where $\mathcal{F}(\cdot)$ represents the fold operation to combine all the residual maps into the anomaly map $M \in \mathbb{R}^{H\times W}$. Since overlaps exist between the residual maps, we use the average operation to deal with the overlapped areas. This strategy can suppress the blocking effect [52] in the final output and establish the information interaction between residual maps.

However, the pixel mixing phenomenon is obvious at the boundaries between the background and anomalies, which decreases the spectral differences between the background and anomalies. Therefore, it is challenging to detect the anomalies near the boundaries, which results in low residual responses, as shown in Fig. 5(c). We use a residual postprocessing method called residual diffusion (RD) to address this issue.

For a voxel in the residual cube r_i, we recalculate its value, i.e., averaging the residuals within the 3-D space centered around this voxel. In this way, within the region of abnormal targets, high residuals diffuse toward the areas with low residuals, which increases the residual response in those areas, as shown in Fig. 5(d). To implement the RD, a 3-D average pooling filter with a kernel size of $3 \times 3 \times 5$ is utilized

$$e_i = \text{AvgPool3d}(r_i) \tag{11}$$

where AvgPool3d represents the 3-D average pooling. Then, we sum the recalculated residual cube $e_i \in \mathbb{R}^{\mathcal{H}\times\mathcal{W}\times D}$ along the band dimension to obtain the residual map m_i.

The proposed GT-HAD method is described in detail in Algorithm 1.

Algorithm 1 GT-HAD

Input: HSI cubes $\mathcal{X} = \{x_i\}_{i=1}^N$
Initialize: gating states list $\mathcal{G} = [g_1, \ldots, g_i, \ldots, g_N] = [0, \ldots, 0, \ldots, 0]$
Training of the network:
while training process **do**
1: net forward using (2) - (9);
2: update $\mathcal{G}$;
3: calculate the loss function using (1);
4: net backward;
end while
Output: obtain the anomaly map using (10) - (11);

IV. Experiments and Analysis

In this section, we compare the performance of GT-HAD with ten existing HAD methods on six real HSI datasets. The results indicate that GT-HAD outperforms the existing methods significantly.

A. Datasets Description

Here we introduce the six HSI datasets and these datasets are gathered and available in our project.[1]

1) Los Angeles-1 Dataset: This dataset is acquired using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the region of Los Angeles. It comprises an image of dimensions 100 × 100 pixels, with a spatial resolution of 7.1 m/pixel and 205 spectral bands. The dataset exhibits 144 abnormal pixels originating from several airplanes, which account for 1.44% of the total number of pixels.

2) Los Angeles-2 Dataset: The second dataset is also captured by the AVIRIS sensor from Los Angeles. There exist two airplanes with 87 abnormal pixels to be detected in this image, constituting 0.87% of the total number of pixels. It consists of 205 spectral bands, with a spatial size of 100 × 100 and a spatial resolution of 7.1 m/pixel.

3) Gulfport Dataset: The third dataset is acquired utilizing the AVIRIS sensor from Gulfport, which features a spatial resolution of 3.4 m/pixel. The image has a spatial extent of 100 × 100 and encompasses 191 spectral bands, after excluding unsuitable bands. The anomalies come from three airplanes in various scales with 60 abnormal pixels, which occupy 0.60% of the total number of pixels.

4) Texas Coast Dataset: The fourth dataset is captured by the AVIRIS sensor over the Texas Coast region. It comprises 100 × 100 pixels with a spatial resolution of 17.2 m/pixel and contains 204 spectral bands. The dataset contains several houses exhibiting abnormal characteristics, with 67 abnormal pixels accounting for 0.67% of the total number of pixels.

5) Cat Island Dataset: The fifth dataset pertains to the Cat Island vicinity, which exhibits a spatial resolution of 17.2 m/pixel and a spatial resolution of 150 × 150. Following the exclusion of the noisy bands, 188 spectral bands are retained for analysis. Anomalies consist of a single airplane with 19 abnormal pixels, which account for 0.08% of the total number of pixels. Similar to the previous datasets, the AVIRIS sensor is employed for data acquisition.

6) Pavia Dataset: The final dataset, Pavia, is acquired using the reflective optics system imaging spectrometer (ROSIS) sensor. The image has a spatial extent of 150 × 150 and

[1] https://github.com/jeline0110/GT-HAD

TABLE I
OPTIMAL PARAMETERS OF TEN HAD METHODS ON SIX DATASETS

Methods	Parameters	Los Angeles-1	Los Angeles-2	Gulfport	Texas Coast	Cat Island	Pavia
RX	/	/	/	/	/	/	/
KIFD	ζ	300	300	300	300	300	100
2S-GLRT	(w_{out}, w_{in})	(9, 7)	(19, 15)	(13, 11)	(9, 5)	(25, 3)	(21, 5)
MsRFQFT	(N_d, σ)	(3, 1.2)	(3, 4.0)	(3, 10)	(3, 1.4)	(3, 0.4)	(3, 1.6)
CRD	$(w_{out}, w_{in}, \lambda)$	(15, 7, 10^{-6})	(17, 15, 10^{-6})	(19, 15, 10^{-6})	(9, 7, 10^{-6})	(25, 15, 10^{-6})	(7, 5, 10^{-6})
GTVLRR	(λ, β, γ)	(0.5, 0.2, 0.05)	(0.5, 0.2, 0.05)	(0.5, 0.2, 0.05)	(0.05, 0.2, 0.02)	(0.05, 0.2, 0.02)	(0.05, 0.2, 0.02)
PTA	(μ, r)	(0.01, 10)	(0.01, 10)	(0.01, 10)	(0.1, 0)	(0.001, 15)	(0.0001, 15)
PCA-TLRSR	(d, λ, λ')	(4, 0.06, 0.01)	(5, 0.06, 0.01)	(17, 0.06, 0.05)	(15, 0.06, 0.01)	(15, 0.06, 0.05)	(4, 0.05, 0.01)
Auto-AD	σ	1.0×10^{-5}	1.5×10^{-5}	1.2×10^{-5}	1.0×10^{-5}	1.0×10^{-5}	1.0×10^{-5}
LREN	λ	1.0	0.1	0.001	1.0	1.0	1.0

exhibits a spatial resolution of 1.3 m/pixel, along with 102 spectral bands, following the exclusion of unsuitable bands. The primary anomalies are several vehicles on the bridge, encompassing 68 abnormal pixels, which correspond to 0.30% of the total number of pixels.

B. Experiment Setup

1) Comparison Methods: We compare GT-HAD with ten existing HAD methods, including RX [6], KIFD [14], 2S-GLRT [10], MsRFQFT [17], CRD [18], GTVLRR [25], PTA [26], PCA-TLRSR [29], LREN [42], and Auto-AD [31]. RX, KIFD, 2S-GLRT, and MsRFQFT are statistics-based methods, CRD, GTVLRR, PTA, and PCA-TLRSR are representation-based methods, while LREN and Auto-AD are DNN-based methods. Among these methods, RX and CRD are regarded as the most classical methods for the HAD task, while the other methods have also gained widespread attention in recent literature and have shown competitive performance.

We empirically tune the parameters of all compared methods and select the optimal parameters for comparative experiments. Specifically, in RX, no parameters that need to be tuned. In KIFD, the first ζ principal components of HSI are used as input, and we vary its value from 50 to 300 to determine the optimal detection performance. For 2S-GLRT, we investigate the detection performance under various outer window sizes w_{out} ranging from 5 to 25 and inner window sizes w_{in} ranging from 3 to 15, then the best detection performance under specific window sizes (w_{out}, w_{in}) is collected. In MsRFQFT, the selected band number N_d is fixed at 3, and the tuning range of parameter σ is set to 0.2–10 with the step of 0.2. We tune the parameters (N_d, σ) for each dataset individually to achieve the best performance. Similar to 2S-GLRT, for CRD, we first fix λ at 10^{-6}, then vary w_{out} and w_{in} from 5 to 25 and from 3 to 15, respectively. Finally, the optimal detection performance is achieved using specific window sizes (w_{out}, w_{in}). In GTVLRR, the dictionary is constructed with a category set of 6, and the largest 20 elements in each category are selected as the atoms of the dictionary. Then, the parameters (λ, β, γ) are tuned based on the suggestions in the original article [25]. For PTA, parameters α, τ, and β are set to 1, and parameter μ and truncated LR r are set in the range $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$ and range 0–40, respectively. The optimum detection performance is determined by employing a specific parameter set (μ, r). In PCA-TLRSR, the first d principal components of HSI are used as input, and d is varied from 1 to 20. Meanwhile, parameters λ and λ' are selected from the set {0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5}. Finally, the best detection performance is determined based on a specific parameter list (d, λ, λ'). In Auto-AD, the training process stops when the average variation of the loss falls below a certain threshold σ, which is varied from 1.0×10^{-5} to 1.5×10^{-5} to achieve the optimal detection performance. In LREN, the coefficient λ in the optimization problem requires tuning. The optimal values of λ are obtained for different datasets by varying its value from 10^{-6} to 1. Table I summarizes the optimal parameters of ten HAD methods on six datasets.

For the training of GT-HAD, we apply the adaptive moment estimation (ADAM) to optimize the model. The training process consists of 150 epochs, with a batch size of 64 and a learning rate of 0.001. Besides, all DNN-based methods, including GT-HAD, are implemented on an NVIDIA 3090 GPU. In terms of deep learning framework, GT-HAD and Auto-AD use Pytorch 1.7.0, while LREN utilizes Tensorflow 1.14.0. On the other hand, all non-DNN methods are implemented using MATLAB 2020a on a computer with an Intel i7-10700 CPU and 16 GB of RAM.

2) Evaluation Metrics: We utilize qualitative and quantitative evaluation metrics to evaluate the detection performance of the compared HAD methods. The qualitative evaluation metrics include color anomaly map, box-whisker plot [53], 2-D and 3-D receiver operating characteristic (ROC) curves [54], and the quantitative evaluation metric uses the area under the 2-D ROC curve (AUC). The color anomaly map provides insight into the detection performance of the methods on both background suppression and anomaly response, while the box-whisker plot assesses the separability between the background and anomalies. In addition, 2-D and 3-D ROC curves illustrate the relationship between the probability of detection (PD) and FAR at different thresholds (τ). PD denotes the number of correctly detected abnormal pixels among all the abnormal pixels, while FAR is the number of pixels that are marked as abnormal targets among all the background pixels. A 2-D ROC curve that approaches the upper-left corner of the axis indicates a better detection performance and a larger AUC. The ideal AUC is 1, signifying that all abnormal targets are detected with no false alarms.

C. Comparison Results

In this section, we provide a unified analysis of the detection performance of 11 methods on six datasets.

The color anomaly maps of different methods on each dataset are presented in Fig. 6. We can see that GT-HAD achieves a balance between background suppression and anomaly response on the Los Angeles-1, Los Angeles-2, Gulfport, and Pavia datasets. Taking Los Angeles-1 as an

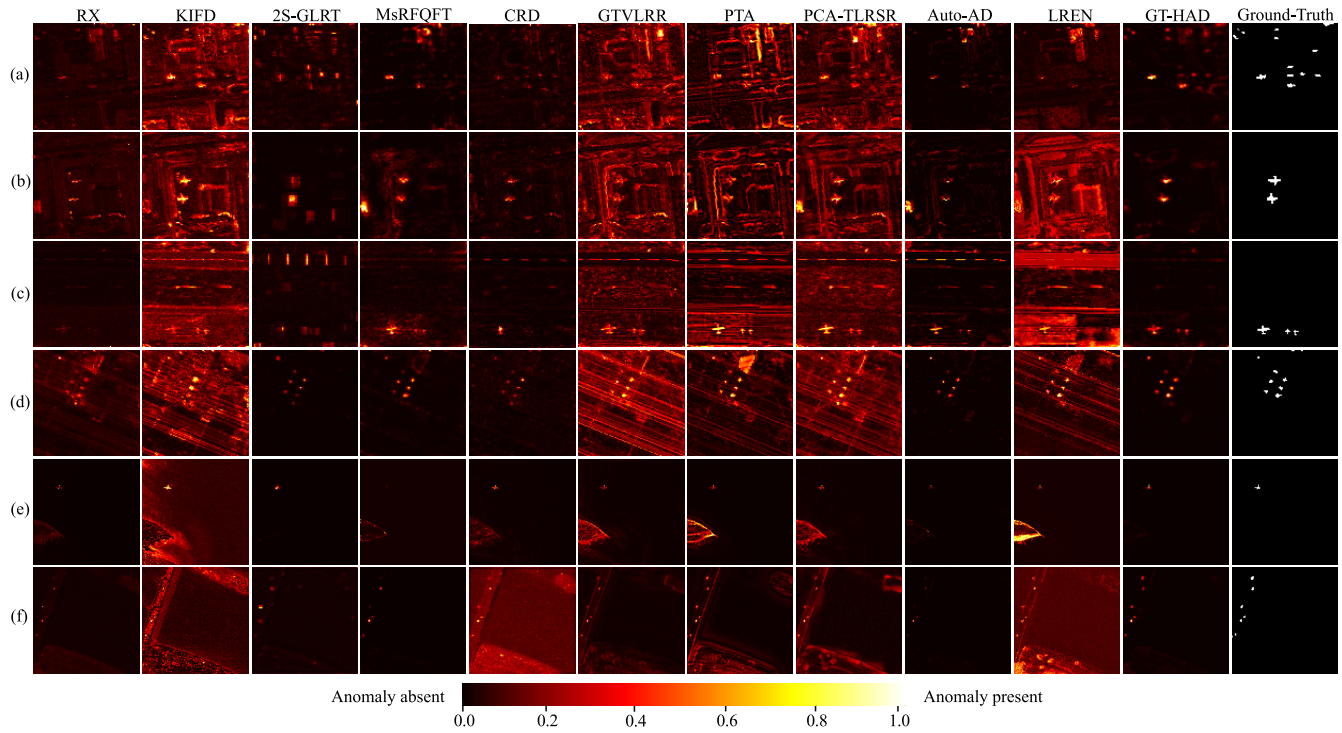

Fig. 6. Color anomaly maps of different HAD methods on six datasets. (a) Los Angeles-1. (b) Los Angeles-2. (c) Gulfport. (d) Texas Coast. (e) Cat Island. (f) Pavia.

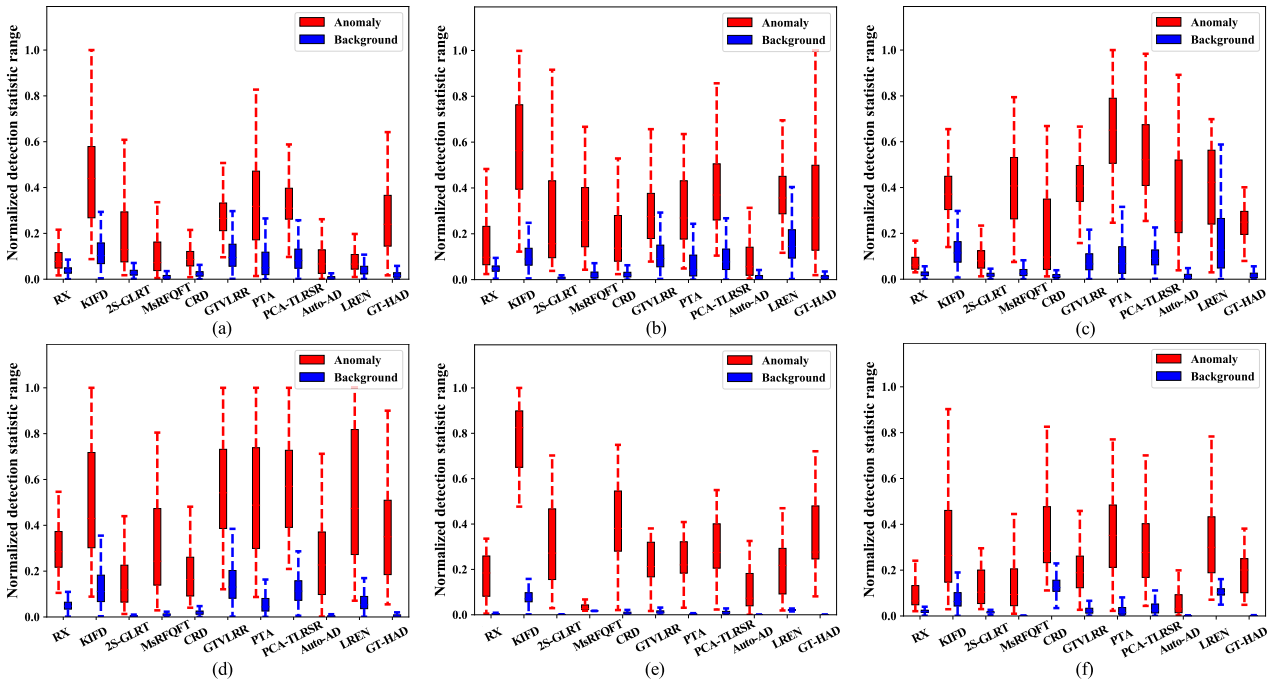

Fig. 7. Box-whisker plots of different methods on six datasets. (a) Los Angeles-1. (b) Los Angeles-2. (c) Gulfport. (d) Texas Coast. (e) Cat Island. (f) Pavia.

example, while KIFD, GTVLRR, and PCA-TLRSR can highlight the anomalies to some extent, they generate a high number of false alarms in the background region. On the other hand, although RX, 2S-GLRT, CRD, Auto-AD, and LREN exhibit lower residual responses in the background region, they miss most anomalies. Conversely, GT-HAD can locate all anomalies effectively and suppress the background with high accuracy. On the Texas Coast dataset, 2S-GLRT, CRD, Auto-AD, and GT-HAD closely resemble the ground-truth map, but the other three methods exhibit limited responses in the abnormal region compared with GT-HAD. Of the 11 methods evaluated across six datasets, only 2S-GLRT on the Cat Island dataset performs similar performance to GT-HAD. These color anomaly maps provide compelling evidence that GT-HAD achieves effective background reconstruction while successfully suppressing the anomaly reconstruction.

Fig. 7 shows the box-whisker plots achieved by 11 methods on six datasets. The box-whisker plots show the abnormal

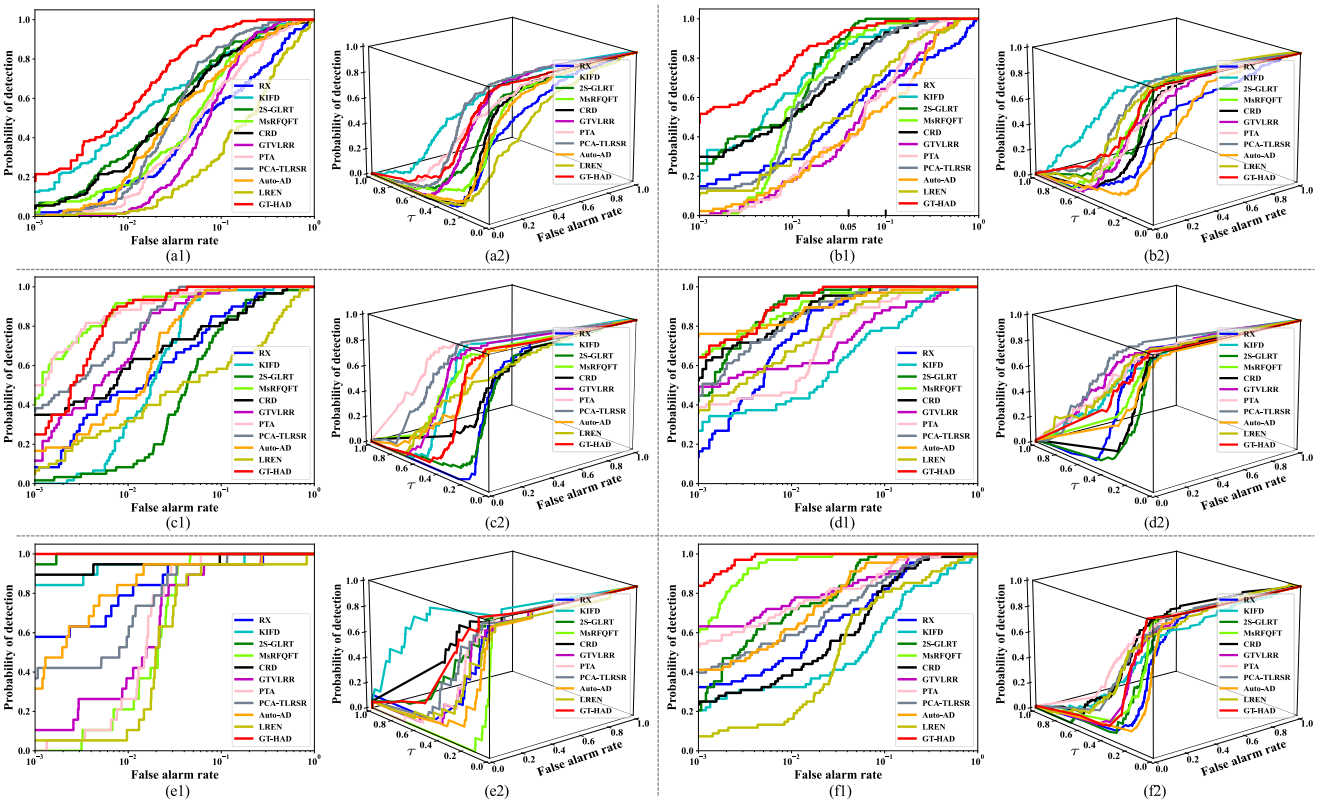

Fig. 8. Two-dimensional and 3-D ROC curves of different methods on six datasets. (a) Los Angeles-1. (b) Los Angeles-2. (c) Gulfport. (d) Texas Coast. (e) Cat Island. (f) Pavia. (a1)–(f1) Two-dimensional ROC curves. (a2)–(f2) Three-dimensional ROC curves.

TABLE II
AUC SCORES OF ELEVEN HAD METHODS ON SIX DATASETS

Datasets	RX	KIFD	2S-GLRT	MsRFQFT	CRD	GTVLRR	PTA	PCA-TLRSR	Auto-AD	LREN	GT-HAD
Los Angeles-1	0.8221	0.9359	0.9236	0.9114	0.9212	0.9004	0.8809	0.9455	0.9087	0.7327	**0.9775**
Los Angeles-2	0.8404	0.9775	0.9873	0.9798	0.9681	0.8840	0.9104	0.9664	0.8626	0.9134	**0.9901**
Gulfport	0.9526	0.9728	0.9183	0.9945	0.9445	0.9874	0.9946	0.9930	0.9810	0.8293	**0.9950**
Texas Coast	0.9907	0.9303	0.9970	0.9953	0.9940	0.9478	0.9757	0.9923	0.9854	0.9783	**0.9976**
Cat Island	0.9807	0.9900	0.9995	0.9798	0.9946	0.9694	0.9831	0.9854	0.9826	0.9343	**0.9996**
Pavia	0.9538	0.8634	0.9867	0.9984	0.9407	0.9747	0.9749	0.9635	0.9822	0.8985	**0.9994**
Average	0.9233	0.9449	0.9687	0.9765	0.9605	0.9439	0.9532	0.9743	0.9407	0.8807	**0.9932**

pixel distributions in red boxes and the background pixel distributions in blue boxes, with the distance between red and blue boxes indicating the background anomaly separability. Especially, the height of the blue box reflects whether the background is well suppressed. As shown In Fig. 7, we can see that GT-HAD balances the background-anomaly separability and background suppression degree on all six datasets. Using Los Angeles-2 as an example, although KIFD has better background-anomaly separability than other methods, it cannot suppress the background nicely. On the other hand, while 2S-GLRT has better background suppression performance than other methods, it cannot separate the background and anomalies well. GT-HAD, by contrast, performs well on both the background-anomaly separability and background suppression degree. In conclusion, the excellence of GT-HAD in enhancing background reconstruction and suppressing anomalies reconstruction provides satisfactory background-anomaly separability and background suppression degree.

The 2-D and 3-D ROC curves obtained by 11 methods on six datasets are illustrated in Fig. 8. We can see that the

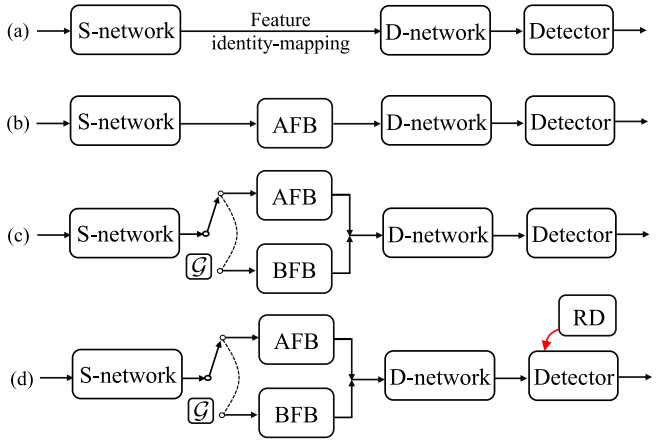

Fig. 9. Network structures of (a) base model, (b) base + AFB model, (c) base + AFB + BFB model, and (d) base + AFB + BFB + RD model, respectively. For simplicity, S- and D-network represent the shallow and deep network layers of the model. $\mathcal{G}$ stores the gating states.

TABLE III
ABLATION STUDY OF GT-HAD ON SIX DATASETS

Model	Los Angeles-1	Los Angeles-2	Gulfport	Texas Coast	Cat Island	Pavia
base	0.8511	0.9021	0.9519	0.9609	0.9653	0.9614
base+AFB	0.9364	0.9399	0.9793	0.9773	0.9815	0.9764
base+AFB+BFB	0.9551	0.9724	0.9900	0.9933	0.9962	0.9951
base+AFB+BFB+RD	**0.9775**	**0.9901**	**0.9950**	**0.9976**	**0.9996**	**0.9994**

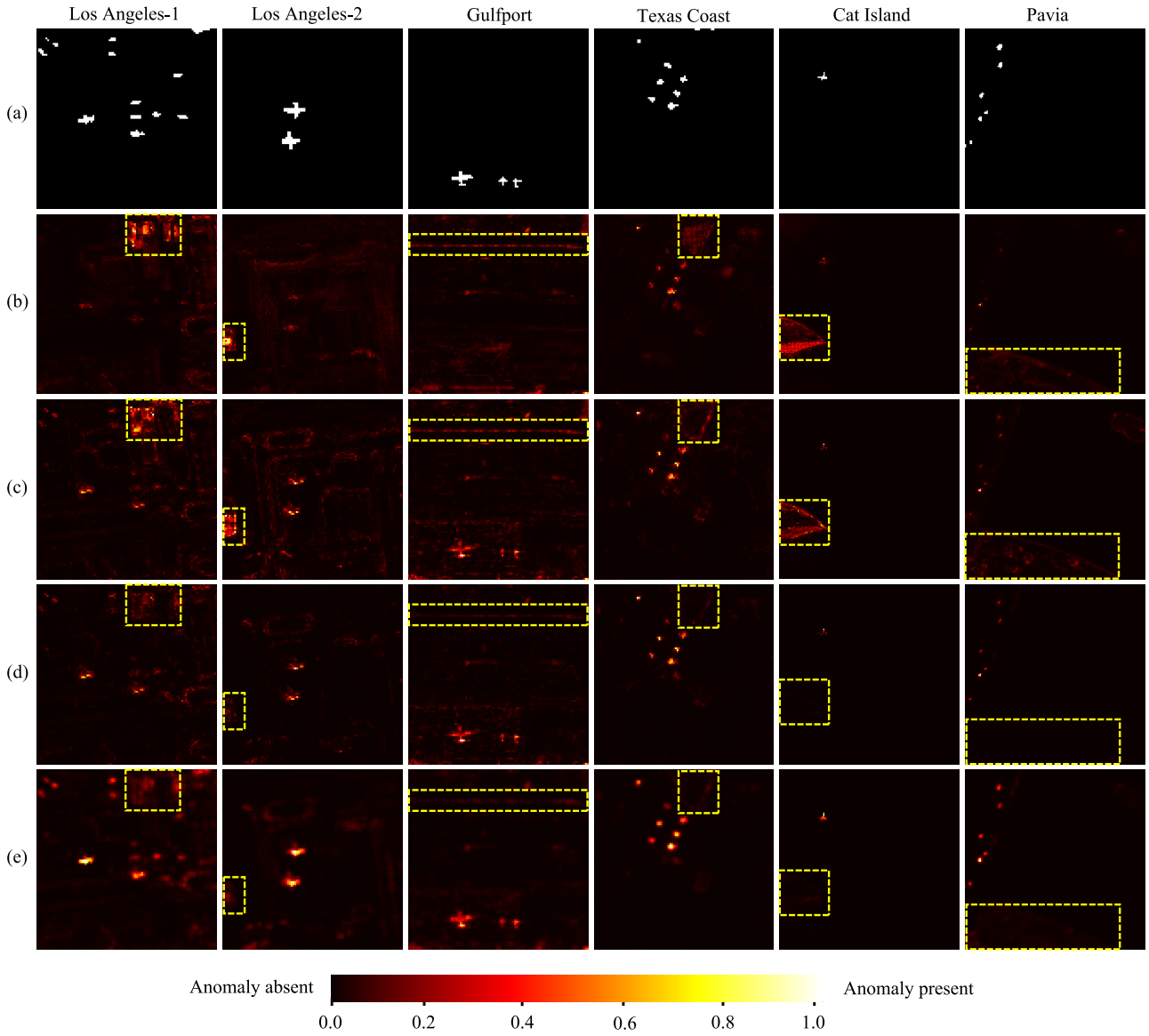

Fig. 10. Ablation visualization results on six datasets. Row 1 represents the (a) ground-truth maps of the six datasets. Rows 2–5 represent the corresponding anomaly maps for (b) base model, (c) base + AFB model, (d) base + AFB + BFB model, and (e) base + AFB + BFB + RD model, respectively. The yellow boxes visible in (b)–(e) locate the background regions that we focus on.

2-D ROC curves of GT-HAD are positioned closest to the upper-left corner of the axis on all the datasets. For example, on the Los Angeles-2 dataset, although 2S-GLRT exhibits slightly higher PDs than GT-HAD when FAR values range from 0.05 to 0.1, GT-HAD shows higher PDs in terms of the overall curve comparison and is closer to the upper-left corner of the axis. The AUC scores of all evaluated methods across six datasets are presented in Table II, with the highest AUC score highlighted in bold for each row. It can be seen that GT-HAD outperforms all other methods on the six datasets in AUC scores. Furthermore, by analyzing the 3-D ROC curves, we can observe that GT-HAD successfully maintains a well-balanced relationship between the PD and FAR at various thresholds on all six datasets. The presented results not only confirm the robustness of GT-HAD across various scenarios but also intuitively present the performance of GT-HAD under different threshold settings. In a word, GT-HAD excels in both 2-D and 3-D ROC analyses, consistently

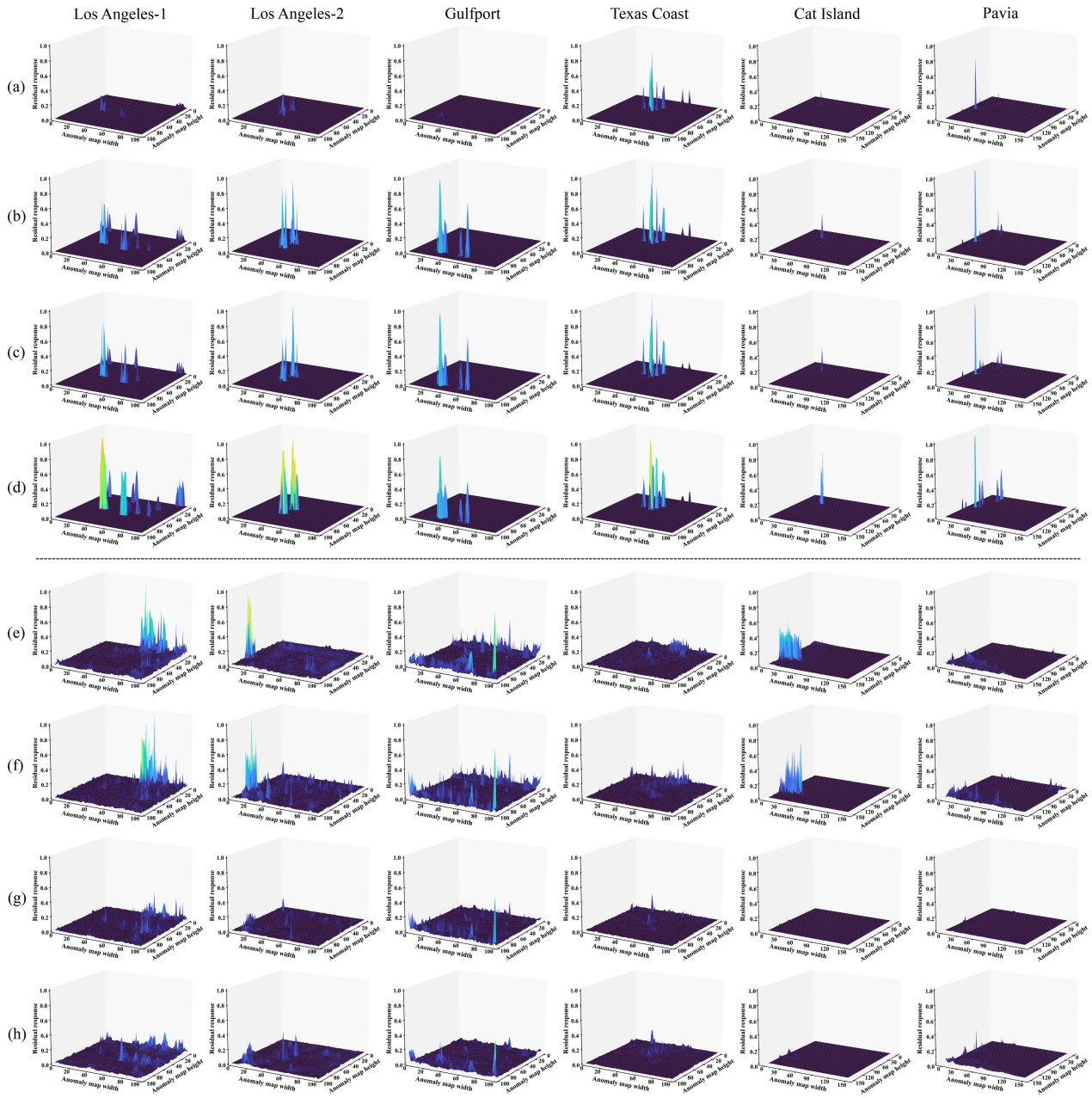

Fig. 11. Three-dimensional residual response maps on six datasets. Rows 1–4 represent the residual responses of **abnormal regions** of (a) base model, (b) base + AFB model, (c) base + AFB + BFB model, and (d) base + AFB + BFB + RD model, respectively. Rows 5–8 represent the residual responses of **background regions** of (e) base model, (f) base + AFB model, (g) base + AFB + BFB model, and (h) base + AFB + BFB + RD model, respectively.

achieving superior AUC scores across all evaluated datasets. Indeed, the integrated capabilities of enhancing background reconstruction and suppressing anomalies reconstruction in GT-HAD ensure the balance between PD and FAR at various thresholds.

D. Ablation Study

This section conducts an ablation study on six datasets to evaluate the impact of AFB, BFB, and RD. First, we establish a base model by replacing AFB and BFB in the GT-HAD with a feature identity-mapping operation while deactivating RD. Then, we progressively integrate AFB, BFB, and RD into the base model. To facilitate a clear discussion, we refer to these models as the base model, base + AFB model, base + AFB + BFB model, and base + AFB + BFB + RD model. Fig. 9 illustrates the corresponding network architectures. All models initialize the same training epochs, batch sizes, and learning rates for a fair comparison. Finally, ablation quantitative and visualization results are reported in Table III and Fig. 10. To better illustrate the detection results, we also present the 3-D residual response maps in Fig. 11, where the residual responses of background and anomalies are intentionally separated for display.

1) Effectiveness of AFB: In Fig. 10(b), the color anomaly map generated by the base model illustrates that the conventional DNN model faces challenges in reconstructing the background with complex data distribution (yellow boxes) and is prone to over-fitting the anomalies with simple data

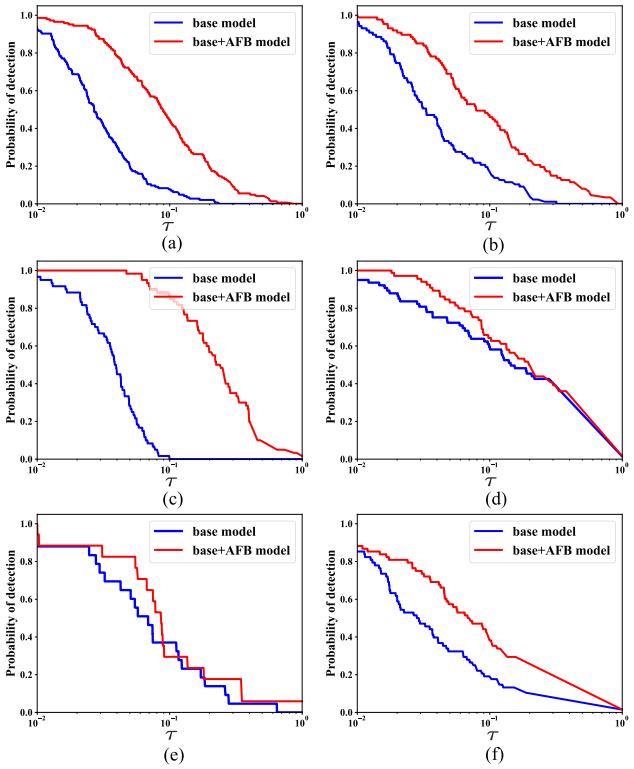

Fig. 12. Two-dimensional ROC curves (PD, τ) of base model and base + AFB model on six datasets. (a) Los Angeles-1. (b) Los Angeles-2. (c) Gulfport. (d) Texas Coast. (e) Cat Island. (f) Pavia.

distribution, which aligns with our previous analysis. We decompose that into two subproblems, i.e., difficulty in reconstructing the complex background and susceptibility to over-fitting the simple anomalies. AFB is exactly designed to address the latter subproblem.

The results in Fig. 10(c) clearly show that the base + AFB model can effectively highlight abnormal regions and generate higher residual responses compared with the base model. These visual results are more intuitive in the 3-D residual response maps presented in Fig. 11(a) and (b). It is worth noting that the base + AFB model exhibits similar background suppression performance to the base model, as shown in Figs. 10(b) and (c) and 11(e) and (f). It suggests that even though AFB is designed for extracting anomalies, it can also effectively suppress the simple background. However, it may not be as effective in suppressing the complex background. Finally, the base + AFB model achieves higher PDs than the base model at various thresholds, which leads to a significant improvement in AUC score, as shown in Fig. 12 and the first two rows of Table III. These experimental results sufficiently validate the effectiveness of AFB in extracting the anomalies.

2) Effectiveness of BFB: The inability of AFB to effectively suppress the complex background promotes the proposal of BFB. By leveraging CMM, BFB is activated and focuses on reconstructing the regions marked by CMM. Given that CMM is the foundation of BFB, conducting a thorough analysis of the marking performance of CMM is crucial. The marking results of CMM are visually presented in Fig. 13(b), with the red areas indicating the marked background pixels.

CMM assesses the presence of background pixels in the HSI cube by gauging whether the HSI cube exhibits the highest similarity to itself after reconstruction. However, in certain smooth regions, adjacent HSI cubes exhibit an exceedingly high degree of similarity, as observed in the grassland and water surface regions of Fig. 13. Hence, when CMM searches for the most similar HSI cube to the reconstructed HSI cube, it is prone to misidentifying the neighboring cube of the target cube as the search result. Ultimately, the pixels within these cubes fail to be marked as background pixels. Although not marked as background pixels, these pixels exhibit high similarity with the surrounding content. Hence, these pixels can be well represented by the surrounding contents, which allows for effective reconstruction without adversely affecting the detection results. This indicates that our method exhibits a high level of fault tolerance. Table IV reports the quantitative measures of pixel overlap between the marked background pixels and the abnormal pixels. These results demonstrate that CMM performs excellently in marking the background and can effectively circumvent most anomalies. This characteristic is particularly advantageous since it ensures that the constraints imposed by AFB on anomalies remain unaltered by BFB.

Fig. 10(c) and (d) depicts the color anomaly maps of the base + AFB model and base + AFB + BFB model, respectively. In addition, Fig. 11(b), (c), (f), and (g) depicts the 3-D residual responses of anomalies and background for the base + AFB model and base + AFB + BFB model, respectively. Based on these visualization results, two conclusions can be drawn. First, BFB outperforms AFB in terms of suppressing the complex background. Second, BFB has minimal negative effects on anomalies, which enables the base + AFB + BFB model to highlight abnormal regions similar to the base + AFB model effectively. The AUC scores of the base + AFB model and base + AFB + BFB model are reported in the second to third rows of Table III. Moreover, Table V reports the increase rate (IR) in the running time of the model after using CMM on each dataset. It can be observed that the running time of the base + AFB + BFB model increases by only 10% compared with the base + AFB model when CMM is used, yet the AUC score improves considerably. The comprehensive analyses provide evidence that BFB exhibits superior performance in suppressing the complex background and that its core module, CMM, is highly effective in marking the background.

3) Effectiveness of RD: Figs. 10(d) and (e) and 11(c), (d), (g), and (h) show that RD can effectively raise the residual responses of the abnormal region while leaving that of the background region nearly unchanged. The fundamental reason is that BFB has effectively suppressed the background region, thereby minimizing the negative impact of the RD on the background region. The residuals flow within the abnormal region and merge, thus increasing the residual response contrast between the abnormal and background regions. Moreover, the RD effect brings the shape of the anomaly response region closer to the actual shape of the abnormal target. Fig. 14 gives the box-wisher plots of the base + AFB + BFB and base + AFB + BFB + RD models. We can see that the background-anomaly separability of the base + AFB + BFB + RD model is larger than that of the base + AFB + BFB model. It indicates that the positive impact of RD on the residual response of the abnormal region is much greater than its negative impact on that of the background region. Finally, the base + AFB + BFB + RD model obtains further improvements in AUC scores than the base + AFB + BFB model, as reported in the third to fourth rows of Table III. Sufficient experimental results demonstrate the efficacy of RD in boosting the detection performance.

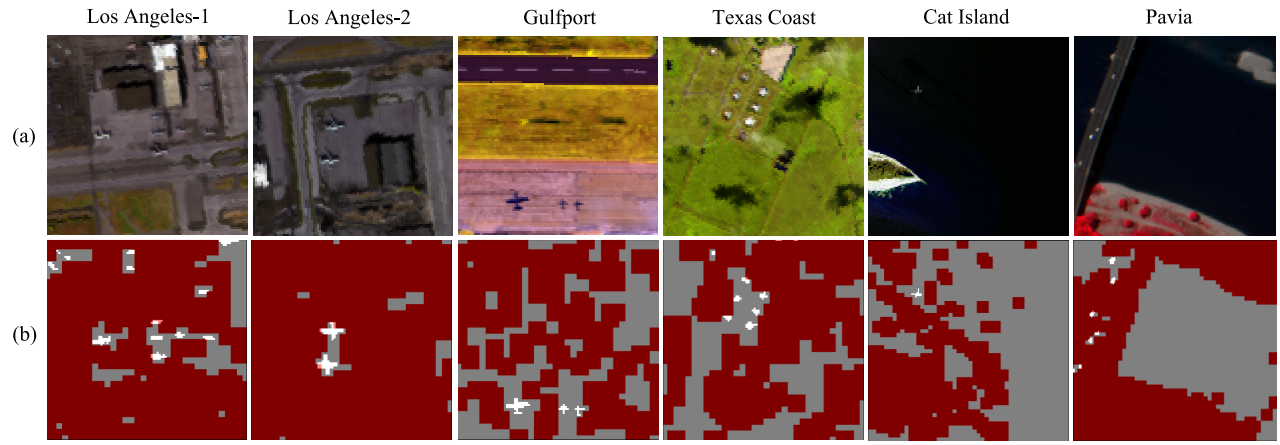

Fig. 13. Background marking results of CMM on six datasets. (a) Display the false-color maps of the six datasets. (b) Display the marking results of CMM, where the red areas indicate the marked background pixels and the white areas indicate the abnormal pixels.

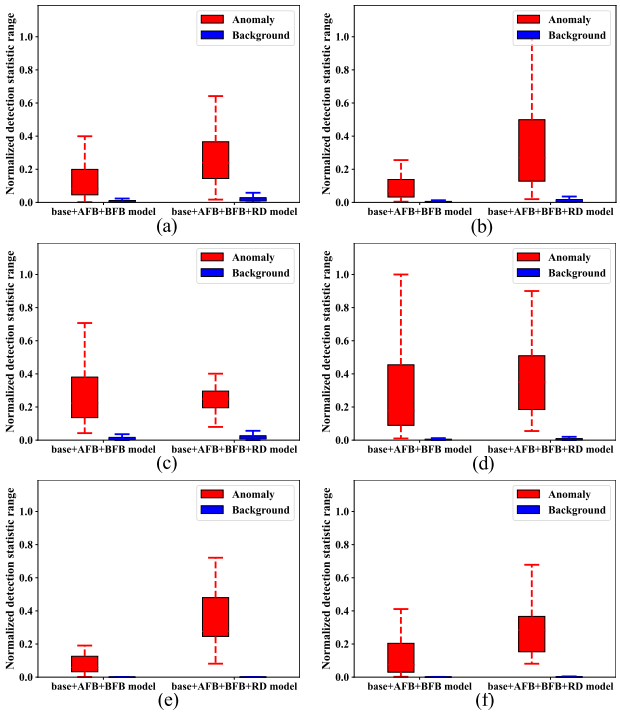

Fig. 14. Box-wisher plots of base + AFB + BFB and base + AFB + BFB + RD models on six datasets. (a) Los Angeles-1. (b) Los Angeles-2. (c) Gulfport. (d) Texas Coast. (e) Cat Island. (f) Pavia.

E. Model Generalization

In this section, we conduct a cross-validation experiment to assess the model's generalization. The number of layer channels in the network is determined by the input HSI data band number. Hence, of the six HSI data, only Los Angeles-1 and Los Angeles-2 are suitable for the cross-validation experiment since these two datasets have the same band number. The experimental results are shown in Table VI. The stability in AUC scores across different testing datasets highlights the generalization behavior of GT-HAD.

F. Model Complexity

In this section, we evaluate GT-HAD in terms of efficiency. Since DNN-based and non-DNN methods are implemented on different hardware platforms and software, we mainly compare GT-HAD with LREN and Auto-AD for the sake of fairness. Table VII gives the model complexity of the involved DNN-based methods, and the experimental data

TABLE IV
OVERLAP (PIXELS) BETWEEN THE ABNORMAL PIXELS AND THE MARKED BACKGROUND PIXELS ON SIX DATASETS

Datasets	Anomaly	Overlap	Proportion
Los Angeles-1	144	9	6.25%
Los Angeles-2	87	8	9.19%
Gulfport	60	0	0.00%
Texas Coast	67	6	8.95%
Cat Island	19	0	0.00%
Pavia	68	6	8.82%

TABLE V
RUNNING TIME (S) OF WHETHER GT-HAD USES CMM ON SIX DATASETS

Datasets	base+AFB model	base+AFB+BFB model	Time IR
Los Angeles-1	26.69	29.54(+2.85)	10.68%
Los Angeles-2	26.97	29.68(+2.71)	10.05%
Gulfport	27.39	30.04(+2.65)	9.67%
Texas Coast	27.21	29.75(+2.54)	9.33%
Cat Island	57.54	63.14(+5.60)	9.73%
Pavia	57.29	63.73(+6.44)	11.24%

TABLE VI
MODEL GENERALIZATION PERFORMANCE ASSESSMENT

Training Dataset	Testing Dataset	AUC
Los Angeles-2	Los Angeles-1	0.9550
Los Angeles-1		0.9775
Los Angeles-1	Los Angeles-2	0.9861
Los Angeles-2		0.9901

TABLE VII
MODEL COMPLEXITY OF THREE DNN-BASED METHODS

Los Angeles-1	Auto-AD	LREN	GT-HAD
Params	3.25M	0.86M	0.26M
FLOPs	11.97G	0.19G	2.64G
Running Time	30.55s	161.85s	29.54s

utilizes the Los Angeles-1 dataset. Compared with LREN and Auto-AD, GT-HAD has the fewest parameters and the shortest running time. The running time of LREN comprises three parts, i.e., spectrum mapping, dictionary construction, and optimized-problem solving. Although LREN has fewer floating point operations (FLOPs) than GT-HAD, the optimized-problem solving stage of LREN significantly increases the running time, which leads to LREN having a longer total time than GT-HAD. In summary, GT-HAD exhibits superior detection performance and demonstrates lower model complexity.

V. Conclusion

In this article, we propose GT-HAD, a novel GT-HAD, which utilizes content similarity to guide the HSI reconstruction. GT-HAD contains two key components, i.e., an AFB that focuses on suppressing the anomalies reconstruction and a BFB that focuses on strengthening the background reconstruction. In addition, a gating unit is developed to regulate the activation states of these two branches based on a CMM. Extensive experimental results demonstrate that GT-HAD outperforms many existing HAD methods and exhibits superior detection performance. Our study introduces a new perspective on integrating transformers into HAD tasks, and it can serve as a stepping stone for further research in this direction. In future work, we would like to devise a streamlined and efficient HAD method. As we delve into this research, the criticality lies not only in advancing the algorithms themselves but also in optimizing the computational efficiency of HAD to meet the stringent requirements of real-time applications.

Acknowledgment

The authors would like to thank the reviewers for their great efforts in reviewing this article.

References

[1] L. Fang, P. Zhou, X. Liu, P. Ghamisi, and S. Chen, "Context enhancing representation for semantic segmentation in remote sensing images," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 9, 2022, doi: 10.1109/TNNLS.2022.3201820.

[2] N. He, L. Fang, S. Li, J. Plaza, and A. Plaza, "Skip-connected covariance network for remote sensing scene classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1461–1474, May 2020.

[3] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[4] M. Shimoni, R. Haelterman, and C. Perneel, "Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag. Replaces Newsletter*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[5] S. Sudharsan, R. Hemalatha, and S. Radha, "A survey on hyperspectral imaging for mineral exploration using machine learning algorithms," in *Proc. Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, Mar. 2019, pp. 206–212.

[6] 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, Oct. 1990.

[7] H. Kwon, "Adaptive anomaly detection using subspace separation for hyperspectral imagery," *Opt. Eng.*, vol. 42, no. 11, p. 3342, Nov. 2003.

[8] 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.

[9] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[10] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, "Multipixel anomaly detection with unknown patterns for hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5557–5567, Oct. 2022.

[11] Y. Gu, Y. Liu, and Y. Zhang, "A selective kernel PCA algorithm for anomaly detection in hyperspectral imagery," in *Proc. IEEE Int. Conf. Acoust. Speed Signal Process. Proc.*, vol. 2, Toulouse, France, May 2006, pp. 725–728.

[12] V. Roth, "Kernel Fisher discriminants for outlier detection," *Neural Comput.*, vol. 18, no. 4, pp. 942–960, Apr. 2006.

[13] X. Miao, Y. Liu, H. Zhao, and C. Li, "Distributed online one-class support vector machine for anomaly detection over networks," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1475–1488, Apr. 2019.

[14] S. Li, K. Zhang, P. Duan, and X. Kang, "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 319–329, Jan. 2020.

[15] Y. Zhang, Y. Dong, K. Wu, and T. Chen, "Hyperspectral anomaly detection with Otsu-based isolation forest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9079–9088, 2021.

[16] C.-I. Chang, H. Cao, and M. Song, "Orthogonal subspace projection target detector for hyperspectral anomaly detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4915–4932, 2021.

[17] B. Tu, X. Yang, W. He, J. Li, and A. Plaza, "Hyperspectral anomaly detection using reconstruction fusion of quaternion frequency domain analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 31, 2023, doi: 10.1109/TNNLS.2022.3227167.

[18] 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.

[19] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.

[20] R. Wang, H. Hu, F. He, F. Nie, S. Cai, and Z. Ming, "Self-weighted collaborative representation for hyperspectral anomaly detection," *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107718.

[21] N. Ma, Y. Peng, and S. Wang, "A fast recursive collaboration representation anomaly detector for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 588–592, Apr. 2019.

[22] H. Su, Z. Wu, A.-X. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogramm. Remote Sens.*, vol. 169, pp. 195–211, Nov. 2020.

[23] S. Wang, X. Wang, Y. Zhong, and L. Zhang, "Hyperspectral anomaly detection via locally enhanced low-rank prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6995–7009, Oct. 2020.

[24] Y. Xu, B. Du, L. Zhang, and S. Chang, "A low-rank and sparse matrix decomposition-based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1248–1252, Jul. 2020.

[25] 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.

[26] 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.

[27] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, "Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 4, 2022, doi: 10.1109/TNNLS.2022.3152252.

[28] S. Sun, J. Liu, Z. Zhang, and W. Li, "Hyperspectral anomaly detection based on adaptive low-rank transformed tensor," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jan. 24, 2023, doi: 10.1109/TNNLS.2023.3236641.

[29] 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.

[30] 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.

[31] 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.

[32] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Deep low-rank prior for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527017.
[33] X. Wang, L. Wang, and Q. Wang, "Local spatial–spectral information-integrated semisupervised two-stream network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535515.
[34] N. Huyan, X. Zhang, D. Quan, J. Chanussot, and L. Jiao, "AUD-net: A unified deep detector for multiple hyperspectral image anomaly detection via relation and few-shot learning," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 27, 2022, doi: 10.1109/TNNLS.2022.3213023.
[35] 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.
[36] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS^{3}LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504218.
[37] K. Li et al., "Spectral–spatial deep support vector data description for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5522316.
[38] K. Li et al., "Spectral difference guided graph attention autoencoder for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 72, 2022, Art. no. 5001817.
[39] Y. Ma, S. Cai, and J. Zhou, "Adaptive reference-related graph embedding for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5504514.
[40] 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.
[41] C. Zhao, C. Li, S. Feng, and W. Li, "Spectral–spatial anomaly detection via collaborative representation constraint stacked autoencoders for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
[42] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "LREN: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, vol. 35, 2020, pp. 4139–4146.
[43] H. Du, J. Wang, M. Liu, Y. Wang, and E. Meijering, "SwinPA-net: Swin transformer-based multiscale feature pyramid aggregation network for medical image segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Sep. 19, 2022, doi: 10.1109/TNNLS.2022.3204090.
[44] Y. Wang et al., "Learning oriented object detection via naive geometric computing," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 10, 2023, doi: 10.1109/TNNLS.2023.3242323.
[45] H. Chen, G. Yang, and H. Zhang, "Hider: A hyperspectral image denoising transformer with spatial–spectral constraints for hybrid noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 19, 2022, doi: 10.1109/TNNLS.2022.3215751.
[46] I. O. Tolstikhin et al., "MLP-Mixer: An all-MLP architecture for vision," in *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, 2021, pp. 24261–24272.
[47] M. Chen, H. Peng, J. Fu, and H. Ling, "AutoFormer: Searching transformers for visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 12270–12280.
[48] D. Lian, Z. Yu, X. Sun, and S. Gao, "AS-MLP: An axial shifted MLP architecture for vision," 2021, *arXiv:2107.08391*.
[49] T. Plötz and S. Roth, "Neural nearest neighbors networks," in *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, 2018, pp. 1087–1098.
[50] Z. Zha, X. Yuan, J. Zhou, C. Zhu, and B. Wen, "Image restoration via simultaneous nonlocal self-similarity priors," *IEEE Trans. Image Process.*, vol. 29, pp. 8561–8576, 2020.
[51] C. Mou, J. Zhang, and Z. Wu, "Dynamic attentive graph learning for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4328–4337.
[52] Z. Zhang, Y. Liu, J. Liu, F. Wen, and C. Zhu, "AMP-net: Denoising-based deep unfolding for compressive image sensing," *IEEE Trans. Image Process.*, vol. 30, pp. 1487–1500, 2021.
[53] D. F. Williamson, R. A. Parker, and J. S. Kendrick, "The box plot: A simple visual method to interpret data," *Ann. Intern. Med.*, vol. 110, no. 11, pp. 916–921, 1989.
[54] 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. 2021.

Jie Lian received the B.S. degree from the Southwest University of Science and Technology, Mianyang, China, in 2017, and the M.S. degree from North China Electric Power University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing.

His research interests include hyperspectral and medical image processing.

Lizhi Wang (Member, IEEE) received the B.S. and Ph.D. degrees from Xidian University, Xi'an, China, in 2011 and 2016, respectively.

He is currently a Professor with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing, China. His research interests include computational photography and image processing.

Dr. Wang received the Best Paper Runner-Up Award of ACM MM 2022 and the Best Paper Award of IEEE VCIP 2016.

He Sun received the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2020.

He is currently an Assistant Researcher with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China. His research interests include hyperspectral target detection and band selection.

Hua Huang (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1996 and 2006, respectively.

He is currently a Professor with the School of Artificial Intelligence, Beijing Normal University, Beijing, China. His main research interests include image and video processing, computational photography, and computer graphics.

Dr. Huang received the Best Paper Award of ICML 2020/EURASIP2020/PRCV2019/ChinaMM2017.