Hyperspectral anomaly detection based on variational background inference and generative adversarial network

Zhiwei Wang [a,b,1], Xue Wang [a,b,1], Kun Tan [a,b,*], Bo Han [c], Jianwei Ding [d], Zhaoxian Liu [d]

[a] *Key Laboratory of Geographic Information Science, East China Normal University, Shanghai, 200241, China*
[b] *Key Laboratory of Spatial-Temporal Big Data Analysis and Application of Natural Resources in Megacities (Ministry of Natural Resources), East China Normal UniversityShanghai, 200241, China*
[c] *Institute of Remote Sensing Satellite, China Academy of Space Technology, Beijing, 100081, China*
[d] *Second Surveying and Mapping Institute of Hebei, Shijiazhuang, 050037, China*

ARTICLE INFO

ABSTRACT

Hyperspectral anomaly detection is aimed at detecting targets with significant spectral differences from their surroundings. Recently, deep generative models have been applied to anomaly detections, while the existing generative adversarial network (GAN)-based methods have difficulty in accurately modeling the background and achieving spectrum reconstruction. In this article, a hyperspectral anomaly detection network based on variational background inference and generative adversarial framework (VBIGAN-AD) is proposed. The proposed VBIGAN model can learn the background distribution characteristics of HSIs and enhance the detection performance by the use of reconstruction errors. Specifically, the VBIGAN framework consists of sample and latent GANs, which establishes the relationship between data samples and latent samples through two sub-networks to capture the data distribution. Furthermore, the variational inference method is introduced and the hyperspectral background distribution can be converged to a multivariate normal distribution. To accurately learn the background distribution characteristics and reconstruct the background spectra, the coupling loss is conducted by enforcing feature match in the two discriminators on the basis of composite loss, and the results show that the additional loss can promote the detection performance. As a result, the reconstruction errors generated by the VBIGAN-AD method is utilized to detect abnormal targets. The experiments conducted on five datasets proved the robustness and applicability of the proposed VBIGAN-AD method.

1. Introduction

Hyperspectral image (HSI) is a three-dimensional cube in which two dimensions provide the spatial features of the materials, and the other one offers continuous spectral reflectance vectors [1]. Accordingly, hyperspectral imagery has received with great attention in various application areas, such as image classification [2,3], dimensionality reduction [4], and target/anomaly detection [5]. Hyperspectral anomaly detection, without any prior spectra of the targets, is aimed at locating the abnormal targets with significant spectral differences from the background [6].

During the past decades, anomaly detection algorithms with different background modeling approaches have been developing. The most classical algorithm of the statistical-based methods is the Reed-Xiaoli (RX) detector [7]. The RX algorithm assumes that the background satisfies a multivariate normal distribution and uses the Mahalanobis distance to detect anomalies. There are two kinds of RX algorithm, which are the global RX (GRX) and local RX (LRX) [8].The difference between them is whether use the entire image or the local double window for modeling. Zhang et al. [9] transformed HSI into frequency domain only by fractional Fourier transform and then used the tensor RX algorithm to detect anomalies. Moreover, researchers improve the background reliability of RX algorithm from the perspective of kernel method and spatial information [10,11]. Nevertheless, the background distribution in the real world is quite complex, which usually leads to lower detection accuracy [11].

Differing from the aforementioned statistical distribution approaches, representation based background modeling algorithms have been rapidly developing. Because the background of the HSI has low-rank characteristics, a low-rank representation (LRR) theory was introduced into anomaly detection [12]. The low-rank and sparse matrix decomposition based Mahalanobis dis-

tance (LSMAD) model detected abnormal targets by obtaining purer background information and employing the Mahalanobis distance [13]. Zhang et al. [14] used spectral difference low-rank dictionary representation learning for global background modeling, which accurately constructs the pure background dictionary. Wang et al. [15] proposed a detector based on a tensor low-rank sparse representation and constructed a dictionary learning algorithm to effectively characterize the multiple subspace properties of complex backgrounds. Chang et al. [16] proposed the unsupervised OSP-AD algorithm based on OSP-GoDec, and the generated low-rank matrix and sparse matrix can be utilized as the background and target subspaces respectively. Alternatively, Li et al. [17] proposed a collaborative representation based detector (CRD), which uses ensemble learning by the combination of neighboring pixels and the linear representation principle, and finally uses the spectral residual values to determine anomalies. Tan et al. [18] combined inverse distance weighting and a multiple-window sliding filter algorithm based on the CRD algorithm, which adequately incorporates multi-scale spatial features to enhance the adaptation to complex backgrounds. Zhao et al. [19] combined a weighted Cauchy distance graph and local adaptive CRD to make full use of spatial and spectral information, which is reliable and robust in complicated background conditions. However, for the conventional detection methods mentioned above, the background of the hyperspectral imagery becomes complex, and often involves more parameter settings, which makes the traditional methods less powerful of target detection and applications.

Deep learning based methods have been widely used for anomaly detection due to their ability of capturing the distributional properties of complex data and deep learning features [6]. Unsupervised deep learning methods, such as autoencoders (AEs), are widely employed for hyperspectral anomaly detection because they can learn deep features and reconstruct original spectra without label information [20–22]. Zhao et al. [21] used a stacked denoising autoencoder (SDA) to extract high-level features of the spectra and improve the targets detection accuracy. Fan et al. [22] investigated robust graph AE (RGAE) detector by embedding a super pixel segmentation-based graph regularization term into AE, which was demonstrated to preserve the spatial structure of HSI. An autonomous anomaly detection (Auto-AD) method was proposed by Wang et al. [23], which was the first to use full convolutional AE to extract spatial features and thus improved the model detection capability. Furthermore, Wang et al. [24] combined model-driven low-rank prior and data-driven full convolutional AE to propose the deep low rank prior-based method (DeepLR), in which the network parameters can be updated with low-rank background through an iterative optimization framework. Deep generative models express the probability distribution of multivariate data in a certain way, and some scholars have been working on this direction [25–27]. For example, Xie et al. [25] added a spectrally constrained strategy into an AAE to learn the latent representations, and then used a two-layer structure to achieve anomaly detection. In spatial and spectral constraints (SASCs) [26], Wasserstein distance and spatial filtering (AD-WDSF) [28], authors introduced spatial feature extraction methods to improve detection performance. In addition, the concept based on weakly supervised learning is applied in background distribution learning, which is mainly used to improve the reliability of target detection by constructing a background sample extraction strategy [6,27,29].

The deep learning-based anomaly detection methods above have significantly improved the anomaly detection performance from the perspectives of spectral feature extraction, weakly supervised learning, etc. However, the current deep learning methods for anomaly detection still have the following drawbacks:

(1) Since the traditional AE-based algorithms are basically deterministic mappings, the detection algorithms based on AE are hard to handle variations in background and abnormal samples. Second, the various variants of GAN networks suffer from training instability, and low quality of generated samples, resulting in high false positives and low detection accuracy in anomaly detection [28].

(2) Due to the lack of prior knowledge of the targets and backgrounds, the existing deep learning-based algorithms cannot fully take the statistical distribution characteristics of backgrounds and anomalies into consideration. Meanwhile, the training samples contain both background and anomaly samples, and the anomaly samples will impact the training process of the network, so the accuracy of anomaly detection in reconstruction error is reduced [29].

In this article, to deal with the aforementioned problems, a novel variational background inference based on generative adversarial network (VBIGAN-AD) is proposed, which relies on the reconstruction errors generated by the network to detect anomalies. Specifically, the complex distribution properties of the background with the low probability properties of the anomalies inspired to construct a robust network that accurately learns the background distribution and effectively improves the detection accuracy. Thus, the variational inference method is introduced into the VBIGAN model to make the hyperspectral background distribution converge to a multivariate normal distribution, which effectively avoids the interference of the abnormal samples on the network background learning. The VBIGAN framework consists of latent GAN and sample GAN. The latent GAN is used to discriminate whether the latent variable is from hyperspectral background distribution or a multivariate normal distribution. From the perspective of the latent variables, we design multi-component samples fed into the sample discriminator to accurately reconstruct background spectra over anomalies. Furthermore, we optimize the spectral reconstruction loss and the adversarial loss, and design the discriminative feature matching loss to ensure the stable training of the network model. The main contributions of this article are as follows:

(1) We investigate a novel VBIGAN-AD anomaly detection algorithm consisting of a sample GAN and a latent GAN, and each sub-network has independent discriminator. Two GAN subnetworks establish the relationship between data samples and latent samples to better improve detection performance.

(2) We integrate the variational inference and adversarial learning to make the hyperspectral background distribution converge to multivariate distribution, which can effectively capture the distribution characteristics of background and anomaly.

(3) We devise the composite loss function and discriminative feature matching loss to effectively ensure the stable learning of background distribution characteristics and effective generation of spectra.

The rest of this article is organized as follows. Section 2 is a brief introduction to the AE and GAN networks. Section 3 details the basic principles of the proposed VBIGAN-AD algorithm. Section 4 depicts the experimental results. Section 5 concludes the article.

2. Related work

Recently, deep learning has been applied in the fields of anomaly detection, because of its advantage of being able to automatically learn the data distribution [30]. The basic networks structures commonly used in anomaly detection are shown in Fig. 1. As a typical unsupervised network, the autoencoder (AE) [20] shows great application value for the accurate learning of data
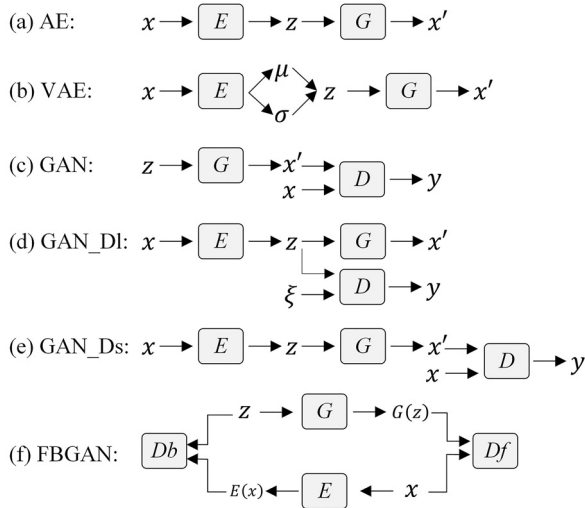
Fig. 1. Illustration of the structure of AE, VAE, GAN, GAN_Dz, GAN_Ds, and FBGAN networks, where x and x' are the input and generated data, z is the latent vector, and y represents the data real or fake. E, G, and D are the encoder, generator, and discriminator, respectively.

distributions and spectral reconstruction. Deep generative model is one that can learn the underlying feature distribution of high-dimensional data by observing the existing data samples and generating data with the same distribution as the real data. Among the most representative network models are the variational autoencoders (VAEs) [31] and GANs [32]. Therefore, the next section focuses on AE and GAN networks.

2.1. Autoencoder

As shown in Fig. 1(a), the AE [20] is composed of an encoder E and a generator G The E projects the inputs x into the latent space z, and then the G uses z to reconstruct x' reversely. The network is trained using the reconstruction loss $L = \|x - x'\|_2$ to force the output to be equal to its input. The AE is calculated as follows:

$$x' = G(E(x)) \tag{1}$$

However, the AE has difficulty in accurately characterizing the background distribution and learning weak abnormal information, which makes it difficult for AE to detect anomalies. As an advanced version of AE, VAE shows great potential for generating complex data [31]. VAE uses the encoder to learn a variance function $q(z)$ with parameters μ and σ, which maps the observed variables to the latent variables and then converges the distribution to a fixed distribution. Thus, the E can be modeled based on the mapping results and the G can generate data with the same distribution as the observed variables. Moreover, the hyperspectral data is mainly of multivariate distribution, while the sampling process of VAE is based on single-peaked distribution, which makes the generated data blurred.

2.2. Generative adversarial network

The GAN framework [32], as illustrated in Fig. 1(c), which is composed of a generator G and a discriminator D. The task of the G is to map the input $z \sim p_z(z)$ to a given real $x \sim p(x)$ as much as possible. Meanwhile, the goal of the D is to determine whether the data is true or false. The objective function of the GAN network is as follows:

$$\min_{G} \max_{D} E_{x \sim p(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{2}$$

Afterwards, many new variants of GAN have been developed [33,34]. The GAN to accurately capture the distributional properties of the data to generate new expected spectral data, which coincides with background modeling in anomaly detection [35]. For example, with the assumption that the background of the HSI obeys a multivariate normal distribution, a latent discriminator Dl is utilized to adversarial training with the E, which is named the GAN_Dl framework, as shown in Fig. 1(d). The GAN_Dl is used to match the distribution of latent variable data to the prior distribution $\xi = \mathcal{N}(0, 1)$, which facilitates stable training of the model [29]. However, the above GAN_Dl framework has difficulty in maintaining the detailed features of the original data, which further decreases the detection performance. The GAN_Ds is used to match the generated samples of the generator by adding a sample discriminator Ds, as shown in Fig. 1(e), which enables the Ds to help the G competitively generate stable spectral samples [29]. As shown in Fig. 1(f), the forward-backward GANs (FBGANs) consists of encoder, generator, forward discriminator and backward discriminator, which can better capture the data distribution and are applied in MINST dataset anomaly detection [36]. In this study, we chose a GAN network as the basic structure, and then build a new adversarial network framework that can effectively applied to hyperspectral anomaly detection.

3. Proposed method

The real hyperspectral image is mainly of multivariate distributions, while the sampling process of VAE is based on single-peaked distribution, which results in the generated data often being blurred. By contrast, GAN generates samples randomly with multivariate sampling, and the generated samples can reflect clearer and richer details than VAE. The GAN_Dl and GAN_Ds have shown good performances in anomaly detection, whereas the GAN_Dl tends to generate blurred spectral data. The GAN_Ds attaches a sample discriminator that produces stable spectral samples with more accurate spectral detail features. The combination of GAN and VAE has been explored in recent studies [37]. VAE-GAN combines the advantages of VAE and GAN to form an unsupervised generative model, which can improve the shortcomings of the base model. Moreover, FBGAN can better capture the data distribution by learning the probability distribution of normal and abnormal examples [36].

Therefore, inspired by the combination of VAE-GAN [37] and FBGAN [36], a variational background inference and GAN (VBIGAN) is proposed to address the problems of inaccurate background modeling and spectral reconstruction. Mathematically, we define $X = \{x_1, x_2, \ldots, x_{m \times n}\} \in \mathbb{R}^{B \times mn}$, where B represents the number of bands, m and n represent the number of image rows and columns. Fig. 2 shows the flow chart of the VBIGAN framework, which is composed of three parts: data preparation, network learning, and reconstruction error generation for anomaly detection. Firstly, we utilize all pixels $x_i \in \{i = 1, \ldots, mn\}$ for training in each iteration. The connection between real data sampling and latent sampling is established and the distribution characteristics of the background are learned during the learning process of the VBIGAN. After the iterative training process, we can obtain the reconstruction error of the generated and the original HSI based on the trained VBIGAN to detect abnormal targets. Next, we will introduce the VBIGAN framework in detail.

3.1. VBIGAN for anomaly detection

The proposed VBIGAN framework contains encoder E, generator G, sample discriminator Ds, and latent discriminator Dl, as shown in Fig. 2. In the following, we focus on the network learning part of the VBIGAN framework, which consists of a sample GAN and a
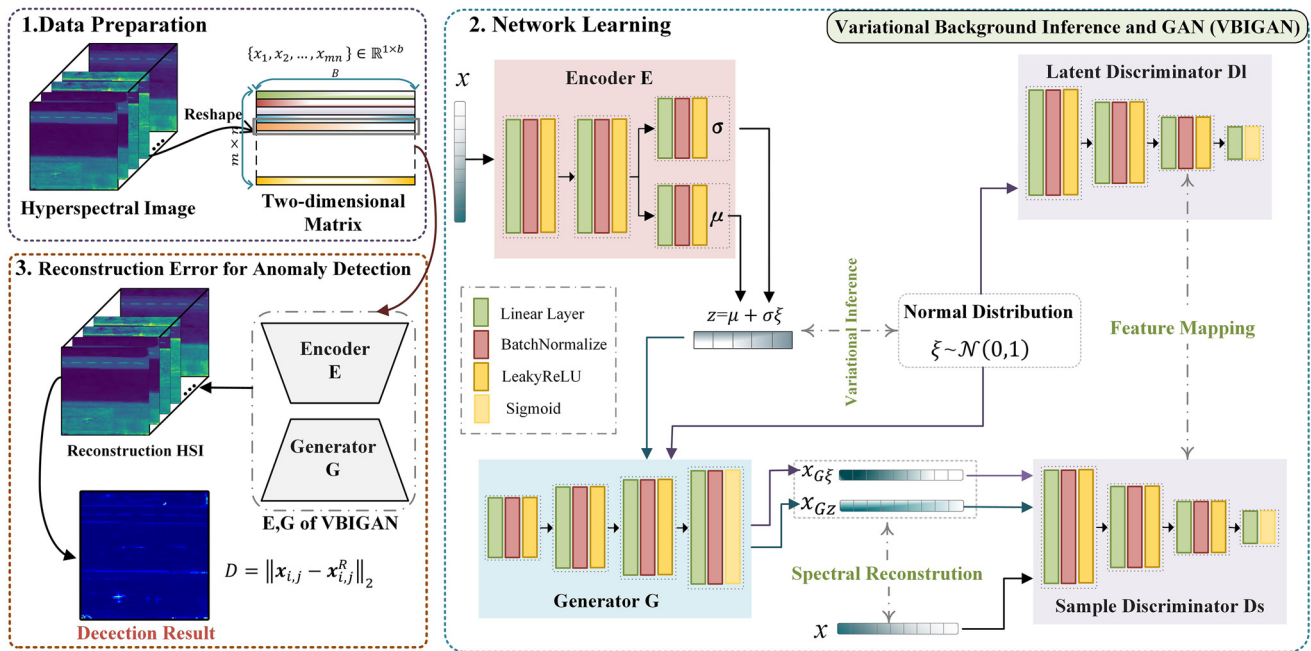
Fig. 2. Process overview of the VBIGAN-AD framework.

latent GAN. In the latent GAN, the mapping from spectra to latent samples is firstly created, and then the variational inference is introduced to learn the distributional characteristics of backgrounds and anomalies in latent space. The sample GAN is utilized as the spectra generation process, which generates the counterfeit spectra by the normal distribution samples and latent variable samples. After that, the adversarial learning is conducted to ensure accurate reconstruction of the spectra.

We first introduce the variational inference approach, where the E learns to map the observed variables to the latent variables, and then converges the distribution to a multivariate normal distribution. Specifically, the variance and mean of the latent variables are first obtained by passing the observed variables through the E, and then the sampled data is obtained by reparameterization, which is consistent with the forward propagation of VAE. Sampling from multivariate normal distribution can ensure the latent samples are better mapped to the background spectra. For the Dl, the input contains the normal distribution sample ξ and the latent variable sample z, which facilitates the learning of the background distribution characteristics. In this process, the reparametrized sampled data transformed by the E is considered as generated fake data, while the data sampled in the fixed distribution is considered as real data. The adversarial training process guides the reparametrized generated data to match the prior distribution, and the encoder aims to fool the Dl into trusting that the data sampled by the hidden variables is from the true distribution. The Dl is included to determine whether the data is from the true or fixed distribution, which enhances the detailed characteristics of the generated hyperspectral data.

Moreover, we use the reconstruction property of background and anomaly samples for detection from the perspective of spectral generation. A large number of background samples can be extracted based on the latent variable z, which enhances the reliability of background learning. For the anomaly samples extracted from the latent distribution, the large variance makes the generated data vastly different from the true anomaly samples. The normal distribution sample ξ and the latent variable sample z are input into the G to generate two generated samples $x_{G\xi}$ and x_{Gz}, respectively. x_{Gz} denotes the spectral feature obtained from the E

and then reconstructed by the G. x_{Gz} denotes the spectral feature obtained from the E and then reconstructed by the G. $x_{G\xi}$ represents a faked spectral generated by the G from the fixed distribution sample. The fixed distribution has been fixed as normal distribution, which can generate HSIs that correspond to the real background distribution. For the Ds, the multi-component input contains not only the samples x but also samples $x_{G\xi}$ and x_{Gz}. The real sample x can enable the Ds to distinguish between true and false samples. $x_{G\xi}$ and x_{Gz} serve the same purpose of discriminating the sample true and false, the spectra are accurately reconstructed, and the background distribution is learned. Therefore, the Ds can help the G to generate more accurate background spectra and increase the detail features.

Therefore, the purpose of the E is to obtain the latent variable sample z using variational inference, and the purpose of the Dl is to discriminate whether the data are from the latent variable sample z or the normally distributed sample ξ. Therefore, the training of the E and the Dl are adversarial relationship. The G ensures that the spectra are accurately generated by the sample ξ and z, while the aim of the Ds is to discriminate whether the data is fake or real, so the G performs adversarial learning with the Ds. Moreover, the two discriminators independently implement adversarial learning, and at the same time, the two discriminators are coupled with each other to better capture the data distribution. The VBIGAN framework is able to fuse the respective advantages of the sample GAN and the latent GAN, couple the feature matching of the two discriminators to keep the input and output consistent.

The four networks structures of the VBIGAN model is shown in Fig. 3 The E consists of three fully connected layers, where μ and σ come from two different fully connected layers, and the leaky ReLU is utilized. The G consists of four fully connected layers whose output dimension is consistent with the number of bands, and the last layer is activated using a sigmoid function. The Ds and the Dl are also consisting of four fully connected layers, and the last layer of both discriminators utilizes a sigmoid function to output the discriminative probabilities for loss calculation during the adversarial training. In addition, the design outputs feature vectors in the third layer of the network for coupling the two GAN sub-networks.
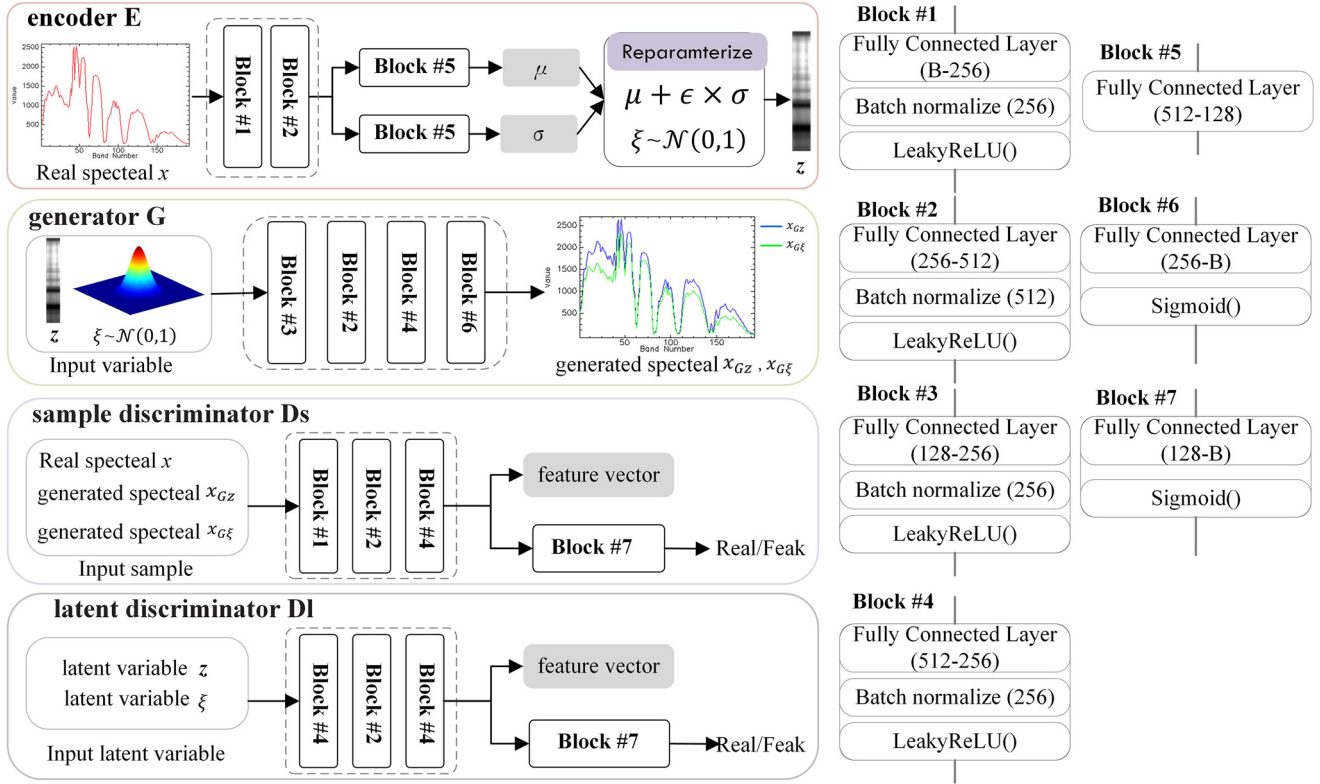
Fig. 3. The implementation of each network in the VBIGAN framework.

3.2. Loss terms of the VBIGAN framework

(1) Variational Inference Loss: The purpose of our introduced variational inference process in the VBIGAN framework is to learn the latent variable distribution to sample the real data distribution. The variational inference process can be described as $z = \mathcal{N}(\mu, \sigma^2)$ and the latent variable z can be reparameterized by $z = \mu + \sigma\xi$, where $\xi = \mathcal{N}(0, 1)$ is an auxiliary variable, μ and σ are encoded by the E. Using the Kullback-Leibler divergence penalty, the sample process for z is based on a reparameterization of the variable ξ such that the latent variable z resembles a prior distribution. Thus, the variational inference process can be transformed by reparameterization as follows:

$$L_{KL} = D\left(N\left(\mu, \sigma^2\right) \| \mathcal{N}(0, 1)\right) = 0.5\left(1 + \log \sigma^2 - \mu^2 - \exp\left(\log \sigma^2\right)\right) \tag{3}$$

Therefore, the VBIGAN framework can learn the background distribution characteristics by imposing variational inference loss better, as well as help the G to generate more stable and homogeneous background samples.

(2) Spectral Reconstruction Loss: Firstly, the spectral vectors are reconstructed spectra generated by the E and the G, so the training goal of the VBIGAN model is to make the generated spectra more similar to the original spectra. Therefore, the mean squared error (MSE) loss function is used to calculate the difference between the generated spectra and the original spectra. The MSE is calculated as follows:

$$L_{MSE} = \|x - G(z)\|_2 + \|x - G(\xi)\|_2 \tag{4}$$

where x is the real spectra and the generated spectra Gz and $G\xi$.

Because of the detailed feature consistency of the original and generated spectra, we introduce an additional constraint of spectral angle mapper (SAM) [25] to ensure the accurate reconstruction of the generated spectra. The SAM is used to estimate the similarity

of two spectra and control the direction error effectively. The spectral constraint is defined as:

$$L_{SAM} = 2 - \frac{1}{\pi}\left[\arccos\left(\frac{xG(z)}{\|x\|_2 \|G(z)\|_2}\right) + \arccos\left(\frac{xG(\xi)}{\|x\|_2 \|G(\xi)\|_2}\right)\right] \tag{5}$$

The SAM constraints are utilized to ensure better results and faster convergence of the generated reconstructed spectra. Therefore, the combined use of MSE and SAM constitutes the total spectral loss, which makes the spectra consistent in both global and local.

(3) Adversarial Loss: To improve the shortcomings of the GAN framework, the Wasserstein GAN with gradient penalty (WGAN-GP) preserves the advantages of the WGAN network in terms of improved gradient updates while enabling a significant increase in training speed [38].Thus, we introduce the WGAN-GP for both discriminators to solve the gradient disappearance problem during learning. The objective function for the traditional WGAN-GP is defined as follows:

$$L_D = \mathbb{E}_{z \sim p_z(z)}[D(G(z))] - \mathbb{E}_{x \sim p(x)}[D(x)] + \lambda \mathbb{E}_{z \sim p_{\text{penalty}}(z)}$$
$$\left[\left(\left\|\nabla_z D(G(z))\right\|_2 - 1\right)^2\right] \tag{6}$$

where $Z \sim p_{\text{penalty}}(Z)$ is the penalty sample distribution and λ is the penalty coefficient.

The VBIGAN model contains the Ds and Dl. Firstly, the E is adversarial learning with the Dl to match the prior distribution $\xi = \mathcal{N}(0, 1)$ in the latent distribution $z \sim p(z)$, which facilitates the background distribution learning. For the Dl, the input contains two distribution samples from the sample $\xi = \mathcal{N}(0, 1)$ and $z \sim p(z)$ encoded by the E. Therefore, the objective function of the adversarial loss in the Dl is transformed as follows:

$$L_{Dl} = \mathbb{E}_{z \sim p(z)}[Dl(z)] - \mathbb{E}_{\xi \sim \mathcal{N}(0,1)}[Dl(z)] + \lambda \mathbb{E}_{z \sim p_{\text{penalty}}(z)}$$

$$\left[\left\| \nabla_z Dl(z) \right\|_2 - 1 \right] \tag{7}$$

$$L_{Gl} = -\mathbb{E}_{z \sim p(z)}[Dl(z)] \tag{8}$$

By minimizing E and maximizing Dl on L_{Dl}, the E can learn to represent the hyperspectral background distribution accurately, while generate homogeneous background information.

We designed the Ds, which can help the G to reconstruct more accurate background spectra and increase the detail features. For the Ds, the input contains not only the real samples but also two generated samples $x_{G\xi}$ and x_{Gz}. Therefore, the objective function of the adversarial loss in the Ds is transformed as follows:

$$
\begin{aligned}
L_{Ds} = \quad & \mathbb{E}_{z \sim p(z)}[Ds(G(z))] + \mathbb{E}_{\xi \sim \mathcal{N}(0,1)}[Ds(G(\xi))] - \mathbb{E}_{x \sim p(x)}[Ds(x)] + \\
& \lambda \mathbb{E}_{z \sim p_{\text{penalty}}(z)}\left[\left\| \nabla_z Ds(G(z)) \right\|_2 - 1 \right] + \\
& \lambda \mathbb{E}_{\xi \sim \mathcal{N}(0,1)}\left[\left\| \nabla_z Ds(G(\xi)) \right\|_2 - 1 \right]
\end{aligned}
\tag{9}
$$

$$L_{Gs} = -\mathbb{E}_{z \sim p(z)}[Ds(G(z))] - \mathbb{E}_{\xi \sim \mathcal{N}(0,1)}[Ds(G(\xi))] \tag{10}$$

where x_{Gz}, $x_{G\xi}$ represents the generated samples of the latent variables z, ξ. By minimizing G and maximizing Ds on L_{Ds}, the G can generate more realistic spectra. Therefore, the Ds discriminates the real data x and the two fake data x_{Gz} and $x_{G\xi}$. The G should use both x_{Gz} and $x_{G\xi}$ to cheat the Ds.

(4) Feature Mapping Loss: To obtain more stable generated data, we utilize Euclidean distance measures between discriminative features in the middle layer of the two discriminators, which is called feature matching losses [36]. The VBIGAN model is expected to be tightly coupled, which can reduce the risk of mode collapse. The data sample x and latent variables z must have the same discrimination score in both the sample and latent discriminators: $Dl(z) = Ds(x)$. Likewise, the latent variables z, ξ, and generated samples x_{Gz}, $x_{G\xi}$ must have the same discrimination score in both the two discriminators: $Dl(z) = Ds(G(z)) = Ds(G(\xi))$. However, the sample GAN and the latent GAN are somewhat independent, so the individual discriminant scores do not provide sufficient information for the proper coupling of the two GANs. Thus, we define the feature matching losses as follows:

$$L_{fm\_E} = \mathbb{E}_{x \sim p(x)} \left\| Ds^h(x) - Dl^h(z) \right\|_2^2 \tag{11}$$

$$L_{fm\_G} = \mathbb{E}_{z \sim p(z)} \left\| Ds^h(G(z)) - Dl^h(z) \right\|_2^2 + \mathbb{E}_{\xi \sim \mathcal{N}(0,1)} \left\| Ds^h(G(\xi)) - Dl^h(z) \right\|_2^2 \tag{12}$$

where Ds^h is the last hidden layer of the Ds, and Dl^h is that of the Dl. L_{fm_E} is the Euclidean distance between the extracted features of data after the Ds and the features of latent variables z via the Dl. L_{fm_E} is an expected Euclidean distance between the discriminative features of a data sample x and latent variables z. L_{fm_G} is that of the latent variables z, ξ and generated samples x_{Gz}, $x_{G\xi}$. As shown in Fig. 3(c) and (d), we compute the feature matching loss using the third layer output feature of the Ds, Dl, where the output features Ds^h and Dl^h have the same feature dimension. The E uses L_{fm_E} minimization to learn the inverse mapping of the G, prompting the learning of accurate spectral features. For the latent samples, the Dl uses L_{fm_G} minimization to learn feature matches with the Ds, which prompts more accurate learning of background distribution. Thus, we construct feature matching losses in the middle layer of the two discriminators to tightly couple the two GAN subnetworks, to obtain more stable generated data.

Up to this point, the composite loss function of the VBIGAN model can be shown as follows:

$$L_{\theta_G} = L_{Gf} + \lambda_1 L_{MSE} + \lambda_2 L_{SAM} \tag{13}$$

$$L_{\theta_E} = L_{Gb} + L_{KL} + \lambda_1 L_{MSE} + \lambda_2 L_{SAM} + \lambda_3 L_{fm\_E} \tag{14}$$

$$L_{\theta_{Ds}} = L_{Ds} \tag{15}$$

$$L_{\theta_{Dl}} = L_{Dl} + \lambda_4 L_{fm\_G} \tag{16}$$

where each part is given the explicit expression above. λ_1, λ_2, λ_3, and λ_4 are set to 1, 0.8, 0.5, and 0.5, respectively, based on experience and the results of several experiments. L_{θ_G} represents the total loss of the G, L_{θ_E} represents the total loss of the encoder E and to learn the background distribution, and $L_{\theta_{Ds}}$ and $L_{\theta_{Dl}}$ are related to the total loss of the sample discriminator Ds and the latent discriminator Dl.

3.3. Reconstruction error of the VBIGAN framework

The VBIGAN framework undergoes an iterative training process, so we can use the trained model to detect abnormal targets. Firstly, the reconstructed background image $\mathbf{X}^R \in \mathbb{R}^{B \times mn}$ can be obtained through the E and G trained network, with the input of all the image pixels using (17):

$$\mathbf{X}^R = G(E(\mathbf{X})) \tag{17}$$

By the aforementioned reconstruction process of hyperspectral images, the VBIGAN model can reconstruct the background spectra accurately, while it is difficult to reconstruct the abnormal spectra. In addition, Fig. 4 shows the spectral reconstruction results on the Salinas and Gulfport datasets to visually represent the spectral generation capability of the VBIGAN network. Eight representative abnormal pixels and eight representative background pixels were selected from each of the two datasets. For the background samples, the VBIGAN network model could accurately reconstruct the original spectra, and its generated spectra remained highly similar to the original spectral profiles. For the abnormal samples, the generated spectra of the VBIGAN network maintained the discrepancy with the original spectra. Therefore, the proposed network is able to learn the background distribution properties more effectively, while making the background and anomaly targets become more separated.

Overall, we can use the l_2 norm between the original spectra and the generated spectra to detect the targets, i.e., the reconstruction error. The detection results are calculated as follows:

$$d = \left\| x_{i,j} - x_{i,j}^R \right\|_2 \tag{18}$$

where d is detection result, $x_{i,j}^R$ is the spectral at location (i, j) of $\mathbf{X}^R$. Therefore, the VBIGAN model uses reconstruction errors to construct an auxiliary target detector.

Finally, to illustrate the VBIGAN network training and detection process, Table 1 shows the pseudo-code description of the proposed VBIGAN-AD algorithm.

4. Experiments

4.1. Datasets

Five kinds of HSI datasets were used to verify the superiority of the proposed VBIGAN-AD algorithm, one of which was a synthetic hyperspectral dataset and the other were real hyperspectral datasets. The sensors, image sizes, anomaly pixels and other parameters of the four data sets are shown in Table 2. Salinas was acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The main background classes are plants, soils,
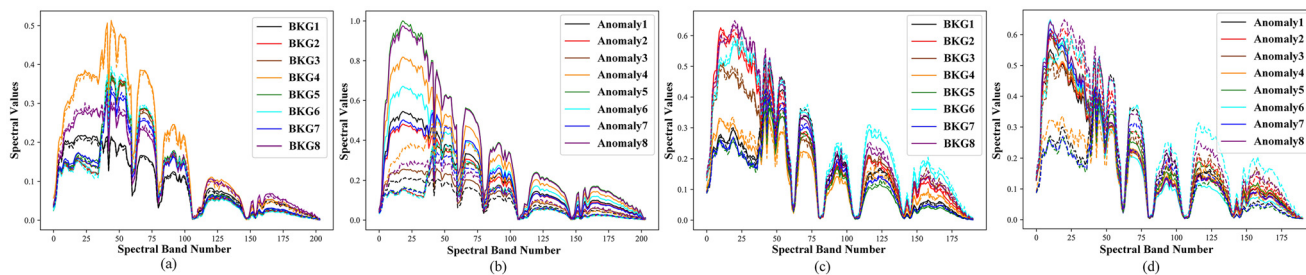
Fig. 4. Reconstructed results for different pixels in the Salinas and Gulfport datasets, where the solid lines are the original spectra and the dashed lines are the generated spectra. (a) and (b) are spectral curves of background and abnormal samples in the Salinas dataset, respectively. (c) and (d) are spectral curves of the background and abnormal samples in the Gulfport dataset, respectively.

Table 1
Pseudocode for the proposed algorithm.

Algorithm VBIGAN-AD
Input: HSI $\mathbf{X} = R^{(mn,b)}$
Initialize: Network: encoder E, generator G, sample discriminator Ds, latent discriminator Dl
Training of the VBIGAN Network:
input x to E and encode to obtain z;
input z, ξ to G and generate x_{Gz}, $x_{G\xi}$;
calculate the L_{MSE}, L_{SAM} using (4)–(5);
input x, x_{Gz}, $x_{G\xi}$ to Ds to obtain the discriminative probability, also obtain the feature vector;
calculate the L_{Ds}, L_{Gs} using (9)–(10), update Ds using (15);
input z, ξ to Dl to obtain the discriminative probability, also obtain the feature vector;
calculate the L_{Dl}, L_{Gl} using (7)–(8), calculate the feature matching using (11)–(12);
update Dl using (16);
update G with the gradient using (13);
update E with the gradient using (14).
End
the anomaly detection result using (17)–(18)
Output: anomaly detection map

Table 2
Parameters Related on the Five Datasets.

Dataset	Sensor	Spatial resolution	Image Size	Bands	Wavelengths	Anomaly pixels	Anomaly ratio
Salinas	AVIRIS	3.7m	120×120	204	400–2500 nm	25	0.17%
Pavia	ROSIS	1.3m	100×100	102	430–860 nm	43	0.29%
Gulfport	AVIRIS	3.4m	100×100	191	400–2500 nm	60	0.60%
Los Angeles	AVIRIS	7.1m	100×100	205	400–2500 nm	87	0.87%
Urban	HYDICE	1m	80×100	162	370–2510 nm	21	0.26%

etc., while the abnormal targets are buildings. We randomly selected 25 target locations. Synthetic abnormal spectra z is generated by $z = f \times t + (1 - f) \times b$, where f is the abundance fraction, t and b are the anomaly and background spectra, respectively. The synthetic data and ground-truth map are shown in Fig. 5(a). The Pavia dataset consists of images from the Reflection Optical System Imaging Spectrometer (ROSIS). The main background categories are bridges, water and bare soil, and the anomalies are vehicles. For the Gulfport dataset, the main background categories are airfield and plants, and the targets are three different scales aircraft [11]. The main background categories on the Los Angeles dataset are buildings, runways, vegetation, water, and bare soil, while the anomaly is two aircraft [11]. For the Urban dataset, the main background categories are asphalt, plant and soil, and the abnormal targets are vehicles.

4.2. Experimental setup

(1) Comparison Methods: Eleven popular methods of anomaly detection have been utilized as contrast methods, including GRX [8], LSMAD [13], CRD [17], OSP-AD [16], SDA-SF [21], RGAE [22], Auto-AD [23], DeepLR [24], BiGAN [39], FBGAN [36], and SC-AAE [25]. The GRX, LSMAD, CRD and OSP-AD algorithms use different

background modeling algorithms. The SDA-SF, RGAE, Auto-AD and DeepLR are deep learning algorithms. The BiGAN, FBGAN and SC-AAE are GAN-based anomaly detection algorithms.

(2) Parameter Settings: All algorithms for traditional detection algorithms were computed on Python 3.8.0, where the deep learning algorithms are carried out on a GeForce RTX 3070, with Torch-GPU 1.9.0 and CUDA 11.3. We used an Adam optimizer to optimize the VBIGAN network. Also, we defined the initial learning rate as 0.0001. The batch size was the total number of image pixels and the number of epochs was 1000. The penalty coefficient in (7) and (9) was chosen as $\alpha = 0.5$ empirically.

For the baseline methods, optimal parameters were required for each algorithm. For the LSMAD method, the rank r was set to 28, and the cardinality k was set to 0.002. For the CRD algorithm, the regularization parameter was set to 10^{-6} and the inner and outer double window sizes were set as shown in Table 3. In the OSP-AD algorithm, the two key parameters p and m values are shown in Table 3. The network structure parameters of the Auto-AD algorithm were set as recommended in [23]. The hyperparameters of BiGAN, FBGAN and SC-AAE are consistent with the VBIGAN network.

(3) Evaluation Metrics: 3-D receiver operating characteristic(ROC) curves [40], the area under the curve (AUC)scores [40],
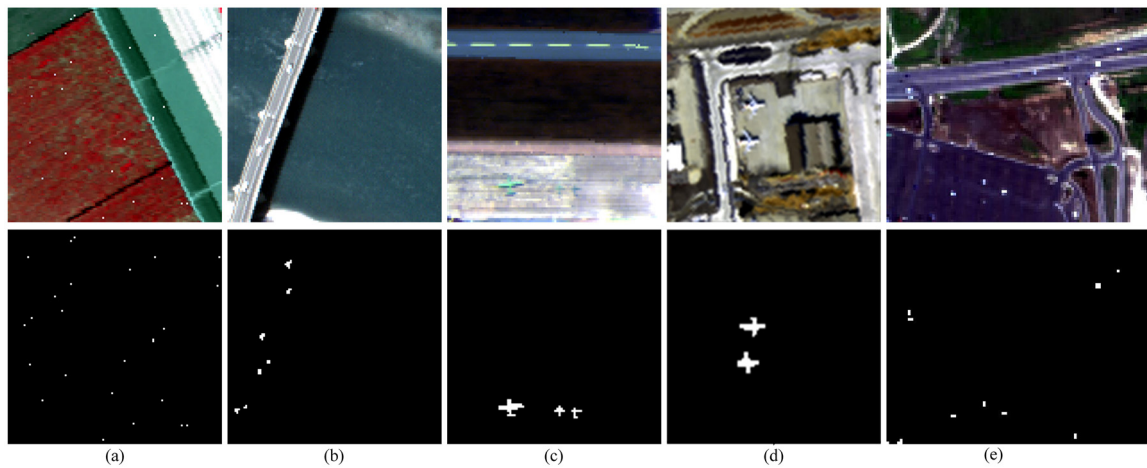
Fig. 5. Five hyperspectral datasets, the first row are the pseudo-color maps and the second row are the corresponding ground-truth maps of each datasets. (a) Salinas. (b) Pavia. (c) Gulfport. (d) Los Angeles. (e) Urban.

Table 3
Parameters Related on the Five Datasets.

| Dataset	CRD	OSP-AD	RGAE	SDA-SF	DeepLR
Salinas	$w_{out} = 9, w_{in} = 7$	$p = 8, m = 6$	$\lambda = 10^{-2}$, S = 50	$pca = 20$ $\gamma = 10$ $lr = 3$	$\lambda = 1.0$ $\mu = 0.5$
Pavia	$w_{out} = 11, w_{in} = 3$	$p = 5, m = 4$	$\lambda = 10^{-3}$, S = 100		
Gulfport	$w_{out} = 13, w_{in} = 3$	$p = 14, m = 10$	$\lambda = 10^{-3}$, S = 150		
Los Angeles	$w_{out} = 15, w_{in} = 13$	$p = 8, m = 5$	$\lambda = 10^{-2}$, S = 300		
Urban	$w_{out} = 7, w_{in} = 5$	$p = 9, m = 8$	$\lambda = 10^{-2}$, S = 150		
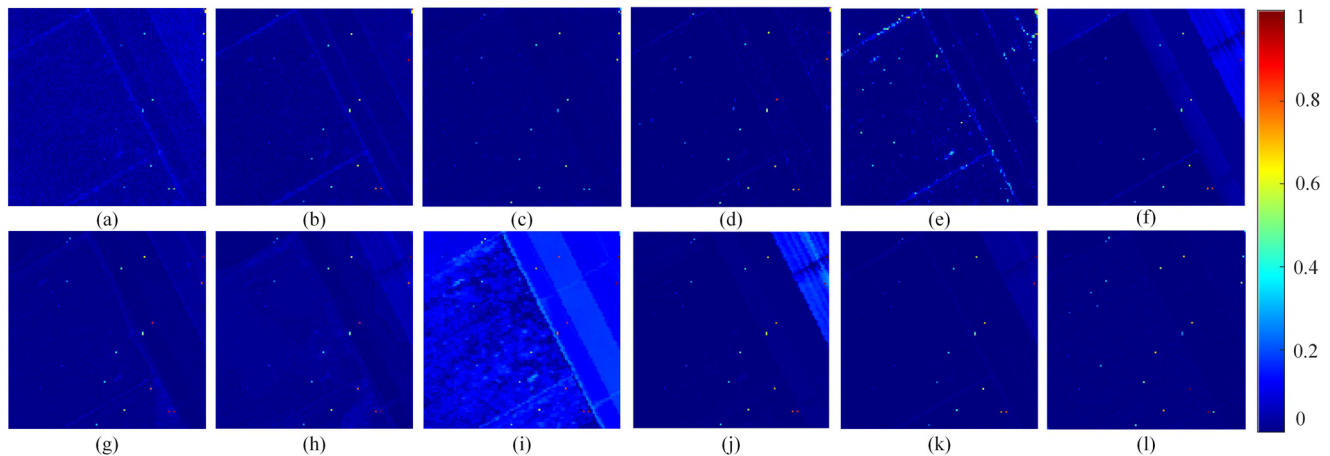
Fig. 6. The anomaly detection maps on the Salinas dataset. (a) GRX. (b) LSMAD. (c) CRD. (d) OSP-AD. (e) SDA-FC. (f) RGAE. (g) Auto-AD. (h) DeepLR. (i) BiGAN. (j) FBGAN. (k) SC-AAE. (l) VBIGAN-AD.

and box-whisker plot [13] are employed to evaluate the performance of the algorithm. The 3-D ROC curve is plotted by the probability of detection P_D, the false alarm rate P_F, and threshold τ. The 2D-ROC curves (P_D, P_F), (P_D, τ) and (P_F, τ) are utilized to evaluate the anomaly detection algorithm. Meanwhile, $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, $AUC_{(F,\tau)}$, the target detectability AUC_{TD}, the background suppressibility AUC_{BS}, the overall detection probability AUC_{ODP} are utilized to quantitatively evaluate. A higher value of $AUC_{(D,F)}$, $AUC_{(D,\tau)}$ and AUC_{TD} means a better detection performance. A smaller value of $AUC_{(F,\tau)}$ or a higher value of AUC_{BS} indicates a better ability to suppress the background. A higher value of AUC_{ODP} indicate more robustness and stability of the algorithm. Therefore, the $AUC_{(D,F)}$, $AUC_{(F,\tau)}$ and AUC_{ODP} are the preferred evaluation criterion, followed by the other AUC values. Box-whisker plot is primarily used to evaluate the separation of the anomaly and background.

4.3. Detection performance

We quantitatively analyzed and evaluated the afore-mentioned anomaly detection algorithms. As shown in Figs. 6–10, the two-dimensional detection maps of the afore-mentioned algorithms in five datasets are visualized. The detection maps on the Salinas dataset are shown in Fig. 6. The GRX, LSMAD and RGAE algorithms had difficulty detecting anomalies with low fraction of abundance. Although SDA-FC and BiGAN were able to detect abnormal pixels with different proportions of abundance fractions, they contained more background false positives. The CRD, OSP-AD, Auto-AD, DeepLR and SC-AAE algorithms showed better background suppression performance while failed to detect the anomalies. The proposed VBIGAN-AD algorithm could effectively suppress the background and accurately reflected abnormal pixels with different proportions of abundance fractions.
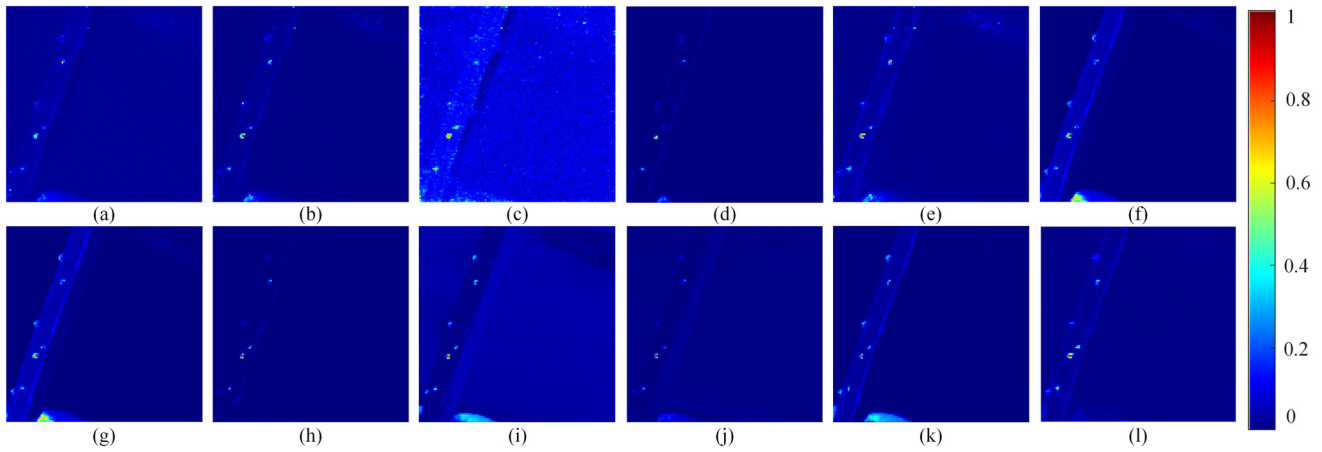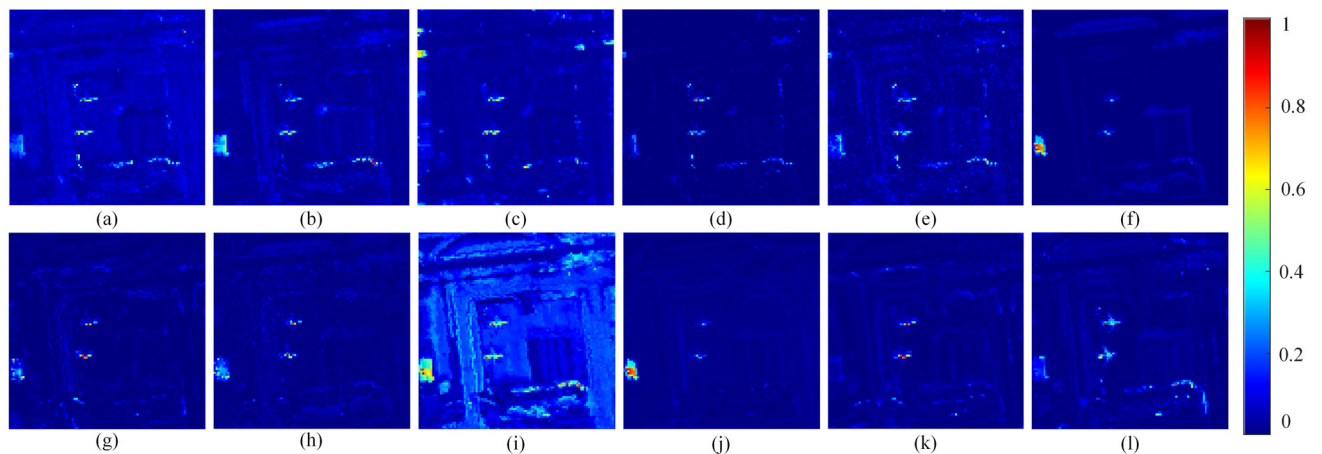
Fig. 7. The anomaly detection maps on the Pavia dataset. (a) GRX. (b) LSMAD. (c) CRD. (d) OSP-AD. (e) SDA-FC. (f) RGAE. (g) Auto-AD. (h) DeepLR. (i) BiGAN. (j) FBGAN. (k) SC-AAE. (l) VBIGAN-AD.
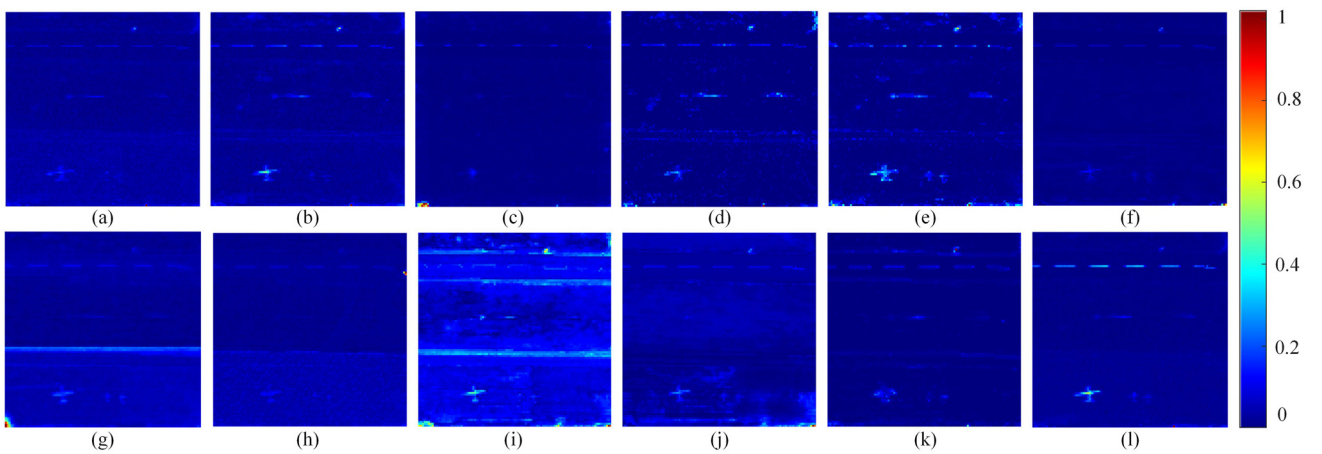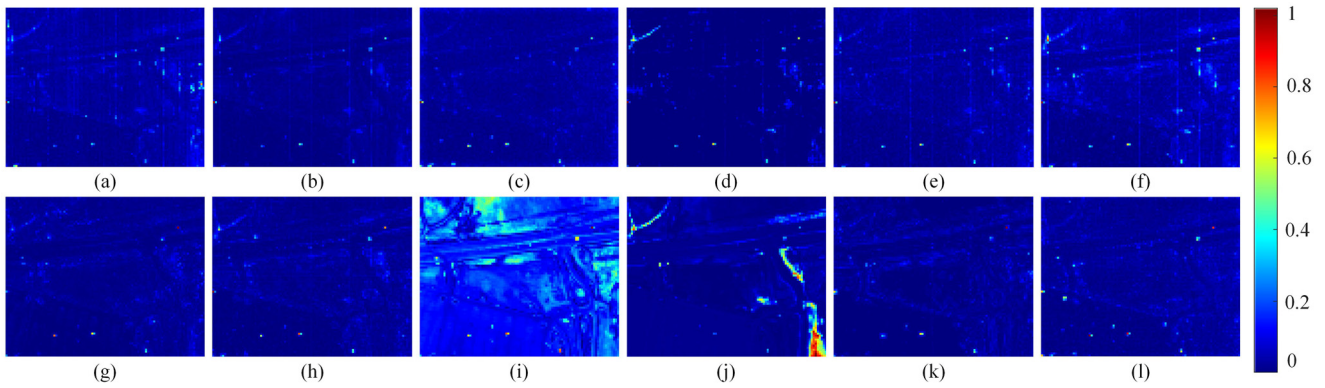

Fig. 8. The anomaly detection maps on the Gulfport dataset. (a) GRX. (b) LSMAD. (c) CRD. (d) OSP-AD. (e) SDA-FC. (f) RGAE. (g) Auto-AD. (h) DeepLR. (i) BiGAN. (j) FBGAN. (k) SC-AAE. (l) VBIGAN-AD.

Fig. 9. The anomaly detection maps on the Los Angeles dataset. (a) GRX. (b) LSMAD. (c) CRD. (d) OSP-AD. (e) SDA-FC. (f) RGAE. (g) Auto-AD. (h) DeepLR. (i) BiGAN. (j) FBGAN. (k) SC-AAE. (l) VBIGAN-AD.

Fig. 7 showed the detection results on the Pavia dataset which proved that the CRD, SDA-FC, RGAE, BiGAN and SC-AAE were capable to accurately identify the vehicle targets. The targets were mixed with the complex background and noise. There were missed targets in the results of GRX, LSMAD, OSP-AD and FBGAN. The visualization results demonstrated that the VBIGAN-AD had better

suppression of the background. The detection maps on the Gulfport and Los Angeles datasets are shown in Figs. 8 to 9. For the Gulfport dataset, GRX, CRD, RGAE algorithm showed that the background was constrained at low levels, but the three aircrafts could not be detected distinctly. Although the LSMAD and VBIGAN-AD could detect three aircrafts on different scales, the detection abil-

Fig. 10. The anomaly detection maps on the Urban dataset. (a) GRX. (b) LSMAD. (c) CRD. (d) OSP-AD. (e) SDA-FC. (f) RGAE. (g) Auto-AD. (h) DeepLR. (i) BiGAN. (j) FBGAN. (k) SC-AAE. (l) VBIGAN-AD.
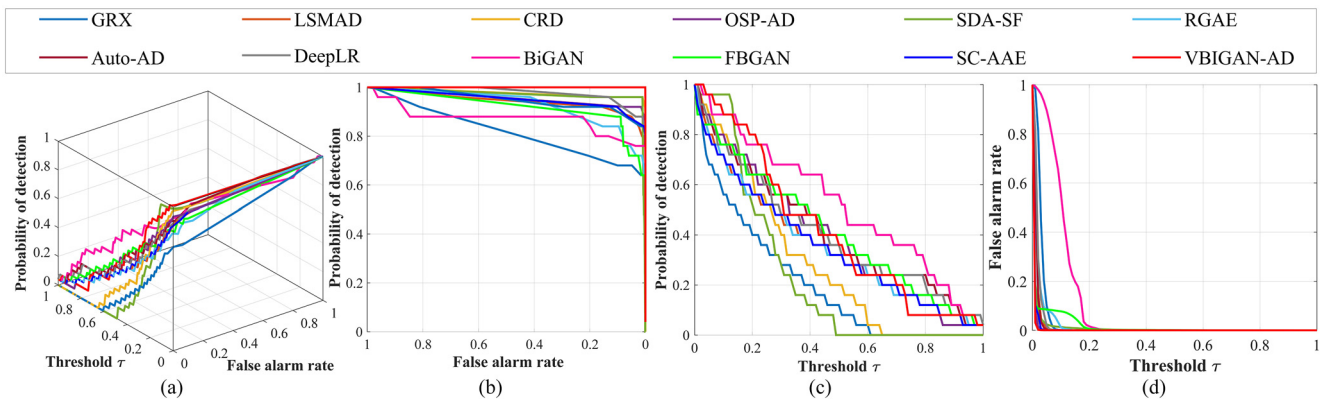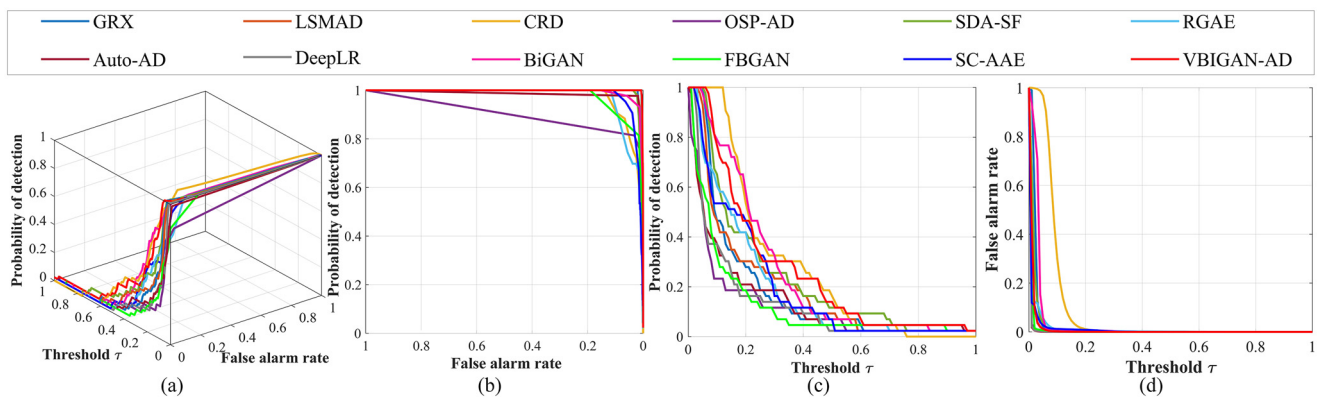

Fig. 11. 3-D ROC curves and 2-D ROC curves of the Salinas dataset. (a) 3-D ROC curves. (b) 2-D ROC curves (P_D, P_F). (c) 2-D ROC curves (P_D, τ). (d) 2-D ROC curves (P_F, τ).

Fig. 12. 3-D ROC curves and 2-D ROC curves of the Pavia dataset. (a) 3-D ROC curves. (b) 2-D ROC curves $((P_D, P_F)$. (c) 2-D ROC curves (P_D, τ). (d) 2-D ROC curves (P_F, τ).

ity for small-scale aircrafts was slightly weaker. The VBIGAN-AD detected the edges of the aircraft and maintained target integrity on the Los Angeles dataset. The OSP-AD, RGAE and FBGAN could better suppress the background, while the building roofs had been wrongly identified as anomalies, and the target pixels could not be detected effectively. The BiGAN had difficulty suppressing the background in two aircraft datasets. For the detection results on the Urban dataset, the GRX, SDA-FC, RGAE and BiGAN had false alarm targets, and VBIGAN-AD algorithm could effectively suppress the background and accurately detect anomalies.

The ROC curves on the five datasets are displayed in Figs. 11–15. The ROC curves of (P_D, P_F) and (P_D, τ) are near the upper right corner, which indicates a higher detection performance;the ROC curves of (P_F, τ) is near the lower left corner,

which indicates a better ability to suppress the background. From the ROC curves of (P_D, P_F) on the five datasets, the VBIGAN-AD algorithm significantly outperformed the other algorithms. In addition, the VBIGAN-AD achieved a detection probability of 80% or more with a much lower false alarm rate. For the Gulfport and Los Angeles datasets, although the ROC curves of (P_D, P_F) for VBIGAN-AD and the other algorithms cross over, the ROC curve for VBIGAN-AD remained in the upper right corner overall. The five deep learning-based detection algorithms showed similar performance in terms of the ROC curves of (P_D, τ), which were essentially on top of the traditional algorithms. From the representation of the ROC of (P_F, τ) on the five datasets, the VBIGAN-AD was located at the optimal position in the lower left corner, which indicated the algorithm was able to keep the background at a rela-
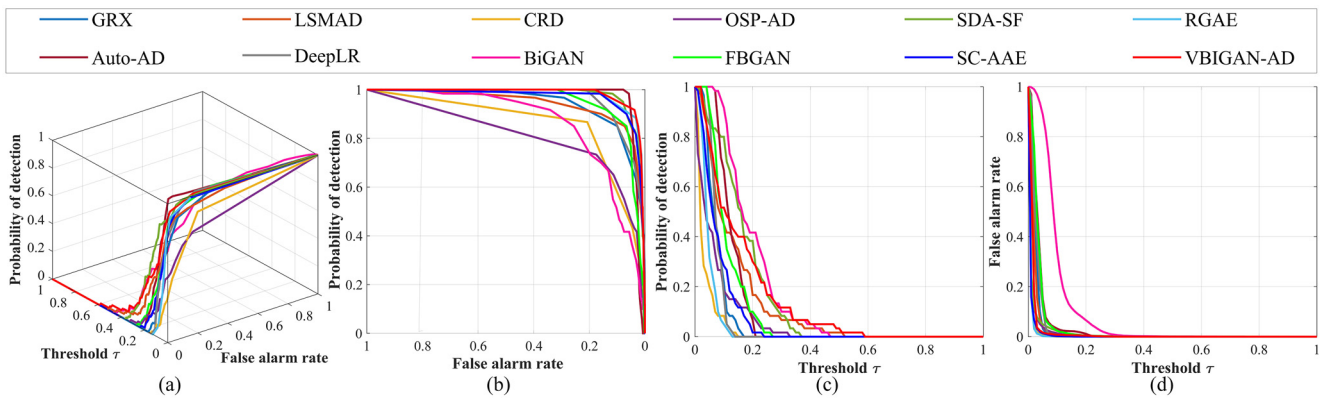
Fig. 13. 3-D ROC curves and 2-D ROC curves of the Gulfport dataset. (a) 3-D ROC curves. (b) 2-D ROC curves (P_D, P_F). (c) 2-D ROC curves (P_D, τ). (d) 2-D ROC curves (P_F, τ).
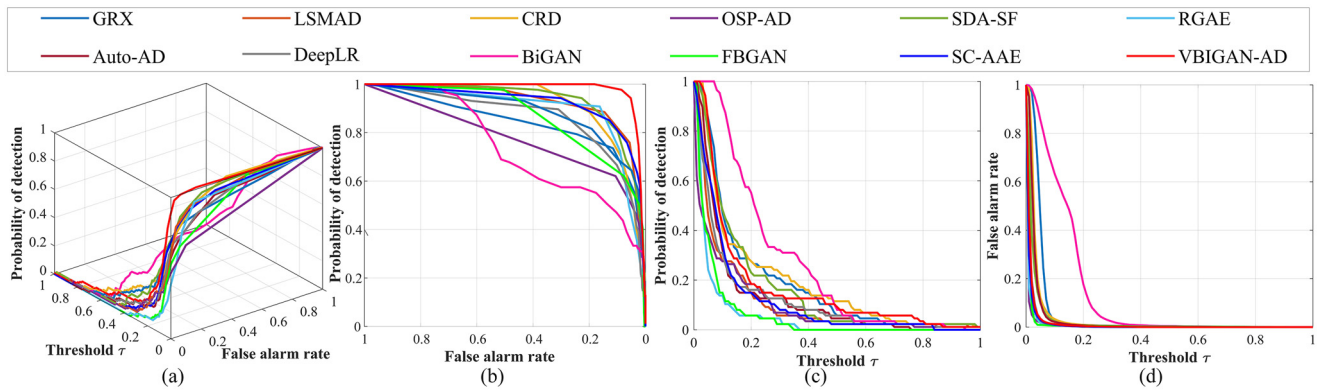

Fig. 14. 3-D ROC curves and 2-D ROC curves of the Los Angeles dataset. (a) 3-D ROC curves. (b) 2-D ROC curves (P_D, P_F). (c) 2-D ROC curves (P_D, τ). (d) 2-D ROC curves (P_F, τ).
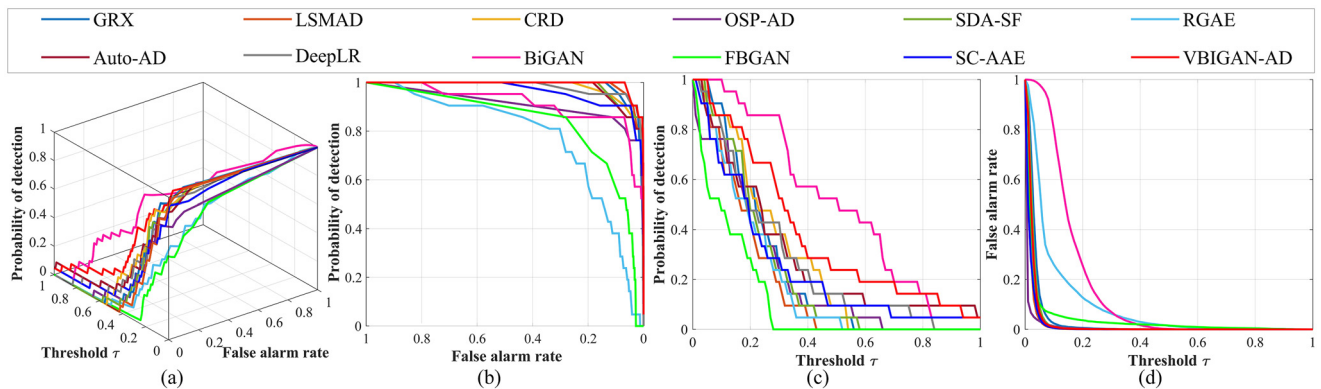

Fig. 15. 3-D ROC curves and 2-D ROC curves of the Urban dataset. (a) 3-D ROC curves. (b) 2-D ROC curves (P_D, P_F). (c) 2-D ROC curves (P_D, τ). (d) 2-D ROC curves (P_F, τ).

tively low level. For the RGAE, Auto-AD, DeepLR and SC-AAE, significant differences existed on the five datasets. Although the ROC curves of (P_D, τ) for BiGAN obtained the best performance, the ROC curves of (P_F, τ) verified that BiGAN had difficulty in suppressing background. All in all, the proposed VBIGAN-AD is equally good at suppressing the background while detecting the targets, further demonstrating the excellent robustness of the proposed algorithm.

As shown in Fig. 16, the red and cyan boxes of the box-whisker plots indicate anomalies and the background respectively, and the gap between the two boxes indicates the degree of separation from targets and background. For the Salinas dataset, the GRX, LSMAD and BiGAN had wider background ranges, and the background of the GRX algorithm overlapped with the anomalies. The VBIGAN-AD showed a better separation between the background and the anomalies. The box-whisker plots from the Pavia dataset shown

that the algorithms could suppress the background well besides CRD, and the VBIGAN-AD exhibits better detection performance. The cyan boxes were crossed with red boxes for the GRX, OSP-AD and FBGAN on the Gulfport and Los Angeles datasets, which indicated that these methods had difficulty in distinguishing the background and anomalies. The LSMAD, SDA-SF, RGAE, and VBIGAN-AD achieved separation. From the results of the Urban dataset, the background of the RGAE overlapped with the outliers, and the others performed well.

Furthermore, the AUC scores and time consumption for each detection algorithm of the five datasets are shown in Tables 5–8. For the Salinas dataset, VBIGAN-AD obtained the highest $AUC_{(D,F)}$, and AUC_{TD} with 0.9996, 1.4337, respectively. The $AUC_{(D,\tau)}$ of BiGAN was the highest, but the $AUC_{(F,\tau)}$ and AUC_{BS} value of BiGAN were the poorest. Although the proposed algorithm obtained subopti-
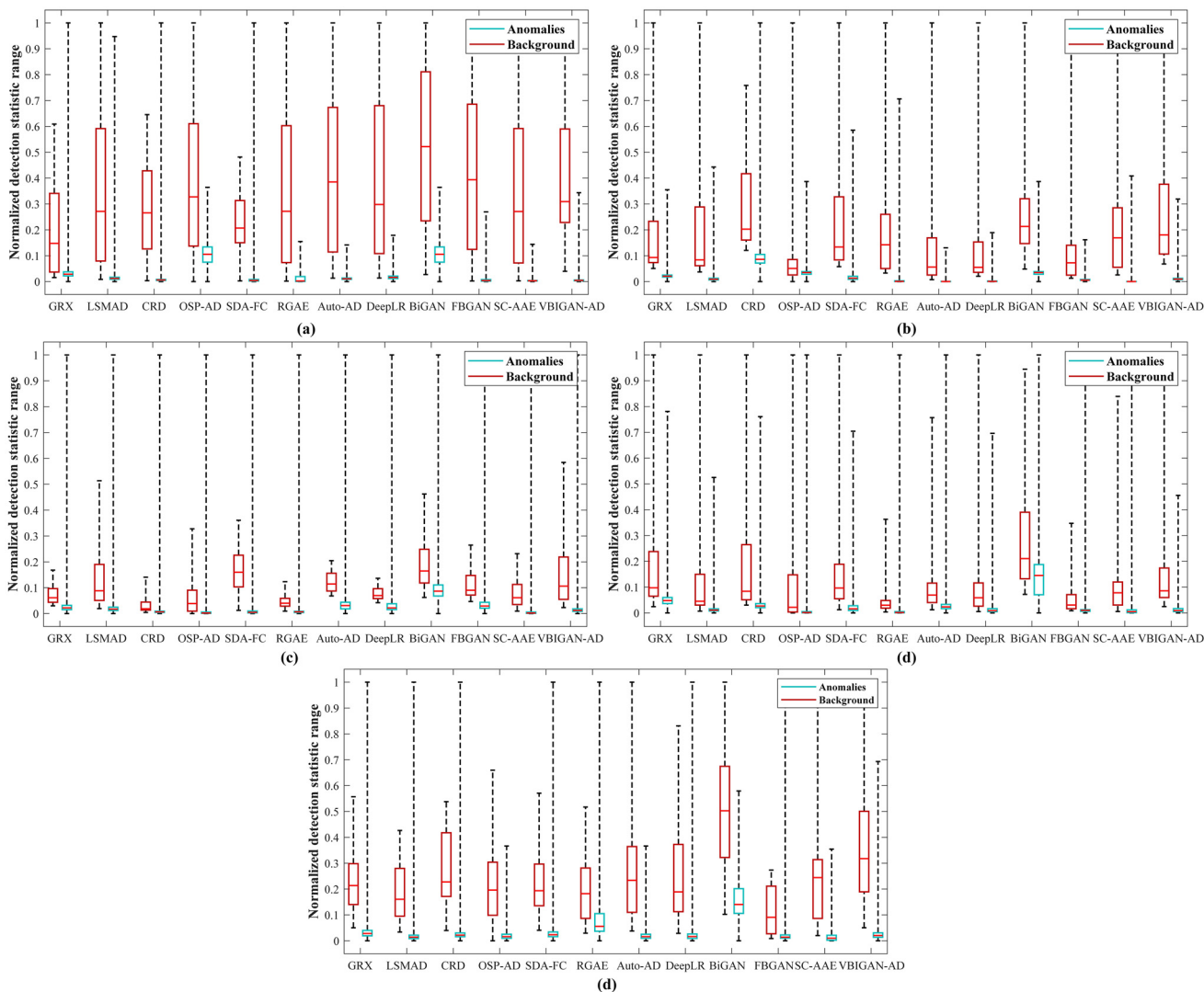
Fig. 16. Box-whisker plots for the different methods. (a) Salinas. (b) Pavia. (c) Gulfport. (d) Los Angeles. (e) Urban.

Table 4
AUC scores and time consumption of the comparison methods on the Salinas dataset.

Method	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	Inference time/s	FLOPs/G
GRX	0.8073	0.2143	0.0314	1.0216	0.7759	1.1829	0.81	-
LSMAD	0.9501	0.3692	0.0144	1.3193	0.9357	1.3547	10.27	-
CRD	0.9638	0.3012	0.0069	1.2647	0.9566	1.2942	3.48	-
OSP-AD	0.9560	0.4098	**0.0032**	1.3658	0.9528	1.4066	790.43	-
SDA-SF	0.9675	0.2408	0.0101	1.2083	0.9573	1.2307	1.49	1.71
RGAE	0.9290	0.3689	0.0157	1.2980	0.9134	1.3532	1.75	0.58
Auto-AD	<u>0.9836</u>	0.4307	0.0126	<u>1.4143</u>	<u>0.9709</u>	1.4181	5.61	8.59
DeepLR	0.9558	0.4212	0.0184	1.3770	0.9375	1.4028	2.84	3.92
BiGAN	0.8720	**0.5372**	0.1077	1.4092	0.7644	<u>1.4295</u>	1.41	3.17
FBGAN	0.9348	0.4291	0.0173	1.3639	0.9175	1.4118	1.44	3.30
SC-AAE	0.9656	0.3723	0.0050	1.3379	0.9606	1.3672	18.52	0.50
VBIGAN-AD	**0.9996**	<u>0.4338</u>	<u>0.0049</u>	**1.4337**	**0.9951**	**1.4299**	3.84	1.89

mal results for the $AUC_{(F,\tau)}$) score, the AUC_{BS} obtained the highest score of 0.9951, which proved the superiority of the background suppression. The proposed algorithm on the Pavia dataset achieved the highest values of $AUC_{(D,F)}$ and AUC_{TD}, which were 0.9990 and 1.2756, respectively, while AUC_{BS} got the suboptimal value. For the two airfield datasets, the VBIGAN-AD obtained the highest AUC score for $AUC_{(D,F)}$, $AUC_{(D,\tau)}$, AUC_{TD}, AUC_{ODP}. The optimal or suboptimal value of AUC_{BS} was obtained by the proposed algorithm respectively, which demonstrated the ability of the proposed algo-

rithm to suppress the background. Moreover, the proposed algorithm achieved suboptimal values for $AUC_{(D,\tau)}$ and AUC_{BS}, and the highest values for the $AUC_{(D,F)}$, AUC_{TD}, AUC_{ODP} scores on the Urban dataset. Experimental results demonstrated the outperformance of the proposed algorithm to handle hyperspectral anomaly detection with anomalies embedded background. In summary, the VBIGAN-AD shows a superior detection performance and can suppress the background well, achieving effective separation of anomalies and background.

Table 5
AUC scores and time consumption of the comparison methods on the Pavia dataset.

Method	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	Inference time/s	FLOPs/G
GRX	0.9944	0.1858	0.0232	1.1802	0.9712	1.1627	0.72	-
LSMAD	0.9946	0.1999	0.0118	1.1945	0.9828	1.1881	4.59	-
CRD	0.9756	**0.2948**	0.0918	<u>1.2704</u>	0.8838	1.2030	16.71	-
OSP-AD	0.9725	0.1276	<u>0.0016</u>	1.1001	0.9709	1.1260	687.91	-
SDA-SF	0.9934	0.2425	0.0179	1.2359	0.9755	1.2247	1.53	1.71
RGAE	0.9721	0.1982	0.0095	1.1703	0.9626	1.1886	0.99	0.29
Auto-AD	0.9972	0.1589	**0.0010**	1.1560	**0.9961**	1.1579	5.35	7.79
DeepLR	<u>0.9980</u>	0.1437	0.0018	1.1417	**0.9961**	1.1419	2.79	3.72
BiGAN	0.9929	0.2673	0.0339	1.2602	0.9590	<u>1.2334</u>	1.34	2.23
FBGAN	0.9826	0.1405	0.0079	**1.2756**	0.9746	1.1325	1.35	2.36
SC-AAE	0.9839	0.2051	0.0068	1.1890	0.9771	1.1983	19.97	0.21
VBIGAN-AD	**0.9990**	<u>0.2765</u>	0.0120	**1.2756**	<u>0.9870</u>	**1.2645**	3.83	1.23

Table 6
AUC scores and time consumption of the comparison methods on the Gulfport dataset.

Method	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	Inference time/s	FLOPs/G
GRX	0.9526	0.0736	0.0248	1.0262	0.9278	1.0489	0.60	-
LSMAD	0.9515	0.1302	0.0552	1.0818	0.9304	1.1090	6.15	-
CRD	0.9097	0.0325	0.0092	0.9423	0.9007	1.0235	17.82	-
OSP-AD	0.8429	0.0642	0.0099	0.9072	0.8332	1.0544	221.71	-
SDA-SF	0.9781	0.1554	0.0318	<u>1.1350</u>	0.9653	<u>1.1293</u>	1.10	1.19
RGAE	<u>0.9801</u>	0.0461	<u>0.0077</u>	1.0263	**0.9726**	1.0384	0.82	0.38
Auto-AD	0.9700	0.1243	0.0369	1.0944	0.9333	1.0875	4.82	5.90
DeepLR	0.9661	0.0762	0.0281	1.1424	0.9382	1.0482	2.64	2.60
BiGAN	0.8647	**0.1944**	0.0983	1.0592	0.7666	1.0962	1.05	2.12
FBGAN	0.9568	0.1106	0.0348	1.0674	0.9221	1.0758	1.07	2.21
SC-AAE	0.9756	0.0771	**0.0062**	1.0528	0.9696	1.0710	13.14	0.33
VBIGAN-AD	**0.9872**	<u>0.1589</u>	0.0157	**1.1462**	<u>0.9715</u>	**1.1433**	2.80	1.25

Table 7
AUC scores and time consumption of the comparison methods on the Los Angeles dataset.

Method	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	Inference time/s	FLOPs/G
GRX	0.8404	0.1841	0.0516	1.0245	0.7888	1.1325	0.84	-
LSMAD	0.9317	0.1116	0.0146	1.0433	0.9171	1.0970	6.61	-
CRD	0.9216	<u>0.1906</u>	0.0356	<u>1.1121</u>	0.8860	<u>1.1550</u>	63.94	-
OSP-AD	0.8105	0.0978	**0.0054**	0.9083	0.8051	1.0924	256.58	-
SDA-SF	0.9348	0.1641	0.0294	1.0989	0.9100	1.1394	1.08	1.19
RGAE	0.9309	0.0547	<u>0.0083</u>	0.9856	0.9227	1.0464	0.89	0.41
Auto-AD	0.8642	0.1237	0.0313	0.9879	0.8329	1.0924	4.97	5.98
DeepLR	0.8926	0.1235	0.0145	1.0161	0.8781	1.1089	2.67	2.64
BiGAN	0.7385	**0.2688**	0.1385	1.0053	0.6001	1.1283	1.03	2.21
FBGAN	0.9069	0.0585	0.0133	0.9655	0.8936	1.0452	1.06	2.30
SC-AAE	<u>0.9398</u>	0.1175	0.0106	1.0531	<u>0.9292</u>	1.1069	12.64	0.34
VBIGAN-AD	**0.9803**	0.1728	0.0151	**1.1531**	**0.9652**	**1.1577**	2.81	1.32

Table 8
AUC scores and time consumption of the comparison methods on the Urban dataset.

Method	$AUC_{(D,F)}$	$AUC_{(D,\tau)}$	$AUC_{(F,\tau)}$	AUC_{TD}	AUC_{BS}	AUC_{ODP}	Inference time/s	FLOPs/G
GRX	0.9848	0.2383	0.0344	1.2231	0.9505	1.2040	0.43	-
LSMAD	<u>0.9897</u>	0.2020	0.0230	1.1916	<u>0.9667</u>	1.1790	3.67	-
CRD	0.9864	0.2873	0.0260	1.2737	0.9604	1.2613	2.81	-
OSP-AD	0.9053	0.2240	**0.0074**	1.1292	0.8979	1.2166	150.78	-
SDA-SF	0.9837	0.2395	0.0281	1.2232	0.9556	1.2114	0.84	0.95
RGAE	0.7723	0.1954	0.0965	0.9677	0.6758	1.0989	0.21	0.26
Auto-AD	0.9843	0.3111	0.0207	1.2954	0.9635	1.2904	4.06	4.59
DeepLR	0.9837	0.2866	0.0214	1.2703	0.9623	1.2652	2.42	2.61
BiGAN	0.9121	**0.5082**	0.1621	**1.4303**	0.7523	<u>1.3537</u>	0.86	1.54
FBGAN	0.8305	0.1183	0.0369	0.7937	0.7937	1.0814	0.84	1.62
SC-AAE	0.9688	0.2625	0.0159	1.3555	0.9530	1.2466	9.96	0.24
VBIGAN-AD	**0.9937**	<u>0.3979</u>	<u>0.0158</u>	<u>1.3916</u>	**0.9764**	**1.3821**	2.25	0.90

Moreover, Tables 5 to 8 listed the inference time and the FLOPs (floating point operations) of the detection algorithms.It is cleared that the CRD, LSMAD and OSP-AD algorithms took longer inference time, making these algorithms much less applicable.The SDA-FC and RGAE based on AE had less computational complexity and fewer model parameters. Auto-AD obtained the highest FLOPs and SC-AAE had the longest inference time. The inference phase in the BiGAN, FBGAN, VBIGAN just included the encoding and generating which had less time consuming. Overall, the time consumption of VBIGAN-AD is in a reasonable range of computational costs and has an excellent detection performance, which has important application prospects.

Table 9

Comparison of $AUC_{(D,F)}$ scores on each part of the VBIGAN framework.

Subnetwork	Parts					Salinas	Pavia	Gulfport	Los Angeles	Urban
	E	G	Ds	Dl	VI					
AE	√	√	×	×	×	0.9642	0.9956	0.9643	0.9448	0.9901
VAE	√	√	×	×	√	0.9562	0.9940	0.8808	0.9055	0.9133
GAN_Ds	√	√	√	×	×	0.9859	0.9964	0.9553	0.9590	<u>0.9925</u>
GAN_Dl	√	√	×	√	×	0.9657	0.9950	0.9125	0.9620	0.9856
GAN_Ds_VI	√	√	√	×	√	<u>0.9882</u>	0.9968	<u>0.9692</u>	<u>0.9763</u>	0.9893
GAN_Ds_ Dl	√	√	√	√	×	0.9714	<u>0.9974</u>	0.9528	0.9706	0.9912
VBIGAN	√	√	√	√	√	**0.9989**	**0.9990**	**0.9872**	**0.9803**	**0.9937**

Table 10

Comparison of $AUC_{(F,\tau)}$ scores on each part of the VBIGAN framework.

Subnetwork	Parts					Salinas	Pavia	Gulfport	Los Angeles	Urban
	E	G	Ds	Dl	VI					
AE	√	√	×	×	×	0.0326	0.0347	0.0327	0.0789	0.0719
VAE	√	√	×	×	√	0.0685	0.0370	0.0591	0.0723	0.0855
GAN_Ds	√	√	√	×	×	<u>0.0171</u>	**0.0116**	0.0202	0.0669	**0.0136**
GAN_Dl	√	√	×	√	×	0.0227	0.0382	0.0289	0.0268	0.0395
GAN_Ds_VI	√	√	√	×	√	0.0175	0.0129	**0.0134**	<u>0.0155</u>	0.0241
GAN_Ds_ Dl	√	√	√	√	×	0.0198	0.0145	0.0213	0.0189	0.0162
VBIGAN	√	√	√	√	√	**0.0049**	<u>0.0120</u>	<u>0.0157</u>	**0.0151**	<u>0.0158</u>

Table 11

Comparison of AUC scores for VBIGAN networks with different losses.

Loss Term		Salinas		Pavia		Gulfport		Los Angeles		Urban	
SAM	FM	$AUC\,(D,F)$	$AUC\,(F,\tau)$	$AUC\,(D,F)$	$AUC\,(F,\tau)$	$AUC\,(D,F)$	$AUC\,(F,\tau)$	$AUC\,(D,F)$	$AUC\,(F,\tau)$	$AUC\,(D,F)$	$AUC\,(F,\tau)$
×	×	0.9897	0.0152	0.9941	0.0131	0.9765	0.0185	0.9781	0.0266	0.9851	0.0241
√	×	0.9944	0.0120	<u>0.9974</u>	0.0160	0.9811	**0.0136**	0.9793	**0.0138**	<u>0.9902</u>	<u>0.0185</u>
×	√	<u>0.9961</u>	<u>0.0085</u>	0.9968	<u>0.0129</u>	<u>0.9842</u>	0.0183	**0.9852**	0.0204	0.9894	0.0192
√	√	**0.9996**	**0.0049**	**0.9990**	**0.0120**	**0.9872**	<u>0.0157</u>	<u>0.9803</u>	<u>0.0151</u>	**0.9937**	**0.0158**

4.4. Ablation experiments

1) Influence of the different subnetworks of VBIGAN framework

To verify the background learning ability and anomaly detection performance of each part in the VBIGAN framework, seven different sets of network structures are utilized to analyze the different combinations of encoder E, generator G, sample discriminator Ds, latent discriminator Dl, and variational inference VI. The subnetworks of VBIGAN are named as AE, VAE, GAN_Ds, GAN_Dl, GAN_Ds_VI and GAN_Ds_ Dl, which are consisted of different modules and listed in Table 11. Moreover, the settings of epochs, optimizer and learning rate are fixed all networks. $AUC_{(D,F)}$ and $AUC_{(F,\tau)}$ are utilized to evaluate the detection performance of the model.

$AUC_{(D,F)}$ and $AUC_{(F,\tau)}$ of the various networks are reported in Tables 2–11. Firstly, the $AUC_{(D,F)}$ of the AE was obvious higher than that of the VAE. The detection accuracy of the model could be further enhanced in GAN_Ds_VI, and suboptimal values of $AUC_{(D,F)}$ were obtained for the Salinas, Gulfport, and Los Angeles datasets. The GAN_Ds_ Dl achieved the secondly highest detection accuracy on Pavia dataset. Combining the network framework of E, G, Ds, Dl and the variational inference, the VBIGAN obtained the highest $AUC_{(D,F)}$ values on the five datasets. From the $AUC_{(F,\tau)}$ scores obtained by the basic network model, VBIGAN, with a higher detection accuracy, showed its superiority on the abnormal discrimination and background suppression.

2) Influences of the different losses in VBIGAN network

To verify the effects of spectral angle mapping (SAM) and feature matching (FM) loss on the anomaly detection performance of the VBIGAN network, we define four forms of VBIGAN network models. The first one is the prototype VBIGAN network without spectral angle mapping and feature matching loss. The second and third models are VBIGAN networks with the introduction of spectral angular distance and feature matching loss, respectively. The fourth model is a prototype VBIGAN network that contains both spectral angle mapping and feature matching loss. We only used $AUC_{(D,F)}$ and $AUC_{(F,\tau)}$ to evaluate the detection performance and the suppression background performance of the model, respectively. The fourth model is a prototype VBIGAN network that contains both spectral angle mapping and feature matching loss.

As shown in Table 11, taking the prototype VBIGAN network as the benchmark, the $AUC_{(D,F)}$ and $AUC_{(F,\tau)}$ could be further improved by using the networks with spectral angular distance and feature matching losses, respectively. In addition, after adding spectral angle mapping loss, the detection accuracy was slightly improved by introducing feature matching loss. Further, the use of spectral angle mapping and feature matching common constrained helps the VBIGAN network model to achieve better detection performance. Overall, the VBIGAN model is the best structure that meets the anomaly detection expectations by adding a common constraint of spectral angle distance and feature matching.

5. Conclusion

In this article, a variational background inference based on GAN (VBIGAN) framework is proposed for hyperspectral anomaly detection. We develop a background modeling research in both model design and loss construction to address the problems of inaccurate background learning and spectral reconstruction. Firstly, to avoid the turbulence of abnormal samples while enhancing the stability of network learning, we propose the VBIGAN framework based on variational inference. The VBIGAN framework establishes two coupled GAN sub-networks, and separates the anomaly from the background using reconstruction error autonomously. Then, the composite loss function is designed to effectively ensure the stable learning of background distribution characteristics. Experi-

ments showed that the VBIGAN method achieves excellent detection performance on all hyperspectral datasets.The results of the VBIGAN various subnetworks from ablation experiments indicate that VBIGAN is of more robust to anomalies with a lower false alarm rate. From the contrast experiments on five datasets, the detection performance demonstrates the superiority of VBIGAN in detecting anomalies and suppressing background.

In the future, hyperspectral clustering and image segmentation methods can be investigated to extract pure background sample elements for network training, which can improve the performance of detecting backgrounds. In addition, a spatial convolution module will be introduced to extract the spatial-spectral features, and to improve the recognition of abnormal targets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

References

[1] H. Su, Z. Wu, H. Zhang, Q. Du, Hyperspectral anomaly detection: a survey, IEEE Geosci. Remote Sens. Mag. 10 (1) (2021) 64–90.

[2] X. Wang, K. Tan, Q. Du, Y. Chen, P. Du, CVA 2 E: a conditional variational autoencoder with an adversarial training process for hyperspectral imagery classification, IEEE Trans. Geosci. Remote Sens. 58 (8) (2020) 5676–5692.

[3] C. Xing, C. Duan, Z. Wang, M. Wang, Binary feature learning with local spectral context-aware attention for classification of hyperspectral images, Pattern Recognit. 134 (2023) 109123.

[4] A.-A. Dalal, M.A.A. Al-qaness, Z. Cai, E.A. Alawamy, IDA: improving distribution analysis for reducing data complexity and dimensionality in hyperspectral images, Pattern Recognit. 134 (2023) 109096.

[5] W. Xie, J. Lei, S. Fang, Y. Li, X. Jia, M. Li, Dual feature extraction network for hyperspectral image analysis, Pattern Recognit. 118 (2021) 107992.

[6] H. Qin, W. Xie, Y. Li, K. Jiang, J. Lei, Q. Du, Weakly supervised adversarial learning via latent space for hyperspectral target detection, Pattern Recognit. 135 (2023) 109125.

[7] I.S. Reed, X. Yu, Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, IEEE Trans. Acoust. 38 (10) (1990) 1760–1770.

[8] J.M. Molero, E.M. Garzon, I. Garcia, A. Plaza, Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 6 (2) (2013) 801–814.

[9] L. Zhang, J. Ma, B. Cheng, F. Lin, Fractional fourier transform-based tensor RX for hyperspectral anomaly detection, Remote Sens. (Basel) 14 (3) (2022) 797.

[10] J.A.P. Hidalgo, A. Pérez-Suay, F. Nar, G. Camps-Valls, Efficient nonlinear RX anomaly detectors, IEEE Geosci. Remote Sens. Lett. 18 (2) (2020) 231–235.

[11] X. Kang, X. Zhang, S. Li, K. Li, J. Li, J.A. Benediktsson, Hyperspectral anomaly detection with attribute and edge-preserving filters, IEEE Trans. Geosci. Remote Sens. 55 (10) (2017) 5600–5611.

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

[13] Y. Zhang, B. Du, L. Zhang, S. Wang, A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 54 (3) (2015) 1376–1389.

[14] X. Zhang, X. Ma, N. Huyan, J. Gu, X. Tang, L. Jiao, Spectral-difference low-rank representation learning for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 59 (12) (2021) 10364–10377.

[15] M. Wang, Q. Wang, D. Hong, S.K. Roy, J. Chanussot, Learning tensor low-rank representation for hyperspectral anomaly detection, IEEE Trans. Cybern. (2022).

[16] C.-I. Chang, H. Cao, M. Song, Orthogonal subspace projection target detector for hyperspectral anomaly detection, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14 (2021) 4915–4932.

[17] W. Li, Q. Du, Collaborative representation for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 53 (3) (2014) 1463–1474.

[18] K. Tan, Z. Hou, F. Wu, Q. Du, Y. Chen, Anomaly detection for hyperspectral imagery based on the regularized subspace method and collaborative representation, Remote Sens. (Basel) 11 (11) (2019) 1318.

[19] X. Zhao, W. Li, C. Zhao, R. Tao, Hyperspectral target detection based on weighted cauchy distance graph and local adaptive collaborative representation, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13.

[20] L. Zhang, B. Cheng, A combined model based on stacked autoencoders and fractional fourier entropy for hyperspectral anomaly detection, Int. J. Remote Sens. 42 (10) (2021) 3611–3632.

[21] C. Zhao, X. Li, H. Zhu, Hyperspectral anomaly detection based on stacked denoising autoencoders, J. Appl. Remote Sens. 11 (4) (2017) 042605.

[22] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, J. Ma, Hyperspectral anomaly detection with robust graph autoencoders, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–14.

[23] S. Wang, X. Wang, L. Zhang, Y. Zhong, Auto-AD: autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–14.

[24] S. Wang, X. Wang, L. Zhang, Y. Zhong, Deep low-rank prior for hyperspectral anomaly detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–17.

[25] W. Xie, J. Lei, B. Liu, Y. Li, X. Jia, Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection, Neural Netw. 119 (2019) 222–234.

[26] Z. Li, Y. Zhang, J. Zhang, Hyperspectral anomaly detection for spectral anomaly targets via spatial and spectral constraints, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–15.

[27] Y. Li, T. Jiang, W. Xie, J. Lei, Q. Du, Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning, IEEE Trans. Geosci. Remote Sens. 60 (2021) 1–11.

[28] X. Cheng, M. Wen, C. Gao, Y. Wang, Hyperspectral anomaly detection based on Wasserstein distance and spatial filtering, Remote Sens. (Basel) 14 (12) (2022) 2730.

[29] T. Jiang, W. Xie, Y. Li, J. Lei, Q. Du, Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection, IEEE Trans. Neural Netw. Learn. Syst. 33 (11) (2021) 6504–6517.

[30] F. Zhang, S. Kan, D. Zhang, Y. Cen, L. Zhang, V. Mladenovic, A graph model-based multiscale feature fitting method for unsupervised anomaly detection, Pattern Recognit. (2023) 109373.

[31] A. Nazabal, P.M. Olmos, Z. Ghahramani, I. Valera, Handling incomplete heterogeneous data using VAEs, Pattern Recognit. 107 (2020) 107501.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, Commun. ACM 63 (11) (2020) 139–144.

[33] D. Chen, L. Yue, X. Chang, M. Xu, T. Jia, NM-GAN: noise-modulated generative adversarial network for video anomaly detection, Pattern Recognit. 116 (2021) 107969.

[34] A.T. Gnanha, W. Cao, X. Mao, S. Wu, H.-S. Wong, Q. Li, The residual generator: an improved divergence minimization framework for GAN, Pattern Recognit. 121 (2022) 108222.

[35] J. Zhu, F. Deng, J. Zhao, J. Chen, Adaptive aggregation-distillation autoencoder for unsupervised anomaly detection, Pattern Recognit. 131 (2022) 108897.

[36] Y. Kim, S. Choi, Forward-backward generative adversarial networks for anomaly detection, in: Asian Conference on Machine Learning, PMLR, 2019, pp. 1142–1155.

[37] W. Xu, K. Shawn, G. Wang, Toward Learning a Unified Many-to-many Mapping For Diverse Image Translation, Vol. 93, Elsevier, 2019, pp. 570–580.

[38] X. Yu, X. Zhang, Y. Cao, M. Xia, VAEGAN: A collaborative filtering framework based on adversarial variational autoencoders, in: IJCAI, 2019, pp. 4206–4212.

[39] M.O. Kaplan, S.E. Alptekin, An Improved biGAN Based Approach for Anomaly Detection, Vol. 176, Elsevier, 2020, pp. 185–194.

[40] C.-I. Chang, An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis, IEEE Trans. Geosci. Remote Sens. 59 (6) (2020) 5131–5153.

Zhiwei Wang received the BS degree in Surveying and Mapping Engineering and the MS degree in photogrammetric and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2018 and 2021, respectively. He is currently pursuing the PhD degree in cartography and geographic information system with the Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China. His research interests include hyperspectral image processing, anomaly detection and object detection.

Xue Wang received the BS degree in geographic information system and the PhD degree in photogrammetric and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2014 and 2019, respectively. He is currently a Post-Doctoral Researcher with the Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China. His research interests include hyperspectral imagery processing, deep learning, and ecological monitoring.

Kun Tan (Senior Member, IEEE) received the BS degree in information and computer science from Hunan Normal University, Changsha, China, in 2004, and the PhD degree in photogrammetric and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2010. From September 2008 to September 2009, he was a Joint PhD Candidate of remote sensing with Columbia University, New York, NY, USA. From 2010 to 2018, he was with the Department of Surveying, Mapping, and Geoinformation, China University of Mining and Technology. He is currently a Professor with the Key Laboratory of Geographic Information Science (Ministry of Education), East China Normal University, Shanghai, China. His research interests include hyperspectral image classification and detection, spectral unmixing, quantitative inversion of land surface parameters, and urban remote sensing