Characterization of Background-Anomaly Separability With Generative Adversarial Network for Hyperspectral Anomaly Detection

Jiapeng Zhong, Weiying Xie, Member, IEEE, Yunsong Li, Member, IEEE, Jie Lei, Member, IEEE, and Qian Du, Fellow, IEEE

Abstract—Hyperspectral images (HSIs) have unique advantages in distinguishing subtle spectral differences of different materials. However, due to complex and diverse backgrounds, unknown prior knowledge, and imbalanced samples, it is challenging to separate background and anomaly. In this article, we present a novel characterization of background-anomaly separability with a generative adversarial network (BASGAN) for hyperspectral anomaly detection. The key contribution is the proposal to explicitly constrain the background and anomaly separability by characterizing background spectral samples while avoiding anomaly reconstruction. First, we use a class saliency map extraction algorithm to obtain pseudobackground and anomaly samples for adversarial training. To further mitigate the suffering of anomaly contamination in background distribution estimation, we introduce background-anomaly separability constrained loss function to enhance the reconstruction of the background while weakening the anomaly reconstruction in a semisupervised way. Additionally, a discriminator is induced into the latent space to make the encoded representation resemble Gaussian distribution during adversarial training. The other is adversarial training in the reconstruction space so that the background estimation can be improved. Experiments conducted on real data sets illustrate the superior background-anomaly separability of the proposed method.

Index Terms—Background-anomaly separability, generative adversarial networks (GANs), hyperspectral anomaly detection.

I. INTRODUCTION

W

ith the development of spectroscopic technology, hyperspectral images (HSIs) can provide informative spectral information about physical properties of different materials [1]–[3]. Due to the rich spectral information with hundreds of adjacent spectral channels, HSIs are widely applied in various fields, in which anomaly detection plays a pivotal role [4]–[6]. As a special method of target detection [7], [8], hyperspectral anomaly detection attempts to locate pixels that are significantly different in the spatial or spectral domain from their surrounding pixels without any prior information [9]–[11], which has aroused great interest in remote sensing [12].

Among a diversity of anomaly detection methods, statistical-based methods are the most widely used and well-known methods that distinguish anomalies from background by statistical differences. A significant body of prior work, such as the benchmark algorithm Reed–Xiaoli (RX) [13] and its various extended versions [14]–[18], are dedicated to establishing a background model to suppress the background. The RX algorithm assumes that the background model obeys the multivariate Gaussian distribution and constructs the background statistical model by estimating the mean value and covariance of samples. Because of the sparsity of the anomaly component, the influence of anomalies on the background parameter estimation is considered negligible. Then anomalies are detected by comparing the Mahalanobis distance between each test pixel and the background. The global RX detection (GRXD) and the local RX detection (LRXD) involve all pixels and neighboring pixels, respectively [15]. However, it may not be reasonable to assume the background obeying Gaussian distribution in these methods due to the complexity of real HSI data. To solve this problem, the kernel RX (KRX) aims to model the background in a high-dimensional space through a kernel function. The choice of an appropriate kernel function and its complexity have led to huge computational costs [16], [18].
The representation-based algorithm is another developing branch, which usually includes collaborative representations [19], sparse representations [20], [21], and low-rank representations [22]. Among them, the collaborative representation-based detector (CRD) constructs a predictable background based on the concept that each background pixel can be well represented by its spatial neighborhood, while anomaly pixels cannot. The advantage is the ability to adaptively model the background even when anomaly pixels are involved. The low-rank and sparse matrix decomposition (LRaSMD)-based Mahalanobis distance method for hyperspectral anomaly detection (LSMAD) takes full advantage of the LRaSMD technique to set the background apart from anomalies and explores the low-rank prior knowledge of the background to compute the background statistics [22]. Then it applies the Mahalanobis distance differences to detect the probable anomalies. The aforementioned methods are based on original reflectance spectrum, and none of them consider a spectral transform. In the fractional Fourier entropy (FrFE)-based hyperspectral anomaly detection, the fractional Fourier transform (FrFFT) is employed as preprocessing to obtain features in an intermediate domain between an original reflectance spectrum and its Fourier transform with complementary strengths [23]. However, the construction and estimation of the background model in the talked background modeling methods are overcomplicated, and the astringent distribution assumption is required.

To solve these inherent problems, a method based on graph theory constructs a vertex- and edge-weighted graph, and then performs a pixel selection process to locate the anomaly targets, which makes the method more adaptive and improves the detection performance [24]. Another method proposes a high-order 2-D crossing approach to find the regions of rapid change in the spectrum and design a low-complexity discrimination framework for fast hyperspectral anomaly detection, which runs without any a priori assumption [25]. Since spatial information also has an important role in anomaly detection, different from the above methods, in the Attribute and Edge-preserving filtering-based Detection (AED) [26] algorithm, the property and distinct spatial signatures are detected with attribute filtering for the pixels with the specific area and a Boolean map-based fusion approach to obtain an initial pixel-wise detection result.

Recently, the emerging and attractive deep learning-based methods, such as autoencoder (AE) and generative adversarial networks (GANs) [27]–[29], have advantages in capturing deep features with nonlinear properties and approaching complex functions through layers, which broaden the prospects of hyperspectral anomaly detection. Generally, an AE-like architecture without noise prior is as feature extractor or applied in anomaly detection by the residual between the reconstructed background image and the original image, which means that anomalies are expected to have large reconstruction errors and the background has small reconstruction errors [30]–[32]. A notable example is shown in [32], where an unsupervised anomaly detection model employs an encoder–decoder convolutional neural network with skip connections to learn the normality distribution for anomaly detection. Because of the evident error gap between anomalies and the reconstructed background, the residual obtained by the reconstructed background data and the original data is also applied. The GAN-based anomaly detection frameworks [33]–[36] are instrumental to effectively detect anomalies on high-dimensional and complex data sets. For example, inspired by BiGAN [37], AnoGAN [35], and EGBAD [38], Akcay et al. [39] proposed the GANomaly approach to train a generator network on normal samples to learn their distribution. Meanwhile, only a generator and a discriminator are required in this approach as a standard GAN architecture. As far as we know, most of the traditional semisupervised based anomaly detection studies require hand-designed features and pure manually labeled background labels [7], [40]. The recently proposed semisupervised anomaly detection approach [41] learns a neural-network mapping of inputs that minimizes the volume of data around a predetermined point. Although the application of GANs is suitable for HSI processing, it has not been well applied to hyperspectral anomaly detection, and it is difficult for unsupervised anomaly detection methods to make a breakthrough in performance. Besides, a supervised method yields encouraging performance, but it is limited by samples that are of insufficient sizes to be effectively modeled, and lack of sufficient samples with prior information makes it difficult for application [42], [43].

To address these problems, we establish a novel semisupervised network architecture specially designed for hyperspectral anomaly detection to achieve the balance between the performance and sample limitations. Unlike previous semisupervised work requiring manual tagging to collect labeled training data, our proposed method is fully adaptive and needs no human labeling. Based on the fundamental work of GAN, the primary purpose of this research is to explore the inferences inside the GAN by constraining the separability between anomalies and the background spectral vectors to find an accurate representation of the background data distribution, which can be distinguished statistically from anomaly samples. The main contributions of this work can be concluded as follows.

1) **GAN-Based Semisupervised Anomaly Detection:** Based on GAN, we transform the unsupervised hyperspectral anomaly detection problem into a semisupervised perspective that can achieve an adaptive balance between the performance and sample limitations. On the GAN architecture with higher generalization ability, we introduce latent representation and image reconstruction adversarial loss into the proposed semisupervised adversarial training method.

2) **Background-Anomaly Separability:** We adaptively construct pseudo background samples for training, which overcomes the difficulty of manually and accurately labeling. For the first time, a background-anomaly separation constraint is imposed into the GAN using pseudo anomaly samples to modify the suspected anomaly samples generated in the background estimation process with a large separability gap from the background.

3) **Background Suppression Capability:** We theoretically and experimentally discuss the ability to enhance the difference between the highlighted anomalies and the suppressed background of the proposed BASGAN.
Experiments on the real data sets including both point and structural anomalies illustrate that our method is superior to the existing state-of-the-art methods.

The remainder of this article is divided into five sections. Section II reviews the related work about deep learning methods in anomaly detection. Section III describes the proposed method. Section IV is devoted to experiments and results. Section V draws the conclusion.

II. RELATED WORK

Based on competition of two networks, the idea of GANs was initially proposed in [33] within a zero-sum game framework in game theory. There are two models in adversarial iterative training: the generator model $G$ and the discriminator model $D$. The generator is fed with input samples $x$ and is optimized to generate the latent feature that fools the discriminator with $z \sim p(z)$ until $z \sim q(z)$ can be viewed as coming from the imposed prior distribution $z \sim p(z)$ in the latent feature space. At the same time, the discriminator is fed with the potential samples $z \sim q(z)$ from the output of the generator and the samples $z \sim p(z)$ which obey the Gaussian aggregated posterior distribution. The discriminator, on the other hand, is trained to correctly predict whether the samples are from the imposed prior distribution or the generated latent feature, and then gives the judgment of truth and falsity to update the parameters of the generator. Through continuous iterative training and learning, $G$ generates more data. With optimization on a zero-sum game framework, each network strengthens its prediction capability until they reach an equilibrium. The following min–max objective can describe the competitive training process:

$$\min_G \max_D E_{z \sim q(z|x)}[\log D(z)] + E_{z \sim p(z)}[\log(1 - D(G(z)))]$$

(1)

where $z$ is output samples of the generator. $z \sim p(z)$ denotes a target Gaussian probability distribution. $D(z)$ is the discriminative model. $q(z|x)$ represents both the encoding model and generative model.

GAN’s strong ability to model complex high-dimensional data distributions, especially image distributions, can help generate data with the most advanced performance. Derived from the basic framework of GANs, many variants of GANs have been proposed through changing objective function and architecture. Most existing GAN-based methods are trained on normal vectors and even normal and anomalous vectors, which are not well applied to hyperspectral anomaly detection. Therefore, without any normal or anomalous training samples, their applications are limited [44]. To approximate the performance of the supervised method while eliminating the limitation of training samples, we first propose a BASGAN for hyperspectral anomaly detection.

III. PROPOSED APPROACH

In this section, we construct a semisupervised hyperspectral anomaly detection architecture based on GAN to meet the requirement that background samples should be well represented while anomaly samples should perform poorly to find an accurate representation of background data distribution and distinguish statistically and clearly from small anomaly samples.

A. Problem Definition

The pseudo background and anomalous vectors are firstly obtained from the saliency map after coarse searching as the input for training. Then we perform adversarial training on the proposed AE architecture, where the separability constrained loss function and two discriminators are imposed. The conceptual model is trained only on background spectral vectors and tested on both anomalous and background spectral vectors. Mathematically, the definition and formulation of the model are as follows.

The background and anomalous vectors of HSI can be denoted as $H = \{h_1, h_2, \ldots, h_{MN}\}$. According to the class saliency map, we can obtain train data set $D_{Tn}$ and test data set $D_{Tt}$, in which the train data set $D_{Tn} = \{(h_1, l_1), (h_2, l_2), \ldots, (h_q, l_q)\}$ contains $q$ pseudobackground spectral vectors, where the $i$th label $l_i = 0$ represents background class. The testing data set $D_{Tt} = H = \{h_1, h_2, \ldots, h_{MN}\}$ comprises all the pseudobackground and anomalous spectral vectors, where $h_i \in [0, 1]$ represents the $i$th spectral vector after normalization.

Based on the data set defined above, we will train model $W$ based on $D_{Tn}$ and evaluate its performance on $D_{Tt}$. The training objective $T$ of the model $W$ is to capture the distribution of $D_{Tn}$ in the image space and the hidden vector space. Moreover, one would expect a higher loss in the reconstruction of the output image for anomalous spectral vectors. The training target $T$ will produce the minimum reconstruction loss for background spectral vectors, but higher loss for anomalous spectral vectors. Thus, we can judge whether the input sample is an anomalous vector from the loss.

B. Pipeline

Fig. 1 illustrates the flowchart of the proposed method, which comprises a generator $G$ and two discriminators $D_1$ and $D_2$. The network $G$ is composed of an encoder and a decoder. Through mapping a high-dimensional image to a low-dimensional latent representation, the encoder network captures the distribution of the input data. The hidden layer of the encoder consists of two fully connected layers, and the activation function is LeakyRelu. Compared to ReLU, LeakyRelu can overcome the dead nodes caused by training with a positive response maintained and setting the negative response to zero. Instead, LeakyRelu sets a small negative slope (such as $0.2$). Being symmetrical to the encoder, the network structure of the decoder consists of two fully connected layers, the activation functions LeakyRelu and Sigmoid.

As for the two discriminators shown in Fig. 1, they contain fully connected layers and use LeakyRelu as the activation function, respectively. We set the learning rate to $10^{-3}$, and the training batch size to $q$, which is the number of pseudobackground spectral vectors. The loss function of the entire network includes the background-anomaly separability constrained loss, latent representation adversarial loss, image
reconstruction adversarial loss, and samples learning and generative loss. Then, Adam’s algorithm is used to implement the optimization process. $D_l$ is trained to distinguish the latent samples from the probabilistic encoder conditioned on the input samples. Since there are no anomaly samples presented during training, it is difficult to enforce the corresponding images not being anomalous. Instead, we make sure that all background spectral vectors generated from latent samples are from the same spectral space distribution as the given distribution. To maximize the reconstruction ability of the network, we use $D_{II}$ to achieve accurate background distribution estimation. When $D_{II}$ cannot distinguish whether the input is from the real distribution vectors or the generated false vectors, the output-generated vectors will look similar to the background vectors.

C. Training Objective

As mentioned above, the motivation is to reconstruct the training pseudobackground spectral vectors in the image or latent space correctly while reconstructing anomalies incorrectly because it has never been trained on anomalous vectors. Therefore, for anomalous vectors, we expect a higher loss in the output reconstruction or potential representation. To achieve this, we combine the four discussed loss functions, and each loss value has its contribution within the entire training objective.

To obtain the pseudobackground and anomaly samples, a class saliency map is generated first. Because anomalies always appear as small, specific area attributes and have significant gray values, attribute filtering is applied to the spatial structure by preserving or removing connected components. The closing and opening operations are used to remove dark and bright connected components in a small area, respectively.

Therefore, the differential map to preserve the objects in small areas is calculated

$$X = |U - V^o(U)| + |U - V^c(U)|$$  \hspace{1cm} (2)

where $V^o(U)$ and $V^c(U)$ represent the attribute openings and closings extracted from $X$, respectively. The first and second terms in (2) preserve the bright connected components in a small area and the dark connected components, respectively. Thus a class saliency map $X$ can be obtained. According to $X$, coarse searching is conducted to find the pseudobackground and anomaly spectral vectors, which can be expressed as

$$D_{Trn} = \{(b_1, l_1), (b_2, l_2), \ldots, (b_q, l_q)\}$$  \hspace{1cm} (3)

where $D_{Trn}$ is the set of background spectral vectors, and $b_i$ is the background spectral vector with $l_i = 0$. In addition

$$D_{Ano} = \{(a_1, k_1), (a_2, k_2), \ldots, (a_n, k_p)\}$$  \hspace{1cm} (4)

where $D_{Ano}$ is the set of anomalous vectors, and $a_i$ is obtained according to the class saliency map with $k_i = 1$. $D_{Trn} \cup D_{Ano} = D$ and $D_{Trn} \cap D_{Ano} = \emptyset$, $p + q = MN$.

1) Background-Anomaly Separability Constrained Loss: To explicitly constrain the background-anomaly separability, a distance constraint is imposed on background and anomalous vectors, and then the network is trained to obtain and save the parameters. To explicitly constrain the background-anomaly separability, a distance constraint is imposed on the background and anomalous vectors, and then the network is trained to obtain and save the parameters. Considering the discrepancy between background and anomaly is expected to be as evident as possible, we add the following constrain in BASGAN for better performance. The distance between the anomaly $(a_i, k_i)$ and reconstructed background $\hat{b}_i$ is designed to be as
large as possible during optimization. Hence, the background-anomaly separability can be enhanced through the proposed background-anomaly separability constrained loss function

\[
\text{LOSS}_{\text{Ac}} = \sum_{i=1}^{q} \| \mathbf{b}_i - \hat{\mathbf{b}}_i \|_2^2 - \lambda \| \mathbf{b} - \mathbf{\Phi} \|_2^2. \tag{5}
\]

Given a training pseudo data \( \mathbf{b}_i \), the loss function aims to minimize the reconstruction error between \( \mathbf{b}_i \) and the output \( \hat{\mathbf{b}}_i \). Meanwhile, it intends to maximize the reconstruction error between \( \mathbf{b}_i \) and the pseudo anomaly \( \mathbf{a} \). To minimize the background-anomaly separability constrained loss function, the first term \( \sum_{i=1}^{q} \| \mathbf{b}_i - \hat{\mathbf{b}}_i \|_2^2 \) must be minimized, and the second term \( \lambda \| \mathbf{b} - \mathbf{\Phi} \|_2^2 \) must be maximized.

2) Latent Representation Adversarial Loss: To ensure that the hidden features in the latent space follow the multivariate Gaussian distribution, the encoder \( \text{En} \) and the discriminator \( D_I \) form a GAN in the latent space learning stage. The encoder is converted into a generator at this stage to guarantee the generated hidden features can fool the discriminator. At the same time, the purpose of the discriminator \( D_I \) is to discriminate input from encoder \( \text{En} \) or a priori Gaussian distribution. The latent representation adversarial loss is

\[
\text{LOSS}_z = \sum_{i=1}^{q} E_{\mathbf{b}_i \sim \text{p}_{\mathbf{b}_i}} [\log D_I(\mathbf{b}_i)] + E_{\mathbf{b}_i \sim \text{p}_{\mathbf{b}_i}} [\log(1 - D_I(\text{En}(\mathbf{b}_i)))] \tag{6}
\]

where \( \text{En}(\mathbf{b}_i) \) is the output of the encoder. As the input, \( \mathbf{b}_i \) is the background sample which obeys the prior distribution \( \mathbf{b}_i \sim \text{p}_{\mathbf{b}_i} \). We train \( D_I \) along with the AE network using \( \min_{\text{En}} \max_{D_I} \text{LOSS}_z \). Since the latent space is a hypercube with \( \lambda \| \mathbf{b} - \mathbf{\Phi} \|_2^2 \) must be maximized.

3) Image Reconstruction Adversarial Loss: The decoder \( \text{De} \) aims to generated images from latent samples while \( D_{II} \) intends to distinguish between images of a given class and reconstructed from the decoder \( \text{De} \). When \( D_{II} \) is fooled, randomly selected fake vectors often look similar to a given distribution. The image reconstruction adversarial loss is evaluated as

\[
\text{LOSS}_{II} = \sum_{i=1}^{q} E_{\mathbf{b}_i \sim \text{p}_{\mathbf{b}_i}} [\log D_{II}(\mathbf{b}_i)] + E_{\mathbf{z'} \sim \text{N}(0, I)} [\log(1 - D_{II}(\text{De}(\mathbf{z'})))] \tag{7}
\]

where \( \text{De}(\mathbf{z'}) \) is the output of the decoder. When \( D_{II} \) is fooled, fake vectors chosen at random in general will look similar to examples from the given class. \( D_{II} \) is learned together with the AE network using \( \min_{\text{De}} \max_{D_{II}} \text{LOSS}_{II} \).

4) Samples Learning and Generative Loss: With the input of a set of images, GAN performs a two-player game between generative and discriminative network. The generator network \( G \) attempts to generate samples (false samples) based on a given real Gaussian distribution \( \mathbf{z'} \sim \text{N}(0, I) \), and the generator network learns the distribution of given background samples

\[
\text{LOSS}_G = \sum_{i=1}^{q} E_{\mathbf{z'} \sim \text{N}(0, I)} [\log(1 - \text{De}(\mathbf{z'}))] \tag{8}
\]

where \( \text{De}(\mathbf{z'}) \) is the output of the decoder.

Finally, the total training goal is the sum of the above losses, so the advantages of an individual loss are exerted in the overall objective function. We set the weighting parameters of each item in the overall objective function as 1, which is empirically shown to yield optimal performance

\[
\text{Loss} = \text{LOSS}_{\text{Ac}} + \text{LOSS}_z + \text{LOSS}_{II} + \text{LOSS}_G. \tag{9}
\]

D. Background Suppression Approach for Detection

As demonstrated in Fig. 1, the original HSI is reconstructed with the learned models in the overall architecture. Given a test data \( \mathbf{H} \) as input, we reload the parameters that have been learned with the reconstruction of the background spectral vectors in the training model. We get an output HSI that can reconstruct the background spectral vector very well, but it does not perform well on anomaly reconstruction. Then the Mahalanobis distance is used to detect anomalies in the suppressed output as

\[
\mathbf{R} = (\hat{\mathbf{h}}_i - \mu)^T \Gamma^{-1} (\hat{\mathbf{h}}_i - \mu) \tag{10}
\]

where \( \mathbf{R} \) is the calculation result of Mahalanobis distance, \( \mathbf{h}_i \in L \times 1 \) is the i-th vector in reconstructed HSI \( \mathbf{H} \), \( \mu \in L \times 1 \) and \( \Gamma \in L \times L \) are the mean and the covariance matrix of \( \mathbf{H} \), respectively.

On this basis, we perform the background suppression algorithm is applied to increase the difference between anomaly and background further as

\[
\mathbf{R}' = (1 - e^{-\beta \mathbf{X}}) \mathbf{R} \tag{11}
\]

where \( \mathbf{R}' \) is the obtained detection map by the proposed method, which suppresses the background of \( \mathbf{R} \), and \( \mathbf{X} \) is calculated by (2).

To confirm the validity of the designed model, we take the Hyperspectral Digital Imagery Collection Experiment (HYDICE) data set as an example and compare the spectral curves of the pseudo background and anomalous vectors before and after reconstruction. From Fig. 2, it can be seen

![Fig. 2. Comparison of the reconstructed and original spectral curves. (a) Anomaly. (b) Background.](image-url)
from the spectral curves that the reconstructed and the input background vectors have definite similarities. The trend of the spectral curves are consistent, and the reconstruction effect tends to be more uniform. However, for an anomalous vector, since it has not been well trained, the curve exhibits extremely unstable characteristics and has a large difference from the original one. To sum up, the network structure designed in this article can improve background-anomaly separability markedly.

IV. EXPERIMENTAL EVALUATION

A. Data Set Description

1) San Diego Data Set: The first data set was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the San Diego airport area, CA, USA, which is widely used for anomaly detection. The size of the image is $100 \times 100 \times 189$ with the noisy bands removed. The San Diego data set and its corresponding ground truth are obtained from [45].

2) HYDICE Data Set: The second data set was recorded by the HYDICE sensor over an urban area, CA, USA. The image size is $80 \times 100 \times 175$, in which the noisy bands have been removed. There are 21 anomalous pixels in the ground truth of this data, which are cars and roofs.

3) Airport-Beach-Urban (ABU) Data Set: The third data set was captured by the AVIRIS sensor. Among airport, beach, and urban scenes, two of the urban scenes are used because of different types of anomalies when compared with other data sets designed in the experiment. The size of the image is $100 \times 100 \times 224$. The noisy bands in the original images have been removed. Due to different heights of flights, spatial resolutions of the images differ from each other. The ABU data set and its corresponding ground truth are made available on the home page.

4) El Segundo Data Set: The fourth data set was captured by the AVIRIS sensor, covering the El Segundo region of CA, USA, with a wavelength range of 366–2496 nm, 224 spectral channels, and spatial size of $250 \times 300$. The ground resolution of each pixel is 7.1 m. The image data set mainly consists of an oil refinery area, several residential areas, a park, and a campus. Structures such as oil storage tanks and towers are considered as anomalous targets.

5) Grand Island Data Set: The fifth data set was also acquired by the AVIRIS sensor at the location of Grand Island on the Gulf Coast, part of Jefferson Parish, LA, USA. The data set contains $300 \times 480$ pixels and 224 spectral channels with a wavelength range of 366–2496 nm and a spatial resolution of about 4.4 m. The materials in the main background scene are islands and water. Artificial objects in the water are selected as anomalies to be detected.

B. Comparative Methods and Evaluation Criterion

As performance-leading detectors, five state-of-the-art anomaly detection methods listed below are used as comparison methods. These methods are called benchmarks for hyperspectral anomaly detection and are often cited in various kinds of literature, including RX [13], AED [26], CRD [19], LSMAD [22], and FrFE [23]. The RX algorithm performs well in data sets representing simple scenes. Based on collaborative representation, the CRD method is suitable for anomaly detection in complex scenes. The window size affects the detection performance a lot. For the CRD method, the optimal regularization parameter is set to $10^{-6}$ [19]. Meanwhile, the sizes of the two windows are selected as 7 and 13, respectively. The comparison methods also include the space-based algorithm AED that performs well on real data sets. The optimal parameters are described in [26]. The LSMAD method extracts background and anomaly information to alleviate anomaly contamination and the inverse covariance matrix calculation problem. The rank values for different data sets are selected empirically. The FrFE-based hyperspectral anomaly detection method is also performed for signal enhancement and noise suppression. The optimal parameters are shown in [23].

The most commonly used criteria, the receiver operating characteristic (ROC) curve and the area under the ROC curve.
Fig. 4. Background-anomaly separation analysis of the compared methods on (a) San Diego, (b) HYDICE, (c) ABU-urban1, (d) ABU-urban2, (e) EI Segundo, and (f) Grand Island.

(AUC), are used for quantitative assessment. The ROC curve of \((P_d, P_f)\) can effectively illustrate the relationship between the true positive rate and false positive rate by changing different thresholds on the output of a detector.

**C. Parameters Analysis**

The parameters’ impact on the experimental results including the weight of separability constraint \((\lambda)\) and the number of hidden layers are analyzed. It can be seen from Fig. 3(a) that when the coefficient \(\lambda\) of the background-anomaly separability constrained loss function is changed from 0.025 to 0.4 at two-time intervals, the AUC score of \((P_d, P_f)\) illustrate different performance. By conducting experiments on different data sets with controlled variables only changing the separability constrained loss function, it can be seen that when \(\lambda\) is set to 0.05, the experimental data set can achieve the best experimental results. As shown in Fig. 3(b), the number of fully connected layers is varied from 1 to 5 to obtain the optimal network architecture. With the increase of the layers, the AUC score of \((P_d, P_f)\) tends to increase as well. When the number of layers are 3, 4, and 5, the AUC score of \((P_d, P_f)\) becomes stable. Thus, the number of layers is set to 3 in consideration of the balance of the performance and the computation costs.

**D. Experimental Results**

In Fig. 4, different approaches are compared in terms of background-anomaly separability on the real data sets. As seen from this figure, the anomaly distribution range is located at a relatively high part while that of the background is relatively low. Compared with the RX, AED, CRD, LSMAD, and FrFE methods, the distribution range of the anomaly and background value of the proposed method is more concentrated, and the gap between the distribution ranges is larger. This phenomenon shows that the proposed method can separate a small number of anomalies from a large number of background to a greater extent, which is beneficial to the detection. The second-best AED algorithm can also distinguish background and anomaly well, but its separability is not as good as the proposed method, and for some specific data sets, such as Grand Island and EI Segundo, the separation is not significant enough, which may result in misdetection or false alarm. Overall, the BASGAN method can make the separability between anomaly and background more prominent, and the analysis from both subjective and objective results agrees with the conclusion.

Tables I and II show the AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\) for different methods on six real hyperspectral data sets, which demonstrate the competitive performance of the BASGAN method in terms of the AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\). To quantitatively investigate and intuitively represent the detection performance, we plot the corresponding ROC curves of \((P_d, P_f)\) of different methods on each data set.
TABLE I

<table>
<thead>
<tr>
<th>Images</th>
<th>BASGAN</th>
<th>AED</th>
<th>RX</th>
<th>CRD</th>
<th>LSMAD</th>
<th>FrFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Diego</td>
<td>0.99542</td>
<td>0.98500</td>
<td>0.94041</td>
<td>0.96254</td>
<td>0.97597</td>
<td>0.97232</td>
</tr>
<tr>
<td>HYDICE</td>
<td>0.99975</td>
<td>0.99580</td>
<td>0.97637</td>
<td>0.99271</td>
<td>0.99706</td>
<td>0.99259</td>
</tr>
<tr>
<td>ABU-urban1</td>
<td>0.99900</td>
<td>0.99800</td>
<td>0.99964</td>
<td>0.99964</td>
<td>0.97811</td>
<td>0.98918</td>
</tr>
<tr>
<td>ABU-urban2</td>
<td>0.99260</td>
<td>0.98450</td>
<td>0.96930</td>
<td>0.93074</td>
<td>0.92798</td>
<td>0.96463</td>
</tr>
<tr>
<td>Grand Island</td>
<td>0.99996</td>
<td>0.99973</td>
<td>0.99999</td>
<td>0.99931</td>
<td>0.99995</td>
<td>0.99988</td>
</tr>
<tr>
<td>El Segundo</td>
<td>0.99148</td>
<td>0.96658</td>
<td>0.97826</td>
<td>0.85944</td>
<td>0.96919</td>
<td>0.95536</td>
</tr>
</tbody>
</table>

TABLE II

<table>
<thead>
<tr>
<th>Images</th>
<th>BASGAN</th>
<th>AED</th>
<th>RX</th>
<th>CRD</th>
<th>LSMAD</th>
<th>FrFE</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Diego</td>
<td>0.00638</td>
<td>0.03330</td>
<td>0.05889</td>
<td>0.26227</td>
<td>0.01335</td>
<td>0.04244</td>
</tr>
<tr>
<td>HYDICE</td>
<td>0.00397</td>
<td>0.01650</td>
<td>0.03798</td>
<td>0.04160</td>
<td>0.01388</td>
<td>0.04493</td>
</tr>
<tr>
<td>ABU-urban1</td>
<td>0.00305</td>
<td>0.02290</td>
<td>0.05552</td>
<td>0.03678</td>
<td>0.02665</td>
<td>0.04260</td>
</tr>
<tr>
<td>ABU-urban2</td>
<td>0.00262</td>
<td>0.03060</td>
<td>0.04370</td>
<td>0.04430</td>
<td>0.02544</td>
<td>0.01959</td>
</tr>
<tr>
<td>Grand Island</td>
<td>0.00262</td>
<td>0.01821</td>
<td>0.00738</td>
<td>0.00885</td>
<td>0.00153</td>
<td>0.00584</td>
</tr>
<tr>
<td>El Segundo</td>
<td>0.00158</td>
<td>0.00652</td>
<td>0.00952</td>
<td>0.00291</td>
<td>0.00260</td>
<td>0.025829</td>
</tr>
</tbody>
</table>

Fig. 5(a) shows the visual inspection results of the San Diego data set. In the first lines of Tables I and II, the corresponding AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\) are listed. The AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\) obtained by BASGAN are 0.99542 and 0.00638, respectively, which are the optimal detection results when compared with the other five methods. It can also be seen from Fig. 5(a) that the RX and CRD methods only detect a part of anomalies, especially for the San Diego data set compared to other data sets. There are relatively more anomalies of false alarms in AED, CRD, and FrFE methods. The detection results of AED and LSMAD have lost some shape information about anomalies. Similarly, FrFE and CRD methods can avoid false positives, but they miss some anomalies and reduce the rate of detection. Both the subjective and objective results validate the effectiveness and advantages of the BASGAN.

For the HYDICE data set, the detection results are shown in Fig. 5(b). According to the data in Tables I and II and Fig. 5(b), the RX and CRD cannot preserve the shape of the anomalies and omit some of them. The AED method can provide better detection results but will produce false positives. The result of FrFE method contains too many background samples though its performance is the closest to BASGAN. The proposed method can effectively highlight anomaly and suppress background.

For the two urban scenes in ABU data sets, the detection maps are presented in Fig. 5(c) and (d). Tables I and II show that the AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\) almost reach the ideal values for the first scene. The AUC scores of
Fig. 6. ROC curves of the compared methods on (a) San Diego, (b) HYDICE, (c) ABU-urban1, (d) ABU-urban2, (e) El Segundo, and (f) Grand Island.

\((P_d, P_f)\) obtained by the BASGAN in the first urban scene is 0.99900, which is much higher than those five typical methods. The BASGAN method offers a trade-off performance in consideration of the average AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\). Though the AED algorithm performs robustly and has similar detection accuracy with BASGAN according to the results for the second scene, the AUC score of \((P_f, \tau)\) of the BASGAN method is 0.00305 which is almost 10% as small as 0.02290 obtained by the AED method. Moreover, the results of the BASGAN method are visually most similar to the map compared to RX, CRD, AED, LSMAD, and FrFE methods.

For the Grand Island data set, Fig. 5(e) shows visual inspection results. Though FrFE, LSMAD, and AED methods can detect most of the anomalies, the proposed method gains the best performance according to Tables I and II. The LSMAD method yields a smaller false alarm rate (0.00153) when compared to the BASGAN method (0.00262), but it also has a lower AUC score of \((P_d, P_f)\) when compared to the proposed method (0.99996) and messier visual representation. The RX, FrFE, and AED methods with good performance can detect almost all anomalies and retain the shape of the target anomalies, but they also detect some wrong anomalies, while the proposed method can produce detection results with better background suppression.

For the EI Segundo data set, from the detection results shown in Fig. 5(f), it is not difficult to find that all the methods are close to the reference map. Compared with AED, RX, LSMAD, and CRD, the AUC score of \((P_d, P_f)\) of the BASGAN method is 0.99148 and the AUC score of \((P_f, \tau)\) is 0.00158, which means it can detect more anomalies with a lower false alarm rate that is crucial to evaluate the performance of anomaly detection. Among all the results, the AUC score of \((P_f, \tau)\) obtained by the BASGAN method is closer to the ideal value 0, and the AUC scores of \((P_d, P_f)\) is the best of all the methods. Visually, the LSMAD and CRD methods successfully retain the background and obtain satisfactory false alarm rates, but it sacrifices the detection accuracy of the anomalies and only detect part of the anomalies. The AED and FrFE methods can almost completely detect anomalies, but cannot suppress background interference. The LRX method can avoid false alarms, but the detection rate is low. The subjective and objective results are consistent with each other, and the effectiveness and advantages of the method are further verified.

Furthermore, the ROC curves of all the data sets are shown in Fig. 6 with the best parameters for each method. The area under these curves are consistent with the AUC scores reported in Tables I and II. Since all the ROC curves of \((P_d, P_f)\) lie nearer the upper left corner, it is noticeable that the performance of the BASGAN algorithm is outstanding. It can also be concluded that the BASGAN method can obtain a high probability of detection with a low false alarm rate, which further confirms that the proposed BASGAN method outperforms.

All experiments were performed on MATLAB (R2019a) environment on a server with CPUs: Intel (R) Core (TM) i5-7200U CPU at 2.70 GHz with 8 GB of RAM and a GeForce
RTX 2080 graphics card based on a system running Python 3.6.0, TensorFlow-GPU 1.10.0 and CUDA 10.0.

E. Ablation Study

To analyze the effect of each component on the final detection result (i.e., $D_I$, $D_{II}$, and separability constraint), we carry an ablation study on six real HSI data sets over the AUC scores of $(P_d, P_f)$ and $(P_f, r)$ by replacing the BASGAN with the different models. In specific, there are four scenarios for analysis. In the first scenario, we only consider the AE as the network architecture. It is noted that the AUC value obtained by the AE reaches at 0.96696 for the El Segundo data set. When $D_I$ is introduced into the AE in the second scenario, the performance of the system improves marginally by 1.935%. When adding both $D_I$ and $D_{II}$ on top in the third scenario, the detection results become better and improves by 0.440%, which indicates that the introduction of $D_I$ and $D_{II}$ step is quite efficient. In the final scenario, we perform on the separability constrained loss function added BASGAN model, and the performance is further improved by 0.147%. The observation demonstrates that the separability constraint indeed helps with anomaly detection, especially for the San Diego, HYDICE, ABU-urban2, and El Segundo data sets. From the average point of view, the successively increasing percentages for the other scenarios of the six real HSIs are 0.678%, 0.198%, and 0.065%, respectively. It is obvious that the average AUC score of $(P_d, P_f)$ of the methods without $D_I$, $D_{II}$, and separability constraint are smaller than the proposed method, and the average AUC score of $(P_f, r)$ of the compared method is higher than the proposed method, which confirms that the combination of adversarial-learning and separability constraint with AE in the proposed method contributes effectively to improve detection performance.

V. CONCLUSION

In this article, we propose a novel characterization of the background-anomaly separability model with the GAN for hyperspectral anomaly detection. Several aspects of the model deserve highlights. We obtain the saliency map by coarse searching to roughly generate pseudo anomalies and background vectors for training. The BASGAN method characterizes background samples by learning from pseudo background samples while avoiding the reconstruction of anomaly samples in a semisupervised way without the requirement of manual labels. To reduce the interference from anomalies on background modeling, we introduce a background-anomaly separability constrain loss function for the first time. The AUC scores of $(P_d, P_f)$ is improved from 0.99555 to 0.99620 on average when the separability constraint is imposed in the model. At a relatively high range, even a slight increase means great improvement of performance. Moreover, we induce a discriminator into the latent space and conduct adversarial training, making the encoded representation obey Gaussian distribution. Then, another adversarial training is employed in the input space so that all randomly drawn potential samples can generate examples closer to realistic. The two discriminators can improve the learning ability and strengthen the separability. Experimental results demonstrate that the proposed method is superior in the case of the discussed data sets.

ACKNOWLEDGMENT

The authors would like to thank Prof. X. Kang for providing the HYDICE, El Segundo, and Grand Island data sets for the experiment.

REFERENCES


<table>
<thead>
<tr>
<th>Component</th>
<th>San Diego</th>
<th>HYDICE</th>
<th>ABU-urban1</th>
<th>ABU-urban2</th>
<th>Grand Island</th>
<th>El Segundo</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>0.99260</td>
<td>0.98147</td>
<td>0.99780</td>
<td>0.98266</td>
<td>0.99987</td>
<td>0.96696</td>
<td>0.98689</td>
</tr>
<tr>
<td>AE with $D_I$</td>
<td>0.99475</td>
<td>0.99845</td>
<td>0.99839</td>
<td>0.98415</td>
<td>0.99986</td>
<td>0.98586</td>
<td>0.99338</td>
</tr>
<tr>
<td>AE with $D_I$ and $D_{II}$</td>
<td>0.99491</td>
<td>0.99951</td>
<td>0.99878</td>
<td>0.99124</td>
<td>0.99986</td>
<td>0.99002</td>
<td>0.99555</td>
</tr>
<tr>
<td>AE with $D_I$, $D_{II}$ and separability constraint</td>
<td>0.99542</td>
<td>0.99875</td>
<td>0.99900</td>
<td>0.99260</td>
<td>0.99996</td>
<td>0.99148</td>
<td>0.99620</td>
</tr>
</tbody>
</table>


Jiaping Zhong received the B.E. degree in telecommunications engineering from Jilin University, Changchun, China, in 2018. She is pursuing the M.S. degree with the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an, China. Her research interests include hyperspectral image processing, deep learning, and anomaly detection.

Weiying Xie (Member, IEEE) received the B.S. degree in electronic information science and technology from the University of Jinan, Jinan, China, in 2011, the M.S. degree in communication and information systems from Lanzhou University, Lanzhou, China, in 2014, and the Ph.D. degree in communication and information systems from Xidian University, Xi’an, China, in 2017. She is an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. She has published more than 20 articles in refereed journals, including the IEEE TRANSACTIONS ONGEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, Neural Networks, and Pattern Recognition. Her research interests include neural networks, machine learning, hyperspectral image processing, and high-performance computing.
Yunsong Li (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 1999 and 2002, respectively. He joined the School of Telecommunications Engineering, Xidian University, in 1999, where he is a Professor. He is the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks. His research interests focus on image and video processing and high-performance computing.

Jie Lei (Member, IEEE) received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 2006 and 2010, respectively. He was a Visiting Scholar with the Department of Computer Science, University of California, Los Angeles, CA, USA, from 2014 to 2015. He is an Associate Professor with the School of Telecommunications Engineering, Xidian University, is a member of Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks, and is also with the Science and Technology on Electro-Optic Control Laboratory, Luoyang, China. His research interests focus on image and video processing, computer vision, and customized computing for big-data applications.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000. She is the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. She is also an Adjunct Professor with the College of Surveying and Geo-informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a fellow of the International Society for Optics and Photonics (SPIE). She received the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was the Co-Chair of the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society from 2009 to 2013 and the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, in 2012. She has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the Journal of Applied Remote Sensing, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.