KDAD: Knowledge Distillation-Based Anomaly Detection for Thermal Infrared Hyperspectral Image

Enyu Zhao, *Member, IEEE*, Hao Zhang, Nianxin Qu, *Student Member, IEEE*, Yulei Wang, *Member, IEEE*, and Yongguang Zhao

***Abstract*—Autoencoder (AE) is extensively utilized in hyperspectral anomaly detection (HAD) tasks owing to its robust feature extraction and image reconstruction capabilities. However, AE lacks constraints on anomaly samples during the training process, leading to the reconstruction part of some anomalies alongside background features, which ultimately diminishes the detection accuracy; additionally, most existing HAD algorithms have been specifically designed for data captured within the visible and near-infrared bands, resulting in a notable gap in methodologies tailored for thermal infrared hyperspectral image (TI_HSI). To address these issues, a knowledge distillation-based anomaly detection (KDAD) model is proposed in this study, aimed at thermal infrared hyperspectral data. KDAD constructs a spatial information map utilizing a dual-window model through the spectral-spatial fusion module, thereby enabling simultaneous fusion of spectral and spatial features via a collaborative stacked AE with dual branches; a residual enhancement module (REM) is introduced based on transfer learning techniques to achieve background purification while forming a distillation AE model comprising an efficient student AE and an intricate teacher AE; meanwhile, REM incorporates a clustering weight generation mechanism that facilitates pixel density-aware category division through dimensionality reduction and clustering processes, and constructs a background-enhanced weight matrix by integrating Mahalanobis distance tensor analysis with dynamic threshold adjustment strategy in order to enforce prior constraints on anomalies; finally, the anomaly detection module formulates an anomaly detection process grounded in clustering techniques and cosine similarity metrics to facilitate high-precision anomaly detection within TI_HSIs. Experimental results on thermal infrared hyperspectral datasets indicate that KDAD markedly enhances background suppression capability and improves anomaly localization accuracy. Furthermore, its detection performance across various scenarios outperforms that of existing algorithms.**

***Index Terms*—Anomaly detection, autoencoder (AE), knowledge distillation, thermal infrared hyperspectral image (TI_HSI).**

Received 3 August 2025; revised 16 September 2025; accepted 13 October 2025. Date of publication 15 October 2025; date of current version 27 October 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 42271355, and in part by the Fundamental Research Funds for the Central Universities under Grant 3132025270. *(Corresponding authors: Yulei Wang; Yongguang Zhao.)*

Enyu Zhao, Hao Zhang, Nianxin Qu, and Yulei Wang are with the Center for Hyperspectral Imaging in Remote Sensing, Information Science and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: zhaoenyu@dlmu.edu.cn; 3377518143@qq.com; qvnx@dlmu.edu.cn; wangyulei@dlmu.edu.cn).

Yongguang Zhao is with the Key Laboratory of Quantitative Remote Sensing Information Technology, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: zhaoyg@aircas.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3622117

I. Introduction

Thermal infrared hyperspectral image (TI_HSI) encompasses electromagnetic wave radiation information across multiple continuous bands, with a wavelength range of 8–14 μm. This type of imagery contains rich thermal radiation information and simultaneously offers the geometric spatial relationships and spectral characteristics of ground objects. Furthermore, it demonstrates advantages such as the capability for night imaging [1], [2], [3], [4], [5], [6], [7]. Hyperspectral anomaly detection (HAD) is a technique that capitalizes on the abundant spectral and spatial information present in hyperspectral images to identify anomalous targets that differ from surrounding background pixels [8]. HAD plays a crucial role in various applications, including agricultural production management [9], forest fire monitoring [10], [11], and mineral resource exploration [12]. As such, it holds promising prospects for widespread application [13].

HAD techniques have led to the development of a variety of algorithms specifically designed for analyzing visible and near-infrared hyperspectral images. The traditional algorithms can be categorized into three main types: statistics-based algorithms, representation-based algorithms, and matrix decomposition-based algorithms [14]. Statistics-based HAD methods operate under the assumption that the background pixels conform to a specific statistical model. Consequently, pixels that deviate from this distribution are identified as anomalies [15], [16], [17]. The RX (Reed–Xiaoli) algorithm proposed by Reed et al. posits that the background portion of an image adheres to a multivariate Gaussian distribution, while anomalous targets exhibit significant differences from it. Anomalies are then distinguished from other image elements by calculating the Mahalanobis distance between the pixel of interest and the background pixels [18], [19], [20]. While statistics-based algorithms offer advantages such as straightforward calculation and convenient implementation, they often yield satisfactory detection results in scenarios with relatively simple backgrounds. However, their efficacy tends to diminish in more complex environments [21]. Representation-based methods assume that background pixels can be effectively represented via spatial neighborhoods or a background dictionary. In contrast, anomalous pixels present challenges in representation due to their inherent uniqueness, which facilitates the differentiation between background and anomalies within images [22], [23], [24], [25]. The collaborative-representation-based detector (CRD) introduced by Li and Du [26] leverages the inherent characteristics of collaborative representation among pixels to facilitate anomaly detection. This is accomplished through the reconstruction of a target using its neighboring pixels, followed by the assessment of reconstruction error. Xiao et al. [27] propose a tensor low-rank sparse representation learning anomaly detection method for HAD. Nevertheless, these methodologies are often vulnerable to noise and outliers during the construction of the background dictionary, which may lead to inaccurate representations and consequently diminish the quality of the final detection results [28]. Building upon matrix decomposition, HAD algorithms decompose hyperspectral data into low-rank background components and sparse anomaly components, thereby enabling a decoupled analysis of background structures and anomaly features. Given that anomalous targets in hyperspectral data typically occupy only a limited number of pixels while exhibiting a sparse spatial distribution, anomalies can be detected utilizing low-rank matrix decomposition techniques [29], [30]. Xu et al. [31] developed an approach founded on low-rank and sparse representation (LRSR), employing low-rank matrix representations of the background dictionary to depict background pixels while uncovering local spectral features via sparse constraints; this methodology has yielded favorable outcomes. Yu et al. [32] introduce a novel HAD method grounded in graph regularized low-rank representation (GLR), effectively harnessing both global and local information present within hyperspectral images. Although such algorithms successfully differentiate between backgrounds and anomalies by leveraging low-rank and sparse properties in complex scenes, they often face challenges related to high computational complexity as well as inadequate performance when addressing noise and nonsparse anomalies [33].

In recent years, the application of deep learning models in HAD has made considerable advancements and has gradually emerged as one of the predominant methodologies. The central premise involves the automatic extraction of abstract and hierarchical features from data through multilayer neural network architectures to facilitate effective identification of anomalous targets [34], [35], [36], [37], [38], [39], [40], [41], [42]. Li et al. [43] are pioneers in introducing the convolutional neural networks (CNNs) into HAD, proposing a CNN-based detector model. This model incorporates both differential pixel pairs and similar pixel pairs for training within the CNN framework. Subsequently, it inputs both the pixel under investigation and the mean values of its surrounding pixels into the trained CNN to ascertain whether such a pixel is anomalous or not. Fu et al. [44] present an anomaly detection approach termed DeCNN-AD (denoising CNN-anomaly detection), which capitalizes on spatial correlations in representation coefficients via a CNN denoiser, integrates prior knowledge within a low-rank representation framework, and optimizes background dictionary construction using clustering techniques to enhance anomaly detection. Additionally, deep learning networks' capability for image reconstruction serves as an advantageous tool for detecting anomalies. Notably, autoencoders (AEs) have been employed in HAD tasks due to their robust data reconstruction abilities; they are predicated upon the principle that reconstruction errors associated with background pixels are substantially lower than those corresponding to anomalous pixels [45]. Wang et al. [46] propose an auto-AD (autonomous anomaly detection) model, which integrates an adaptive weight loss function into a fully convolutional AE, thereby enhancing its capability to distinguish anomalies from the background in complex scenes. Ma et al. [47] employ a memory AE that introduces storage modules at various hidden layers of AE to facilitate multiscale reconstruction of both backgrounds and anomalous pixels within the spectral domain. Cao et al. [48] introduce an AiANet (adaptive interactive attention network) that incorporates a low-rank module into an AE, effectively improving the accuracy of background modeling across different scenarios. In HAD tasks, the reconstruction error generated by AE for images can serve as an indicator of the anomaly degree of anomaly present in pixels; thus, utilizing this reconstruction error during detection can enhance the accuracy. However, due to the powerful generalization capacity of AEs, there are instances where they may effectively reconstruct both background points and anomaly points within hyperspectral images simultaneously. As such, it becomes difficult to separate anomaly points from background points based on reconstruction errors obtained from these models [49], [50].

Current HAD technologies predominantly utilize data derived from visible and near-infrared bands, with an absence of anomaly detection algorithms specifically tailored for TI_HSI [5]. Hyperspectral imaging in the visible and near-infrared regions is contingent upon prevailing illumination conditions, with the radiation captured by imaging sensors primarily consisting of the reflected radiation from various substances [51]. In contrast, imaging within thermal infrared bands relies fundamentally on the self-thermal radiation emitted by ground objects, rendering it independent of lighting conditions. Consequently, even under low-light scenarios such as nighttime, relevant target information remains accessible [52]. Due to inherent disparities in their respective imaging mechanisms, existing anomaly detection methods are not directly applicable to TI_HSIs [4]. Furthermore, when compared to visible and near-infrared hyperspectral data, thermal infrared data tends to exhibit lower resolution and more complex radiation characteristics, which render it difficult to visually distinguish anomaly points from background points directly. This situation underscores an urgent need for the development of anomaly detection algorithms that are specifically designed for TI_HSI.

To address the aforementioned challenges, this study proposes a knowledge distillation-based anomaly detection (KDAD) model specifically designed for TI_HSI, comprising three main components: a spectral-spatial fusion module (SSFM), a residual enhancement module (REM), and an anomaly detection module (ADM). The SSFM employs a dual-window approach to extract a spatial information map and utilizes a dual-branch collaborative stacked AE (DBCSAE) framework to capture spectral and spatial information, ultimately generating a spectral-spatial fusion image. Within the REM, a knowledge distillation-based weighted AE (KDWAE) is introduced to reconstruct the fused image, which yields an enhanced reconstruction residual map. Finally, the ADM integrates clustering and cosine similarity

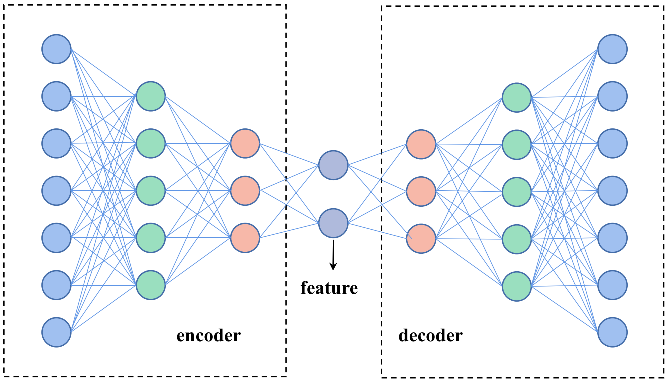


Fig. 1. Schematic diagram of the AE structure.

measures to conduct anomaly detection on the cascaded image. The main contributions of this article are as follows.

1) This study proposes a KDAD model that integrates spatial information from an image with an enhanced reconstruction residual map, thus addressing a significant gap in thermal infrared HAD. It mitigates challenges associated with distinguishing between anomalous and background pixels caused by the low resolution and radiation complexity of thermal infrared data, achieving commendable detection performance on thermal infrared hyperspectral datasets.
2) In SSFM, the dual-window model captures spatial neighborhood relationships, facilitating the acquisition of a spatial information map that enhances the differences in spatial features between anomalies and the background. Additionally, the dual-branch stacked AE can process both spectral and spatial information concurrently, enabling feature complementarity through cascaded fusion that enhances the accuracy of anomaly detection.
3) The REM is presented to tackle the challenge wherein AE reconstructs both anomalies and backgrounds simultaneously. It establishes a KDWAE. Through the knowledge distillation transfer learning mechanism, KDWAE allows the student network to learn the background feature representation from the teacher network while implementing a background-enhanced weight matrix constraint. This approach strengthens the modeling capability for background features while suppressing reconstruction activities related to anomalous pixels. By employing hierarchical feature interaction and weight optimization techniques, KDWAE efficiently models backgrounds while selectively diminishing anomalous signals. Consequently, it enhances sensitivity in reconstruction residuals toward anomalous targets, resulting in an improved reconstruction residual map.
4) The ADM constructs background clusters based on the size of clustering results and selects the background dictionary by evaluating the Mahalanobis distance. Subsequently, it calculates the cosine similarity between each pixel and the background dictionary to obtain detection outcomes. The ADM accommodates the detection of original images, spectral-spatial fusion images, and enhanced images, thus providing an efficient and flexible detection process.

The rest of this article is structured as follows. Section II reviews the relevant literature. A detailed description of the proposed method is provided in Section III. Section IV systematically presents the thermal infrared HAD datasets, along with the comparative experiments between the proposed method and other algorithms, as well as ablation experiments on key modules. Finally, Section V concludes with a summary of the work accomplished and outlines potential directions for future research.

II. Related Works

A. Autoencoder

AE is an unsupervised learning neural network model (as shown in Fig. 1), which typically consists of two main components: an encoder and a decoder [53]. The encoder transforms input data into the latent space representation, while the decoder reconstructs the original data from this latent representation. The latent space serves as the core feature of the AE; it provides a compressed representation of the input data via encoding and encapsulates the abstract features inherent in that data. For an AE with (M-1) hidden layers, consider a sample $x_i \in R^L$ with L features as an example, the output ${h_i}^{(m)}$ of the mth layer of the AE corresponding to x_i can be expressed as follows:

$$ {h_i}^{(m)} = g\left({h_i}^{(m-1)^T} + b^{(m)}\right)^T, \ m = 1, \ldots M \tag{1} $$

where $g(\cdot)$ is the encoder network, T denotes matrix transposition, and $b^{(m)}$ represents the bias vector of the mth layer.

B. Knowledge Distillation

Knowledge distillation is a technique employed for model compression and transfer learning. This method facilitates the transfer of knowledge from a complex teacher model to a lightweight student model, thereby enabling the latter to achieve high performance at a relatively low computational cost [54]. The fundamental concept underlying this approach is to utilize the outputs or features produced by the teacher model as guidance during the training of the student model. Consequently, this reduces overall model complexity while retaining, and potentially enhancing, task performance. A knowledge distillation system comprises the following two main components.

1) *Teacher Network:* Typically characterized by its deep architecture and complexity, this network possesses robust feature extraction capabilities that allow it to capture intricate high-level semantic features and subtle differences in patterns within data. By doing so, it establishes a benchmark of knowledge that can be utilized by the student model.
2) *Student Network:* This is a lightweight network designed to replicate either the outputs or intermediate layer features of the teacher model via an appropriately formulated distillation loss function. In this manner, it learns efficient representations of features while maintaining computational efficiency.

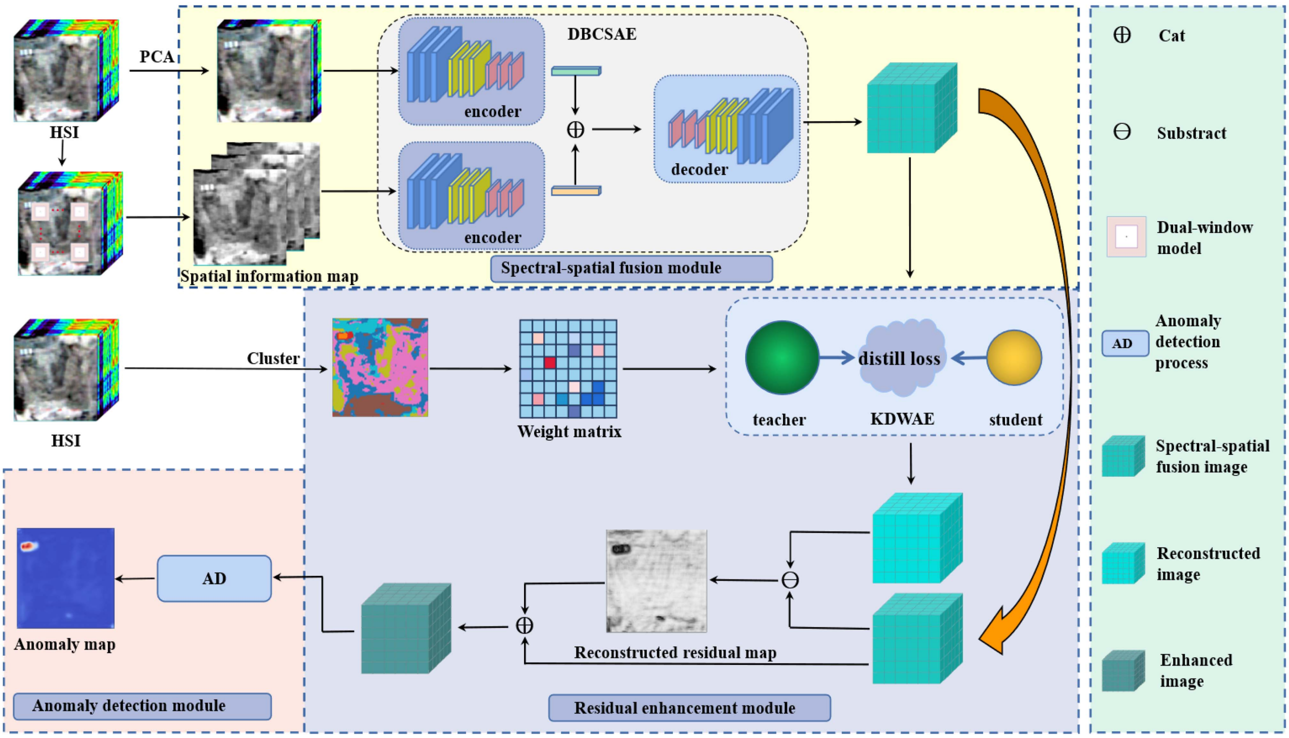


Fig. 2. Flowchart of the proposed KDAD method.

Feature-level knowledge distillation emphasizes the feature representations found in the intermediate layers of the teacher model. By aligning the spatial structures, channel dependencies, or attention mechanisms, this approach facilitates the student model's ability to grasp the semantic relevance of deep features [55]. Specifically, feature-level distillation preserves high-level abstract information from the data by enforcing consistency between the intermediate layer feature maps of both the student and teacher models. The feature Loss function L_{feat} between the teacher and the student model can be defined as follows:

$$L_{feat} = \frac{1}{N} \sum_{i=1}^{L} \| {f_i}^T - {f_i}^S \|_2^2 \tag{2}$$

where L is the number of feature maps, ${f_i}^T$ and ${f_i}^S$ are the feature maps of the ith layer of the teacher and student model, respectively, and $\| \cdot \|_2^2$ represents the square of the L_2 norm.

III. Methodology

To tackle the challenges associated with distinguishing anomalies from backgrounds in TI_HSI, this study proposes a novel anomaly detection method, referred to as KDAD, that integrates spatial information and knowledge distillation mechanism (as illustrated in Fig. 2). This method consists of three core modules: SSFM captures spatial neighborhood relationships through a dual-window model to generate the spatial information map (*SPAM*), achieving complementary fusion of spectral-spatial features via a dual-branch stacked AE; REM establishes a teacher-student network grounded in the knowledge distillation framework, incorporating principal component analysis (PCA) and K-means clustering to produce an enhanced background weight matrix based on pixel density classification; ADM filters background classes according to cluster sizes, constructs a dynamic background dictionary via Mahalanobis distance, and employs cosine similarity to quantify the degree of anomaly.

A. Spectral-Spatial Fusion Module

1) Acquisition of SPAM: TI_HSIs encompass a wealth of spectral information across multiple bands, and their spatial characteristics play a vital role in anomaly detection. Given that the target size of anomalous pixels is typically small in such a detection process, this study adopts a dual-window model (as shown in Fig. 3) to extract spatial information by leveraging the similarity relationships between each pixel and its neighboring pixels.

A hyperspectral image is classified as a form of three-dimensional structured data, where the image can be represented as $X \in \mathbb{R}^{H \times W \times C}$ (H, W, and C represent the height, width, and number of spectral channels, respectively). The cosine similarity, named s, between one pixel vector and its neighborhood pixel vector is computed based on (3). The average values of s for far- and near-neighborhood regions are referred to as μ_1 and μ_2, respectively, which are calculated using (4) and (5)

$$s = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \tag{3}$$

$$\mu_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} s_i \tag{4}$$

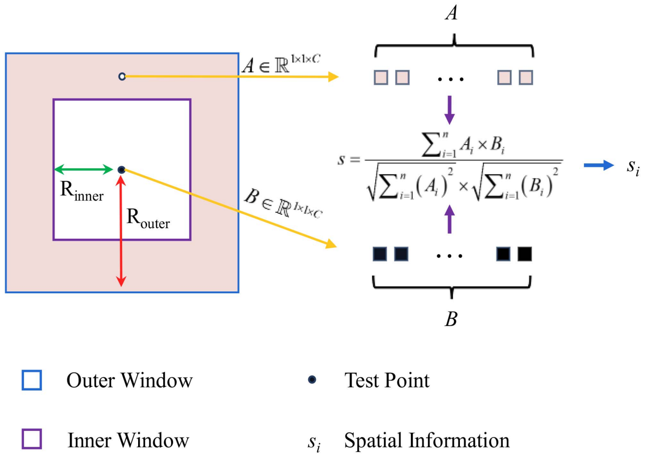

$A \in \mathbb{R}^{1\times1\times C}$
A
$B \in \mathbb{R}^{1\times1\times C}$
B
$\mathrm{R}_{\mathrm{inner}}$
$\mathrm{R}_{\mathrm{outer}}$
$s = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}}$
s_i
Outer Window
Test Point
Inner Window
s_i Spatial Information

Fig. 3. Schematic diagram of the dual-window model.

$$\mu_2 = \frac{1}{N_2}\sum_{j=1}^{N_2} s_j \tag{5}$$

where A_i and B_i represent the spectral vectors of the ith pixel in the far-neighborhood region and the ith pixel in the near-neighborhood region, respectively, N_1 and N_2 are the total number of pixels in the far- and near-neighborhood, respectively; $s_i(s_j)$ is the cosine similarity for the ith(jth) pixel in the far-(near-) neighborhood region.

In this section, s_i of each pixel is regarded as its spatial information. Assuming there are N_1 pixels within the far-neighborhood region, the *SPAM* with dimension (H, W, N_1) can be constructed by calculating all values of s_i for each pixel. The dual-window model effectively capitalizes on the characteristic that anomalous pixels are relatively rare in hyperspectral images. This results in distinct representations of background and anomalous pixels in *SPAM*, which can be classified into the following three cases.

1) For background pixels, there exists a considerable number of background points in both their far- and near-neighborhoods; consequently, both μ_1 and μ_2 are relatively large.
2) In the case of centrally distributed anomalous pixels, numerous background pixels populate their far-neighborhood while an abundance of anomalous pixels is present in their near-neighborhood, resulting in a small μ_1 and a large μ_2.
3) For a single-point anomaly or small-scale anomalies, there are few similar pixels in both far- and near-neighborhoods; hence, both μ_1 and μ_2 remain small.

2) Dual-Branch Stacked Autoencoder: DBCSAE is designed based on a dual-branch collaborative learning framework to extract and fuse spectral and spatial features. Initially, PCA is performed on the original X, retaining 95% of the cumulative variance percentage of the original image to obtain the dimensionality-reduced image HSI_{spec}; then, a stacked convolutional AE is employed to extract the original spectral features of the image and the features of the spatial information map *SPAM*, respectively.

The encoder network architecture consists of four convolutional layers (Conv) interleaved with rectified linear activation layers (ReLU), employing a convolution kernel size of 3 × 3 and a stride of 1. The encoder processes HSI_{spec} and *SPAM* separately, capturing their corresponding features into latent feature spaces. A cascading fusion process occurs in this latent space, leading to the generation of spectral-spatial fusion features. The spectral features extracted by the encoder are represented as Y_{spe}, while the spatial features are denoted as Y_{spa}. Ultimately, these spectral and spatial features are concatenated along the channel dimension to produce the combined spectral-spatial fusion features Y_{fusion}

$$Y_{fusion} = cat\left(Y_{spe}, Y_{spa}\right). \tag{6}$$

In the fusion training process, the reconstruction loss for both the spectral encoder and spatial encoder is computed using mean square error

$$L = \frac{1}{N}\sum_{\mathrm{i}=1}^{N}\left(\hat{X}_i - X_i\right)^2 \tag{7}$$

where L is the loss function of the encoder, N is the total number of image pixels, X_i and $\hat{X}_i$ are the input and output of AE during the training.

The architecture of the decoder consists of four transposed convolutional layers (ConvT) that correspond to the encoder, with alternating ReLU activations. Each layer employs a convolution kernel size of 3 × 3 and a stride of 1. The procedure for decoding is as follows:

$$X_{fusion} = D\left(Y_{fusion}\right) \tag{8}$$

where $D(\cdot)$ is the decoder network, $X_{fusion} \in \mathbb{R}^{H\times W\times C'}$ is the spectral-spatial fusion image, $C^{'} = C + N_1$.

B. Residual Enhancement Module

The reconstruction error serves as a crucial metric for distinguishing background and anomalous pixels. To enhance the distinguishability of reconstruction errors between anomalies and backgrounds, this section integrates feature-level distillation constraints within stacked convolutional AEs, thereby introducing a knowledge distillation weighted AE submodule. This establishes a collaborative learning architecture comprised of teacher-student encoder networks, which is displayed in Fig. 4. The KDWAE submodule exploits the distillation loss of the teacher and student encoders to force the student network to prioritize modeling the background distribution during training, ultimately driving it to focus more on reconstructing the backgrounds and ignore the anomalies. Concurrently, reconstruction loss constrains the student network to remain aligned with its image reconstruction objectives. The submodel's ratio parameters α and β for both distillation loss and reconstruction loss are dynamically adjusted by an adaptive parameter network, facilitating dynamic weight optimization that enhances feature differentiation and adapts throughout the training process. On this basis, this model employs a weight matrix to impose simultaneous constraints on both teacher and student networks.

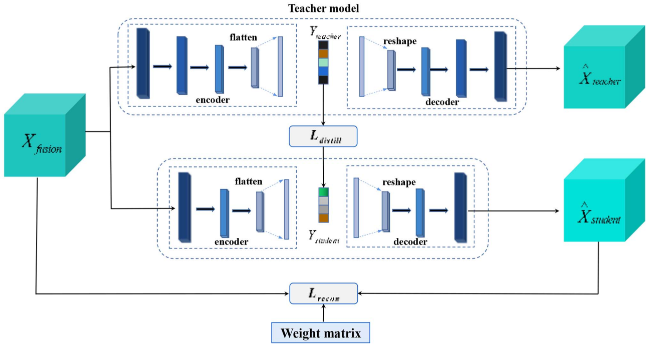


Fig. 4. Structure diagram of knowledge distillation weighted AE.

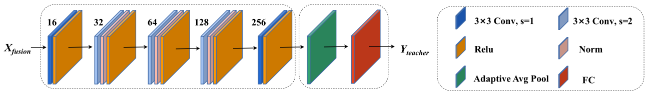


Fig. 5. Network structure of the teacher encoder.

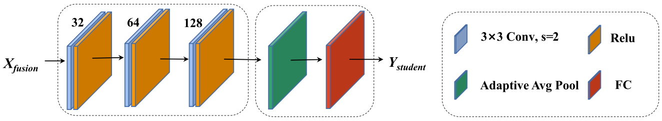


Fig. 6. Network structure of the student encoder.

Through hierarchical interaction between teacher and student features alongside weighted restrictions concerning different image regions' characteristics, it bolsters robust representations of background features while suppressing the model's capacity for reconstructing anomalous features.

1) Teacher Encoder: The teacher encoder utilizes a deep convolutional network, which comprises five convolutional layers, activation layers, normalization layers, and adaptive average pooling layers (as expressed in Fig. 5). This architecture is capable of extracting low-dimensional features from the fused image (X_{fusion}) into the latent feature space, standardizing the size of the feature map via the adaptive average pooling layer, and ultimately mapping it to a p-dimensional feature vector ($Y_{teacher}{}^{p} \in \mathbb{R}^{p}$) through the fully connected layer.

$$Y_{teacher}{}^{p} = f_{teacher}(X_{fusion}; \Theta_{teacher}) \tag{9}$$

where p denotes the dimension of the feature vector, $f_{teacher}(\cdot)$ signifies the nonlinear mapping of the teacher encoder, and $\Theta_{teacher}$ contains the parameters associated with both the convolutional layers and fully connected layers.

2) Student Encoder: In contrast to the teacher encoder, the student encoder adopts a lightweight architecture, including three convolutional layers, activation layers, and an adaptive average pooling layer (as shown in Fig. 6). This design compresses the input of the spectral-spatial fused image into the latent feature space. Within this framework, when the fused image X_{fusion} is fed into the student encoder, the low-dimensional features of the fused image are extracted and transformed into the latent feature space through the convolutional layers. Finally, the size of the feature map is standardized by the adaptive average pooling layer before being mapped to a q-dimensional feature vector ($Y_{student}{}^{q} \in \mathbb{R}^{q}$) through the fully connected layer.

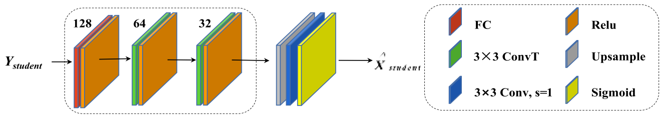


Fig. 7. Network structure of the decoder.

$$Y_{student}{}^{q} = f_{student}(X_{fusion}; \Theta_{student}) \tag{10}$$

where q is the dimension of the feature vector, $f_{student}$ represents the nonlinear mapping of the student encoder.

3) Decoder: The decoder is designed to accurately reconstruct the original image by leveraging both teacher and student features. To ensure that the dimensions of the teacher features align with those of the student features, a feature mapping layer denoted as g_{map} has been introduced. This layer functions as a fully connected network, serving to map the p-dimensional feature vector $Y_{teacher}{}^{p}$, produced by the teacher encoder, to a q-dimensional feature vector that matches the dimensions of the output from the student encoder. The processed teacher feature $Y_{teacher}{}^{q} \in \mathbb{R}^{q}$ following the application of this mapping layer can be expressed as follows:

$$Y_{teacher}{}^{q} = FC_{m}\left(Y_{teacher}{}^{p}\right) \tag{11}$$

where $FC_{m}(\cdot)$ is the function of the feature mapping layer.

In the following section, a decoder is employed to decode the mapped teacher feature $Y_{teacher}{}^{q}$ and student feature $Y_{student}{}^{q}$, respectively, in order to reconstruct the original image. The decoder can be represented as a nonlinear mapping $g_{decoder}$, which encompasses deconvolution layers, upsampling layers, and fully connected layers aimed at dynamically restoring the input image size (as shown in Fig. 7).

Additionally, the mean squared error is used to represent the reconstruction loss of the teacher model [see (12)]

$$L_{recon}{}^{t} = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{X}_{teacher_{j}} - X_{j}\right)^{2} \tag{12}$$

where $L_{recon}{}^{t}$ represents the reconstruction loss of the teacher model, N denotes the number of image pixels, $\hat{X}_{teacher_{j}}$ refers to the reconstructed image from the teacher features, and X_{j} is the input image.

For the teacher features that have been mapped, the resulting decoded output is

$$\hat{X}_{teacher} = g_{decoder}\left(Y_{teacher}{}^{q}; \Theta_{decoder}\right). \tag{13}$$

The loss function of the student model is represented by reconstruction loss and distillation loss

$$L_{recon}{}^{s} = \frac{1}{N}\sum_{j=1}^{N}\left(\hat{X}_{student_{j}} - X_{j}\right)^{2} \tag{14}$$

$$L_{distill} = \frac{1}{H}\sum_{i=1}^{H}\left(Y_{student_i} - Y_{teacher_i}\right)^2 \tag{15}$$

$$L_{total}{}^{s} = \alpha L_{recon}{}^{s} + \beta L_{distill} \tag{16}$$

where $L_{recon}{}^{s}$ represents the reconstruction loss of the student model, $L_{distill}$ denotes the distillation loss of the student model, $L_{total}{}^{s}$ is the total loss, H is the dimension of the feature vector, and α and β are proportional parameters regulated by the Adaptive Parameter Network (APN).

For the student features, the decoded output is

$$\hat{X}_{student} = g_{decoder}\left(Y_{student}{}^{q}; \Theta_{decoder}\right). \tag{17}$$

4) Adaptive Parameter Network (APN): The APN is capable of dynamically generating weights, α and β, which are utilized to balance the distillation loss $L_{distill}$ and reconstruction loss $L_{recon}{}^{s}$. At the core of the APN architecture are two fundamental subnetworks.

1) The difference network, referred to as $\mathrm{AN}_{\mathrm{diff}}(\cdot)$, assesses the distinguishability between backgrounds and anomalies by analyzing the feature difference between the teacher and student models. Its input consists of a concatenation of teacher and student features, while its output is represented by a difference ratio value, denoted as r_{diff}

$$r_{diff} = \mathrm{AN}_{\mathrm{diff}}(Y_{teacher}{}^{q} \oplus Y_{student}{}^{q}),\ r_{diff} \in [0.1, 0.9]. \tag{18}$$

2) The step network $\mathrm{AN}_{\mathrm{step}}(\cdot)$ is designed to dynamically adjust the weight update step size based on the training epoch. Its input consists of the current training epoch e, while its output is a corresponding step ratio value r_{step}.

$$r_{step} = \mathrm{AN}_{\mathrm{step}}\left(e\right),\ r_{step} \in [0.5, 1.0]. \tag{19}$$

Therefore, parameters α and β can be expressed as follows:

$$\alpha = r_{diff} \times r_{step} \tag{20}$$

$$\beta = \left(1 - r_{diff}\right) \times r_{step}. \tag{21}$$

The APN modifies the training parameters α and β in accordance with the dynamic variations in features and the progression of training steps. When there is a significant disparity between teacher and student features, the distillation parameter α increases while β decreases. Conversely, when the feature difference between these two is minimal and the number of training epochs is substantial, the APN raises the value of the reconstruction parameter β and reduces that of α.

5) Weight Matrix: In the context of anomaly detection, the proportion of anomalous pixels relative to the entire image is typically small. Consequently, clustering techniques can be employed to segment the original image, enabling a preliminary assessment of each pixel's degree of anomaly based on category size, thereby facilitating the generation of a weight matrix. The specific steps involved are as follows.

1) The PCA is employed on the original TI_HSI to alleviate the computational burden of subsequent data processing and obtain a dimension-reduced image X'.
2) The K-means cluster algorithm is used to categorize the image, resulting in K clusters$\{C_1, C_2, \ldots, C_K\}$.
3) Establish the background threshold b and the anomaly threshold a ($a > b$) for preliminary judgment. The criteria for judgment are outlined as follows:

$$C_i \in \begin{cases} B,\ |C_i| \le b \cdot N \\ A,\ |C_i| \le a \cdot N \end{cases} \tag{22}$$

where the total number of image pixels is N, $|C_i|$is the number of pixels in cluster C_i, $B = \{B_1,\ B_2,\ \ldots,\ B_n\}$ denotes the set of background candidate classes, $A = \{A_1,\ A_2,\ \ldots,\ A_m\}$ represents the set of anomalous candidate classes, and n and m are, respectively, the number of classes of background candidate and anomaly candidate.

4) The Mahalanobis distance from pixel x to its cluster center U_i is calculated from (23)

$$D\left(x, U_i\right) = \sqrt{(x - U_i)^{T} \sum\nolimits_{i}^{-1} \left(x - U_i\right)} \tag{23}$$

where $\sum_i$ is the covariance matrix. For the background candidate classes, 1 minus those Mahalanobis distances is normalized to the interval [0.5, 1], and placed in the background distance set D_b; for the anomalous candidate classes, Mahalanobis distances are normalized to the interval [0, 0.5], and placed in the anomalous distance set D_a.

5) $W \in \mathbb{R}^{H\times W\times 1}$ is a weight matrix, and (i, j) represents the pixel position in the matrix, and initialize $W(i,j) = 0.5(\forall(i,j))$. Fill it according to the following rules:

$$W\left(i,j\right) = \begin{cases} D_b\left(i,j\right), & X'(i,j) \in B' \\ D_a\left(i,j\right), & X'(i,j) \in A' \\ 0.5, & X'(i,j) \notin (A' \cup B') \end{cases} \tag{24}$$

where B' and A' are the sets of all pixels in the background and anomalous candidate classes, respectively.

The loss function of the student model can be improved through the weight matrix W obtained above as follows:

$$L_{total}{}^{s} = \left(\alpha L_{recon}{}^{s} + \beta L_{distill}\right) \times W. \tag{25}$$

6) Enhanced Image Acquisition: The reconstruction residual map derived from the image reconstructed by KDWAE, in comparison to the input, effectively captures the degree of anomalies. The enhanced reconstruction residual map $R_E \in \mathbb{R}^{H\times W\times 1}$ can be obtained as follows:

$$R_E = \hat{X}_{student} - X_{fusion}. \tag{26}$$

Then, the acquired X_{fusion} is cascaded and spliced with R_E to generate a spectral-spatial residual fusion enhanced map from the following equation:

$$E = cat\left(X_{fusion}, R_E\right) \tag{27}$$

where $E \in \mathbb{R}^{H\times W\times C''}$ denotes the enhanced image, $C'' = C' + 1$, C' is the dimension of the spectral-spatial fusion image.

C. Anomaly Detection Module

The ADM is employed to perform the anomaly detection process on the obtained enhanced image E obtained from prior processing. The primary design objective of the ADM is to achieve accurate localization of anomalies by efficiently leveraging the

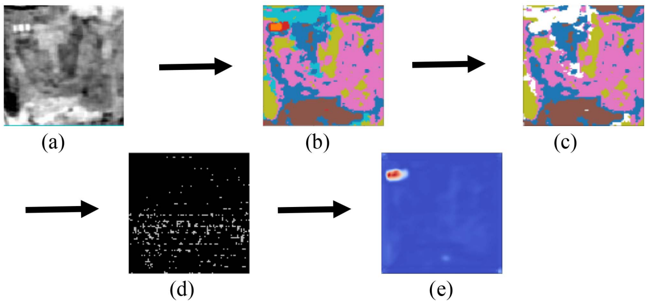


Fig. 8. Workflow and intermediate results of the anomaly detector. (a) TI_HSI. (b) Clustering result. (c) Pure clustering. (d) Background dictionary. (e) AD result.

spectral-spatial features and residual features in the enhanced image, all while employing a lightweight computing approach (as introduced in Fig. 8). The input to the ADM is a TI_HSI [Fig. 8(a)]. Initially, it undergoes K-means clustering, and the clustering result map is depicted in Fig. 8(b). Subsequently, the background class is derived using (22), which can be seen in Fig. 8(c). To assess pixel similarities, we calculate the Mahalanobis distance from each pixel within pure clusters to their respective cluster centers as outlined in (23). These distances are then sorted, with pixels corresponding to the smallest 10% of distances selected to form the background dictionary *BD*, as displayed in Fig. 8(d). The anomaly degree for any given pixel x within the input image is defined as the average cosine similarity between the pixel x and those identified within the background dictionary *BD*. Let $BD = \{b_1, b_2 \dots b_m\}$, then the anomaly degree of x is obtained by (28). The anomaly detection result, demonstrated in Fig. 8(e), can be obtained from calculating the anomaly degrees of all pixels.

$$AD_x = \frac{1}{M}\sum_{m}^{M} \cos\left(x, b_m\right) \tag{28}$$

where AD_x is the anomaly degree of the pixel x, and cos(x, b_m) is the cosine similarity between x and the background point b_m.

IV. Experiments and Results

In this section, the performance of KDAD is compared with that of eight existing methods across three thermal infrared HAD datasets. Furthermore, ablation experiments are conducted to validate the effectiveness of each module introduced in KDAD.

A. Experimental Datasets

The TI_HSIs utilized in this study were acquired in Hengdian Town, Dongyang City, Zhejiang Province, China. These images process a spatial resolution of 1 m and cover a spectral range from 8.061 to 11.217 μm, encompassing a total of 110 channels. Three regions suitable for TI_HSI anomaly detection categories are selected from the original image. These regions are classified into background and anomaly via the supervised support vector machine technique within ENVI software, thereby serving as the datasets for anomaly detection purposes. The hyperspectral images, along with ground truth maps corresponding to the three datasets, are presented in Fig. 9.

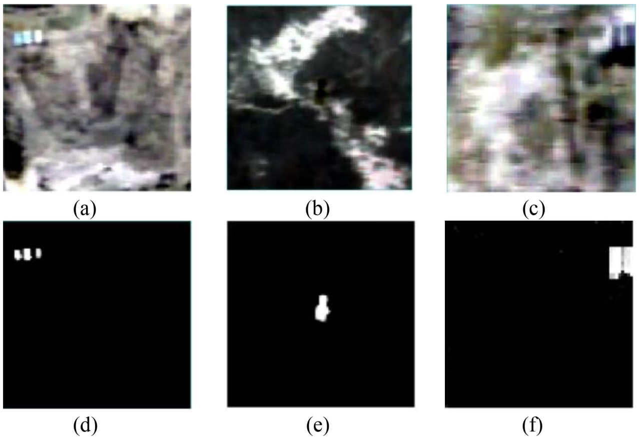


Fig. 9. Hyperspectral images and corresponding ground truth maps used in the experiment. Datasets: (a) House. (b) Tower. (c) Truck. Ground truth maps: (d) House. (e) Tower. (f) Truck.

1) House Dataset: This dataset comprises a TI_HSI with dimensions of 96 × 96 pixels. It features various scenes, including ground, wasteland, and houses, with the houses identified as the anomalies. The image contains a total of 9216 pixels, among which there are 48 anomalous pixels, resulting in an anomaly proportion of 0.52%.

2) Tower Dataset: This dataset consists of a TI_HSI measuring 100 × 100 pixels. It depicts a forested area that includes trees, land, and signal towers; the latter are classified as anomalies within this context. This image encompasses 10 000 pixels in total, with 74 identified as the anomalous pixels, yielding an anomaly proportion of 0.74%.

3) Truck Dataset: It features an image sized at 64 × 64 pixels depicting two trucks parked in an open space; these trucks are considered anomalies for the purposes of analysis. The image is composed of a total of 4096 pixels and includes 80 anomalous pixels, leading to an anomaly proportion of approximately 1.95%.

B. Compared Methods and Evaluation Criteria

1) Compared Methods: The proposed KDAD is compared against eight existing methods, including RX [15], MsRFQFT [56], CRD [26], LRSR [31], Auto-AD [46], BS^{3}LNet [35], SSCADE [57], and GT-HAD [58]. Among these methods, the first four are traditional algorithms: RX is an anomaly detection technique predicated on the assumption of Gaussian distribution; MsRFQFT integrates random forest with frequency domain analysis for anomaly detection; CRD and LRSR are representation-based anomaly detection algorithms. The latter four methods employ deep learning techniques: Auto-AD, BS^{3}LNet, and SSCADE enhance the background feature representation via self-supervised learning or AE structure; GT-HAD utilizes gated Transformer models to capture spectral-spatial similarities and dynamically adjust the activation states of both background and anomaly branches.

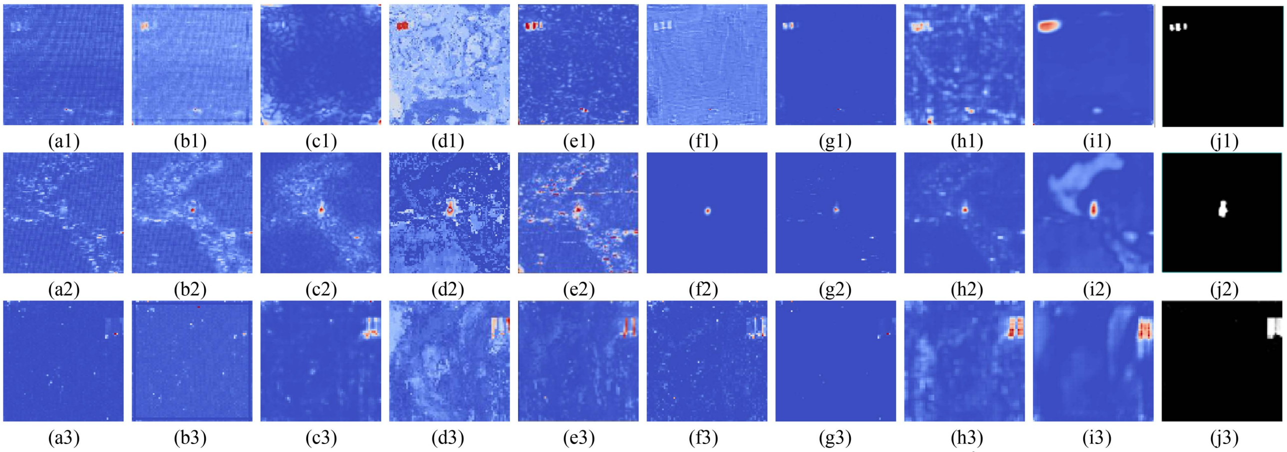


Fig. 10. Color anomaly maps of different methods. (a) RX. (b) CRD. (c) MsRFQFT. (d) LRSR. (e) Auto-AD. (f) BS^{3}LNet. (g) SSCADE. (h) GT-HAD. (i) KDAD. (j) Groundtruth.

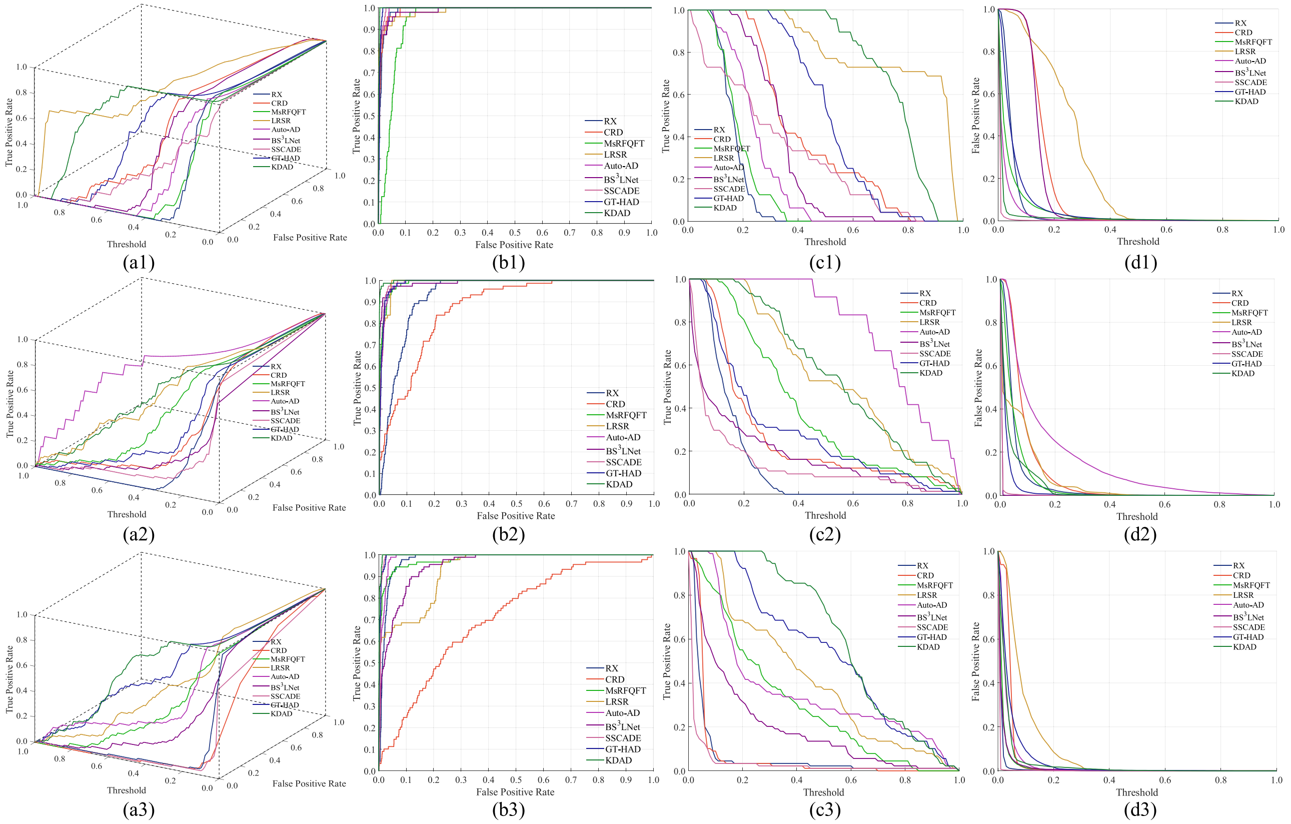
2) Evaluation Criteria: To clearly illustrate the detection performance of each algorithm, this study employs three-dimensional receiver operating characteristic (3-D ROC) and area under the curve (AUC) to quantitatively analyze the detection results across various algorithms. The 3-D ROC curve intuitively reflects the comprehensive performance of an algorithm across multiple dimensions by plotting the three-dimensional relationships among detection probability (PD), false alarm rate (PF), and decision threshold (τ). Furthermore, the 3-D ROC can be decomposed into three two-dimensional curves: (PD, PF), (PD, τ), and (PF, τ). Among these, the (PD, PF) curve represents the traditional two-dimensional ROC curve (2-D ROC), which quantifies the algorithm's ability to balance the detection rate and false alarm rate under different thresholds. A curve that is closer to the upper left corner indicates superior performance, suggesting that the algorithm effectively suppresses false alarms while maintaining a high detection rate. The (PD, τ) curve illustrates how detection rates vary with changes in the threshold values, reflecting an algorithm's sensitivity to anomalous targets; a curve nearer to the upper right corner signifies better performance in detecting anomalies across varying thresholds. The (PF, τ) curve depicts how false alarm rates fluctuate with changing thresholds and demonstrates an algorithm's capability to mitigate background noise; a position closer to the lower left corner indicates enhanced performance through low false alarm rates at diverse thresholds. The AUC serves as a quantitative measure for evaluating 2-D ROC performance within a range of 0–1; values approaching 1 signify a superior algorithm. Additionally, this study produces color anomaly maps and a separability map to qualitatively demonstrate all methods' performance on the three datasets.

C. Detection Performance

The detection results for all employed methods across the three datasets are illustrated in Fig. 10 as color anomaly maps. The color gradient of the anomaly map transitions from blue to red, representing the degree of anomalous behavior associated with each pixel. A higher degree of anomaly corresponds to a color closer to red, indicating an increased likelihood that the pixel is classified as anomalous. The first nine columns of the color anomaly maps display the detection outcomes obtained through various models specifically designed for anomaly detection, while the final column depicts the ground truth map corresponding to the datasets. Fig. 11 presents the 3-D ROC and 2-D ROC curves for different detection models.

The relevant color anomaly maps for the House dataset (as shown in the first row of Fig. 10) indicate that RX and MsRFQFT algorithms exhibit limited capacity for anomaly detection, demonstrating weak responses in identifying anomalous targets within the images, coupled with a tendency to misclassify background pixels as anomalies. CRD, LRSR, and BS^{3}LNet algorithms display inadequate background suppression capabilities, leading to a considerable number of false alarms, a concern further illustrated in Fig. 11(d1). Both Auto-AD and GT-HAD methods are capable of correctly detecting anomalous pixels; they also incorrectly classify some background pixels as anomalies, thereby compromising overall detection efficacy. SSCADE suppresses background and reduces the false alarm rate; however, it is prone to missing detections of certain anomalous pixels. In contrast, the proposed KDAD demonstrates robust performance by accurately detecting anomalies while simultaneously suppressing background pixels. As evidenced by Fig. 11(c1) and (d1), KDAD achieves high detection accuracy across various thresholds with a low false alarm rate; its results align closely with the ground truth of the dataset.

The detection results for the Tower dataset are presented in the second row of Fig. 10. The scene depicted in the Tower dataset is relatively complex [as illustrated in Fig. 9(b)]. From the detection maps, it can be observed that RX struggles to identify any anomalies; CRD detects only incomplete anomalous targets and demonstrates insufficient background suppression. Notably, Auto-AD displays a strong anomaly response [as shown in Fig. 11(c2)]; however, upon examining Fig. 11(d2), it becomes evident that the false alarm associated with Auto-AD remains



(a1) (b1) (c1) (d1)

(a2) (b2) (c2) (d2)

(a3) (b3) (c3) (d3)

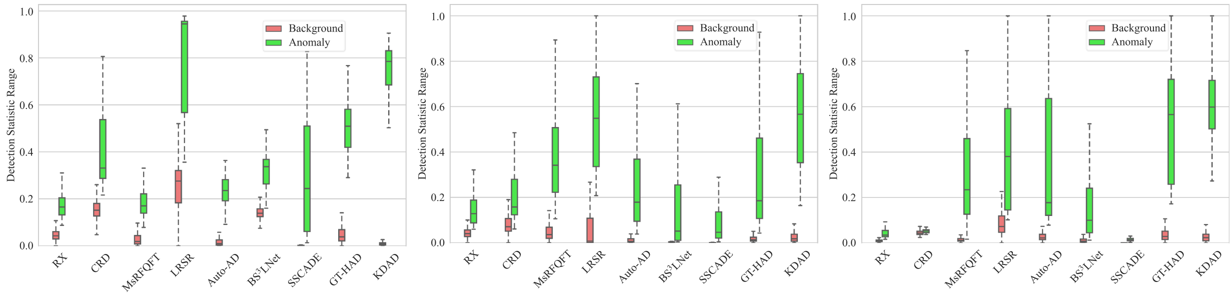
Fig. 11. ROC curves of different detection methods on three datasets. The first row is the House dataset, the second row is the Tower dataset, and the third row is the Truck dataset. (a) 3-D ROC curve. (b) 2-D ROC curve. (c) 2-D (PD, τ) curve. (d) 2-D (PF, τ) curve.

high, resulting in suboptimal overall detection performance. BS^{3}LNet, SSCADE, and GT-HAD exhibit instances of missed detections concerning anomalous targets; both MsRFQFT and LRSR show inadequate background suppression capabilities, leading to an elevated number of false alarms. By comparison, the detection result produced by our proposed KDAD closely aligns with ground truth, the identified anomalies correspond accurately to target positions while maintaining a low false alarm rate.

The detection results for the Truck dataset are displayed in the third row of Fig. 10, which highlights that this dataset contains numerous interferences [as displayed in Fig. 9(c)]. Within this dataset, RX and CRD encounter difficulties when attempting to differentiate anomalies from the background within the detection maps. Auto-AD, MsRFQFT, BS^{3}LNet, and SSCADE demonstrate severe instances of missed detections; Conversely, LRSR, GT-HAD, and KDAD effectively identify anomalous targets. In particular, while GT-HAD exhibits excellent target fidelity, it falls short of KDAD regarding background suppression. As shown in Fig. 11(c3) and (d3), KDAD surpasses GT-HAD in terms of accuracy while achieving a lower false alarm rate. In addition, the separability map is utilized to visually represent the differentiability of detection results from various methods. The background-anomaly separation map evaluates the algorithm's capacity to distinguish between background and anomalous pixels by illustrating the distribution differences between anomalous pixels and distant background pixels. In these separability maps, a greater distance between the anomalous cylinder and its corresponding background cylinder indicates superior separability of the algorithm. Fig. 12 presents the separability maps of detection obtained from different models on the three datasets. It can be observed that LRSR, GT-HAD, and KDAD exhibit relatively good separability. Notably, KDAD demonstrates superior separability performance across all these datasets.

Table I presents the AUC scores for each model across the three datasets. The KDAD method demonstrates optimal performance on the House, Tower, and Truck datasets: In the House scenario, its AUC(PD, PF) of 0.9980 surpasses that of SSCADE (0.9972); while its AUC(PF, τ) is lower at 0.0045 compared to SSCADE's score of 0.0096; however, its AUC(PD, τ) of 0.9903 outperforms those of other methods evaluated. In the Tower scenario, KDAD achieves an AUC(PD, PF) and an AUC(PD, τ) of 0.9988 and 0.9729, respectively, exceeding BS^{3}LNet's figures (0.9963 and 0.8784); additionally, its AUC(PF, τ) attains a notably low value of 0.0009; Finally, in the Truck scenario, KDAD's AUC(PD, PF) reaches a score of 0.9978—surpassing GT-HAD's score of 0.9954; it also exhibits superior performance in both the AUC(PD, τ) (with a score of 0.9144) and AUC(PF, τ) (at 0.0085), which are among the best metrics observed across all models analyzed.

The results from these experiments indicate that the proposed KDAD exhibits considerable robustness across different scenarios. The 3-D ROC curve provides an elaborate view of the



(a) (b) (c)

Fig. 12. Separability maps of different detection methods on three datasets. (a) House dataset. (b) Tower dataset. (c) Truck dataset.

TABLE I
AUC SCORES OF VARIOUS MODELS ON THREE DATASETS

Dataset	AUC	RX	CRD	MsRFQFT	LRSR	Auto-AD	BS^{3}LNet	SSCADE	GT-HAD	KDAD
House	AUC(PD, PF)	0.9889	0.9903	0.9492	0.9887	0.9920	0.9894	0.9972	0.9949	**0.9980**
	AUC(PF, τ)	0.0299	0.0219	0.0906	0.0238	0.0153	0.0315	0.0096	0.0101	**0.0045**
	AUC(PD, τ)	0.6041	0.7708	0.0416	0.8542	0.8125	0.8333	0.8958	0.8750	**0.9903**
Tower	AUC(PD, PF)	0.9324	0.9021	0.9904	0.9893	0.9850	0.9963	0.9912	0.9922	**0.9988**
	AUC(PF, τ)	0.1372	0.2822	0.0276	0.0404	0.0269	0.0122	0.0183	0.0199	**0.0009**
	AUC(PD, τ)	0.0811	0.2027	0.7297	0.6351	0.6216	0.8784	0.6892	0.6757	**0.9729**
Truck	AUC(PD, PF)	0.9772	0.7514	0.9796	0.9294	0.9891	0.9507	0.9932	0.9954	**0.9978**
	AUC(PF, τ)	0.0435	0.6278	0.0399	0.2230	0.0286	0.1215	0.0119	0.0093	**0.0085**
	AUC(PD, τ)	0.3371	0.0562	0.7207	0.6067	0.6404	0.4269	0.8764	0.9047	**0.9144**

*The optimal results are displayed in bold, while the second-optimal results are marked with underlined.

model's performance under different thresholds in detail. It is evident that KDAD successfully maintains an effective balance between the detection rates and the false alarm rates across all datasets at different thresholds; furthermore, it not only exhibits strong separability but also delivers outstanding performance regarding AUC scores.

D. Ablation Experiment

To evaluate the effectiveness of each module in the KDAD model, a series of ablation experiments is performed in this section. The baseline model is referred to as "base," which contains only the ADM component, with AUC(PD, PF) serving as the primary evaluation metric. The ablation studies focus on ADM, SSFM, and KDWAE within REM. The ablation experiments concerning the internal components of SSFM are categorized into two types: For spatial information extraction, the dual-window model is replaced by a sliding single-window one; For image reconstruction, SPAM is eliminated, and the dual-branch AE is replaced by a single-branch variant, with the SSFM modules in these two scenarios denoted as SWSSFM and SBSFM, respectively. The exploration of ADM involves substituting it with the RX detector. In addition, separate ablation analyses are performed on both the knowledge distillation module and weight matrix module within KDWAE. The detailed procedures are outlined as follows.

1) Eliminate the Knowledge Distillation Module: The proportion parameter for the distillation component within the loss function of KDWAE is set to 0, which is denoted as WAE

$$L_{total}{}^{s} = (\alpha L_{recon}{}^{s} + 0L_{distill}) \times W. \tag{29}$$

2) Eliminate the Weight Matrix Module: The weight matrix from the loss function of KDWAE has been removed, resulting in its revised formulation as KDAE

$$L_{total}{}^{s} = (\alpha L_{recon}{}^{s} + \beta L_{distill}). \tag{30}$$

3) Substitute the Proposed Weight Matrix W With a Weight Matrix Derived From Reconstruction Errors: Instead of utilizing the weight matrix W proposed in this article, adopt the weight matrix construction approach of the Auto-AD algorithm, which is based on reconstruction errors. This alternative variant is referred to as KDRWAD.

Table II presents the results of integrating SSFM, WAE, KDAE, and KDWAE into the model. As indicated in the table, the incorporation of SSFM and KDWAE modules leads to a significant enhancement in AUC(PD, PF) values across all three datasets, thereby highlighting their essential roles within the detection algorithm. Notably, the inclusion of the KDWAE module yields an average increase of 0.9% in AUC (PD, PF) values, affirming its critical role in enhancing background features while suppressing anomalies. Within the SSFM module framework, the dual-window model achieves an average AUC(PD, PF) value that is 0.26% higher than that of the single-window model;

TABLE II
ABLATION EXPERIMENTS [AUC(PD, PF)] OF THE PROPOSED METHOD ON THREE DATASETS

	House dataset	Tower dataset	Truck dataset
base	0.9882	0.9786	0.9185
base+SSFM	0.9904	0.9913	0.9889
base+KDWAE	0.9894	0.9863	0.9692
base+SSFM+WAE	0.9932	0.9922	0.9929
base+SSFM+KDAE	0.9928	0.9979	0.9978
base+SWSSFM+KDWAE	0.9975	0.9937	0.9956
base+SBSFM+KDWAE	0.9938	0.9980	0.9932
base+SSFM+KDRWAE	0.9972	0.9950	0.9943
RX+SSFM+KDWAE	0.9917	0.9890	0.9901
base+SSFM+KDWAE	**0.9980**	**0.9988**	**0.9979**

The bold values indicate the optimal AUC(PD, PF) values under different ablation configurations.

moreover, the dual-branch AE demonstrates a considerable advantage over its single-branch equivalent. In relation to the KDWAE module's configuration, omitting the weight matrix component has only a negligible overall effect on final performance; nonetheless, it markedly influences specific datasets. For instance, excluding this component improves detection outcomes by 0.52% on the House dataset when compared with results obtained with it retained. In contrast, removal of the knowledge distillation module results in a notable decrease in AUC(PD, PF) values; this underscores the significance of background feature transfer within the teacher-student networks for enhancing reconstruction residual quality. The comparison between the proposed weight matrix construction method and an alternative based on reconstruction error emphasizes additional benefits associated with utilizing the weight matrix W. The table further illustrates superior performance exhibited by ADM, which realizes an average improvement of 0.8% relative to findings from employing RX detector methodology.

These experimental findings reveal that integration of any module into the model results in substantial improvements in detection performance, thereby affirming both the necessity and efficacy of each component within the architectural framework.

E. Parameter Analysis

The proposed KDAD method incorporates several key parameters that necessitate analysis to enhance the performance of anomaly detection. These parameters include the inner and outer window sizes of the dual-window model, as well as the anomaly threshold a and background threshold b. In the experimental setup, the inner window size varies from 3 to 9, while the outer window size ranges from 5 to 11, ensuring that the outer window properly encompasses the inner window within an appropriate spatial context. Concurrently, the thresholds a and b are adjusted within the intervals [0.2, 0.3] and [0.05, 0.15], respectively. The corresponding AUC(PD, PF) values under different parameter configurations are displayed in Tables III and IV.

To assess the impact of inner and outer window sizes on performance, this study maintains all other parameters at constant levels while evaluating various combinations within their specified ranges. As illustrated in Table III, the AUC(PD, PF) scores for the House, Tower, and Truck datasets exhibit variability across distinct pairings of window size. Notably, when configured with an inner window size of 5 paired with an outer window size of 7, the model achieves optimal AUC performance across all datasets.

To determine the optimal values for parameters a and b, a variety of threshold pairs within their specified ranges are evaluated. As depicted in Table IV, the AUC(PD, PF) values vary with different combinations of (a, b). Notably, however, the combination where $a = 0.25$ and $b = 0.15$ consistently yields the highest performance across the House, Tower, and Truck datasets. This pair demonstrates superior efficacy in distinguishing anomalies from background data.

F. Running Time

In this section, the computation time of various detection methods is analyzed. All experiments reported in this article are conducted on a computer equipped with a 12th Gen Intel Core i7-12700 processor and a 64-bit operating system, featuring a main frequency of 2.10 GHz and 16.0 GB of RAM. The running times for these models are presented in Table V. The traditional algorithm RX is the fastest, exhibiting an average running time of 0.03 s across the three datasets; in contrast, the average running times of LRSR and BS^{3}LNet increase significantly to 303.25 and 1211.48 s, respectively, which may be attributed to the complex model structure and substantial processing demand for the prolonged time; the average running time of Auto-AD is recorded at 139.02 s, which indicates considerable consumption of computing resources during feature extraction and model inference processes; the proposed KDAD demonstrates an average running time of 88.24 s, revealing notable efficiency advantages compared to most other comparative methods due to its optimized computational approach that avoids excessive iterative calculations and extensive parameter training. Although SSCADE has a shorter average running time of only 58.82 s, KDAD strikes a better balance between computational efficiency and detection accuracy overall. Notably, on the Truck dataset, KDAD achieves a remarkably lower running time of 40.43 s compared to many competitive methods evaluated in this study.

G. Effectiveness Analysis of the Adaptive Parameter Network

To evaluate the effectiveness of APN in dynamically adjusting the weights of knowledge distillation loss and reconstruction loss, this study designs an experiment that compares static weight assignments for these losses with dynamic weight assignments managed by the APN. For the static weight assignment experiment, nine sets of fixed combinations of (α, β) are defined, encompassing a gradual range where α varies from 0.1 to 0.9 and β ranges from 0.9 to 0.1. The experimental results are summarized in Table VI.

Under static weight assignment conditions, the AUC(PD, PF) values corresponding to different (α, β) combinations exhibit

TABLE III
EXPERIMENT ON THE ANALYSIS OF INNER AND OUTER WINDOW SIZE PARAMETERS [AUC(PD, PF)] ON THREE DATASETS

Dataset	(3, 5)	(3, 7)	(3, 9)	(3, 11)	(5, 7)	(5, 9)	(5, 11)	(7, 9)	(7, 11)	(9, 11)
House	0.9954	0.9967	0.9945	0.9945	**0.9980**	0.9970	0.9963	0.9966	0.9939	0.9979
Tower	0.9969	0.9971	0.9980	0.9984	**0.9988**	0.9968	0.9946	0.9976	0.9945	0.9933
Truck	0.9919	0.9894	0.9872	0.9940	**0.9979**	0.9914	0.9877	0.9893	0.9938	0.9942

* The first row represents the combination of inner and outer window sizes. The bolded entries indicate the optimal values.

TABLE IV
EXPERIMENT ON THE ANALYSIS OF THRESHOLD PARAMETERS [AUC(PD, PF)] *a* AND *b* ON THREE DATASETS

Dataset	(0.3, 0.05)	(0.3, 0.1)	(0.3, 0.15)	(0.25, 0.05)	(0.25, 0.1)	(0.25, 0.15)	(0.2, 0.05)	(0.2, 0.1)	(0.2, 0.15)
House	0.9955	0.9952	0.9955	0.9948	0.9941	**0.9980**	0.9978	0.9977	0.9974
Tower	0.9951	0.9967	0.9974	0.9946	0.9970	**0.9988**	0.9914	0.9976	0.9890
Truck	0.9939	0.9918	0.9909	0.9951	**0.9979**	**0.9979**	0.9952	0.9938	0.9940

* The first row represents the combination of thresholds *a* and *b* (*a*, *b*). The bolded entries indicate the optimal values.

TABLE V
RUNNING TIME (S) OF VARIOUS MODELS ON THREE DATASETS

Dataset	RX	CRD	MsRFQFT	LRSR	Auto-AD	BS^{3}LNet	SSCADE	GT-HAD	KDAD
House	0.05	10.36	0.5	427.76	136.79	1342.43	68.15	319.4	123.27
Tower	0.03	11.39	0.52	343.45	168.61	1444.3	79.09	162.37	101.02
Truck	0.02	4.48	0.28	138.54	111.67	847.7	29.22	130.05	40.43
Average	0.03	8.74	0.43	303.25	139.02	1211.48	58.82	203.94	88.24

TABLE VI
COMPARATIVE EXPERIMENTS [AUC(PD, PF)] BETWEEN APN AND STATIC WEIGHT ASSIGNMENT ON THREE DATASETS

Dataset	(0.1, 0.9)	(0.2, 0.8)	(0.3, 0.7)	(0.4, 0.6)	(0.5, 0.5)	(0.6, 0.4)	(0.7, 0.3)	(0.8, 0.2)	(0.9, 0.1)	APN
House	0.9973	0.9974	0.9977	**0.9981**	0.9974	0.9976	0.9959	0.9964	0.9962	0.9980
Tower	0.9948	0.9950	0.9945	0.9955	0.9975	0.9922	0.9954	0.9955	0.9968	**0.9988**
Truck	0.9915	0.9937	0.9955	0.9968	0.9945	0.9947	0.9930	0.9969	0.9955	**0.9979**

* The first row represents the combination of weight ratio parameters α and β. The bolded entries indicate the optimal values.

significant fluctuations depending on varying parameter pairings across datasets: on the House dataset, only the combination (0.4, 0.6) achieves an AUC(PD, PF) score of 0.9981—the optimal result for this dataset—whereas most other pairs yield notably lower AUC(PD, PF) scores than this optimal benchmark. Similarly, within both Tower and Truck datasets, only a limited number of (α, β) pairs demonstrate optimal performance; conversely, other combinations reveal substantial performance gaps when compared to these top-performing pairs. In contrast, under APN-based dynamic weight assignment conditions, our model consistently reaches AUC(PD, PF) values across all three datasets—House, Tower, and Truck—that not only exceed those achieved by most static (α, β) combinations but also exhibit significantly enhanced stability in performance.

V. CONCLUSION

In this study, a thermal infrared HAD model based on dual-window spectral-spatial information fusion and KDAD is proposed. The anomaly detection is performed on three hyperspectral thermal infrared datasets, offering a novel approach for identifying anomalies in TI_HSIs. KDAD accurately extracts the spatial information map via a dual-window model and integrates these with a dual-branch stacked AE to achieve the deep fusion of spectral and spatial features, thereby significantly enhancing the capability to differentiate between backgrounds and anomalies. Subsequently, through the knowledge distillation weighted AE framework, our approach employs dynamic adjustments within the teacher-student network alongside the weight matrix. Additionally, an adaptive parameter network is utilized to fine-tune both reconstruction loss and distillation loss throughout the network's operation. This process strengthens the robust modeling of background features while reducing anomalous reconstruction, thereby improving the sensitivity of reconstruction residuals to anomalies. Finally, a lightweight anomaly detector is designed to rely on clustering techniques combined with cosine similarity analysis. This facilitates efficient integration of spectral-spatial fusion images along with an enhanced reconstruction residual map, thus ensuring a balance between detection accuracy and operational efficiency. Experimental results indicate that the KDAD model surpasses existing algorithms in detection performance across the three thermal infrared hyperspectral datasets, demonstrating enhanced background suppression capability and improved accuracy in anomaly localization. Nevertheless, the current algorithm has certain limitations when it comes to detecting single-point anomalies with an extremely low pixel ratio or small targets with anomalous distributions.

In future work, we will further explore more effective anomaly detection frameworks to address this issue.

References

[1] E. Zhao, N. Qu, Y. Wang, C. Gao, and J. Zeng, “TEBS: Temperature–emissivity–driven band selection for thermal infrared hyperspectral image classification with structured state-space model and gated attention,” *Int. J. Appl. Earth Observ. Geoinformation*, vol. 142, 2025, Art. no. 104710.

[2] J. Yang, J. Zhao, L. Chen, H. Geng, and P. Zhang, “Learning nonconvex tensor representation with generalized reweighted sparse regularization for hyperspectral anomaly detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 14718–14737, 2025.

[3] E. Zhao, Y. Su, N. Qu, Y. Wang, C. Gao, and J. Zeng, “Self- and cross-attention enhanced transformer for visible and thermal infrared hyperspectral image classification,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 13408–13422, 2025.

[4] E. Zhao et al., “Thermal infrared hyperspectral band selection via graph neural network for land surface temperature retrieval,” *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5003414.

[5] Y. Wang, X. Chen, E. Zhao, and M. Song, “Self-supervised spectral-level contrastive learning for hyperspectral target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510515.

[6] S. Mei, X. Chen, Y. Zhang, J. Li, and A. Plaza, “Accelerating convolutional neural network-based hyperspectral image classification by step activation quantization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502012.

[7] Y. Wang, S. Mei, M. Ma, Y. Liu, T. Gao, and H. Han, “Hyperspectral object tracking with context-aware learning and category consistency,” *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5508716.

[8] Y. Liu, W. Xie, Y. Li, Z. Li, and Q. Du, “Dual-frequency autoencoder for anomaly detection in transformed hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523613.

[9] K. Gkountakos, K. Ioannidis, K. Demestichas, S. Vrochidis, and I. Kompatsiaris, “A comprehensive review of deep learning-based anomaly detection methods for precision agriculture,” *IEEE Access*, vol. 12, pp. 197715–197733, 2024.

[10] M. Coca, I. C. Neagoe, and M. Datcu, “Hybrid DNN-Dirichlet anomaly detection and ranking: Case of burned areas discovery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4414116.

[11] I. Masari, G. Moser, and S. B. Serpico, “Manifold learning and deep generative networks for heterogeneous change detection from hyperspectral and synthetic aperture radar images,” *IEEE Geosci. Remote Sens. Lett.*, vol. 22, 2024, Art. no. 5500105.

[12] T. Yu, T. Han, P. Lin, Z. Xu, and R. Shao, “Four typical variation patterns of mid-infrared spectra of the felsic mineral anomalies for fault zone identification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5619120.

[13] M. Belgiu and L. Dragut, “Random forest in remote sensing: A review of applications and future directions,” *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.

[14] Y. Mi, B. Tu, Y. Chen, Z. Cao, and A. Plaza, “Hyperspectral anomaly detection via anchor generation,” *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 5038614.

[15] 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.

[16] R. Zhao, B. Du, L. Zhang, and L. Zhang, “Beyond background feature extraction: An anomaly detection algorithm inspired by slowly varying signal analysis,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1757–1774, Mar. 2016.

[17] 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.

[18] I.-S. Reed and X. Yu, “Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution,” *Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Aug. 1990.

[19] S. Liu, M. Song, B. Xue, C.-I. Chang, and M. Zhang, “Hyperspectral real-time local anomaly detection based on finite Markov via line-by-line processing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5503520.

[20] L. Ren, L. Zhao, and Y. Wang, “A superpixel-based dual window RX for hyperspectral anomaly detection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1233–1237, Jun. 2020.

[21] 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.

[22] D. Ma, Y. Hou, M. Chen, B. Li, Z. Wang, and M. Li, “S2G2HAD: A graph-guided Siamese reconstruction network for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526121.

[23] M. Wang, D. Hong, B. Zhang, L. Ren, J. Yao, and J. Chanussot, “Learning double subspace representation for joint hyperspectral anomaly detection and noise removal,” *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5507517.

[24] B. Fu, X. Sun, C. Cui, J. Zhang, and X. Shang, “Structure-preserved and weakly redundant band selection for hyperspectral imagery,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 12490–12504, 2024.

[25] R. Zhao and L. Zhang, “GSEAD: Graphical scoring estimation for hyperspectral anomaly detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 725–739, Feb. 2017.

[26] W. Li and Q. Du, “Collaborative representation for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Aug. 2015.

[27] Q. Xiao, L. Zhao, and S. Chen, “Tensor low-rank sparse representation learning for hyperspectral anomaly detection,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 7356–7359.

[28] H. Li, C. Wei, Y. Yang, Z. Zhong, M. Xu, and D. Yuan, “Unified dynamic dictionary and projection optimization with full-rank representation for hyperspectral anomaly detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 18, pp. 4032–4049, 2025.

[29] 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.

[30] R. Zhao, B. Du, and L. Zhang, “Hyperspectral anomaly detection via a sparsity score estimation framework,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.

[31] 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.

[32] Q. Yu, G. Yan, X. Li, J. Xu, and X. Yang, “Graph regularized low-rank representation for hyperspectral anomaly detection,” in *Proc. 3rd Int. Symp. Comput. Technol. Inf. Sci.*, Aug. 2023, pp. 1162–1165.

[33] W. Qin, H. Wang, F. Zhang, J. Wang, X. Cao, and X.-L. Zhao, “Tensor ring decomposition-based generalized and efficient nonconvex approach for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5539818.

[34] 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.

[35] 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.

[36] M. Feng, Y. Yang, X. Shao, and Q. Shu, “Hyperspectral anomaly detection based on multicomplementary prior-guided tensor decomposition,” *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 5035117.

[37] D. Ma, Z. Liu, and Z. Jiang, “Variation autoencoder of spatial-spectral joint mask for hyperspectral anomaly detection,” *IEEE Signal Process. Lett.*, vol. 32, pp. 1535–1539, 2025.

[38] J. Hu, W. Zheng, R. Wang, and M. Zhao, “A band-selected and regularized network for hyperspectral anomaly detection,” *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 5037514.

[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] Y. Wang, H. Wang, E. Zhao, M. Song, and C. Zhao, “Tucker decomposition-based network compression for anomaly detection with large-scale hyperspectral images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 10674–10689, 2024.

[41] R. Zhao, Z. W. Yang, X. C. Meng, and F. Shao, “A novel fully convolutional auto-encoder based on dual clustering and latent feature adversarial consistency for hyperspectral anomaly detection,” *Remote Sens.*, vol. 16, no. 4, 2024, Art. no. 717.

[42] Z. W. Yang et al., “A multi-scale mask convolution-based blind-spot network for hyperspectral anomaly detection,” *Remote Sens.*, vol. 16, no. 16, 2024, Art. no. 3036.

[43] 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.

[44] M. L. Brandão Junior, V. C. Lima, T. A. P. P. Teixeira, E. R. de Lima, and R. D. Lopes, “Anomaly detection in hyperspectral images via regularization by denoising,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 7600–7600, 2023.

[45] Z. He, D. He, M. Xiao, A. Lou, and G. Lai, “Convolutional transformer-inspired autoencoder for hyperspectral anomaly detection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5508905.

[46] 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.

[47] D. Gong et al., “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jan. 2019, pp. 1705–1714.

[48] T. Zhang, S. Li, B. Chen, H. Yuan, and C. L. P. Chen, “AIA-Net: Adaptive interactive attention network for text–Audio emotion recognition,” *IEEE Trans. Cybern.*, vol. 53, no. 12, pp. 7659–7671, Dec. 2023.

[49] X. Lu, W. Zhang, and J. Huang, “Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, pp. 1527–1537, 2020.

[50] K. Kayabol, E. B. Aytekin, S. Arisoy, and E. E. Kuruoglu, “Skewed T-distribution for hyperspectral anomaly detection based on autoencoder,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5510705.

[51] R. Gade and T. B. Moeslund, “Thermal cameras and applications: A survey,” *Mach. Vis. Appl.*, vol. 25, no. 1, pp. 245–262, Jan. 2014.

[52] E. Zhao et al., “An operational land surface temperature retrieval methodology for Chinese second-generation Huanjing disaster monitoring satellite data,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1283–1292, 2022.

[53] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, “Unsupervised pixel-wise hyperspectral anomaly detection via autoencoding adversarial networks,” *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5502905.

[54] Z. Tu, X. Liu, and X. Xiao, “A general dynamic knowledge distillation method for visual analytics,” *IEEE Trans. Image Process.*, vol. 31, pp. 6517–6531, 2022.

[55] Z. Li, P. Xu, Z. Dong, R. Zhang, and Z. Deng, “Feature-level knowledge distillation for place recognition based on soft-hard labels teaching paradigm,” *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 2, pp. 2091–2101, Feb. 2025.

[56] 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.*, vol. 35, no. 6, pp. 8358–8372, Jun. 2024.

[57] J. Ma, W. Xie, J. Lei, L. Fanget, and Y. Li, “End-to-end spectral-spatial cooperative autoencoding density estimation model,” *Acta Electronica Sinica*, vol. 51, no. 4, pp. 1006–1020, Apr. 2023.

[58] 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.

Enyu Zhao (Member, IEEE) was born in Dalian, Liaoning Province, China, in 1987. He received the Ph.D. degree in cartography and geographic information systems from the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, in 2017.

From 2014 to 2016, he was a joint Ph.D. Student with Engineering Science, Computer Science, and Imaging Laboratory, University of Strasbourg, Strasbourg, France. He is currently an Associate Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. His research interests include quantitative remote sensing and hyperspectral image processing.

Hao Zhang was born in Linyi, Shandong Province, China, in 2000. He received the B.S. degree in computer science and technology from Yantai University, Yantai, China, in 2023. He is currently working toward the M.S. degree in computer technology with Dalian Maritime University, Dalian, China.

His research interests include hyperspectral image anomaly detection and deep learning.

Nianxin Qu (Student Member, IEEE) was born in Liaoyang, Liaoning Province, China, in 1999. He received the M.S. degree in computer science and technology in 2024 from Dalian Maritime University, Dalian, China, where he is currently working toward the Ph.D. degree in computer science and technology.

His research interests include hyperspectral image processing and deep learning.

Yulei Wang (Member, IEEE) was born in Yantai, Shandong Province, China, in 1986. She received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

From 2011 to 2013, she was a joint Ph.D. student with the Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County. From 2011 to 2013, she was a Research Assistant with the Shock, Trauma and Anesthesiology Research Organized Research Center (STAR-ORC), School of Medicine, University of Maryland. She is currently an Associate Professor and Doctoral Supervisor with Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. Her current research interests include hyperspectral image processing, multisource remote sensing fusion, and vital signs signal processing. More details could be found at https://YuleiWang1.github.io/.

Yongguang Zhao received the B.S. degree in geographic information systems from Central South University, Changsha, China, in 2009, and the Ph.D. degree in signal and information processing from the Academy of Opto-Electronics, Chinese Academy of Sciences (CAS), Beijing, China, in 2015.

He is currently an Associate Professor with the Aerospace Information Research Institute, CAS. His research interests include radiometric calibration and quantitative remote sensing applications.