

Article

Background Suppression by Multivariate Gaussian Denoising Diffusion Model for Hyperspectral Target Detection

Weile Han ¹, Yuteng Huang ¹, Jiaqi Feng ², Rongting Zhang ¹  and Guangyun Zhang ^{1,*}

¹ School of Geomatics Science and Technology, Nanjing Tech University, Nanjing 211800, China; hwl@njtech.edu.cn (W.H.); huangyt0812@njtech.edu.cn (Y.H.); zrt@njtech.edu.cn (R.Z.)

² The School of Astronautics, Beihang University, Beijing 100191, China; fengjiaqi@buaa.edu.cn

* Correspondence: zgy@njtech.edu.cn

Highlights

What are the main findings?

- A diffusion model-based hyperspectral target detection (HTD) method is developed, which introduces multivariate Gaussian-distributed background noise into the diffusion process for effective background suppression.
- A spatial-spectral centre-weighted strategy combined with superpixel segmentation is proposed to accurately extract background noise and enhance spectral-spatial feature representation.

What is the implication of the main finding?

- The proposed method significantly improves target detection accuracy and background suppression in complex scenes.
- Experimental results on four real hyperspectral datasets demonstrate its superior performance and practical applicability compared with state-of-the-art hyperspectral target detection methods.

Abstract

Hyperspectral image (HSI) target detection plays a critical role in both military and civilian applications, including military reconnaissance, environmental monitoring, and precision agriculture. However, the complex background of the scene severely restricts the further improvement of hyperspectral target detection performance. To address this challenge, we propose a diffusion model hyperspectral target detection method based on multivariate Gaussian background noise. The method constructs multivariate Gaussian-distributed background noise samples and introduces them into the forward diffusion process of the diffusion model. Subsequently, the denoising network is trained, the conditional probability distribution is parameterised, and a designed loss function is used to optimise the denoising performance and achieve effective suppression of the background, thus improving the detection performance. Moreover, in order to obtain accurate background noise, we propose a background noise extraction strategy based on spatial-spectral centre weighting. This strategy combines with the superpixel segmentation technique to effectively fuse the local spatial neighbourhood information of HSI. Experiments conducted on four publicly available HSI datasets demonstrate that the proposed method achieves state-of-the-art background suppression and competitive detection performance. The evaluation using ROC curves and AUC-family metrics demonstrates the effectiveness of the proposed background-suppression-guided diffusion framework.



Academic Editor: Salah Bourennane

Received: 7 November 2025

Revised: 12 December 2025

Accepted: 22 December 2025

Published: 25 December 2025

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: hyperspectral image (HSI); target detection; diffusion model; background noise generation; background suppression

1. Introduction

Hyperspectral image is a three-dimensional data cube that contains both rich spatial and spectral information and is the basis for the success of many advanced vision tasks. These tasks cover the key areas of target detection [1–3], anomaly detection [4–6], land cover classification [7–9], and change detection [10–12]. In hyperspectral target detection, it effectively utilizes discriminative information from prior spectra, facilitating precise identification of potential targets, even in complex backgrounds, with pixel-level or sub-pixel accuracy. As a result, the development of hyperspectral target detection methods has attracted extensive and significant attention in a wide range of practical applications, including remote sensing monitoring, military reconnaissance, land and marine target observation, mineral exploration, and criminal investigation [13]. However, the practical effectiveness of hyperspectral target detection is fundamentally constrained by two intertwined challenges: the severe scarcity of labeled training samples, as manually annotating pixel-level targets across vast hyperspectral scenes is prohibitively expensive and often infeasible; and the complex, heterogeneous, and dominant nature of background clutter, which makes isolating faint target signals extraordinarily difficult. To address these challenges, wide methods has been explored.

Traditional methods distinguish between target and background by mathematically exploring the differences between spectral vectors and a priori spectra. Constrained Energy Minimization (CEM) is one of the widely favored methods in this domain. It imposes constraints on both the target and background, constructing a finite impulse response filter to minimize the filter output energy [14]. In this process, the output energy of the filter is minimized to effectively suppress background responses. Given its notable advantages in speed and accuracy, researchers have proposed numerous variants of CEM to more precisely capture the inherent characteristics of the target and background. Hierarchical CEM is one of the most representative improvements [15], which utilizes a structure with CEM detectors at different layers. Through a layer-by-layer filtering process, the target is preserved and the background suppressed, gradually improving detection performance. Another insightful variant is the ensemble CEM [16], which innovatively introduces cascaded detection and multi-scale scanning strategies, enhancing the generalization ability and non-linear discriminative capacity of the hyperspectral target detector, resulting in higher detection accuracy and stability. Chang [17] recently found that iterative kernel CEM (IKCEM) has even better performance. In the multi-directional CEM method [18], an adaptive neighborhood feature aggregation strategy is employed to comprehensively and precisely assess the importance of neighborhood information from different directions. Another commonly used target detection method is Orthogonal Subspace Projection (OSP) [19], which reduces the dimensionality of hyperspectral image and suppresses insignificant spectral features. Specifically, it projects each pixel vector onto the background's orthogonal subspace and then projects the residual onto the target vector of interest, thereby enhancing the signal-to-noise ratio. By taking advantage of OSP and CEM, Ren and Chang [20] combined OSP and CEM to derive TCIMF, which used OSP to annihilate undesired targets to increase target detectability and in the meantime used CEM to improve Target background suppression simultaneously. To better cope with background, Chang [21] develops a background-annihilated TCIMF (BA-TCIMF). Spectral Angle Mapper (SAM) [22] mainly detects targets by evaluating the spectral angle between each pixel's spectrum and the spec-

tral signature of the target of interest. Additionally, some methods based on linear mixture models have successfully distinguished between the target and background by performing spectral decomposition on hyperspectral image [23]. Other methods such as the Matched Subspace Detector (MSD) [24], Adaptive Cosine Estimator (ACE) [25,26], and Matched Filter [27] have also demonstrated strong applicability in hyperspectral target detection tasks. While these methods are computationally efficient and require no training data, they often fail to adequately model the complexity and heterogeneity of real hyperspectral backgrounds due to oversimplified statistical assumptions and limited capacity to capture nonlinear spatial–contextual information. Consequently, their performance degrades in scenes with highly variable backgrounds.

Deep learning–based hyperspectral target detection methods can be broadly divided into supervised and unsupervised categories. In the domain of supervised approaches, Li et al. [28] pioneered a convolutional neural network (CNN) framework for hyperspectral target detection, laying a foundation for subsequent studies. Another representative method is HTD-net [29], which employs an improved autoencoder to extract target features and then uses a linear prediction strategy to distinguish background spectra from target samples. However, they are inherently constrained by the scarcity of high-quality labeled samples, as acquiring sufficient pixel-level annotations is costly and time-consuming, particularly when targets are small or rare. Unsupervised models, in contrast, aim to enhance the network’s ability to learn latent target representations by applying specific constraints, thereby alleviating the difficulty of collecting large-scale labeled hyperspectral datasets [30,31]. The Variational Autoencoder (VAE) [32] is one of the most representative and scalable unsupervised architectures. Xie et al. [31] proposed an enhanced VAE to capture more complex nonlinear structures in hyperspectral data, while Shi et al. [33] integrated residual learning to construct a macro–micro residual autoencoder tailored for hyperspectral target detection. In addition, several studies emphasize the role of background modeling. For instance, adversarial autoencoders in [34] estimated background distributions through adversarial learning, and the method in [35] used a VAE with orthogonal-subspace constraints to disentangle background components. Although these methods attempt to model background information, their modeling is typically indirect and remains limited in addressing the intrinsic complexity, heterogeneity, and noise characteristics of real-world hyperspectral backgrounds. Most approaches rely solely on spectral-domain statistics without explicitly characterizing background noise distributions or spatial–spectral dependencies. As a result, when the background exhibits high spectral variability, nonlinear mixing, or locally correlated clutter, the learned models often become insufficiently robust, leading to reduced target–background separability.

In contrast, the proposed method explicitly models background noise by estimating a multivariate Gaussian distribution from the scene and incorporating it into a forward diffusion process. This enables the diffusion model to learn both the generation and suppression mechanisms of background noise in a principled, physically grounded manner. Furthermore, we introduce a centre-weighted strategy that integrates spatial–neighborhood information to accurately capture local spectral variations. By combining noise-aware diffusion and spatial–spectral weighting, our method achieves stronger background suppression and more reliable target enhancement, particularly in highly complex, heterogeneous, or spatially correlated backgrounds. Additionally, the method naturally addresses the scarcity of labeled samples by employing background-only self-supervised training, requiring only a single hyperspectral image and a single prior target signature without pixel-level annotations. Note: In this study, “limited labeled samples” refers specifically to scenarios where the number of available positive samples is extremely scarce. Overall, the main contributions of this work are threefold.

(1) We introduce the diffusion model into the field of hyperspectral target detection and improve the noise addition process of the diffusion model. By incorporating multivariate Gaussian background noise, the denoising network can learn the distribution of background noise, thereby acquiring the capability to suppress the background. This method significantly mitigates background interference, contributing to improved detection rates and reduced false alarm rates.

(2) To obtain an accurate background noise distribution, we propose a spatial-spectral centre-weighted multivariate Gaussian background noise generation (SSBNG) strategy. By leveraging superpixel segmentation to consider local spatial neighborhood information, the spatial-spectral correlations within hyperspectral images are fully exploited. This provides high-quality and reliable noise samples for the subsequent noise addition stage, effectively enhancing the overall accuracy and reliability of the study.

(3) To validate the effectiveness of the proposed method, comprehensive experiments and evaluations were conducted. The experimental results demonstrate that our method outperforms existing state-of-the-art methods in hyperspectral target detection. Notably, the proposed model exhibits remarkable adaptability in real-world scenarios with complex background information, particularly in terms of background suppression, showcasing robust background suppression capabilities.

The remainder of this paper is organized as follows. Section 2 details the proposed novel hyperspectral target detection method. Section 3 reports and analyzes the experimental results obtained on four real-world hyperspectral datasets. Section 4 discussion. Section 5 concludes the paper.

2. Proposed Method

In this section, we introduce a novel framework for target detection. The core idea of the framework is to train the diffusion model by generating multivariate Gaussian background noise, so as to construct a denoising network that can effectively suppress the background interference. Considering that the performance of the denoising network is highly dependent on the accuracy of the background noise distribution, we design a concise and efficient spatial-spectral centre-weighted background noise generation (SSBNG) strategy. This strategy makes full use of the spatial-spectral properties of hyperspectral data to generate high-quality background noise samples. After the training is completed, the denoising network performs background suppression on the hyperspectral data, and the final detection results are subsequently obtained by Mahalanobis distance calculation. It should be noted that the proposed background modeling strategy uses a single multivariate Gaussian to represent the background. This simplification is reasonable for scenes where the background is relatively homogeneous or the dominant background type can be captured by a single Gaussian. Empirical checks on the tested datasets indicate that this assumption reasonably approximates the background distribution. However, in highly heterogeneous scenes containing multiple distinct background materials, the performance of the method may degrade. Future extensions could adopt mixture-of-Gaussians models to better capture multi-modal background distributions. Therefore, the proposed method is most suitable for scenarios with approximately homogeneous background structures. Figures 1 and 2 clearly illustrates the overall architecture and workflow of the proposed framework.

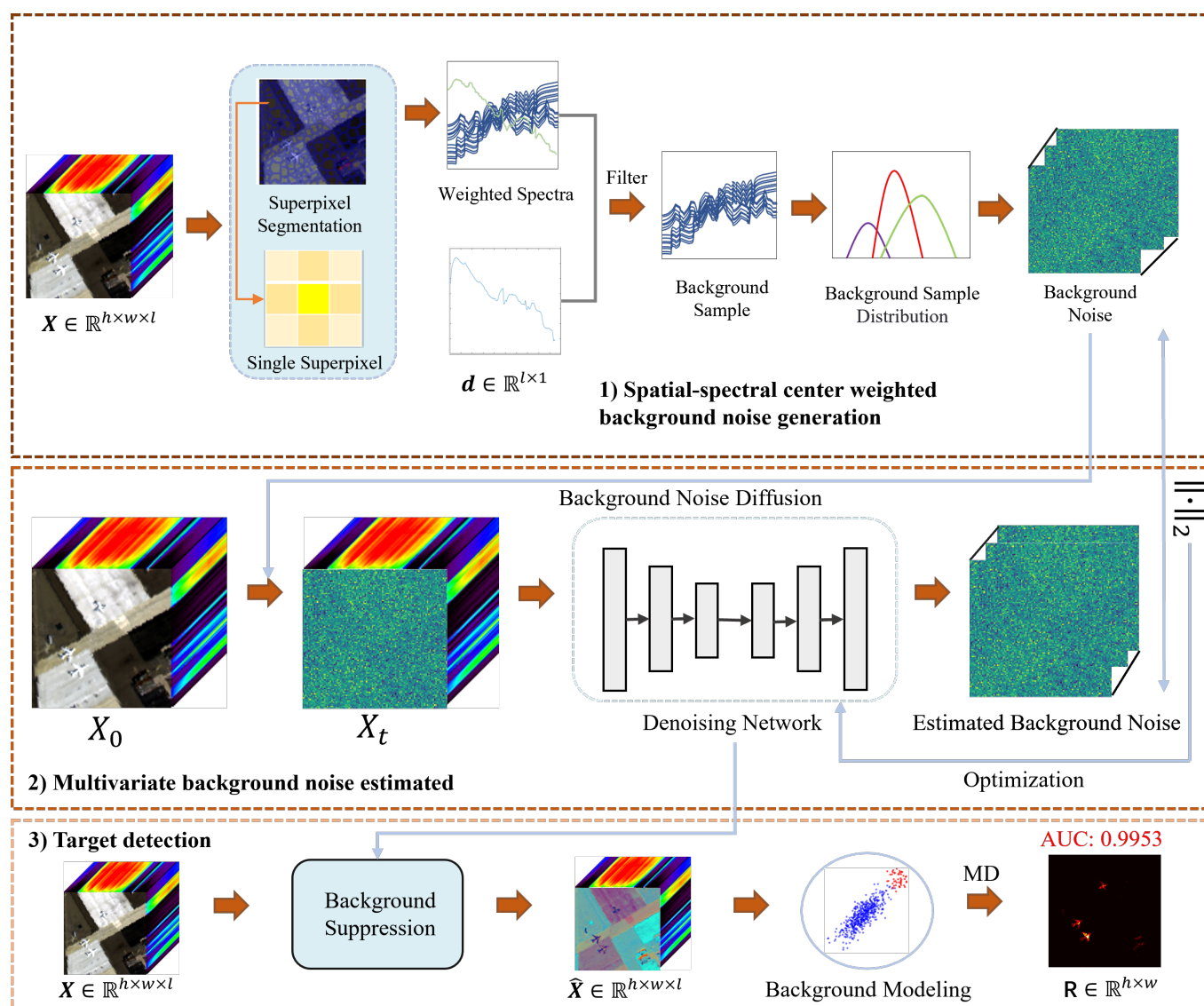


Figure 1. Schematic of the proposed target detection method.

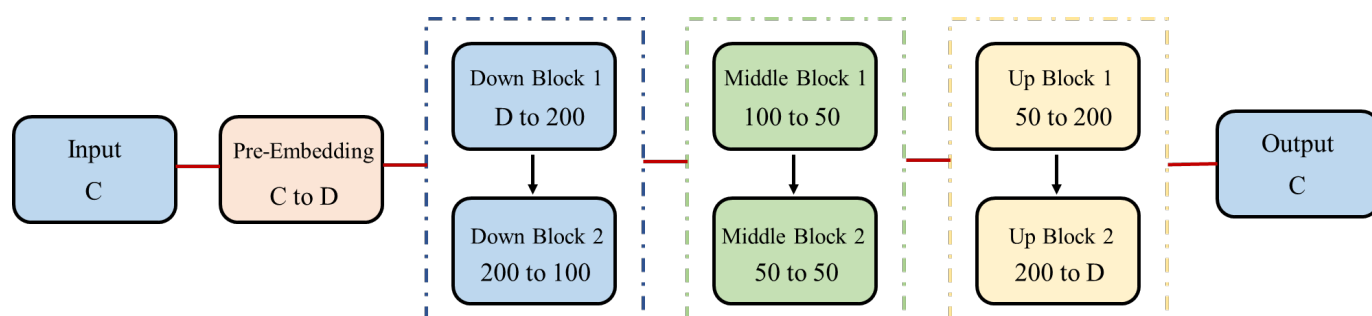


Figure 2. Structure diagram of the denoising network. Pre-Embedding denotes the pre-embedding layer; down block, middle block, and up block represent residual layers; C indicates the number of channels; D represents the pre-embedding dimension; and “200 to 100” means the dimension changes from 200 to 100.

2.1. Brief Review of Diffusion Model

Diffusion models, also known as diffusion probabilistic models [36], are a class of latent variable models (LVMs) [37]. Inspired by nonequilibrium thermodynamics, these models construct a Markov chain to progressively perturb the input sample into a standard Gaussian distribution. The denoising diffusion probabilistic model (DDPM) [36,38] employed

in this work consists of two core processes: the forward diffusion process and the reverse denoising process. In the forward process, noise is added to the original sample $\mathbf{x}_0$ at each time step $t \in \{1, \dots, T\}$, forming a Gaussian Markov chain with the following distribution

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}) \quad (1)$$

where β_t is a predefined noise scheduling parameter. This process can be simplified to directly sample the sample $\mathbf{x}_t$ at a given time step from $\mathbf{x}_0$

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (2)$$

where $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$. The reverse process learns to progressively reconstruct the original image from noise using a neural network (typically a U-Net). The training objective is to minimize the noise prediction error

$$\mathcal{L} = \mathbb{E}_{\mathbf{x}_0, \epsilon, t} [\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t)\|_2] \quad (3)$$

where ϵ is the real noise added, and ϵ_θ is the network's predicted value.

Due to the high dimensionality of hyperspectral image, the use of random noise in each iteration of traditional diffusion models may lead to poor convergence of the denoising network. To address this, we use the same Multivariate Gaussian noise and fixed time step t across all training iterations, which enhances network convergence, stabilizes the training process, and reduces the overall training time.

Although diffusion models have demonstrated outstanding performance in tasks such as natural language processing [39,40], time series forecasting [41,42], molecular graph modeling [43,44] and hyperspectral image classification [45], to the best of our knowledge, no literature has yet applied them to hyperspectral target detection. Unlike classification, target detection requires locating specific spatial targets within an image, so our diffusion model is adapted to suppress background noise while preserving target signals. We present hyperspectral target detection method based on DDPM, treating the background as a modelable noise distribution. Through the diffusion–denoising mechanism, the background interference is effectively suppressed, improving detection accuracy and reducing false alarm rates.

2.2. Spatial–Spectral Centre-Weighted Background Noise Generation

In this study, accurately obtaining background samples and generating background noise is crucial for the subsequent diffusion process. To address this, we propose the Spatial–Spectral Background Noise Generation (SSBNG) strategy, which aims to generate multivariate Gaussian background noise by integrating both spatial and spectral information. This strategy provides high-quality and reliable noise samples for the subsequent noise injection process, thereby effectively improving the accuracy and reliability of the entire methodology.

The specific implementation steps are as follows: First, three channels are selected from the hyperspectral image $\mathbf{X} \in \mathbb{R}^{h \times w \times l}$, and transformed into a pseudo-color image. The SLIC method is then applied to segment the image into N_s superpixel regions. In these regions, adjacent pixels exhibit significant similarity in color, brightness, and texture features. For each superpixel region, the average spectral value of all pixels in the region is computed as the initial background sample spectrum, as shown in the following equation

$$\bar{x}_i = \frac{1}{N_{si}} \sum_{j=1}^{N_{si}} x_{ij}, \quad i = 1, 2, \dots, N_s \quad (4)$$

where x_{ij} represents the spectral vector of the j -th pixel in the i -th superpixel, and N_{si} represents the number of pixels in the i -th superpixel. To measure the difference between each pixel and the average spectrum, the Euclidean distance between each pixel in the superpixel and the average spectrum is calculated, as follows

$$g(x_{ij}, \bar{x}_i) = \|x_{ij} - \bar{x}_i\|_2, \quad j = 1, 2, \dots, N_{si} \quad (5)$$

To further optimize the background sample spectrum and ensure it accurately reflects real background information, we use a weight adjustment function to weight the contribution of pixels within each superpixel. This process aims to emphasize pixels that are closer to the average spectrum, while diminishing the influence of anomalous pixels, thereby obtaining a more reliable background spectrum sample. The weight adjustment function [46] is expressed as

$$f(g(x_{ij}, \bar{x}_i)) = \frac{1}{2} \left(1 - \left(\frac{g(x_{ij}, \bar{x}_i)}{\gamma} \right)^2 \right)^2 \quad (6)$$

where γ is an adjustable weight parameter that controls the sensitivity of the weighting function. In this work, γ is set to the maximum value of $g(x_{ij}, \bar{x}_i)$ within each superpixel to normalize the spatial-spectral deviation and ensure stable weighting. Based on the above weighting results, the background sample spectrum for each superpixel is recalculated using the following equation

$$y_i = \frac{\sum_{j=1}^{N_{si}} x_{ij} f(g(x_{ij}, \bar{x}_i))}{\sum_{j=1}^{N_{si}} f(g(x_{ij}, \bar{x}_i))} \quad (7)$$

However, it is important to note that some superpixel regions may contain target pixels, whose spectra significantly interfere with the purity of the background sample. To further purify the background sample, we introduce a CEM filter, which calculates the similarity between the background sample spectrum and the target prior spectrum, and removes samples that are highly similar to the target spectrum based on the similarity score. This ensures that the generated background sample reflects the pure background characteristics as much as possible. Given a spectral matrix consisting of all superpixel blocks $\mathbf{Y} \in \mathbb{R}^{N_s \times l}$ and a prior spectrum $\mathbf{d} \in \mathbb{R}^{l \times 1}$, the output response of the CEM filter can be expressed as

$$s = (\hat{w})^T \mathbf{Y} = \frac{\mathbf{d}^T \mathbf{R}^{-1}}{\mathbf{d}^T \mathbf{R}^{-1} \mathbf{d}} \mathbf{Y} \quad (8)$$

where $\mathbf{R} = \frac{1}{N_s} \mathbf{Y} \mathbf{Y}^T$ is the correlation matrix, and $\hat{w}$ is the optimal finite impulse response filter obtained through a specific algorithm. By sorting the similarity scores of all samples in descending order, the top 20% of samples are removed from the background sample candidate set, following thresholding strategies commonly adopted in prior CEM-based target suppression studies as well as background-learning methods such as BLTSC and OS-VAE. Once the background samples (X_b) are successfully generated, the process of background noise generation begins. First, we perform statistical analysis on the obtained background samples, calculating the mean μ and covariance matrix Σ . The mean reflects the average characteristics of the background samples, while the covariance matrix captures the correlations between different features. Based on these statistics, background noise

samples are generated using a multivariate Gaussian distribution. The probability density function of the multivariate Gaussian distribution is given by

$$f(\mathbf{m}; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{l}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{m} - \mu)^T \Sigma^{-1}(\mathbf{m} - \mu)\right) \quad (9)$$

where $\mathbf{m}$ is the random variable vector corresponding to the generated background noise sample $\mathbf{N}_b$, l is the dimensionality of the data, which in this study is related to the spectral vector dimensions, $|\Sigma|$ is the determinant of the covariance matrix Σ . Using this distribution, we can generate noise that conforms to the background characteristics, with the distribution of the noise in the feature space matching the statistical properties of the original background sample, thereby completing the entire Spatial–Spectral Background Noise Generation (SSBNG) strategy. In this work, we model the background distribution using a single multivariate Gaussian, which provides a tractable and efficient approximation for background noise generation. This assumption is valid for scenes where the background is relatively homogeneous or dominated by a single material class. To examine the suitability of this assumption, we performed empirical distribution checks on the background samples of the tested datasets, which show that their principal spectral variations can be reasonably captured by a unimodal Gaussian distribution. However, we acknowledge that in highly heterogeneous scenes containing multiple distinct background materials, the background distribution may become multi-modal. In such cases, using a single Gaussian may limit modeling capacity and potentially degrade detection performance. We have included a discussion of this limitation and outline possible extensions, such as replacing the single Gaussian with a mixture-of-Gaussians model to better capture background complexity in multimodal environments.

2.3. Multivariate Background Noise Estimated

In the current mainstream diffusion model framework, standard Gaussian noise is the conventional choice for noise introduction. However, in this study, through an in-depth analysis of the characteristics of hyperspectral data and the requirements of target detection tasks, we break from tradition by focusing the diffusion process specifically on background noise. This innovation arises from a profound understanding of the complex relationship between background and target in hyperspectral data. From the training point of view, the denoising network is able to learn and capture these features more efficiently as the background noise distribution features and their statistical laws in hyperspectral data have been accurately modelled through the pre-generation process. Compared to extracting target noise, the task of extracting background noise is significantly less complex. This allows the denoising network to accurately model the patterns of background noise during training, thus providing a solid foundation for the subsequent inference process. During the inference phase, based on the effective learning of background noise during training, the denoising network is able to accurately identify and treat the background as noise. This precise background extraction mechanism greatly enhances the differentiation between the background and the target, making the target more prominent in the data and providing high-quality data representations for subsequent hyperspectral target detection tasks. It is important to note that since background noise does not follow a standard Gaussian distribution, the noise diffusion equation commonly relied upon in traditional DDPM cannot be directly applied to our method. To address this critical issue, we have carefully designed a specific noise diffusion method tailored to background noise, which accounts for the actual distribution characteristics of the background noise and the complex background structure of hyperspectral data. The details of this approach are described as follows.

2.3.1. Diffusion Process

Under the condition of multivariate Gaussian distribution noise generated from hyperspectral background samples, we define the entire diffusion process. The diffusion process at each step can be expressed as

$$X_t = a_t X_{t-1} + \sqrt{1 - a_t} z_t \quad (10)$$

where X_t represents the data state at the t -th step of the diffusion process, which is the core variable that we track and analyze. a_t serves as a critical control parameter, playing a key role in precisely regulating the proportion of original data retained at each step, directly influencing the trend of the data during the diffusion process. In our implementation, a_t is generated using a linear schedule across the diffusion steps. $z_t \sim \mathcal{N}(\mu_b, \Sigma_b)$, meaning that the noise added at each step is not randomly generated, but strictly follows the distribution characteristics determined by the background samples, with mean μ_b and covariance Σ_b . This noise addition method, based on a specific distribution, provides a solid theoretical foundation for subsequent analysis and model construction. Based on the above formula, we can further derive the transition probability at each step of the diffusion process. The transition probability $q(X_t|X_{t-1})$ fully characterizes the probability distribution of the transition from state X_{t-1} to X_t , and it also follows a multivariate Gaussian distribution, expressed as

$$q(X_t|X_{t-1}) = \mathcal{N}(X_t; \sqrt{a_t}X_{t-1} + (1 - \sqrt{a_t})\mu_b, (1 - a_t)\Sigma_b) \quad (11)$$

In this equation, the mean term $\sqrt{a_t}X_{t-1} + (1 - \sqrt{a_t})\mu_b$ depends not only on the previous state but also integrates the background mean μ_b . The coefficients $\sqrt{a_t}$ and $1 - \sqrt{a_t}$ cleverly balance the influences of the previous data and the background information. The covariance term $(1 - a_t)\Sigma_b$ indicates that, as the diffusion steps progress, the noise fluctuation level is closely related to a_t and the background covariance Σ_b , further reflecting the intrinsic rules and characteristics of the entire diffusion process.

2.3.2. Denoising Process

After completing the diffusion process, we proceed to the denoising process. The core task of this stage is to learn a model $p_\theta(X_{t-1}|X_t)$, which gradually restores the data from the noisy state x_t to a state close to the original data. The conditional distribution $p_\theta(X_{t-1}|X_t)$ in the denoising process is also assumed to follow a multivariate Gaussian distribution, expressed as

$$p_\theta(X_{t-1}|X_t) = \mathcal{N}(X_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(X_t, t)) \quad (12)$$

Here, the mean $\mu_\theta(X_t, t)$ and covariance $\Sigma_\theta(X_t, t)$ are the key parameters of the model, and they are parameterized through a neural network. The powerful learning capability of the neural network allows the model to dynamically adjust the mean and covariance based on the input X_t and the current time step t , adapting to different noise contamination levels and data characteristics. To solve and optimize this model more efficiently, we carefully design and represent the mean $\mu_\theta(X_t, t)$ as follows

$$\mu_\theta(X_t, t) = \frac{1}{\alpha_t} \left(X_t - \frac{1 - \alpha_t}{1 - \bar{\alpha}_t} \epsilon_\theta(X_t, t) \right) + (1 - \alpha_t)\mu_b \quad (13)$$

In this equation, $\epsilon_\theta(X_t, t)$ is the output of the neural network, which plays the key role of predicting the noise component. The cumulative decay factor $\bar{\alpha}_t$ is defined as the product of α_i from $i = 1$ to t , i.e., $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$. This reflects the cumulative decay level of

the data from the initial state to the t -th step due to the influence of α_i at each diffusion step. By using the factor $\frac{1-\alpha_t}{1-\alpha_i}$, the predicted noise component is reasonably computed with the current state X_t . This is then combined with $\frac{1}{\alpha_t}$ and $(1-\alpha_t)\mu_b$, which results in a mean expression that considers both the current noise state and the background mean. This design allows the model to fully leverage the available information during the denoising process, providing a more accurate estimate of the previous state X_{t-1} .

The denoising network contains a pre-embedding layer with 256 channels, followed by three residual layers with 200, 100, and 50 channels. Each layer employs Tanh activation and LayerNorm normalization. A skip connection adds the input to the output. The network predicts the noise term ϵ_θ at each diffusion time step.

2.3.3. Loss Function

To ensure that the denoising model we construct accurately learns to remove background noise, we need to carefully design a suitable loss function. The core goal of this loss function is to make the noise prediction $\epsilon_\theta(X_t, t)$ output by the neural network as close as possible to the true background noise $z \sim \mathcal{N}(\mu_b, \Sigma_b)$. We define the loss function in the form of a weighted norm, expressed as

$$L(\theta) = \mathbb{E}_{(X_0, z, t)} \left[\|\epsilon_\theta(X_t, t) - (z - \mu_b)\|_{\Sigma_b^{-1}}^2 \right] \quad (14)$$

In this expression, $\mathbb{E}_{(\chi_0, z, t)}$ represents the expectation over the initial data, noise, and time steps, ensuring that the loss function can comprehensively account for various scenarios. The term $\|\cdot\|_{\Sigma_b^{-1}}$ represents the weighted norm, where Σ_b^{-1} acts as the weighting matrix. This allows for appropriate weighting of the noise errors across different dimensions according to the covariance characteristics of the background noise, thereby making the loss function more accurate in reflecting the difference between the predicted and true noise. The above loss function is mathematically equivalent to

$$L(\theta) = \mathbb{E}_{(X_0, z, t)} \left[(\epsilon_\theta(X_t, t) - (z - \mu_b))^T \Sigma_b^{-1} (\epsilon_\theta(X_t, t) - (z - \mu_b)) \right] \quad (15)$$

This matrix form expression is more concise and clear, while also facilitating matrix operations during the actual computation and optimization process, thus improving computational efficiency and the effectiveness of model training.

2.4. Target Detection

In the target detection process, the hyperspectral data $X \in \mathbb{R}^{h \times w \times l}$ is fed into the trained diffusion model. The diffusion network treats the complex background as noise and effectively removes it, producing a background-suppressed hyperspectral image $\hat{X} \in \mathbb{R}^{h \times w \times l}$. However, due to the spectral similarity between some targets and background materials in certain bands, the background suppression process may cause partial loss of target spectral information. To address this issue, we employ a Mahalanobis distance-based detection strategy that quantitatively measures the difference between each pixel and the background model.

The background model is constructed from the background-suppressed data $\hat{X} \in \mathbb{R}^{h \times w \times l}$ using the background regions whose spatial locations were determined before the diffusion process. Based on the spectral information of these background regions, the mean and covariance of the background distribution are re-estimated to describe the global spectral characteristics after background suppression. Each pixel is then compared to this background model using the Mahalanobis distance.

$$D_M(x) = \sqrt{(x - \mu_B)^T \Sigma_B^{-1} (x - \mu_B)} \quad (16)$$

where $D_M(x)$ represents the Mahalanobis distance between the pixel spectrum x and the background distribution, μ_B and Σ_B denote the mean vector and covariance matrix of the background, respectively. A larger $D_M(x)$ indicates that the pixel is more dissimilar to the background and thus more likely to belong to a target region. The overall procedure of the proposed target detection framework is summarized in Algorithm 1.

Algorithm 1: Hyperspectral Target Detection

Input: Hyperspectral image tensor X , prior target spectrum d

- 1: Perform superpixel segmentation and compute the average spectrum $\bar{x}_i$ by (4)
- 2: Calculate centre-weighted spectral data y_i by (5), (6), and (7)
- 3: Filter out target spectra to obtain pure background spectra X_b by (8)
- 4: Generate multivariate Gaussian background noise N_b by (9)

For each epoch

- Add background noise z_t by (10) and (11)
- Train the denoising network by (12), (13), and (15)

End for

- 5: Use the trained denoising network for background suppression $\hat{X}$
- 6: Compute Mahalanobis distance through background modeling to get detection results D_M by (16)

Output: Final detection result

3. Experimental Evaluation

In this section, we conduct extensive numerical experiments to validate the effectiveness of the proposed method for hyperspectral target detection tasks. We begin by providing a comprehensive introduction to the experimental setup, followed by a comparison of the results with current state-of-the-art methods. In addition, we perform a detailed analysis of the proposed model, including ablation studies and hyperparameter analysis.

3.1. Experimental Setup

All experiments were conducted on a desktop equipped with an NVIDIA GeForce RTX 4060 GPU (16 GB), using PyTorch 2.7.0 as the deep learning framework. The model was trained using the Adam optimizer with an initial learning rate of 0.0001 for 500 epochs and a batch size of 64. The forward diffusion process employed the gamma diffusion mode, with a maximum time step of $T = 1000$, and inference was repeated $K = 5$ times. All hyperparameters were selected using a held-out validation split constructed from the training data, and no test samples were used during hyperparameter tuning.

3.1.1. Dataset

We used four publicly available datasets commonly used for evaluating hyperspectral target detection methods. Below is a summary of the detailed information of these datasets and the included targets.

- **HYDICE Dataset:** The first dataset was collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) airborne sensor in an urban area in California, USA. This urban scene consists of 80×100 pixels, with 175 spectral bands, covering a wavelength range from 400 to 2500 nm. The spatial resolution of the image is 1 m. The scene mainly includes a vegetation area, a building area, and several roads, with some cars present. These man-made objects, namely cars and rooftops, occupy 19 pixels and are considered the targets.
- **San Diego Dataset:** The second hyperspectral dataset was captured by the AVIRIS sensor, covering the area of San Diego airport in California, USA. The spatial size

is 100×100 pixels, with 224 spectral bands covering a wavelength range from 370 to 2510 nm. The spatial resolution is about 3.5 m. In the experiment, 189 bands were used after removing water absorption and low signal-to-noise ratio bands. This dataset includes two images, one with three airplanes located in the upper-right corner occupying 57 pixels, referred to as San Diego, and the other with three airplanes in the centre occupying 134 pixels, referred to as San Diego2. The airplanes are considered the targets of interest.

- **ABU-beach Dataset:** This dataset was captured by the AVIRIS sensor. It contains multiple hyperspectral image, and we selected Beach-1. The spatial size of Beach-1 is 150×150 pixels, with 188 spectral bands. One man-made object is considered the target of interest.

3.1.2. Baselines

We compare our proposed method with several commonly used hyperspectral image detectors: CEM [14] based on target maximization and output energy minimization constraints, H-CEM [15] is an improved method based on CEM, ACE [26] based on adaptive cosine similarity measures, DSC [47] based on clustering dictionary generation and sparse representation, BLTSC [34] based on target suppression constraints and background learning, and OS-VAE [35] based on orthogonal subspaces and background learning. All codes were downloaded from the authors' public repositories, and the hyperparameters were tuned to the optimal settings according to the instructions in the corresponding papers. Before evaluation, all dataset grayscale values were normalized to the range [0, 1].

3.1.3. Evaluation Metrics

We evaluate the detection performance using visualization detection maps and Receiver Operating Characteristic (ROC) [48] curves generated by varying the threshold τ . In the 3-D ROC curve, the number of target pixels with detection values greater than and less than τ are represented as true positives (TP) and false negatives (FN), respectively. Similarly, the number of background pixels with detection values greater than and less than τ are represented by false positives (FP) and true negatives (TN), respectively. The 2-D ROC curves (P_f, P_d) , (τ, P_d) and (τ, P_f) represent the overall detection efficiency, the target detection effectiveness and the background suppression efficiency of different hyperspectral target detection methods, respectively. Specifically, the detection probability P_d and false alarm probability P_f at each threshold are calculated by

$$P_d = \frac{TP}{TP + FN}, \quad P_f = \frac{FP}{FP + TN} \quad (17)$$

For high-performing methods, the ideal 2-D ROC curves should approach the upper-left corner for (P_f, P_d) , the lower-left corner for (τ, P_f) and the upper-right corner for (τ, P_d) of their respective coordinate axes. The relationship between false alarm rate and detection rate can be described by the (P_f, P_d) curve. The area under the 2-D ROC curve, denoted as $AUC(P_f, P_d)$, $AUC(\tau, P_d)$ and $AUC(\tau, P_f)$, are used for quantitative evaluation. For $AUC(P_f, P_d)$ and $AUC(\tau, P_d)$ a higher value indicates superior results, with 1 being the best. In contrast, for $AUC(\tau, P_f)$, a lower value is preferred, with 0 being the best. Moreover, two combined evaluation metrics are introduced to comprehensively assess detection performance, defined as AUC_{OA} and AUC_{SNPR} , given by:

$$AUC_{OA} = AUC_{(P_f, P_d)} - AUC_{(\tau, P_f)} + AUC_{(\tau, P_d)} \quad (18)$$

$$AUC_{SNPR} = \frac{AUC_{(\tau, P_d)}}{AUC_{(\tau, P_f)}} \quad (19)$$

Higher values of AUC_{OA} and AUC_{SNPR} indicate better detection performance. The ideal values for these two criteria are 2 and $+\infty$, respectively.

3.2. Comparative Results of Hyperspectral Image Target Detection

Figure 3 presents the visual detection results of all competing methods on four real hyperspectral image datasets. Overall, our method shows strong capability in background suppression and provides visually clear target–background separation. Classical detectors such as CEM and H-CEM struggle to suppress complex backgrounds while preserving weak targets, as indicated by residual clutter and false alarms in their detection maps. The DSC method, which is sensitive to intricate image structures, produces a large number of false positives; in contrast, our method achieves stronger background suppression than DSC. Although ACE demonstrates competent background suppression, it occasionally overlooks weak targets. Recent background-learning methods such as BLTSC and OS-VAE yield improved suppression of background responses, yet their detection maps still exhibit background leakage or incomplete target coverage. Compared with these approaches, our method generally produces cleaner backgrounds and more concentrated target responses, benefiting from the ability of the multivariate Gaussian diffusion model to characterize background noise. Overall, our approach provides competitive detection performance, particularly excelling in background suppression, as reflected in the visual comparison. Remark 1: This improved visual performance stems from the design of our method. Traditional approaches (CEM, ACE) and sparse-representation-based methods (DSC) rely primarily on linear spectral matching, which is less effective in handling nonlinear background structures and often results in high false-alarm rates. Deep unsupervised methods such as BLTSC and OS-VAE attempt global background modeling but may not fully capture local spatial–spectral variations. In contrast, our multivariate Gaussian diffusion model explicitly learns the background noise distribution, while the Spatial–Spectral Background Noise Generation (SSBNG) strategy enhances adaptability to local background patterns. Consequently, pixels that deviate significantly from the learned distribution retain higher responses, producing detection maps with more homogeneous backgrounds and clearer target highlights.

In Figures 4 and 5, we provide the ROC and Recall curves corresponding to the detection maps mentioned above. Compared to other methods, our method's 2-D ROC (P_f , P_d) curves on the ABU-Beach-I and San Diego2 datasets are closer to the upper-left corner. However, on the HYDICE and San Diego datasets, our method's performance is slightly lower. Specifically, when P_f is less than 10^{-2} , our method slightly underperforms H-CEM, ACE and DSC in terms of accuracy on this particular metric. However, upon observing the actual detection maps in Figure 3, it is clear that, across all datasets, our method produces more convincing and intuitive detection results than H-CEM, ACE and DSC. From the 2-D ROC (τ , P_f) curves, it can be seen that our method's background suppression ability significantly surpasses other methods on all four hyperspectral datasets. This finding emphasizes the importance of considering multiple aspects when evaluating detection results. Overall, our proposed method consistently maintains a highly competitive average best detection rate across all four datasets. Remark2: The differences in ROC curves reveal distinct trade-offs between target sensitivity and background suppression inherent to different methods. Our method's dominant performance in $AUC(\tau, P_f)$ is direct quantitative evidence of the effectiveness of the background noise modeling and denoising mechanism. The minor fluctuations in $AUC(P_f, P_d)$ on specific datasets may stem from spectral simi-

larity between targets and backgrounds in certain bands, causing the denoising process to slightly attenuate extremely faint target signals. We further explore the background suppression performance, target-background separation performance, and baselines on the four datasets, as shown in Figure 6. It is evident that our method exhibits excellent background suppression capability on all four datasets. The background is well suppressed, and from the performance across the four datasets, it is clear that our method achieves significant results in background suppression, strongly confirming the effectiveness of our background suppression design. Meanwhile, the target and background are well separated, with a significant distance and minimal overlap between them.

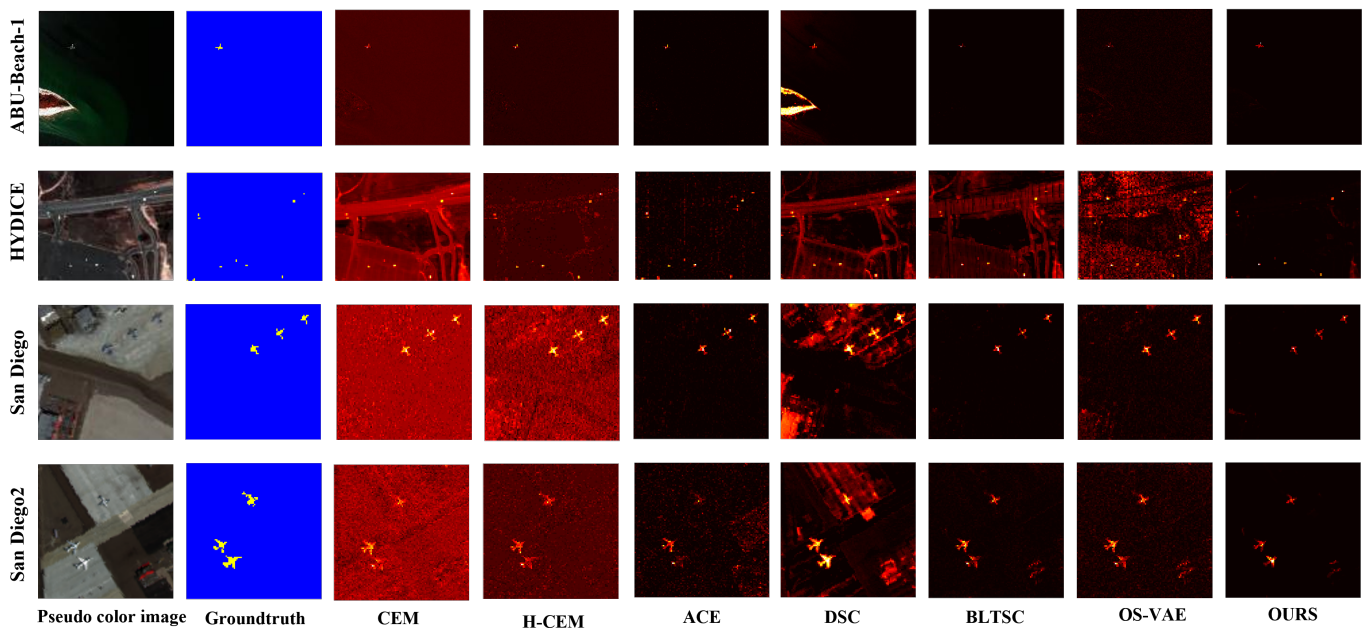


Figure 3. Pseudo color images of the four datasets, along with their corresponding ground-truth images and the detection results from different methods.

Although the ROC curves can provide insights into the detection performance, it is often difficult to accurately evaluate the results due to overlapping results from multiple methods. Therefore, we provide a detailed comparison of the five AUC metrics and AP obtained across the four datasets in Tables 1–4, supporting further quantitative analysis of the detection results. Bold text represents the best results, and underlined text represents the second-best results. Upon analysis, it is clear that, on all four datasets, our method achieves the highest value in the most crucial AUC (P_f , P_d) score, highlighting its competitiveness. Even on the HYDICE and San Diego datasets, where our method presents the second-best ROC curve, it still achieves the highest quantitative score. While DSC demonstrates the best AUC (τ , P_d) results on the four datasets, a closer look at Figure 3 shows that it performs poorly in terms of background suppression. In contrast, despite not achieving the optimal AUC (τ , P_d) value, our method performs excellently in the AUC (τ , P_f) across all four datasets, with quantitative results significantly lower than other comparison methods. This strongly confirms the efficiency of our proposed background suppression method in suppressing complex backgrounds in hyperspectral image. Remark3: The comparison of composite AUC metrics further solidifies our conclusions. The highest AUC (P_f , P_d) and AUC_{SNPR} values indicate that our method achieves the best overall performance and signal-to-noise ratio. The lowest AUC (τ , P_f) unequivocally confirms our unique advantage in background suppression. DSC's advantage in AUC (τ , P_d) likely arises from its tendency to produce globally higher response values (including for background pixels).

While this increases the detection probability of target pixels, it also causes severe false alarms, reflected in its poorer AUC (τ, P_f) and AUC_{SNPR}. By prioritizing an extremely low false alarm rate, our method achieves more reliable and practical detection performance. In conclusion, the comprehensive analysis based on AUC scores again emphasizes the strong competitiveness of our proposed method in hyperspectral target detection, especially in showcasing outstanding advantages in background suppression.

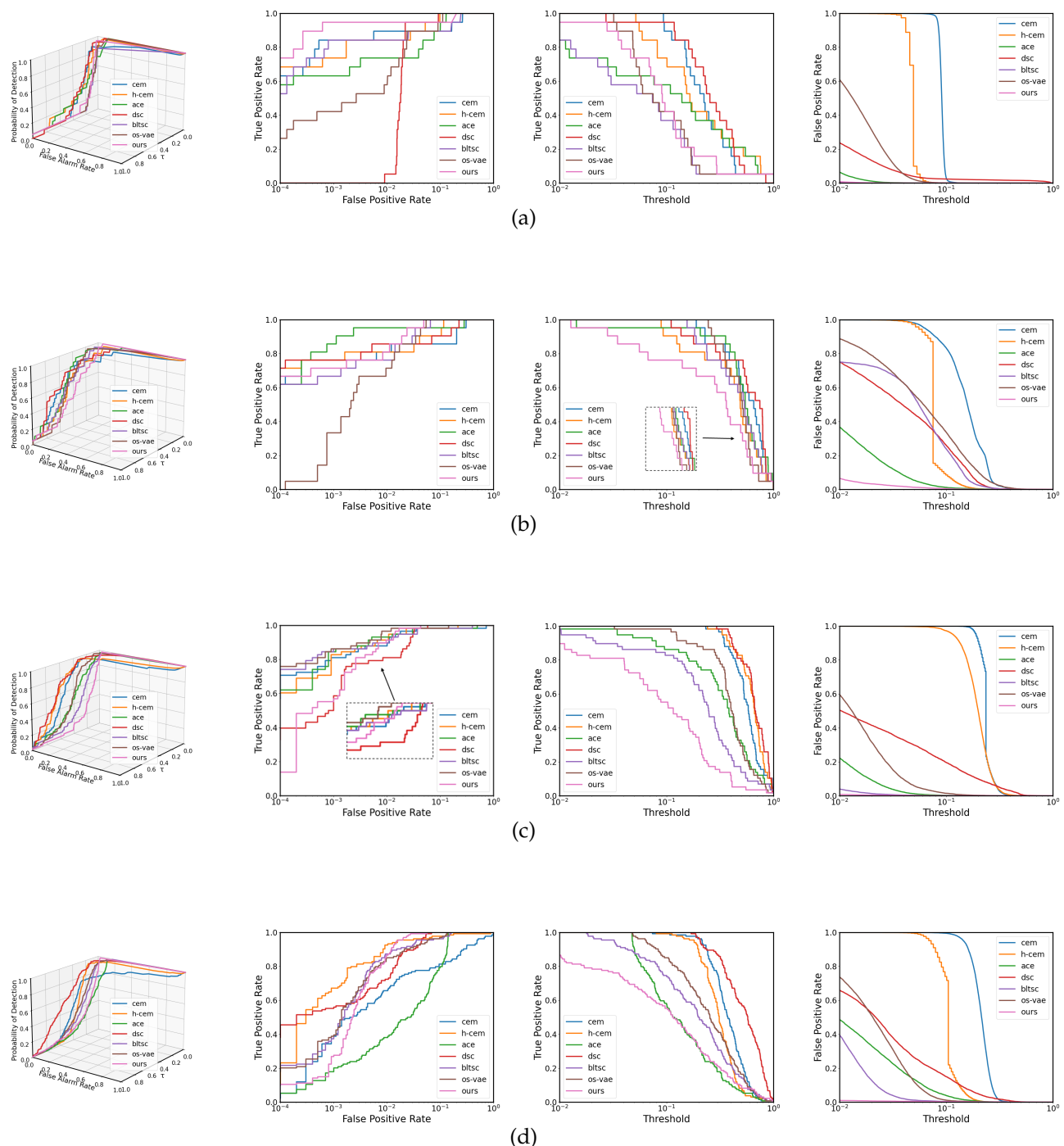


Figure 4. Three-dimensional receiver ROC curves and the corresponding 2-D ROC curves for various comparative methods on four datasets. From left to right are the 3-D ROC curve, 2-D ROC (P_f, P_d) curve, 2-D ROC (τ, P_d) curve and 2-D ROC (τ, P_f) curve, respectively. (a) ABU-beach-1. (b) HYDICE. (c) San Diego. (d) San Diego2.

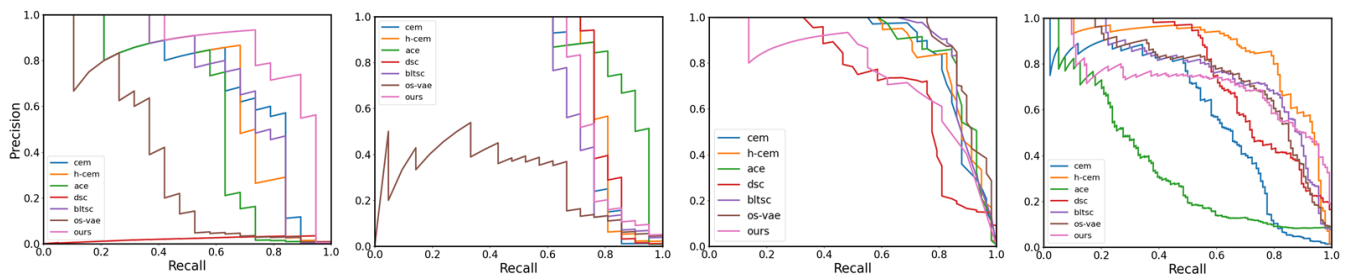


Figure 5. Recall and Precision curves for various comparative methods on four datasets. From left to right: ABU-beach-1, HYDICE, San Diego, San Diego2.

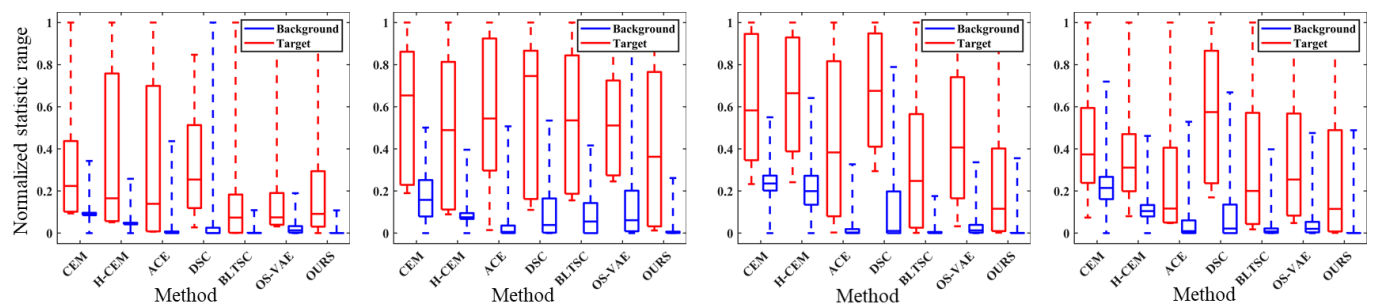


Figure 6. Target-background separation diagrams of competing methods on four datasets. From left to right: ABU-beach-1, HYDICE, San Diego, San Diego2.

Table 1. Quantitative detection results obtained by all methods on **ABU-Beach-1** dataset. Optimal and suboptimal are marked by bold and underlined, respectively.

Method	$AUC_{(P_f, P_d)}$	$AUC_{(\tau, P_d)}$	$AUC_{(\tau, P_f)}$	AUC_{OA}	AUC_{SNPR}	AP
CEM	0.9770	0.2746	0.0909	1.1608	3.0211	0.7066
H-CEM	0.9903	<u>0.2908</u>	0.0484	<u>1.2328</u>	6.0114	0.7052
ACE	0.9789	0.2484	0.0030	1.2243	83.4012	0.5948
DSC	0.9785	0.3064	0.0211	1.2638	14.5034	0.0213
BLTSC	0.9774	0.1234	<u>0.0006</u>	1.1001	<u>222.1560</u>	<u>0.7223</u>
OS-VAE	0.9829	0.1422	0.0161	1.1090	8.8101	0.3557
OURS	<u>0.9902</u>	0.1531	0.0004	1.1429	325.3999	0.8596

Table 2. Quantitative detection results obtained by all methods on **HYDICE** dataset. Optimal and suboptimal are marked by bold and underlined, respectively.

Method	$AUC_{(P_f, P_d)}$	$AUC_{(\tau, P_d)}$	$AUC_{(\tau, P_f)}$	AUC_{OA}	AUC_{SNPR}	AP
CEM	0.9650	<u>0.5992</u>	0.1646	1.3996	3.6402	0.7687
H-CEM	0.9858	0.4867	0.0789	1.3936	6.1703	<u>0.7914</u>
ACE	0.9863	0.5701	<u>0.0140</u>	1.5424	<u>40.8055</u>	0.8284
DSC	0.9778	0.6207	0.0642	<u>1.5343</u>	9.6741	0.7950
BLTSC	<u>0.9923</u>	0.5172	0.0645	1.4450	8.0196	0.7255
OS-VAE	0.9902	0.5079	0.0875	1.4106	5.8020	0.3164
OURS	0.9946	0.3783	0.0045	1.3684	83.1978	0.7616

Table 3. Quantitative detection results obtained by all methods on **San Diego** dataset. Optimal and suboptimal are marked by bold and underlined, respectively.

Method	$AUC_{(P_f, P_d)}$	$AUC_{(\tau, P_d)}$	$AUC_{(\tau, P_f)}$	AUC_{OA}	AUC_{SNPR}	AP
CEM	0.9854	0.5962	0.2376	1.3441	2.5099	0.7646
H-CEM	0.9946	<u>0.6594</u>	0.2026	<u>1.4515</u>	3.2553	0.7667
ACE	0.9897	0.4097	0.0073	1.3921	55.7610	0.7614
DSC	0.9938	0.6959	0.0611	1.6258	11.3451	0.7440
BLTSC	0.9908	0.3002	<u>0.0024</u>	1.2886	<u>124.8318</u>	0.7683
OS-VAE	<u>0.9966</u>	0.4369	0.0191	1.4146	22.9503	<u>0.7722</u>
OURS	0.9971	0.1712	0.0008	1.1674	211.7363	0.7786

Table 4. Quantitative detection results obtained by all methods on **San Diego2** dataset. Optimal and suboptimal are marked by bold and underlined, respectively.

Method	$AUC_{(P_f, P_d)}$	$AUC_{(\tau, P_d)}$	$AUC_{(\tau, P_f)}$	AUC_{OA}	AUC_{SNPR}	AP
CEM	0.9179	<u>0.4003</u>	0.2148	1.1034	1.8632	0.5844
H-CEM	0.9883	<u>0.3310</u>	0.1056	1.2137	3.1358	0.8564
ACE	0.9533	0.1862	0.0226	1.1170	8.2551	0.3499
DSC	<u>0.9917</u>	0.5662	0.0479	1.5100	11.8209	0.7607
BLTSC	0.9905	0.2587	<u>0.0111</u>	1.2382	<u>23.2983</u>	<u>0.7644</u>
OS-VAE	0.9902	0.3007	0.0262	<u>1.2647</u>	11.4640	0.7541
OURS	0.9953	0.1915	0.0012	1.1856	158.9126	0.7456

3.3. Ablation Study and Parameter Analysis

3.3.1. Ablation Study

The core innovations of the proposed hyperspectral target detection method lie in: (1) the effective introduction of spatial neighborhood information through superpixel segmentation, and (2) the adoption of a multivariate Gaussian diffusion model for background suppression. To validate the effectiveness of the proposed method, we designed and conducted a systematic ablation study, with the experimental results presented in Table 5. The experimental design follows a progressive validation strategy: First, the baseline detection results without any optimization methods are used as the control group (first row). Second, the effects of using only spectral information to generate background noise for background suppression (second row) and the effects of solely introducing spatial neighborhood information (third row) are evaluated separately. Finally, the complete scheme, which simultaneously incorporates spatial neighborhood information and employs multivariate Gaussian background suppression, is demonstrated (fourth row). The experimental results indicate that the proposed strategy, which combines spatial-spectral centre-weighted multivariate Gaussian background noise generation with background suppression, offers significant advantages. This finding thoroughly validates the theoretical hypothesis. Specifically, by constructing a deterministic background noise model and integrating it with a denoising network for background suppression, the proposed method significantly enhances the separability between targets and the background, thereby achieving precise target-background separation. The contributions of this study are mainly reflected in two aspects: First, it significantly improves the performance of existing deep learning-based hyperspectral target detection methods. Second, it provides new insights for future research, emphasizing the importance of fully considering background information and implementing effective background suppression strategies to further enhance detection performance. This research direction holds substantial theoretical significance and practical value.

Table 5. Comparison of Target Detection Performance in Hyperspectral Images Under Different Optimization Strategies. ✓ indicates that the method is adopted, × indicates that the method is not adopted; bold values represent the optimal results.

HSIs	Superpixel Segmentation	Background Suppression	$AUC(P_f, P_d)$	$AUC(\tau, P_f)$
ABU-beach-1	×	×	0.9843	0.0095
	×	✓	0.9893	0.0005
	✓	×	0.9884	0.0103
	✓	✓	0.9902	0.0004
HYDICE	×	×	0.9896	0.0265
	×	✓	0.9932	0.0063
	✓	×	0.9918	0.0344
	✓	✓	0.9946	0.0045
San Diego	×	×	0.9802	0.0295
	×	✓	0.9911	0.0023
	✓	×	0.9843	0.0074
	✓	✓	0.9971	0.0008
San Diego2	×	×	0.9778	0.0331
	×	✓	0.9914	0.0009
	✓	×	0.9812	0.0128
	✓	✓	0.9953	0.0012

3.3.2. Parameter Analysis

This section focuses on two key parameters in our proposed method: the time step t and the number of superpixels N_s .

In the multivariate Gaussian diffusion model, the time step t plays a critical role. It not only directly influences the generation of background noise but also significantly affects the performance of the denoising network. When t is relatively small, the addition of background noise is conservative, which may lead to insufficient background suppression. In contrast, an excessively large t can enhance the suppression capability but may introduce excessive noise, thereby degrading target detection accuracy. To systematically evaluate the impact of the time step on detection performance, we designed a series of controlled experiments by varying the value of t and observing its effect on detection accuracy. Based on the statistical analysis of the results presented in Figure 7, we identified the optimal range of t for different datasets.

In the process of background noise generation, we adopt a superpixel-based background noise generation strategy (SSBNG), in which the image is first segmented into multiple superpixels using the SLIC algorithm. The number of superpixels N_s directly affects the quality of the background noise, thereby influencing the final detection performance. Figure 8 illustrates the detection results under different N_s settings across four datasets. It can be observed that, for all datasets, the $AUC(P_f, P_d)$ values initially increase with larger N_s , reach a peak, and then gradually decline. Due to the varying spatial resolutions of the datasets, we selected different values of N_s for each. Specifically, the ABU-beach-1 dataset achieves the highest AUC when $N_s = 2000$; for the HYDICE dataset, the best performance occurs at $N_s = 1000$; for the San Diego dataset, the optimal number of superpixels is $N_s = 1750$; and for the San Diego2 dataset, the best detection performance is observed when $N_s = 1250$.

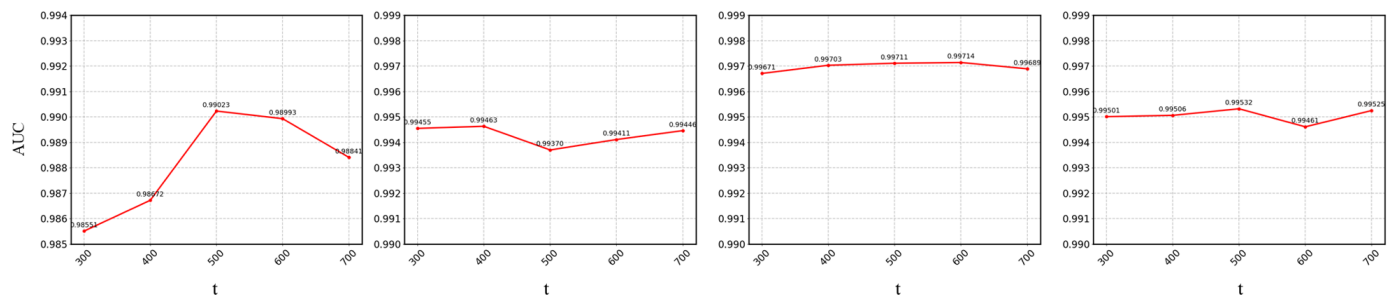


Figure 7. The impact of different time steps t on $AUC(P_f, P_d)$ across four datasets. From left to right: ABU-beach-1, HYDICE, San Diego, San Diego2.

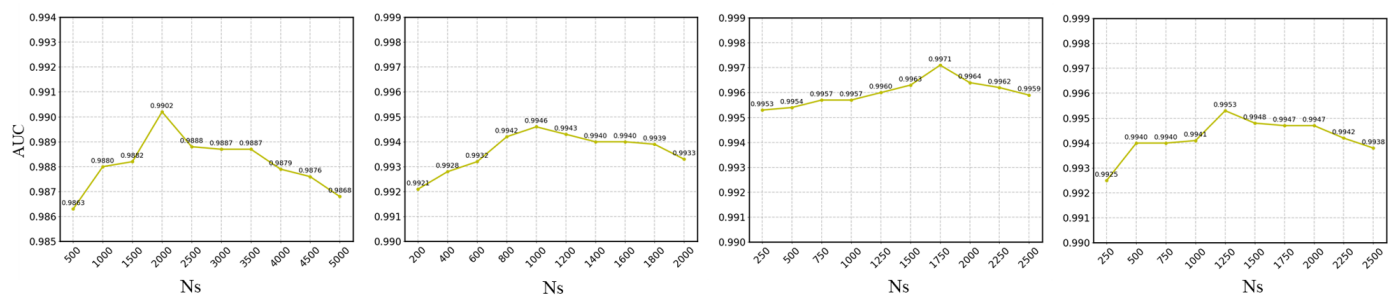


Figure 8. The impact of different number of superpixels N_s on $AUC(P_f, P_d)$ across four datasets. From left to right: ABU-beach-1, HYDICE, San Diego, San Diego2.

4. Discussion

4.1. Spatial–Spectral Fusion and Background Modeling

The experimental results indicate that the proposed diffusion-based hyperspectral target detection method achieves strong background suppression and competitive detection performance across multiple datasets. A key factor underlying this improvement is the integration of spatial and spectral information through the centre-weighted superpixel strategy. Unlike baseline methods such as BLTSC and OS-VAE, which primarily rely on spectral information, our approach explicitly captures local spatial context, allowing the model to differentiate subtle target signatures from complex and heterogeneous background patterns. This spatial–spectral fusion ensures that background suppression is not solely based on spectral differences but also considers the spatial continuity and neighborhood similarity of hyperspectral scenes.

Another important contributor to the performance is the modeling of multivariate Gaussian background noise within the diffusion–denoising framework. By simulating realistic background variability and injecting it into the forward diffusion process, the denoising network can learn a robust representation of the background distribution. This contrasts with deterministic or variational baseline methods, which may not fully account for stochastic background fluctuations across both spatial and spectral dimensions, and therefore may fail to suppress challenging background structures effectively.

4.2. Impact of Hyperparameters

Hyperparameter selection also plays a critical role in our method. The number of noise samples N_s reflects the spatial aggregation of background statistics across superpixel blocks, ensuring that the generated noise accurately represents local background variability. Larger N_s provides more robust background modeling, while smaller N_s may capture finer local differences but could be more sensitive to noise. The diffusion step t governs the degree of spectral perturbation introduced at each diffusion step, influencing how strongly the network is guided to learn spectral patterns associated with the background. A higher

t generally encourages more aggressive suppression of background noise, but excessive t can distort subtle target features, necessitating dataset-specific tuning. These insights explain why optimal hyperparameters vary across datasets, reflecting differences in spatial heterogeneity, spectral variability, and target size.

4.3. Relation to Generative Models

Our findings can be placed in the broader context of generative models for hyperspectral remote sensing. Generative models, including variational autoencoders (VAEs) and diffusion models, have been widely used for background modeling, anomaly detection, and data augmentation. The flexibility of these models enables them to capture complex distributions in hyperspectral data, providing a principled approach to learning background structure and detecting anomalies. Our work demonstrates that combining spatial–spectral contextual information with diffusion-based background modeling leverages the advantages of generative modeling while addressing the specific challenges of hyperspectral target detection, particularly in scenarios with highly heterogeneous backgrounds or small, subtle targets.

4.4. Statistical Variability and Computational Efficiency

Table 6 summarizes the detection performance (AUC) and computational time of DSC, BLTSC, OS-VAE, and our method on four hyperspectral datasets. To quantify statistical variability, we conducted multi-seed experiments for the stochastic algorithms and report mean $\pm$ standard deviation (SD) of the AUC. As shown in the table. Our method achieves consistently high AUC across all datasets with very small standard deviations, indicating stable performance under different random initializations. Regarding computational efficiency, our method is faster than BLTSC and comparable to DSC and OS-VAE, demonstrating favorable trade-offs between performance and runtime. Overall, this table highlights the stability of the proposed method across multiple runs and provides a clear comparison of computational cost, supporting the reliability of the experimental evaluation.

Table 6. Mean $\pm$ Standard Deviation of AUC and Runtime of Different Algorithms in Different Datasets.

Dataset	Metric	DSC	BLTSC	OS-VAE	OURS
ABU-beach-1	AUC	0.9823 $\pm$ 0.0015	0.9769 $\pm$ 0.0019	0.9777 $\pm$ 0.0009	0.9905 $\pm$ 0.0006
	Time(s)	23.33 $\pm$ 0.17	21.39 $\pm$ 0.59	134.69 $\pm$ 0.22	20.68 $\pm$ 0.30
HYDICE	AUC	0.9889 $\pm$ 0.0017	0.9775 $\pm$ 0.0003	0.9917 $\pm$ 0.0011	0.9948 $\pm$ 0.0013
	Time(s)	7.94 $\pm$ 1.11	158.71 $\pm$ 3.16	55.83 $\pm$ 0.57	8.74 $\pm$ 0.24
San Diego	AUC	0.9962 $\pm$ 0.0008	0.9937 $\pm$ 0.0008	0.9901 $\pm$ 0.0014	0.9972 $\pm$ 0.0005
	Time(s)	10.46 $\pm$ 0.07	14.86 $\pm$ 19.85	56.72 $\pm$ 1.44	10.19 $\pm$ 0.07
San Diego2	AUC	0.9895 $\pm$ 0.0014	0.9884 $\pm$ 0.0027	0.9907 $\pm$ 0.0004	0.9954 $\pm$ 0.0004
	Time(s)	10.44 $\pm$ 0.11	12.73 $\pm$ 2.65	41.25 $\pm$ 0.78	12.07 $\pm$ 0.29

5. Conclusions

In this work, we present a diffusion-model-based framework for hyperspectral target detection that leverages multivariate Gaussian background suppression. The proposed model consists of three synergistic components: (1) a spatial–spectral centre-weighted background noise generation module, (2) a background suppression mechanism based on a multivariate Gaussian noise model, and (3) a target detection module. In the noise generation stage, superpixel segmentation and a spatial–spectral weighting function jointly introduce local spatial context, while pure background samples are extracted using a constrained energy minimisation strategy to construct accurate multivariate Gaussian back-

ground noise. In the background suppression stage, the generated noise is incorporated into the diffusion process, enabling the denoising network to learn the background distribution and effectively suppress background interference. These components collectively facilitate more reliable target detection. Experimental results on four public hyperspectral datasets show that the proposed method achieves strong background suppression and competitive overall detection performance.

A limitation of the proposed method is the assumption that the background follows a single multivariate Gaussian distribution, which may not fully capture the complexity of highly heterogeneous scenes. In future work, this assumption can be relaxed by adopting mixture-based or non-parametric background models to better represent multi-modal background distributions.

Author Contributions: Conceptualization, G.Z. and J.F.; methodology, W.H.; validation, Y.H.; writing—original draft preparation, W.H.; writing—review and editing, R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Jiangsu Province (Grant Number BK20230338); National Natural Science Foundation of China (Grant Number 41601365).

Data Availability Statement: Publicly available datasets were analyzed in this study. The datasets can be found here: <https://aviris.jpl.nasa.gov/data/index.html>, <http://xudongkang.weebly.com/datasets.html>, <https://github.com/sxt1996/HYDICE>, accessed on 10 September 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nasrabadi, N.M. Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Process. Mag.* **2013**, *31*, 34–44. [CrossRef]
2. Zare, A.; Jiao, C.; Glenn, T. Discriminative multiple instance hyperspectral target characterization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 2342–2354. [CrossRef]
3. Willett, R.M.; Duarte, M.F.; Davenport, M.A.; Baraniuk, R.G. Sparsity and structure in hyperspectral imaging: Sensing, reconstruction, and target detection. *IEEE Signal Process. Mag.* **2013**, *31*, 116–126. [CrossRef]
4. Ma, J.; Xie, W.; Li, Y.; Fang, L. Bsdm: Background suppression diffusion model for hyperspectral anomaly detection. *arXiv* **2023**, arXiv:2307.09861. [CrossRef]
5. Wang, D.; Zhuang, L.; Gao, L.; Sun, X.; Huang, M.; Plaza, A. BockNet: Blind-block reconstruction network with a guard window for hyperspectral anomaly detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5531916. [CrossRef]
6. Wang, D.; Gao, L.; Qu, Y.; Sun, X.; Liao, W. Frequency-to-spectrum mapping GAN for semisupervised hyperspectral anomaly detection. *CAAI Trans. Intell. Technol.* **2023**, *8*, 1258–1273. [CrossRef]
7. Sun, L.; Wang, X.; Zheng, Y.; Wu, Z.; Fu, L. Multiscale 3-d-2-d mixed cnn and lightweight attention-free transformer for hyperspectral and lidar classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 2100116. [CrossRef]
8. Su, Y.; Gao, L.; Jiang, M.; Plaza, A.; Sun, X.; Zhang, B. NSCKL: Normalized spectral clustering with kernel-based learning for semisupervised hyperspectral image classification. *IEEE Trans. Cybern.* **2022**, *53*, 6649–6662. [CrossRef] [PubMed]
9. Feng, J.; Wang, Q.; Zhang, G.; Jia, X.; Yin, J. CAT: Center Attention Transformer with Stratified Spatial-Spectral Token for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5615415. [CrossRef]
10. Liu, S.; Marinelli, D.; Bruzzone, L.; Bovolo, F. A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 140–158. [CrossRef]
11. Liu, J.; Zhang, W.; Liu, F.; Xiao, L. A probabilistic model based on bipartite convolutional neural network for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 4701514. [CrossRef]
12. Zheng, D.; Wu, Z.; Liu, J.; Xu, Y.; Hung, C.C.; Wei, Z. Explicit change-relation learning for change detection in VHR remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6005005. [CrossRef]
13. Xie, W.; Zhang, X.; Li, Y.; Lei, J.; Li, J.; Du, Q. Weakly supervised low-rank representation for hyperspectral anomaly detection. *IEEE Trans. Cybern.* **2021**, *51*, 3889–3900. [CrossRef] [PubMed]
14. Farrand, W.H.; Harsanyi, J.C. Mapping the distribution of mine tailings in the Coeur d’Alene River Valley, Idaho, through the use of a constrained energy minimization technique. *Remote Sens. Environ.* **1997**, *59*, 64–76. [CrossRef]

15. Zou, Z.; Shi, Z. Hierarchical suppression method for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 330–342. [\[CrossRef\]](#)
16. Zhao, R.; Shi, Z.; Zou, Z.; Zhang, Z. Ensemble-based cascaded constrained energy minimization for hyperspectral target detection. *Remote Sens.* **2019**, *11*, 1310. [\[CrossRef\]](#)
17. Chang, C.I. Constrained energy minimization (CEM) for hyperspectral target detection: Theory and generalizations. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5522921. [\[CrossRef\]](#)
18. Chen, Z.; Lu, Z.; Gao, H.; Zhang, Y.; Zhao, J.; Hong, D.; Zhang, B. Global to local: A hierarchical detection algorithm for hyperspectral image target detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5544915. [\[CrossRef\]](#)
19. Chang, C.I. Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 502–518. [\[CrossRef\]](#)
20. Ren, H.; Chang, C.I. Target-constrained interference-minimized approach to subpixel target detection for hyperspectral images. *Opt. Eng.* **2000**, *39*, 3138–3145. [\[CrossRef\]](#)
21. Chen, J.; Chang, C.I. Background-annihilated target-constrained interference-minimized filter (TCIMF) for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5540224. [\[CrossRef\]](#)
22. Kruse, F.A.; Lefkoff, A.; Boardman, J.; Heidebrecht, K.; Shapiro, A.; Barloon, P.; Goetz, A. The spectral image processing system (SIPS)—Interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [\[CrossRef\]](#)
23. Manolakis, D.; Marden, D.; Shaw, G.A. Hyperspectral image processing for automatic target detection applications. *Linc. Lab. J.* **2003**, *14*, 79–116.
24. Scharf, L.L.; Friedlander, B. Matched subspace detectors. *IEEE Trans. Signal Process.* **1994**, *42*, 2146–2157. [\[CrossRef\]](#)
25. Kraut, S.; Scharf, L.L. The CFAR adaptive subspace detector is a scale-invariant GLRT. *IEEE Trans. Signal Process.* **1999**, *47*, 2538–2541. [\[CrossRef\]](#)
26. Kraut, S.; Scharf, L.L.; McWhorter, L.T. Adaptive subspace detectors. *IEEE Trans. Signal Process.* **2001**, *49*, 1–16. [\[CrossRef\]](#)
27. Fuhrmann, D.R.; Kelly, E.J.; Nitzberg, R. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [\[CrossRef\]](#)
28. Li, W.; Wu, G.; Du, Q. Transferred deep learning for hyperspectral target detection. In Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, USA, 3–28 July 2017; pp. 5177–5180.
29. Zhang, G.; Zhao, S.; Li, W.; Du, Q.; Ran, Q.; Tao, R. HTD-Net: A deep convolutional neural network for target detection in hyperspectral imagery. *Remote Sens.* **2020**, *12*, 1489. [\[CrossRef\]](#)
30. Shen, D.; Ma, X.; Kong, W.; Liu, J.; Wang, J.; Wang, H. Hyperspectral target detection based on interpretable representation network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5519416. [\[CrossRef\]](#)
31. Xie, W.; Yang, J.; Lei, J.; Li, Y.; Du, Q.; He, G. SRUN: Spectral regularized unsupervised networks for hyperspectral target detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 1463–1474. [\[CrossRef\]](#)
32. Kingma, D.P. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
33. Shi, Y.; Li, J.; Yin, Y.; Xi, B.; Li, Y. Hyperspectral target detection with macro-micro feature extracted by 3-D residual autoencoder. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 4907–4919. [\[CrossRef\]](#)
34. Xie, W.; Zhang, X.; Li, Y.; Wang, K.; Du, Q. Background learning based on target suppression constraint for hyperspectral target detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 5887–5897. [\[CrossRef\]](#)
35. Tian, Q.; He, C.; Xu, Y.; Wu, Z.; Wei, Z. Hyperspectral Target Detection: Learning Faithful Background Representations via Orthogonal Subspace-Guided Variational Autoencoder. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5516714. [\[CrossRef\]](#)
36. Nichol, A.Q.; Dhariwal, P. Improved denoising diffusion probabilistic models. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 18–24 July 2021; pp. 8162–8171.
37. Kingma, D.; Salimans, T.; Poole, B.; Ho, J. Variational diffusion models. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 21696–21707.
38. Ho, J.; Jain, A.; Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
39. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. Photorealistic text-to-image diffusion models with deep language understanding. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 36479–36494.
40. Richter, J.; Welker, S.; Lemercier, J.M.; Lay, B.; Gerkmann, T. Speech enhancement and dereverberation with diffusion-based generative models. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2023**, *31*, 2351–2364. [\[CrossRef\]](#)
41. Alcaraz, J.M.L.; Strodthoff, N. Diffusion-based time series imputation and forecasting with structured state space models. *arXiv* **2022**, arXiv:2208.09399.
42. Rasul, K.; Seward, C.; Schuster, I.; Vollgraf, R. Autoregressive denoising diffusion models for multivariate probabilistic time series forecasting. In Proceedings of the International Conference on Machine Learning. PMLR, Virtual, 18–24 July 2021; pp. 8857–8868.
43. Zhang, M.; Qamar, M.; Kang, T.; Jung, Y.; Zhang, C.; Bae, S.H.; Zhang, C. A survey on graph diffusion models: Generative ai in science for molecule, protein and material. *arXiv* **2023**, arXiv:2304.01565. [\[CrossRef\]](#)

44. Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional diffusion for molecular conformer generation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24240–24253.
45. Chen, N.; Yue, J.; Fang, L.; Xia, S. SpectralDiff: A generative framework for hyperspectral image classification with diffusion models. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5522416. [[CrossRef](#)]
46. Tian, J.; Yu, W.Y.; Xie, S.L. On the kernel function selection of nonlocal filtering for image denoising. In Proceedings of the 2008 International Conference on Machine Learning and Cybernetics, Kunming, China, 12–15 July 2008; Volume 5, pp. 2964–2969.
47. Shen, D.; Ma, X.; Wang, H.; Liu, J. A dual sparsity constrained approach for hyperspectral target detection. In Proceedings of the IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 17–22 July 2022; pp. 1963–1966.
48. Chang, C.I. An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5131–5153. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.