Towards Cross-Scene Hyperspectral Anomaly Detection via a One-Step Generative Framework

Cuiwei Liu
School of Computer Science
Shenyang Aerospace University
Shenyang, China
liucuiwei@sau.edu.cn

Yujing Zhao
School of Computer Science
Shenyang Aerospace University
Shenyang, China
zhaoyujing@stu.sau.edu.cn

Zhaokui Li*
School of Computer Science
Shenyang Aerospace University
Shenyang, China
lzk@sau.edu.cn

Huaijun Qiu
School of Computer Science
Shenyang Aerospace University
Shenyang, China
20220071@email.sau.edu.cn

***Abstract*—Hyperspectral Anomaly Detection (HAD) aims to identify small and sparse targets that significantly deviate from the abundant background spectra, without requiring prior knowledge. Most state-of-the-art methods follow a reconstruct-then-detect paradigm, primarily learning scene-specific background distributions. Consequently, the well-trained models suffer from severe performance degradation when applied to unseen scenes. To this end, this study introduces a novel one-step detection paradigm that formulates HAD as a generative conditional diffusion process. We develop a Conditional Diffusion Model-based Hyperspectral Anomaly Detection framework (CDM-HAD), which leverages multi-scale hyperspectral features as conditional guidance to progressively reconstruct anomaly maps from noisy inputs. The iterative denoising process implicitly learns generalizable anomaly-background deviation patterns and alleviates overfitting to training-specific distributions. To overcome data scarcity, the proposed CDM-HAD incorporates an Anomaly Simulation Strategy (ASS) for generating high-quality hyperspectral anomaly data. This approach perceives inherent anomalies in the training scene and adaptively injects cross-scene anomalous spectra, thereby synthesizing diverse and realistic anomaly samples. Experimental results highlight the superior cross-scene generalization of the proposed CDM-HAD, achieving robust anomaly detection across diverse unseen scenes when trained on a single scene.**

***Index Terms*—Hyperspectral anomaly detection, conditional diffusion, anomaly simulation strategy.**

I. Introduction

Hyperspectral Imaging technology [1] acquires radiation information in hundreds of narrow continuous spectral bands, generating a fine and continuous spectral response curve for each spatial pixel. Hyperspectral Anomaly Detection (HAD) aims to identify small, sparse targets exhibiting distinct spectral characteristics against extensive backgrounds in a Hyperspectral Image (HSI), typically without relying on reference spectra [2]. This technology plays a pivotal role in various applications, including military camouflage recognition [3], environmental pollution monitoring [4], unexploded ordnance detection [5], and early warning of forest fire [6].

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant No.62171295, and in part by the Natural Science Foundation of Liaoning Province, China under Grant No. 2025-MSLH-561, and in part by the High Level Talent Research Start-up Fund of Shenyang Aerospace University under Grant No.23YB03.

Corresponding author: Zhaokui Li. E-mail: lzk@sau.edu.cn

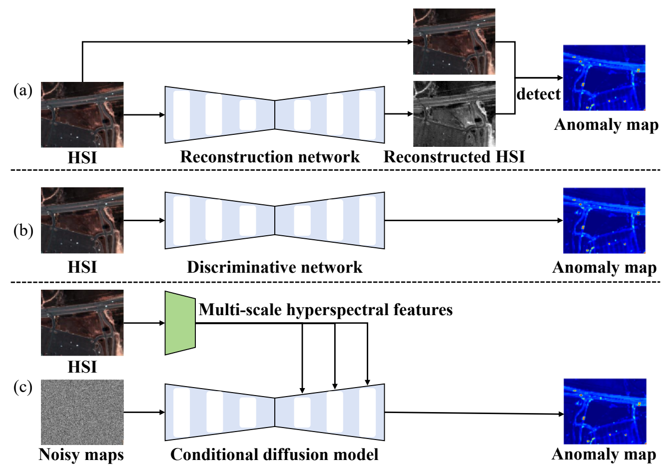


Fig. 1: Comparison of HAD frameworks. (a) Reconstruct-then-detect paradigm. (b) TDD [7]. (c) Our proposed CDM-HAD.

Early work in HAD focus on statistical modeling of background spectra. The Reed-Xiaoli (RX) detector [2] characterizes background spectra with multivariate Gaussian distributions and utilizes distance metrics for anomaly identification. Another line of research [8], [9] adopt representation-based learning to exploit the intrinsic low-dimensional manifold structure of background data.

Recently, deep AutoEncoders (AEs) have emerged as powerful HAD methods, typically following a two-step detection paradigm that first reconstructs the background spectra and then applies post-processing techniques to detect anomalies (see Fig. 1(a)). Among them, Auto-AD [10] trains a fully convolutional AE on HSI data to learn low-dimensional manifold representations, enabling precise reconstruction of the majority background spectra while identifying sparse anomalies through their high reconstruction errors caused by the deviation from the learned manifold. Following works [11], [12] further enforce AEs to reconstruct center pixels exclusively from their surrounding context, thereby significantly amplifying reconstruction errors for anomalies. Despite their effectiveness within individual scenes, Li et al. [7] have demonstrated that the reconstruct-then-detect paradigm exhibits limited generalization capability towards new scenes. This is primarily because they are designed to learn scene-specific background spectral distribution rather than to detect anomalies, making them prone to overfitting to the training data. Since retraining for each new scene is computationally expensive and often impractical, real-world deployment remains challenging.

As shown in Fig. 1(b), Li et al. [7] proposed the Transferred Direct Detection (TDD) model, a one-step detection paradigm that directly projects HSIs into anomaly maps within a discriminative framework. Specifically, TDD first synthesizes HSI-anomaly pairs through random augmentation techniques, then trains a U-Net enhanced with global and local self-attention modules in a supervised manner. Unlike two-step detection methods [10]–[13] that rely on background reconstruction, TDD aims to learn anomaly-background deviation characteristics through discriminative modeling. Despite these advances, the one-step detection paradigm still faces critical challenges. First, HAD networks should learn generalizable anomaly-background spectral deviation representations, avoiding overfitting to specific anomaly-background patterns in the training data. Second, the scarcity of real anomaly data necessitates the synthesis of highly realistic and precisely labeled anomaly data to enable supervised learning from HSIs to anomaly maps. TDD [7] synthesizes anomalies via noise injection and spectral random shuffling, which fundamentally compromises spectral continuity. As a result, the synthesized training samples fail to capture the spectral characteristics of real anomalies.

Early studies [14] have demonstrated that, unlike discriminative models which tend to overfit scene-specific decision boundaries, diffusion models learn the underlying data distribution through iterative denoising and exhibit superior generalization capability across diverse data distributions. Inspired by recent advances in latent diffusion models [15], this study proposes a novel Conditional Diffusion Model-based Hyperspectral Anomaly Detection framework (CDM-HAD), which achieves anomaly detection through a conditional denoising process. As shown in Fig. 1(c), CDM-HAD progressively recover anomaly maps from random noise, conditioned on multi-scale hyperspectral features. By learning anomaly distributions through iterative refinement, CDM-HAD implicitly captures the anomaly-as-background-deviation principle, thus naturally adapting to diverse anomaly-background patterns across scenes. During training, we develop a novel Anomaly Simulation Strategy (ASS) to generate diverse anomaly samples (e.g., HSI-anomaly pairs) that reflect real-world anomaly-background patterns. This approach first identifies salient anomalies that significantly deviate from the background distribution in the training scene, then adaptively injects realistic anomaly spectra from other scenes guided by the anomaly-background spectral deviation principle. Our main contributions are summarized as follows:

1) The hyperspectral anomaly detection problem is reframed as a generative conditional diffusion process, and a novel one-step detection paradigm is proposed for directly generating anomaly maps from a noise input.
2) A conditional diffusion model is designed to implicitly learn generalizable anomaly-background deviation patterns through iterative denoising guided by multi-scale hyperspectral features, thereby effectively mitigating overfitting to scene-specific distributions.
3) An anomaly simulation strategy is proposed, whereby diverse and high-fidelity data for the supervised training of the generative model are synthesized by adaptively implanting cross-scene anomaly spectra while preserving the inherent anomalies of the original scene.

We will make our CDM-HAD method's data and code publicly available upon acceptance.

II. Related Work

A. Traditional methods

As the first HAD method, RX detector [2] assumes a Gaussian background and uses the sample covariance and mean to measure pixel deviations. Following this statistical paradigm, KRX [16] maps spectra into a high-dimensional space to linearize inseparable patterns. Representation-based approaches decompose an HSI into low-rank background and sparse anomaly components. CRD [8] uses ℓ_2-norm constrained collaborative representation for efficiency. Li et al. [17] propose a low-rank and sparse method with a Gaussian mixture model, variational Bayesian inference, and a Manhattan-distance detector for noise-robust detection. Liu et al. [18] present a tensor multi-subspace learning framework with nonconvex low-rank regularization, total variation for spatial smoothness, and a Laplace-based strategy to suppress background and enhance target discrimination.

Despite their considerable advancements, traditional HAD methods often struggle to characterize the complex, non-linear distributions in high-dimensional HSI.

B. Deep learning-based methods

In recent years, a series of deep learning-based HAD methods [19] have emerged by leveraging their powerful feature extraction and representation capabilities. These approaches typically follow a reconstruct-then-detect paradigm. The first stage learns the background distribution by reconstructing the input HSI, while suppressing the reconstruction of anomalies. Subsequently, the residual between the reconstructed HSI and the input HSI is employed as an anomaly indicator under the assumption that anomalies generally lead to higher reconstruction errors.

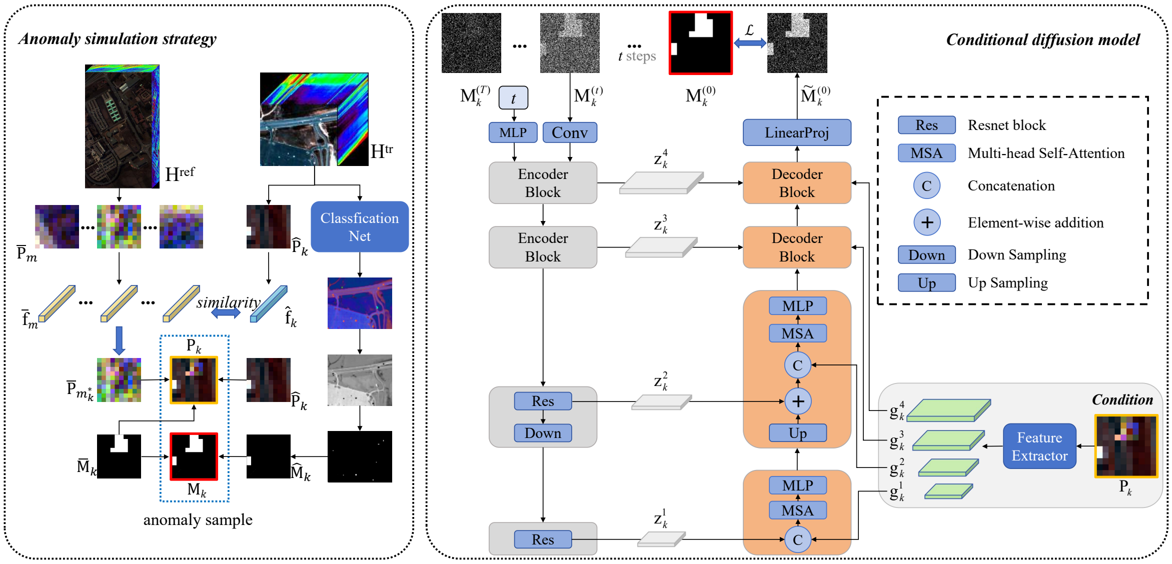


Fig. 2: Training phase of the proposed CDM-HAD.

Fan et al. [20] introduced an $\ell_{2,1}$-norm and superpixel graph into an autoencoder for robustness and spatial consistency. Gao et al. [11], [12] proposed blind-spot reconstruction networks that prevent identity mapping by reconstructing each pixel from its neighbors, and further designed a sliding dual-window network using only background information. Wang et al. [21] extended this with a self-supervised blind-spot network coupling local and non-local features for global-local anomaly suppression. Lian et al. [22] developed a gated transformer with dual branches to dynamically balance anomaly suppression and background amplification.However, these methods focus on learning scene-specific background spectral distribution and exhibit limited generalization capability towards new scenes. Under the reconstruct-then-detect paradigm, the Anomaly Enhancement Transformation Network (AETNet) [13] enhances cross-scene generalization by incorporating diverse training scenes and data augmentations (e.g., random masking). TDD [7] enhances cross-scene generalization by establishing a one-step HAD paradigm, which learns the direct mapping from an input HSI to its corresponding anomaly map through a discriminative network. Our proposed CDM-HAD formulates HAD as a conditional diffusion process and builds a one-step generative framework. It iteratively refines anomaly maps from noise using multi-scale features as guidance, alleviating overfitting and boosting cross-scene generalization.

III. Method

This study addresses the challenge of cross-scene generalization in HAD by proposing a novel CDM-HAD framework. A Conditional Diffusion Model (CDM) is trained on a single source HSI $\mathbf{H}^{\text{tr}}$ and then evaluated on multiple unseen HSIs $\{\mathbf{H}^{\text{te}}_n\}_{n=1}^{N}$ from different scenes without any retraining or fine-tuning. As illustrated in Fig. 2, we first employ ASS to synthesize diverse training samples with anomaly labels. Then the CDM learns to refine noise input into detection maps under multi-scale hyperspectral feature guidance.

A. Anomaly simulation strategy

The training HSI is denoted as $\mathbf{H}^{\text{tr}} \in \mathbb{R}^{W\times H\times B}$, where W, H, and B indicate width, height, and spectral band count, respectively. Our goal is to construct a training dataset $\mathcal{D} = \{(\mathbf{P}_k, \mathbf{M}_k)\}_{k=1:K}$, where $\mathbf{P}_k \in \mathbb{R}^{w\times h\times B}$ is a hyperspectral patch of spatial dimensions $w \times h$, and $\mathbf{M}_k \in \{0,1\}^{w\times h}$ denotes its binary anomaly labels.

Existing anomaly simulation methods [7] rely on random augmentations, which ignore inherent anomalies in the training HSI and produce unrealistic spectral patterns. To address this, we first preliminarily identify potential anomalies in $\mathbf{H}^{\text{tr}}$, then adaptively inject realistic anomalous spectra from other scenes guided by spectral differences. Specifically, we use a pre-trained network [23] to extract deep features from $\mathbf{H}^{\text{tr}}$, normalize them via Z-score, and apply a high threshold to obtain a preliminary anomaly map $\mathbf{M}^{\text{tr}}$. Then, $\mathbf{H}^{\text{tr}}$ is divided into non-overlapping patches $\{\hat{\mathbf{P}}_k \in \mathbb{R}^{w\times h\times B}\}_{k=1:K}$ with corresponding labels $\{\hat{\mathbf{M}}_k \in \{0,1\}^{w\times h}\}_{k=1:K}$.

To ensure spectral diversity in synthetic anomalies, we leverage a reference HSI $\mathbf{H}^{\text{ref}} \in \mathbb{R}^{W'\times H'\times B}$ captured from a distinct scene. From it, a collection of patches $\bar{\mathcal{D}}^a = \{\bar{\mathbf{P}}_m \in \mathbb{R}^{w\times h\times B}\}_{m=1:M}$ are extracted to serve as multiple anomaly sources. For a specific target patch $\hat{\mathbf{P}}_k$, we randomly sample a candidate subset $\bar{\mathcal{D}}^a_k$ from $\bar{\mathcal{D}}^a$ and calculate its cosine similarity to each candidate patch. The least similar candidate $\bar{\mathbf{P}}_{\text{m}^*_k}$ is selected as the anomaly source for $\hat{\mathbf{P}}_k$. This process can be formulated by

$$\text{m}^*_k = \arg\min_{m\in\bar{\mathcal{D}}^a_k} \frac{\hat{\mathbf{f}}_k \cdot \bar{\mathbf{f}}_m}{\|\hat{\mathbf{f}}_k\|_2 \|\bar{\mathbf{f}}_m\|_2}, \quad (1)$$

where $\hat{\mathbf{f}}_k$ and $\bar{\mathbf{f}}_m$ indicate the features of the target patch $\hat{\mathbf{P}}_k$ and a candidate patch $\bar{\mathbf{P}}_m$, respectively.

In the final step, the target patch $\hat{\mathbf{P}}_k$ is augmented by injecting anomalous spectra from the selected candidate $\bar{\mathbf{P}}_{\mathrm{m}_k^*}$. To precisely control the spatial extent of anomaly injection, we create a binary mask $\bar{\mathbf{M}}_k \in \{0,1\}^{w\times h}$, which is initialized with zeros. A randomly positioned square subregion is then activated by setting its values to 1, marking the candidate pixels for anomaly injection. Subsequent application of random affine transformations to this mask generates irregularly-shaped injection regions with realistic geometric variations, thereby enhancing the diversity of synthesized anomalies. Following $\bar{\mathbf{M}}_k$, anomalous spectra are injected into $\hat{\mathbf{P}}_k$, producing a training hyperspectral sample $\mathbf{P}_k$ with realistic anomalies.

$$\mathbf{P}_k = (1-\bar{\mathbf{M}}_k)\odot\hat{\mathbf{P}}_k + \bar{\mathbf{M}}_k\odot\bar{\mathbf{P}}_{\mathrm{m}_k^*}, \tag{2}$$

where $\odot$ denotes element-wise multiplication. The ground-truth anomaly label $\mathbf{M}_k$ corresponding to $\mathbf{P}_k$ is formed by

$$\mathbf{M}_k = \bar{\mathbf{M}}_k \vee \hat{\mathbf{M}}_k, \tag{3}$$

where $\vee$ denotes the logical OR operation. $\mathbf{M}_k$ is the union of $\bar{\mathbf{M}}_k$ and $\hat{\mathbf{M}}_k$, where the former masks the injected anomalies, while the latter annotates inherent anomalies present in the original hyperspectral patch.

B. Conditional diffusion model

With high-quality anomaly samples $\mathcal{D} = \{(\mathbf{P}_k,\mathbf{M}_k)\}_{k=1:K}$, the CDM learns the underlying distribution of the anomaly labels, conditioned on multi-scale hyperspectral features extracted from the HSI samples. This network reverses a forward diffusion process, progressively removing noise from corrupted inputs to restore clean anomaly labels.

1) Forward and reverse diffusion processes: The core of the diffusion model lies in the forward diffusion and reverse denoising processes. The forward process, denoted by q, gradually injects Gaussian noise into the clean anomaly label $\mathbf{M}_k$ (which we denote as $\mathbf{M}_k^{(0)}$) over T discrete timesteps. The state at any timestep t is given by

$$q(\mathbf{M}_k^{(t)}|\mathbf{M}_k^{(0)}) = \mathcal{N}(\mathbf{M}_k^{(t)}|\sqrt{\bar{\alpha}_t}\cdot\mathbf{M}_k^{(0)},(1-\bar{\alpha}_t)\cdot\mathbf{I}), \tag{4}$$

$$\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s, \tag{5}$$

where $\mathbf{M}_k^{(t)}$ is the noisy label at timestep t and $\{\alpha_t\}_{t=1}^T$ is a predefined noise schedule. $\mathcal{N}$ is the normal distribution. This operation enables one-step sampling of $\mathbf{M}_k^{(t)}$ at timestep t.

The reverse process aims to learn the posterior distribution $p_\theta(\mathbf{M}_k^{(t-1)}|\mathbf{M}_k^{(t)},\mathbf{P}_k)$, where the training sample $\mathbf{P}_k$ serves as a powerful condition. A denoising network $\mathcal{F}_\theta$ is optimized to predict the clean label $\tilde{\mathbf{M}}_k^{(0)}$ from the noisy input $\mathbf{M}_k^{(t)}$.

$$\tilde{\mathbf{M}}_k^{(0)} = \mathcal{F}_\theta(\mathbf{M}_k^{(t)},t,\mathbf{P}_k), \tag{6}$$

where θ represents learnable parameters of the network.

2) Network architecture: As shown in Fig. 2, the denoising network $\mathcal{F}_\theta$ uses a U-Net with a 4-level encoder-decoder, injecting multi-scale hyperspectral features as conditional information. For a training sample $\mathbf{P}_k$ with anomaly label $\mathbf{M}_k$, the encoder extracts multi-scale features $\{\mathbf{z}_k^i\}_{i=1:4}$ from the noisy input $\mathbf{M}_k^{(t)}$ via ResNet blocks and downsampling. The decoder recovers the clean anomaly map by integrating $\mathbf{z}_k^i$ and hyperspectral features $\mathbf{g}_k^i$ extracted from $\mathbf{P}_k$. In each decoder block, upsampled inputs are fused with $\mathbf{z}_k^i$ and $\mathbf{g}_k^i$ via addition and concatenation, then enhanced by Multi-head Self-Attention (MSA) and MLP. This hierarchical aggregation progressively reconstructs the anomaly map while preserving structural consistency across resolutions.

$$\mathbf{r}_k^i = \begin{cases}\text{MLP}(\text{MSA}(\text{Concat}(\mathbf{g}_k^i,\mathbf{z}_k^i + UP(\mathbf{r}_k^{i-1})))), if \quad i>1\\ \text{MLP}(\text{MSA}(\text{Concat}(\mathbf{g}_k^i,\mathbf{z}_k^i)), if \quad i=1\end{cases} \tag{7}$$

where $\mathbf{r}_k^i$ denotes output of the i-th decoder block. Finally, a linear projection layer reduces the output of the top-level decoder block to one dimension, producing the predicted anomaly map $\tilde{\mathbf{M}}_k^{(0)}$.

3) Training loss: The denoising network is trained to reconstruct the clean anomaly label $\mathbf{M}_k^{(0)}$ from a noisy input $\mathbf{M}_k^{(t)}$. The standard Binary Cross-Entropy (BCE) loss function is employed to measure the discrepancy between the predicted label map $\tilde{\mathbf{M}}_k^{(0)}$ and the ground-truth label $\mathbf{M}_k^{(0)}$. The overall training objective is to minimize BCE loss over all timesteps, training samples, and noise instances by

$$\mathcal{L} = \mathbb{E}_{t,(\mathbf{P}_k,\mathbf{M}_k)\in\mathcal{D}}\left[BCE(\mathbf{M}_k^{(0)},\mathcal{F}_\theta(\mathbf{M}_k^{(t)},t,\mathbf{P}_k))\right]. \tag{8}$$

C. Inference

During inference, each test HSI is divided into non-overlapping patches. For each patch, CDM-HAD performs a τ-step iterative denoising process that progressively recovers the anomaly map from noisy labels under spatial-spectral feature guidance. The process starts with random Gaussian noise, which the learned denoising network $\mathcal{F}_\theta$ converts into a coarse anomaly map. In subsequent steps, the network refines the map using the previous output and spatial-spectral conditioning. This iterative refinement yields high-quality anomaly detection results.

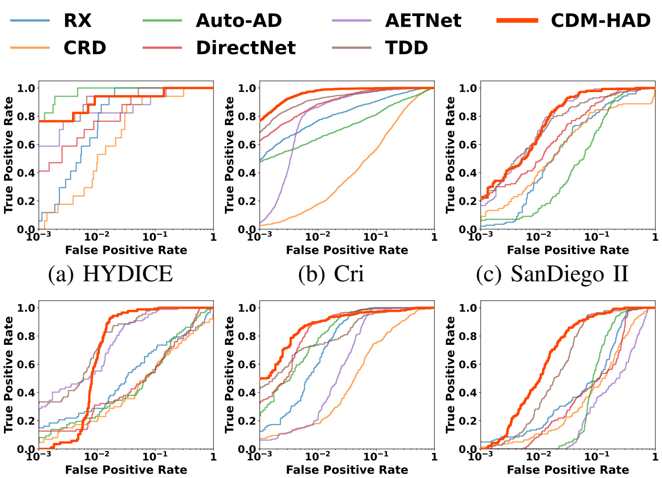
IV. Experiments

A. Datasets and settings

Experiments were conducted on six publicly available datasets: HYDICE, Cri, SanDiego II, ABU-Airport-2 (A2), ABU-Urban-4 (U4), and MUCAD [24]. These datasets exhibit significant diversity in background spectra (e.g., urban, vegetation, ocean) and anomaly targets (e.g., vehicles, buildings, man-made objects), providing solid benchmarks for assessing the HAD performance across different scenes. Specifically, the proposed CDM-HAD is trained on HYDICE and subsequently tested on the other five datasets without fine-tuning. To ensure realistic anomaly simulation, we select the Pavia University dataset as the reference HSI for anomaly sources, leveraging

TABLE I: AUC score of different methods on five datasets.

Method	HYDICE	Cri	SanDiego II	A2	U4	MUCAD	Average(test)
RX [2]	0.9931	0.9675	0.9362	0.8408	0.9887	0.8793	0.9225
CRD [8]	0.9664	0.8549	0.8681	0.7881	0.9065	0.8204	0.8476
Auto-AD [10]	**0.9992**	0.9264	0.9185	0.8516	0.9877	0.8928	0.9154
DirectNet [12]	0.9854	0.9815	0.9582	0.8534	0.9911	0.8538	0.9276
AETNet [13]	0.9827	0.9897	0.9859	0.9823	0.9725	0.7213	0.9303
TDD [7]	0.9960	0.9915	0.9728	0.9882	0.9763	0.9243	0.9706
CDM-HAD	0.9904	**0.9986**	**0.9869**	**0.9889**	**0.9923**	**0.9658**	**0.9865**



(a) HYDICE (b) Cri (c) SanDiego II
(d) A2 (e) U4 (f) MUCAD

Fig. 3: ROC curves on different datasets.

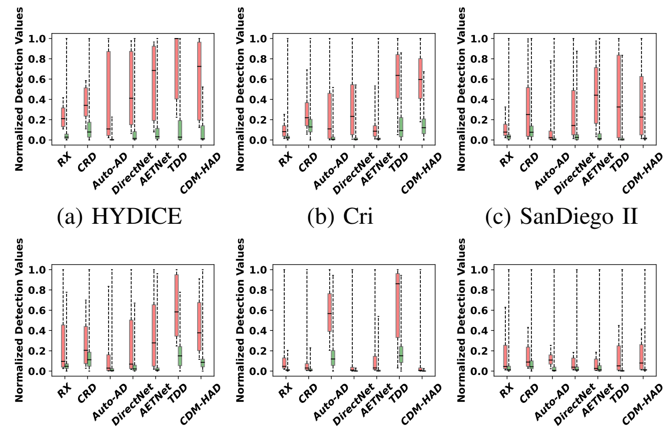
its spectral dissimilarity from the training background and high class diversity.

The HAD performance is evaluated using ROC-AUC, where a higher AUC score indicates superior detection performance. We uniformly adjust all HSIs to $B = 162$ spectral bands via linear interpolation. In the z-score algorithm used to generate the preliminary anomaly map M^{tr}, the top 0.3% of pixels are identified as anomalies. This threshold is consistent with the characteristic definition of anomalies, which assumes a low probability of occurrence and a small spatial proportion (typically less than 1%). Following standard diffusion model configurations, we set the total time step T to 1000 and conduct training over 75 epochs.

B. Experimental results and analysis

Our proposed CDM-HAD is compared with existing HAD methods, spanning three categories: statistical distribution-based [2], representation-based [8], and deep learning-based [7], [10], [12], [13] approaches. Notably, RX [2], CRD [8], Auto-AD [10], and DirectNet [12] require training from scratch on each test scene. In contrast, AETNet [13], TDD [7] and our proposed CDM-HAD directly evaluate the learned model on test scenes without any retraining or fine-tuning. To be more specific, AETNet is trained on a set of clean HSIs, while both TDD and CDM-HAD employ the HYDICE dataset for training.

Table I summarizes the AUC scores of various methods across six datasets. Fig. 3 and Fig. 4 depict ROC curves and separability maps, respectively. Our proposed CDM-HAD achieves a remarkable AUC score of 0.9904 on the training HSI (i.e., HYDICE) and demonstrates superior generalization



(a) HYDICE (b) Cri (c) SanDiego II
(d) A2 (e) U4 (f) MUCAD

Fig. 4: Separability maps on different datasets.

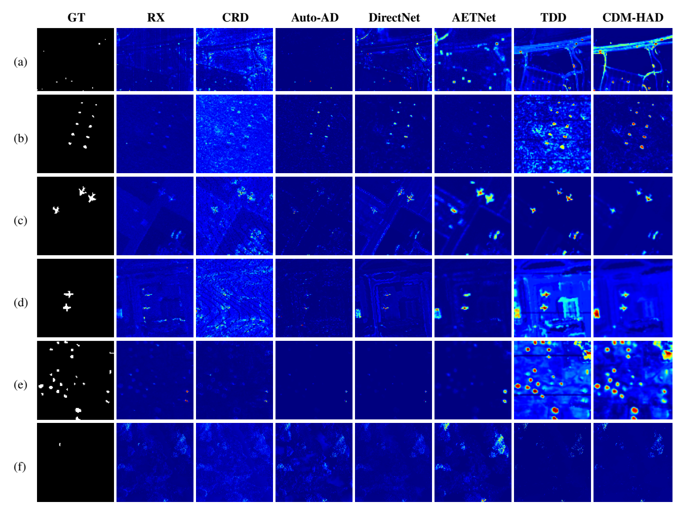


Fig. 5: Anomaly detection maps on (a) HYDICE, (b) Cri, (c) SanDiego II, (d) A2, (e) U4, and (f) MUCAD.

capability in cross-scene evaluation, surpassing AETNet by 5.62% and TDD by 1.59% in average AUC scores. In comparison with HAD methods [2], [8], [10], [12] that require training from scratch on each test scene, our CDM-HAD also demonstrates superior detection performance across all unseen scenes (i.e., Cri, SanDiego II, A2, U4, and MUCAD).

Figure 5 visualizes anomaly detection maps across different scenes. Quantitative results show that CDM-HAD achieves a superior trade-off between detection sensitivity and false alarm suppression in unseen scenes. Its robust performance across diverse scenes indicates that CDM-HAD captures generalizable anomaly-background characteristics rather than overfitting to training-specific distributions.

C. Ablation study

Ablation experiments evaluate two key components: the ASS strategy for generating training samples and the CDM for anomaly detection. Baseline methods are constructed by combining the two data generation strategies (ASS and random spectral shuffling) with the two detection models (CDM

TABLE II: Ablation study (AUC) on two key components.

Model Setting	Cri	SanDiego II	A2	U4	MUCAD	Average
Spectral Shuffling + Discriminative Net	0.9732	0.9538	0.9814	0.8794	0.9243	0.9381
ASS + Discriminative Net	0.9974	0.9678	0.9886	0.9893	0.9635	0.9813
Spectral Shuffling + CDM	0.9958	0.9793	0.9864	0.9863	0.9246	0.9745
ASS + CDM	**0.9986**	**0.9869**	**0.9889**	**0.9923**	**0.9658**	**0.9865**

TABLE III: Ablation study (AUC score) on the number of iteration steps during test.

τ	Cri	SanDiego II	A2	U4	MUCAD	Average
1	0.9971	0.9433	0.9856	0.9907	0.8317	0.9497
5	0.9986	0.9864	0.9889	0.9922	0.9657	0.9864
10	**0.9986**	**0.9869**	**0.9889**	**0.9923**	0.9658	**0.9865**
20	0.9968	0.9859	0.9879	0.9923	**0.9659**	0.9858

and the discriminative net [7]). As shown in Table II, our full approach (ASS + CDM) achieves the best performance. Compared to random shuffling, ASS substantially improves detection by synthesizing realistic and diverse HSI-anomaly pairs. Moreover, CDM learns anomaly-background patterns conditioned on hyperspectral features, outperforming the discriminative net across unseen scenes.

During test, the number of iteration steps (τ) controls denoising iterations for recovering anomaly maps from noise. Table III reports results with varyingτ. A single step already yields satisfactory performance on the Cri, A2, and U4 datasets, confirming strong generalization. Increasing to $\tau = 5, 10$ further refines subtle anomalies and target boundaries, notably improving results on SanDiego II and MUCAD. However, excessive steps raise computational cost and noise sensitivity, increasing false alarms. Thus, τ should be chosen to balance accuracy and efficiency.

V. Conclusion

This study proposes CDM-HAD, a one-step generative framework that formulates hyperspectral anomaly detection as a conditional diffusion process. It addresses two key challenges: cross-scene generalization and scarcity of realistic anomaly samples. By using multi-scale hyperspectral features as conditional guidance, it refines anomaly maps from noise input, alleviating overfitting to scene-specific distributions. The Anomaly Simulation Strategy (ASS) enables realistic and diverse training sample generation. Experiments show state-of-the-art cross-scene performance, and ablation studies confirm the individual effectiveness of both ASS and CDM.

References

[1] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, no. 1, pp. 64–90, 2021.

[2] I. S. Reed and X. Yu, "Adaptive multiple-band cfar detection of an optical pattern with unknown spectral distribution," *IEEE transactions on acoustics, speech, and signal processing*, vol. 38, no. 10, pp. 1760–1770, 1990.

[3] J. Zhao, B. Zhou, G. Wang, J. Ying, J. Liu, and Q. Chen, "Spectral camouflage characteristics and recognition ability of targets based on visible/near-infrared hyperspectral images," in *Photonics*, vol. 9, no. 12. MDPI, 2022, p. 957.

[4] W. Ma, C. Xing, W. Wang, Q. Zhang, Z. Wang, Y. Yang, and C. Liu, "Ozone pollution monitoring using a full-time hyperspectral tomography system for multiple air pollutants," *Nature Communications*, 2025.

[5] F. Samadzadegan, F. Qaderi, T. L. Haghighi, A. Toosi, and S. M. Mousavi, "Unmanned aerial remote sensing for unexploded ordnance detection: Concepts, methods, current status and trends," *Sensing and Imaging*, vol. 26, no. 1, p. 137, 2025.

[6] R. U. Shaik, G. Laneve, and L. Fusilli, "An automatic procedure for forest fire fuel mapping using hyperspectral (prisma) imagery: A semi-supervised classification approach," *Remote Sensing*, vol. 14, no. 5, p. 1264, 2022.

[7] J. Li, X. Wang, S. Wang, H. Zhao, and Y. Zhong, "One step detection paradigm for hyperspectral anomaly detection via spectral deviation relationship learning," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[8] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Transactions on geoscience and remote sensing*, vol. 53, no. 3, pp. 1463–1474, 2014.

[9] Y. Wang, W. Li, Y. Gui, H. Xie, and L. Zhang, "A generalized non-convex surrogated framework for anomaly detection on blurred hyperspectral images," *IEEE Transactions on Image Processing*, 2025.

[10] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-ad: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2021.

[11] 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 Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[12] D. Wang, L. Zhuang, L. Gao, X. Sun, X. Zhao, and A. Plaza, "Sliding dual-window-inspired reconstruction network for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[13] Z. Li, Y. Wang, C. Xiao, Q. Ling, Z. Lin, and W. An, "You only train once: Learning a general anomaly enhancement network with random masks for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–18, 2023.

[14] Z. Kadkhodaie, F. Guth, E. P. Simoncelli, and S. Mallat, "Generalization in diffusion models arises from geometry-adaptive harmonic representations," *arXiv preprint arXiv:2310.02557*, 2023.

[15] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 684–10 695.

[16] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.

[17] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of gaussian for hyperspectral anomaly detection," *IEEE Transactions on Cybernetics*, vol. 51, no. 9, pp. 4363–4372, 2021.

[18] S. Liu, C. Zhu, D. Ran, and G. Wen, "Anomaly detection via tensor multisubspace learning and nonconvex low-rank regularization," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8178–8190, 2023.

[19] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. Plaza, "Bocknet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.

[20] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[21] D. Wang, L. Ren, X. Sun, L. Gao, and J. Chanussot, "Non-local and local feature-coupled self-supervised network for hyperspectral anomaly detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[22] J. Lian, L. Wang, H. Sun, and H. Huang, "Gt-had: Gated transformer for hyperspectral anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 36, no. 2, pp. 3631–3645, 2025.

[23] Z. Zhao, X. Xu, S. Li, and A. Plaza, "Hyperspectral image classification using groupwise separable convolutional vision transformer network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–17, 2024.

[24] T. Hupel and P. Stütz, "Adopting hyperspectral anomaly detection for near real-time camouflage detection in multispectral imagery," *Remote Sensing*, vol. 14, no. 15, p. 3755, 2022.