Background Learning Based on Target Suppression Constraint for Hyperspectral Target Detection

Weiying Xie , *Member, IEEE*, Xin Zhang, Yunsong Li , Keyan Wang , and Qian Du , *Fellow, IEEE*

Abstract—Hyperspectral target detection is critical in both military and civilian applications. However, it is a challenging task due to the complexity of background and the limited samples of target in hyperspectral images (HSIs). In this article, we propose a novel background learning model, called background learning based on target suppression constraint to characterize high-dimensional spectral vectors. Considering insufficient target samples, the model is trained only on the background spectral samples to accurately learn the background distribution. Then the discrepancy between the reconstructed and original HSIs are examined to spot the targets. To obtain a background training dataset, coarse detection is carried out. However, it is quite difficult to retrieve pure background data. Thus, a target suppression constraint is imposed to reduce the impact of suspected target samples on background reconstruction. Experiments on six real HSIs demonstrate that the proposed framework significantly outperforms the current state-of-the-art detection methods and yields higher detection accuracy and lower false alarm rate.

Index Terms—Background learning, hyperspectral image (HSI), target detection, target suppression constraint.

I. INTRODUCTION

HYPERSPECTRAL image (HSI), a 3-D data cube, contains both spectral and spatial information. Benefited from abundant spectral information, HSI has become instrumental in many application scenarios, such as medical diagnosis and treatment [1]–[3], agricultural production [4]–[6] and identification

of the terrain landform [7]–[9]. Among these applications, target detection is considered as a fundamental task and has received a surge of interest [10], [11]. Essentially, target detection can be regarded as a problem of classification and localization [12], [13], which has been widely used spanning from civilian to military.

A significant body of prior work exists within hyperspectral target detection. Traditionally, spectral difference between target and background is exploited to detect targets. Constrained energy minimization (CEM) [14], a classic hyperspectral target detection approach, imposes a constraint on target and background and builds a finite impulse response filter, which minimizes the filters output energy. Many extensions based on CEM have been investigated for hyperspectral target detection [15]–[17]. Among them, hierarchical CEM (hCEM) proposed by Zou and Shi [18] is an improved method based on CEM, which is a hierarchical structure containing different layers of CEM detectors. The hCEM increases detection performance layer by layer through a hierarchical suppression process. Another typical target detection method is orthogonal subspaces projection (OSP) [19], which projects spectral vectors into an orthogonal subspace of background with the aim of maximizing the signal-to-noise ratio (SNR). A combined sparse and collaborative representation (CSCR)-based approach [20] generates a detection output simply by using the difference between the representation residuals of target and background. Other methods, such as adaptive coherence/cosine estimator (ACE) [21], [22], matched filter [23], and matched subspace detectors [24] describe target detection as a hypothesis test problem.

Virtually, an HSI usually covers a large scale of ground scene, which contains various materials with different spectral properties. On this account, traditional target detection algorithms encounter many bottlenecks. For example, CEM estimates background with all pixels of an image, which ignores the influence of target pixels. Moreover, it is found to be difficult for OSP to construct background subspace accurately due to the complexity and diversity of HSIs.

Recent developments in deep learning, shown to be highly capable of characterizing high-dimensional data. The deep neural networks (DNNs) can model complex datasets by extracting multiple layers of representation and abstraction. Thus, DNNs are expected to yield higher detection accuracy for HSIs than those traditional ones using shallow feature extractors [25]–[28]. Generative adversarial networks [29] are promising deep learning models, which provide substantial advantages in learning input distributions with a min–max game. Many variants of

generative adversarial networks (GANs) have been proposed to adapt to different tasks [30]–[32]. For example, by adding the learning of the inverse mapping which maps the data to a latent representation, the original GAN can be extended to bidirectional GAN (BiGAN) [33], which has been successfully used in anomaly detection. Another variant of GAN is known as adversarial autoencoder (AAE) [34], which imposes an arbitrary prior on the latent code to learn the input distribution adversarially. Several contemporary studies demonstrate that AAE is effective in generating an approximated data to the input, and can adapt well to various complex scenes [35]–[37].

A training set containing sufficient samples plays a significant role in the establishment of DNNs. Compared with the small-area and low-probability target samples in HSIs, background occupying a larger area contains abundant samples. In virtue of its superior numbers, background samples are served as training data to establish a background estimation model in a semisupervised way [38]. Among the deep learning models mentioned above, AAE shows its potential in modeling background distribution. In [39], Xie *et al.* detected anomalies by learning the background distribution with an autoencoder and adversarial-learning-based background estimation model that is trained only on the background samples. Jiang *et al.* [40] proposed an adversarial framework to learn a discriminative background reconstruction with anomaly targets being suppressed, such that an initial detection result can be generated by the residual between the original and reconstructed images. In this article, a novel background estimation model based on AAE with a target suppression constraint is proposed, which is conducted to provide an accurate background description, then compare the reconstructed and original HSIs to spot the targets. To obtain a sufficient number of background samples, coarse detection is carried out. However, it is impossible to execute the coarse detection with no deviation which is not conducive to the subsequent target detection. Thus, a target suppression constraint is imposed on the model to reduce the impact of false detection on background reconstruction.

The major contributions of this article can be summarized as follows.

1) The perception of semisupervised learning is introduced to hyperspectral target detection and the constructed background estimation model can adapt well to various complex scenes and overcome the problem of scarce training samples.

2) We especially design a constrained AAE for jointly optimizing adversarial loss, target suppression constraint loss, and autoencoder loss. Compared with the traditional one, the target suppression constrained AAE minimizes the bias introduced by the ambiguous training set while taking full advantage of prior information.

3) Considering it is a tough task to obtain richly accurate background annotation, a simple but powerful detector is executed to display the nonsalient area and get a relatively pure background training set.

The remainder of this article is organized as follows. In Section II, we give a brief introduction of the AAE network. In Section III, we detail our proposed BLTSC method. Experiments and parameter settings are performed in Section IV. Section V draws the conclusion.

II. RELATED WORK

GAN [29], an emerging technology, have shown to be highly capable of learning accurate data distribution and generating new data with the same distribution. Conceptually similar to GAN, AAE, initially proposed by Makhzani *et al.* [34] introduces adversarial training into a traditional autoencoder. AAE consists of a bow-tie shaped generator G and a discriminator D, and its principle can be presented as follows.

The latent feature extracted from an input $\mathbf{x}$ is represented by $\mathbf{z}$. Let the encoding distribution and the decoding distribution be defined by $q(\mathbf{z}|\mathbf{x})$ and $q(\mathbf{x}|\mathbf{z})$, respectively. The aggregated posterior distribution of $q(\mathbf{z})$ on a latent space is formulated by

$$q(\mathbf{z}) = \int_{\mathbf{x}} q(\mathbf{z}|\mathbf{x})p_d(\mathbf{x})d\mathbf{x} \qquad (1)$$

where the $p_d(\mathbf{x})$ represents the data distribution. The AAE aims to minimize the Jensen Shannon divergence between $q(\mathbf{z})$ and $p(\mathbf{z})$. Here, $p(\mathbf{z})$ is the prior distribution we want to impose on the latent space. During the training phase, in addition to minimize the traditional reconstruction error, we also take adversarial training error into account. The adversarial training is realized by two operations, i.e., the generator G with $\mathbf{z} \sim q(\mathbf{z})$ to fool the discriminator D with $\mathbf{z} \sim p(\mathbf{z})$, and the discriminator D to distinguish $\mathbf{z} \sim q(\mathbf{z})$ from $\mathbf{z} \sim p(\mathbf{z})$. In essence, AAE establishes a competition between G and D though a min–max adversarial game, which can be represented as

$$\min_{G} \max_{D} E_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x})}[\log D(\mathbf{z})] \\ + E_{\mathbf{z}_{\text{prior}}\sim p(\mathbf{z}_{\text{prior}})}[\log(1 - D(G(\mathbf{z})))]. \qquad (2)$$

Due to the inherent potential for capturing data distributions, there is a growing body of literature that recognizes the importance of AAE. In [41], Principi *et al.* proposed an acoustic novelty detector based on AAE, and the results showed that the proposed approach provides a relative performance improvement equal to 0.26% compared to the standard autoencoder. A conditional difference AAE (CDAAE) [42] was proposed for facial expression synthesis to handle the problem of disambiguating changes. Moreover, AAE also shows impressive success in learning bilingual dictionary [43], and is more robust and achieves better performance than nonadversarial approaches. However, in the realm of HSI, especially in target detection, research works based on AAE are still inadequate. In this article, a network based on AAE for hyperspectral target detection is developed, and it has been proved that our model significantly outperforms other hyperspectral target detection approaches.

III. PROPOSED METHOD

A semisupervised AAE, subject to a target suppression constraint, is proposed in this work for target detection. Due to scarce target samples, our method resorts to background learning, providing a new perspective to target detection. As depicted in Fig. 1, the proposed background learning based on target suppression constraint (BLTSC) method consists of two main
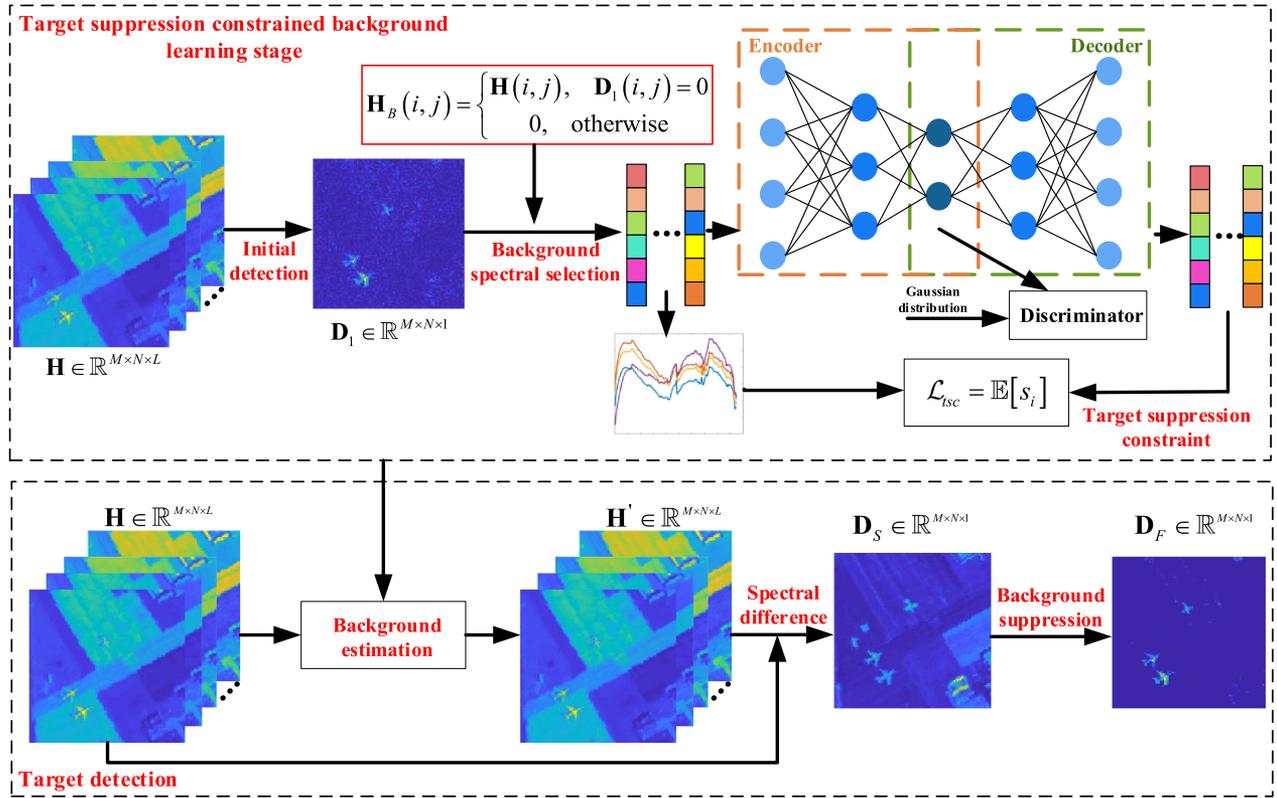
Fig. 1.   Schematic of the proposed BLTSC-based target detection method.

operations. First, the network is trained based on BLTSC to obtain unique background representation. Then, pinpoint targets with the trained model. The BLTSC is adversarially trained on background samples only to reconstruct background as accurately as possible. Let the background training set be denoted as $\mathbf{H}_B$, which can be represented as $\mathbf{H}_B = [\mathrm{h}_1 \mathrm{h}_2 \ldots \mathrm{h}_n]$. After training the network with abundant background samples, our approach can approximate background distribution. Then we test the trained model with an HSI, denoted by $\mathbf{H} \in \mathbb{R}^{M \times N \times L}$, and yield a discriminative reconstruction. The HSI can be divided into background spectra set $\mathbf{B}$ and target spectra set $\mathbf{T}$

$$\mathbf{B} = \left[\mathbf{b}_1^1, \ldots, \mathbf{b}_1^L, \mathbf{b}_2^1, \ldots, \mathbf{b}_2^L, \mathbf{b}_p^1, \ldots, \mathbf{b}_p^L \right]$$

$$\mathbf{T} = \left[\mathbf{t}_1^1, \ldots, \mathbf{t}_1^L, \mathbf{t}_2^1, \ldots, \mathbf{t}_2^L, \mathbf{t}_q^1, \ldots, \mathbf{y}_q^q \right]$$

where $p + q = M \times N$. Similarly, the reconstructed HSI $\mathbf{H}' \in \mathbb{R}^{M \times N \times L}$ can also be divided into two components, reconstructed background $\mathbf{B}'$ and reconstructed target $\mathbf{T}'$

$$\mathbf{B}' = \left[\mathbf{b}_1'^1, \ldots, \mathbf{b}_1'^L, \mathbf{b}_2'^1, \ldots, \mathbf{b}_2'^L, \mathbf{b}_p'^1, \ldots, \mathbf{b}_p'^L \right]$$

$$\mathbf{T}' = \left[\mathbf{t}_1'^1, \ldots, \mathbf{t}_1'^L, \mathbf{t}_2'^1, \ldots, \mathbf{t}_2'^L, \mathbf{t}_q'^1, \ldots, \mathbf{y}_q'^L \right].$$

The trained model concentrates on reconstructing background, therefore, discriminative reconstruction is performed on the background and target samples. In our approach, Leaky Relu acts as the activation function

$$f(x) = \begin{cases} x, & \text{if } x > 0 \\ \lambda x, & \text{if } x \le 0 \end{cases} \tag{3}$$

where $\lambda \sim U(l, u), l < u$ and $u \in [0, 1)$. The background spectrum $\mathbf{b}_i$ and target spectrum $\mathbf{t}_i$ are considered as input, whereas $\mathbf{b}_i'$ and $\mathbf{t}_i'$ are the output. $\mathbf{l}_i^b$ denotes the hidden representation of the background spectrum $\mathbf{b}_i$. Hence, the encoding and decoding process for background samples can be expressed as

$$\mathbf{l}_i^b = f\left(\mathbf{W}_h \mathbf{b}_i + \mathbf{a}_h\right)$$
$$\mathbf{b}_i' = f\left(\mathbf{W}_o \mathbf{l}_i^b + \mathbf{a}_o\right) \tag{4}$$

respectively. Similarly, for targets, the hidden representation of $\mathbf{t}_i$ is denoted by $\mathbf{l}_i^t$. Thus, the encoding and decoding process become

$$\mathbf{l}_i^t = f\left(\mathbf{W}_h \mathbf{t}_i + \mathbf{a}_h\right)$$
$$\mathbf{t}_i' = f\left(\mathbf{W}_o \mathbf{l}_i^t + \mathbf{a}_o\right) \tag{5}$$

respectively, where $\mathbf{W}_h$ and $\mathbf{a}_h$ denote the weight matrix and bias term of encoding layers, respectively, while $\mathbf{W}_o$ and $\mathbf{a}_o$ represent the weight matrix and bias term of decoding layers, respectively. With the optimization goal mentioned in (13), thousands of epochs are conducted until AAE converges. Then the parameters are fixed and used to reconstruct $\mathbf{H}'$.

The network, only trained on background samples, is excepted to fail to reconstruct target samples as it is never trained on such examples. As illustrated in Fig. 2, targets in HSIs are not well recovered, while background pixels are reconstructed with less reconstruction error, which indicates the model enhances discrimination between target and background with background learning. Additionally, a target suppression constraint is imposed
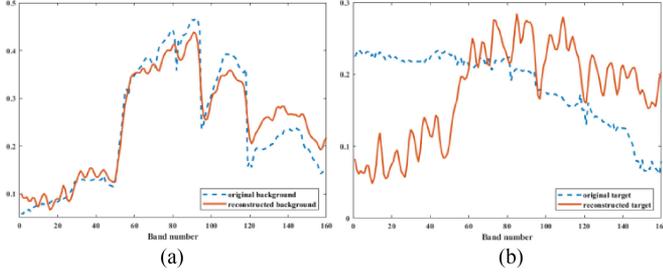
Fig. 2. Reconstructions of background and target for HYDICE dataset. (a) Background. (b) Target.

on the training procedure to approximate background distribution as closely as possible, as mentioned in Section III-B2. In contrast to the original HSI, the reconstructed one has an extremely similar background and more distinguishing target, which indicates that our model obtains a discriminative feature by learning the distribution of the training set. Then $\mathbf{H}$ and $\mathbf{H}'$ are compared pixel by pixel. The output $\mathbf{D}_S$ can be described as $\mathbf{D}_S = [\mathbf{d}_{s1}, \mathbf{d}_{s2}, \cdots \mathbf{d}_{sM \times N}]$, where $\mathbf{d}_{si}$ is the spectral distance between the $\mathbf{i}_{th}$ pixel in $\mathbf{H}$ and $\mathbf{H}'$, calculated by

$$\mathbf{d}_{si} = \cos^{-1}\left(\frac{\mathbf{h}_i^T \mathbf{h}_i'}{(\|\mathbf{h}_i\|)\,(\|\mathbf{h}_i'\|)}\right)$$

$$= \cos^{-1}\left(\frac{\mathbf{h}_i^T \mathbf{h}_i'}{\sqrt{(\mathbf{h}_i')^T \mathbf{h}_i'}\sqrt{\mathbf{h}_i^T \mathbf{h}_i}}\right). \quad (6)$$

Then, a distance map, containing distinguishable targets, is obtained.

Since it is the key point to acquire the BLTSC-based network, we address the problem of obtaining background training set in Section III-A, and describe our training objective detailedly in Section III-B.

A. Background Spectral Selection

A DNN architecture has shown to be highly capable of target detection based on massive volumes of training data. Targets are rare entities, hence it is challenging to obtain their labels. Instead, background samples are far more easy to acquire. With sufficient training samples, the model is capable to reconstruct background as realistically as possible and is conversely expected to have poor performance on target samples. During the training procedure, target suppression constraint is also imposed to aid AAE to learn more powerful background representation.

As noted, the motivation is to maximize the background reconstruction capability. The training set is only composed of background samples. In this work, the CEM is leveraged as a coarse detection approach to obtain the initial training set. Let the original HSI with L bands be denoted as $\mathbf{H} \in \mathbb{R}^{M \times N \times L}$, and the response of CEM filter is given by

$$\mathbf{y} = (\mathbf{w}^*)^T \mathbf{H}$$

$$= \frac{\mathbf{d}^T \mathbf{R}^{-1}}{\mathbf{d}\mathbf{R}^{-1}\mathbf{d}^T}\mathbf{H} \quad (7)$$

where $\mathbf{R} = (1/N)\mathbf{H}\mathbf{H}^T$ represents the correlation matrix and $\mathbf{d}$ is a prior known target spectrum. By analyzing $\mathbf{D}_1 \in \mathbb{R}^{M \times N \times 1}$ reshaped from $\mathbf{y}$, we can locate background pixels roughly.

In the normalized initial detection result $\mathbf{D}_1$, the values of background and target incline to 0 and 1, respectively. To reject targets, a smaller threshold ϵ is configured to binary $\mathbf{D}_1$ while selecting background training samples by

$$\mathbf{H}_D(i,j) = \begin{cases} \mathbf{H}(i,j), & \mathbf{D}_1(i,j) = 0 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where $\mathbf{H}(i,j)$ represents the spectral vector at position (i,j). $\mathbf{H}_D$ is the collection of background samples selected from $\mathbf{H}$ according to $\mathbf{D}_1$. The background training set $\mathbf{H}_B = [\mathrm{h}_1 \mathrm{h}_2 \ldots \mathrm{h}_n]$ randomly samples 75% from $\mathbf{H}_D$.

B. Background Learning Based on Target Suppression Constraint

As mentioned above, the BLTSC aims to learn accurate background distribution. To achieve this, we propose to combine three losses (adversarial, target suppression constraint, autoencoder), each of which contributes to better reconstruction. First, an adversarial loss for matching the distribution of the latent space with the prior distribution. Second, a target suppression constraint loss is adopted to ensure the network is only trained on background samples. Third, we evaluate performance with an autoencoder loss, which calculates the deviation between input samples and reconstructed output.

1) Adversarial Loss: AAE is known as a variant of GANs, which trains model in the adversarial regularization. As shown in (2), AAE samples from the prior distribution to induce $q(\mathbf{z})$ to match $p(\mathbf{z})$. Different distributions of $q(\mathbf{z})$ will result in different kinds of models for various tasks. Each feature of $\mathbf{z}$ is excepted to be independent of each other and obey the known prior distribution $p(\mathbf{z})$. Inspired by existing anomaly detection models, we choose Gaussian distribution as the prior distribution. Two components of AAE, discriminator D and generator G, are jointly trained. D aims to distinguish the true samples from the fake ones generated by the generator. Concurrently, G is trained to confuse D with its generated samples. G tries to minimize the loss in (2), and conversely, D is in an effort to maximize it. Ultimately, $\min_G \max_D \mathcal{L}_{\text{adv}}$ is achieved, formulated as

$$\mathcal{L}_{\text{adv}} = \mathbb{E}[\log D(\mathcal{N}(0,1))] + \mathbb{E}[\log(1 - D(z)] \quad (9)$$

where z is the latent representation of the input. With the minimization of $\mathcal{L}_{\text{adv}}$, an accurate generative reproduction of background distribution is guaranteed.

2) Target Suppression Constraint Loss: It is inevitable that the background samples obtained from coarse detection are ambiguous. To reconstruct the background sufficiently and reliably, the target suppression constraint loss is leveraged. This loss act as a criterion to eliminate the threat of potential target spectrum during the training phase. The spectral angle mapper $\mathbf{s}_i$ for the every reconstructed background spectrum $\mathbf{h}_i'$ and $\mathbf{d}$, is

---

Algorithm 1: Target Detection by the Proposed BLTSC.

Input: HSI $\mathbf{H}$, normalized initial detection result $\mathbf{D}_1$, prior known target spectrum $\mathbf{d}$, parameter λ;
Step:
1: obtain background training set $\mathbf{H}_B$ from $\mathbf{H}$ with (8)
2: establish the background learning model
for each epoch **do**
    minimize adversarial loss $\mathcal{L}_{\text{adv}}$ in (9)
    minimize target suppression constraint loss $\mathcal{L}_{\text{tsc}}$ in (11)
    minimize autoencoder loss $\mathcal{L}_{\text{error}}$ in (12)
end
3: reconstruct $\mathbf{H}'$ from $\mathbf{H}$ with the trained model
4: calculate the distance map $\mathbf{D}_S$ with (6)
5: detect target from $\mathbf{D}_S$ and $\mathbf{D}_1$ with (14)
Output: final detection result $\mathbf{D}_F$.

---

calculated, resulting in $\mathbf{S} = [\mathbf{s}_1 \mathbf{s}_2 \ldots \mathbf{s}_n]$, where $\mathbf{s}_i$ is defined as

$$\mathbf{s}_i = \cos^{-1}\left(\frac{\mathbf{h}_i'^T \mathbf{d}}{(\|\mathbf{h}_i'\|)(\|\mathbf{d}\|)} \right)$$

$$= \cos^{-1}\left(\frac{\mathbf{h}_i'^T \mathbf{d}}{\sqrt{(\mathbf{d})^T \mathbf{d}}\sqrt{\mathbf{h}_i'^T \mathbf{h}_i'}} \right). \tag{10}$$

A smaller $\mathbf{s}_i$ indicates that the $\mathbf{h}_i'$ is most likely reconstructed by a target sample, i.e., in all probability, the $\mathbf{h}_i$ is a target sample. The mean of $\mathbf{s}_i$ which smaller than preset threshold δ should work as the target suppression constraint loss, i.e.,

$$\mathcal{L}_{\text{tsc}} = \mathbb{E}[\mathbf{s}_i], \quad \text{where} \quad s_i < \delta. \tag{11}$$

3) Autoencoder Loss: Additionally, we define autoencoder loss to minimize the reconstruction error between input data $\mathbf{H}_B$ and reconstructed data $\mathbf{H}_B'$, encouraging encoder and decoder to approximate the real distribution, as

$$\mathcal{L}_{\text{error}} = \sum_i \left\| \mathbf{h}_i - \mathbf{h}_i' \right\|_2 \tag{12}$$

where $\mathbf{h}_i'$ is the reconstructed spectrum corresponding to the input spectrum $\mathbf{h}_i$.

Taken together, the total loss is formulated as

$$\mathcal{L} = \mathcal{L}_{\text{adv}} - \mathcal{L}_{\text{tsc}} + \mathcal{L}_{\text{error}}. \tag{13}$$

These three losses perform their respective functions and work together to optimize the model.

C. Target Detection

To finally detect targets, each pixel value $\mathbf{d}_{si}$ is transformed by multiplying a nonnegative number $q(y_i)$ based on output of the coarse detection $\mathbf{y}$ as

$$\mathbf{d}_{fi} = q(y_i)\,\mathbf{d}_{si}. \tag{14}$$

The nonlinear function $q(x)$ is used to suppress the undesired background spectra while keeping the target spectra unchanged. Considering that the target pixels in the distance map are much greater than the background, $q(x)$ should be able to maintain

large values and remove small ones. Here, the nonlinear suppression function is defined as

$$q(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{15}$$

where λ is a positive parameter to adjust the suppression performance. Finally, the detection result $\mathbf{D}_F = [\mathbf{d}_{f1}, \mathbf{d}_{f2}, \cdots \mathbf{d}_{fM \times N}]$ is obtained.

In summary, the BLTSC method is implemented by following steps. First, coarse detection is utilized to obtain initial training data. Second, a background learning-based AAE network is trained to optimize the three losses mentioned in Section III-B. Then the trained model is tested by the entire HSI, and the final target detection is achieved after applying a nonlinear scaling factor. The entire algorithm of BLTSC is summarized in Algorithm 1.

IV. EXPERIMENTS

In this section, experiments are conducted to validate the effectiveness of the proposed method. First, we analysis the effect of parameters to obtain an optimal model. Besides, the component analysis is carried out to verify that both the target suppression constraint and adversarial learning have positive influence on the performance of target detection. Finally, the final performance of the proposed method is evaluated and compared with four typical target detectors: CEM, hCEM, CSCR, and ACE, which are either frequently cited in the literature or have leading performances on several real hyperspectral datasets.

The performance of the proposed BLTSC model is evaluated by the most widely used metrics, i.e., the receiver operating characteristic (ROC), and the area under the curve (AUC) metric [44]. Here, two types of ROC curves are plotted with three parameters: true positive rate (P_D), false positive rate (P_F), and threshold (τ). The ROC curve of (P_D, P_F) describes the tradeoff between P_D and P_F, and the closer the curve is to the upper left corner, the higher the AUC value. The ROC curve of (P_F, τ) illustrates the false alarm rate of the target detection. In general, a larger AUC of (P_D, P_F) and a smaller AUC of (P_F, τ) indicates good detection performance, i.e., more targets are pinpointed with less error.

A. Datasets Description

Our experiments are performed on five datasets, including six HSIs captured over different scenes, and the datasets are described detailedly as follows.

1) HYDICE Dataset: The first dataset was collected by the hyperspectral digital imagery collection experiment (HYDICE) airborne sensor over an urban area, CA, USA. This urban scene consists of 80×100 pixels, with 175 spectral channels in wavelengths ranging from 400 to 2500 nm. The image has a spatial resolution of 1 m. The scene mainly consists of a vegetation area, a construction area, and several roads including some cars. The man-made objects, i.e., cars and roofs which occupy 19 pixels are regarded as targets.

2) San Diego Dataset: The second hyperspectral dataset was acquired by the AVIRIS sensor, covering the San Diego airport
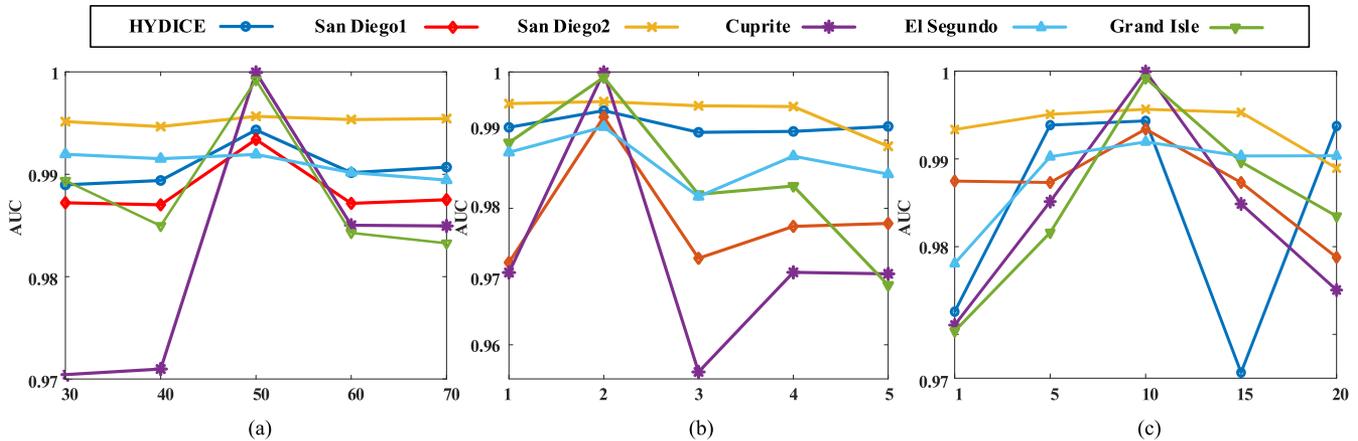
Fig. 3. Parameter sensitivity analysis of influence on different datasets. (a) Number of hidden nodes. (b) Depth. (c) λ.

area, CA, USA. The spatial size is of $text100 \times 100$ pixels and it includes 224 spectral channels in wavelengths ranging from 370 to 2510 nm. The spatial resolution is approximately 3.5 m. In the experiments, a total of 189 bands are used after removing water-absorption and low SNR bands. This dataset includes two images, the one with three airplanes which occupies 57 pixels in the upper right corner is represented by San Diego1 and the other one with three airplanes occupying 134 pixels in the center is represented by San Diego2. The airplanes are considered as targets of interest.

3) Cuprite Dataset: The third hyperspectral dataset was obtained by the AVIRIS sensor, in the Cuprite mining district of Nevada in 1997, while the corresponding ground truth was produced by Tricorder software in 1995. There are about 14 kinds of mineral in this image, including buddingtonite, Na-Montmorillonite, Nontronite (Fe clay), Kaolinite, etc. We use a 250×191 pixel subset of this image to conduct our experiment, which is marked by the red box in this figure. The buddingtonite is selected as the target, which occupies 41 pixels. After removing the low SNR and water absorption bands, 188 bands are left to conduct our experiment.

4) El Segundo Dataset: The forth dataset was captured by the AVIRIS sensor, which has 224 spectral channels in wavelengths ranging from 366 to 2496 nm. This urban scene covers an area of El Segundo, CA, USA, with the spatial size of 250×300. Each pixel has 7.1 m of ground resolution. The image data set is mainly composed of an area of oil refinery, several residential areas, parks, and one school zone. The constructions of oil refinery such as storage tanks and the towers occupy 2048 pixels and are considered as targets.

5) Grand Isle Dataset: The fifth dataset is also an AVIRIS image, which is captured at the location of Grand Isle on the Gulf Coast, part of Jefferson Parish, LA, USA. The dataset consists of 300×480 pixels with 224 spectral channels in wavelengths ranging from 366 to 2496 nm. The spatial resolution is approximately 4.4 m. The main background materials are island and water in the scene. Those man-made objects occupying 279 pixels in the water are selected as the targets to be detected.

TABLE I
PARAMETER SETTINGS OF THE PROPOSED METHOD

Parameters	Number of hidden nodes	Depth	λ	ϵ	δ
Value	50	2	10	0.15	20th $\mathbf{s}_i$

B. Parameter Sensitivity Analysis

Latent code, an effective feature learned by the network, determines the quality of reconstruction. The depth of the network and the number of hidden nodes in the deepest layer are crucial hyperparameters for feature extraction. The more layers and nodes there are, the more accurate information the latent code embodies. However, increasing layers and nodes leads to long training time and overfitting. Besides, the parameter λ in target detection should be determined correctly which enables that the nonlinear transformation will not hurt any target spectra while suppressing background. In this section, different parameters are set to estimate an optimal network structure. When the impact of the number of hidden nodes is analyzed, the depth is set to 2, and the number of nodes for the first hidden layer is set to 200 to obtain sufficient feature information. As depicted in Fig. 3, when the number of hidden nodes is 50, our proposed network can achieve first-rank detection performance. Then, we optimized the depth under the precondition of the aforementioned. It can be observed that 2 is the best depth.

With the fixed hyperparameters, the effect of λ is analyzed. When λ is set to 10, our method has the best detection capacity. The threshold ϵ mentioned in Section III-A is set as 0.15 for all the experimental datasets. Additionally, we sort $\mathbf{s}_i$ and chose the 20th as the preset threshold δ in Section III-B2. The last two parameters are default in our experiments and they can be tuned by user for the optimal results. Our parameter settings are shown in Table I.

C. Component Analysis

1) Effectiveness Evaluation of Target Suppression Constraint: To further validate the influence of the target suppression constraint, a model without it is tested on all the datasets.
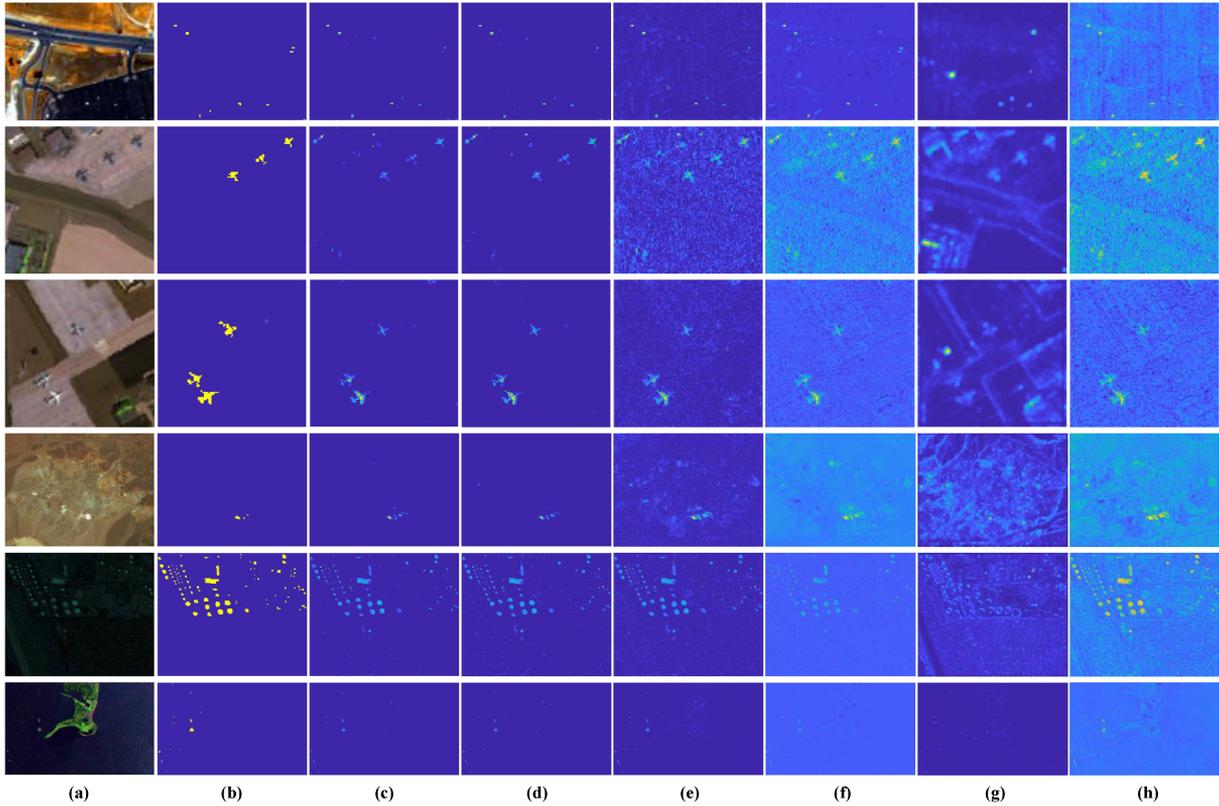
Fig. 4. Detection maps of different methods. (a) Color composites of HYDICE, two HSIs from San Diego, Cuprite, El Segundo, and Grand Isle datasets. (b) Reference map. (c) BLTSC-AAE. (d) BLTSC-SAE. (e) CEM. (f) hCEM. (g) CSCR. (h) ACE.

TABLE II
EVALUATION AUC SCORES OF (P_D, P_F) AND (P_F, τ) FOR THE MAIN COMPONENT (TSC) OF OUR METHOD

HSIs	(P_D, P_F)		(P_F, τ)	
	TSC(Y)	TSC(N)	TSC(Y)	TSC(N)
HYDICE	**0.99433**	0.98435	**0.00001**	0.00003
San Diego1	**0.99340**	0.98763	**0.00036**	0.00076
San Diego2	**0.99566**	0.99467	**0.00061**	0.00149
Cuprite	**0.99995**	0.99986	**0.00014**	0.00023
El Segundo	**0.99193**	0.98936	**0.00800**	0.01594
Grand Isle	**0.99921**	0.98365	**0.00066**	0.00008
Average	**0.99575**	0.98992	**0.00163**	0.00309

The bold entities represent the best performance in each row.

TABLE III
EVALUATION AUC SCORES OF (P_D, P_F) AND (P_F, τ) FOR THE MAIN COMPONENT (ADVERSARIAL LEARNING) OF OUR METHOD

HSIs	(P_D, P_F)		(P_F, τ)	
	AAE	SAE	AAE	SAE
HYDICE	0.99433	**0.99837**	**0.00001**	0.00008
San Diego1	**0.99340**	0.98831	**0.00036**	0.00040
San Diego2	**0.99566**	0.99518	**0.00061**	0.00088
Cuprite	**0.99995**	0.99992	**0.00014**	0.00027
El Segundo	**0.99193**	0.99124	**0.00800**	0.01400
Grand Isle	**0.99921**	0.99919	**0.00066**	0.00081
Average	**0.99575**	0.99543	**0.00163**	0.00274

The bold entities represent the best performance in each row.

The result with the constraint are shown in Table II. The TSC is used to denote this constraint. Y and N in parentheses indicate whether the constraint is employed, respectively. Apparently, the model with TSC is better than the model without TSC in reducing the false alarm rate as well as increasing the detection rate.

2) Effectiveness Evaluation of Adversarial Learning: To figure out how adversarial learning contributes to the reconstruction, a network based on SAE was analyzed. The third and fourth columns in Fig. 4 show BLTSC-AAE and BLTSC-SAE have similar visual inspection effect. However, as reported in Table III, BLTSC-AAE is capable to achieve better detection accuracy on all the datasets except for HYDICE dataset. Although

the AUC value of BLTSC-AAE on HYDICE is 0.99433, down 0.004 from BLTSC-SAE, BLTSC-AAE shows better performance in terms of false alarm rate. Taken together, it can be seen from the average value that BLTSC-AAE behaves better than BLTSC-SAE in representing characteristics and reconstructing spectra, which is due to adversarial learning.

D. Detection Result

In this section, the performance of the proposed BLTSC-based hyperspectral target detection is evaluated and compared with four widely used target detection algorithms, i.e., CEM, hCEM, CSCR, and ACE.
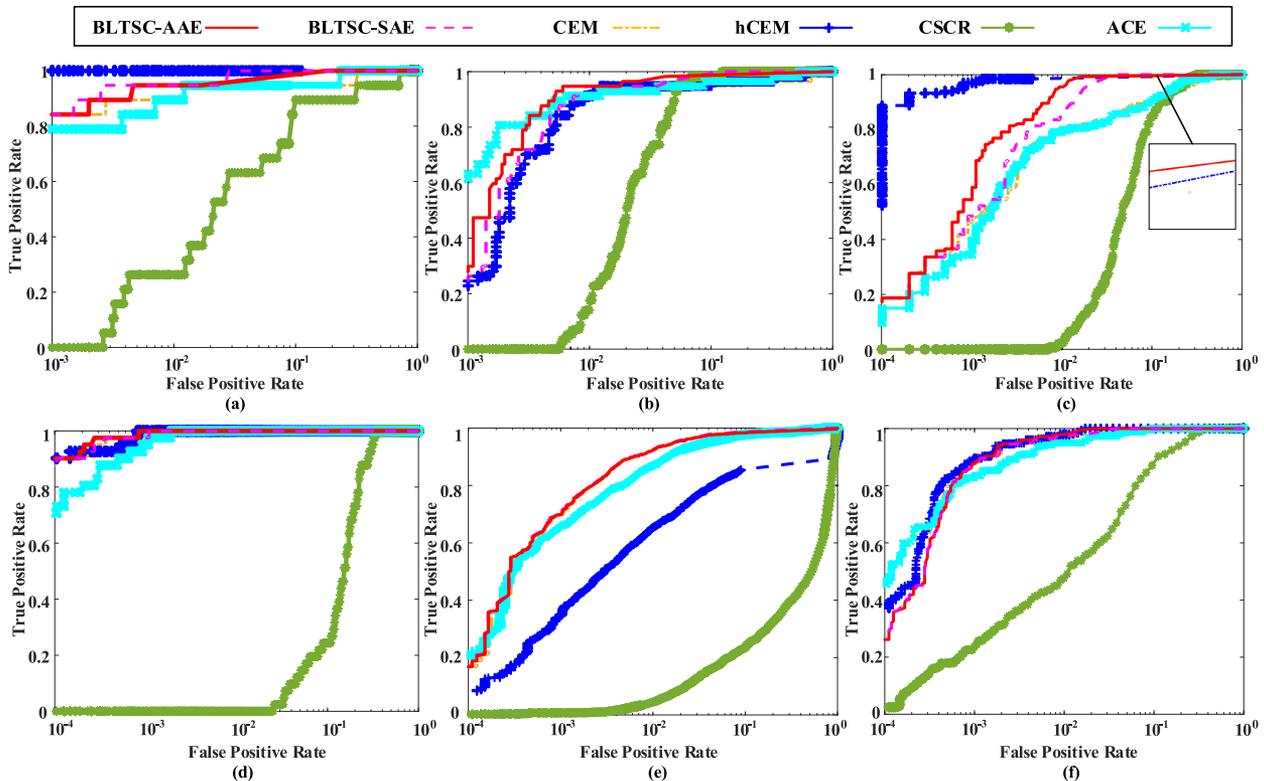
Fig. 5. ROC curves of (P_D, P_F) for the compared methods on the (a) HYDICE, (b) San Diego1, (c) San Diego2, (d) Cuprite, (e) El Segundo, and (f) Grand Isle.

For the datasets mentioned above, the reference maps, i.e., ground truths and detection results are shown in Fig. 4. Obviously, the proposed BLTSC-AAE outperforms according to visual inspection. Besides, our proposed method can adapt to any targets with different sizes. For the typical detection method such as CEM, it is challenging to detect targets precisely while suppressing background. The hCEM is superior to CEM in suppressing background, however, the targets are far smaller and less than the background, which means some targets are easily rejected as uninterested information. CSCR and ACE cannot perform well for the targets containing structural information, and the background noise is unacceptable. On the contrary, the proposed BLTSC provides a more robust detection performance by expanding the difference between background and targets.

For detailed qualitative analysis of detection performance, the log-scale ROC curves of (P_D, P_F) and (P_F, τ) are plotted. As illustrated in Fig. 5, except for the HYDICE, Grand Isle datasets, and San Diego2, our method remains over CEM, hCEM, CSCR, ACE, and BLTSC-SAE. Besides, Fig. 6 shows that the curves of (P_F, τ) of BLTSC-AAE are under other methods for all datasets but San Diego2. These confirm that our method exhibits better detection performance while reducing false alarm rates. The AUC scores of (P_D, P_F) and (P_F, τ) are exploited to offer quantitative analysis. Notably, although the proposed BLTSC-AAE fails to achieve the best intuitive ROC performance on San Diego2, the nonlinearity of abscissa leads to numerical superiority, as seen in the local enlarged drawings of Figs. 5(c) and 6(c). Moreover, we take El Segundo and Grand Isle for example and draw the target-background separability map. The

TABLE IV
EVALUATION AUC VALUES OF THE COMPARED METHODS FOR
EXPERIMENTAL DATASETS

HSIs	AUC values of (P_D, P_F)					
	BLTSC-AAE	BLTSC-SAE	CEM	hCEM	CSCR	ACE
HYDICE	0.99433	0.99837	0.98243	**0.99976**	0.92561	0.98656
San Diego1	**0.99340**	0.98831	0.96837	0.97915	0.92424	0.98177
San Diego2	**0.99566**	0.99518	0.95871	0.99241	0.86947	0.97327
Cuprite	**0.99995**	0.99992	0.99360	0.99986	0.70987	0.99982
El Segundo	**0.99193**	0.99124	0.99155	0.87147	0.71768	0.98726
Grand Isle	**0.99921**	0.99919	0.99881	0.99909	0.98363	0.99755
Average	**0.99575**	0.99543	0.98233	0.97362	0.85512	0.98770
HSIs	AUC values of (P_F, τ)					
	BLTSC-AAE	BLTSC-SAE	CEM	hCEM	CSCR	ACE
HYDICE	**0.00001**	0.00008	0.03010	0.00045	0.07600	0.24164
San Diego1	**0.00036**	0.00040	0.03570	0.27753	0.09832	0.34175
San Diego2	**0.00061**	0.00088	0.05881	0.00103	0.11221	0.19438
Cuprite	**0.00014**	0.00027	0.00124	0.29330	0.27443	0.33809
El Segundo	**0.00800**	0.01400	0.06732	0.17532	0.01607	0.26885
Grand Isle	**0.00066**	0.00081	0.01883	0.17575	0.01369	0.22916
Average	**0.00163**	0.00274	0.03718	0.15390	0.09845	0.26898

The bold entities represent the best performance in each row.

larger distance between the lower bound of the red box and the upper bound of the green box indicates better target-background separability. As shown in Fig. 7, the proposed method behaves better than other methods except for ACE. It is worth noting that BLTSC has the flattest and lowest green box, which indicates the powerful background suppression ability.

As reported in Tables IV, the AUC scores of (P_D, P_F) and (P_F, τ) of the proposed BLTSC-AAE method are approximate to
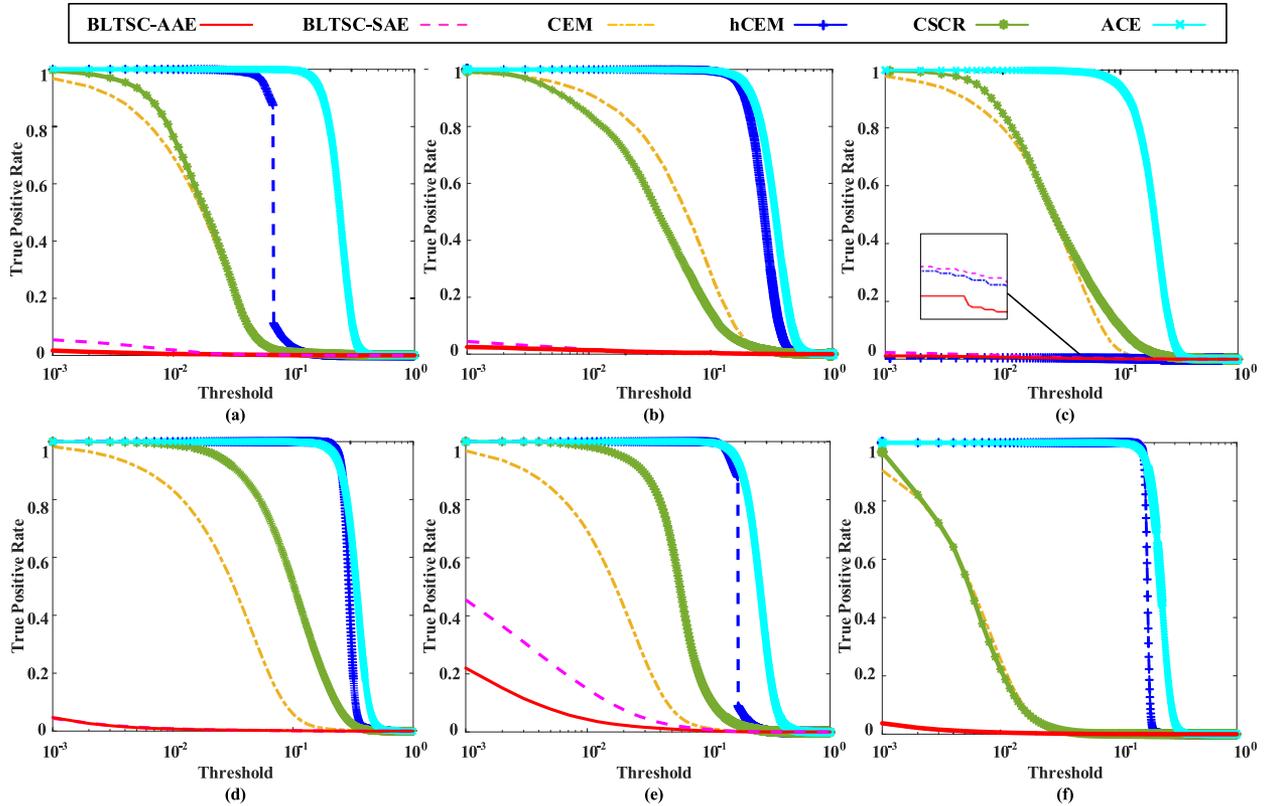
Fig. 6. ROC curves of (P_F, τ) for the compared methods on the (a) HYDICE, (b) San Diego1, (c) San Diego2, (d) Cuprite, (e) El Segundo, and (f) Grand Isle.
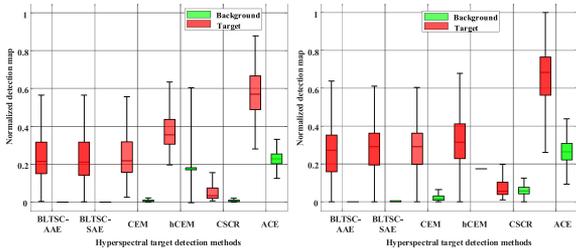

Fig. 7. Target-background separability map for the compared methods on El Segundo and Grand Isle.

TABLE V
DETECTION PERFORMANCE FOR THE COMPARED METHODS WITH NONLINEAR SUPPRESSION

HSIs	AUC values of (P_D, P_F)					
	BLTSC-AAE	BLTSC-SAE	CEM	hCEM	CSCR	ACE
HYDICE	0.99433	0.99837	0.97230	**0.99920**	0.97130	0.98684
San Diego1	**0.99340**	0.98831	0.97212	0.98081	0.97514	0.98177
San Diego2	**0.99566**	0.99518	0.96780	0.98863	0.97778	0.97730
Cuprite	**0.99995**	0.99992	0.99937	0.99935	0.98783	0.99982
El Segundo	**0.99193**	0.99124	0.99039	0.87190	0.99019	0.98659
Grand Isle	**0.99921**	0.99919	0.98848	0.99920	0.99592	0.99288
Average	**0.99575**	0.99543	0.98174	0.97318	0.98303	0.98753

The bold entities represent the best performance in each row.

TABLE VI
AVERAGE COMPUTING TIME (IN SECONDS) OF THE COMPARED METHODS

HSIs	Computing Time (in seconds)					
	BLTSC-AAE	BLTSC-SAE	CEM	hCEM	CSCR	ACE
HYDICE	0.07123	0.09811	**0.04357**	1.32478	3.87727	0.11514
San Diego1	0.08009	0.08789	**0.06001**	1.66199	6.07100	0.19633
San Diego2	0.08238	0.08584	**0.04543**	1.48438	5.06353	0.09634
Cuprite	0.37436	0.27172	**0.18829**	3.49000	22.85086	0.41291
El Segundo	0.64658	0.85098	**0.35202**	6.80602	39.20438	0.81599
Grand Isle	1.22301	1.19254	**0.63296**	9.19997	93.44258	1.51130
Average	0.41294	0.43118	**0.22038**	3.99453	28.41827	0.52467

The bold entities represent the best performance in each row.

ideal values 1 and 0 for all the scenes. The average AUC scores of (P_D, P_F) and (P_F, τ) are 0.99575 and 0.00163, and they are superior to other methods numerically. Besides, to evaluate the compared methods more scientifically and rigorously, the nonlinear suppression mentioned in Section III-C is imposed on all the compared methods. As shown in Table V, our method remains superior and there is no significant performance improvement after nonlinear process, which validates the effectiveness of the proposed model.

Additionally, the computing time of different methods is presented in Table VI. All the experiments are carried out on an Inter (R) Core (TM) i5-8400 CPU machine with 32 GB of RAM. Our proposed method is performed on a system running Python 3.7.6 and TensorFlow 1.14.0, while all compared methods are executed with MATLAB R2018b. Due to the introduction of

network, it takes a few minutes to model the background. Therefore, we measure the running time of detection after the training stage, and the results are recorded in seconds. As shown

TABLE VII
TIME COMPLEXITY FOR THE COMPARED METHODS

Compared methods	Time complexity
BLTSC-AAE	$O(M \times N \times L \times b)$
BLTSC-SAE	$O(M \times N \times L \times b)$
CEM	$O(M \times N \times L^2)$
hCEM	$O(iter \times M \times N \times L^2)$
CSCR	$O(M \times N \times (\omega_{out}^2 - \omega_{in}^2)^2 \times L)$
ACE	$O(M \times N \times L^2)$

in Table VI, the computational complexity of the two BLTSC models is almost the same, which is less than hCEM, CSCR, and ACE. CEM takes less time, but the detection performance is not as good as BLTSC. BLTSC achieves an outstanding detection performance at the cost of tolerable time. More formally, the time complexity for all methods in comparison are calculated and listed in Table VII. The image fed into all methods in comparison is of size $M \times N \times L$. b denotes max hidden nodes number in BLTSC. $iter$, the number of layers in hCEM, is equal to 10 on average. $(\omega_{out}, \omega_{in})$ frames a sliding dual-windows in CSCR, and is set as $(11, 3)$. The time complexity of BLTSC, CEM, and ACE is in the same level while hCEM achieves better performance than CEM at the cost of time. The time complexity of CSCR is more than 60 times that of BLTSC. All the above results tallies with the computing time of the compared methods listed in Table VI.

V. CONCLUSION

In this article, we propose a novel semisupervised AAE model based on background learning named BLTSC for HSIs. It is composed of three complementary workflows: background spectral selection, BLTSC, and testing with the original HSI. In background spectral selection, we exploit the CEM to obtain initial background training data. To retrieve pure background data, target suppression constraint loss is imposed on the model to reduce the impact of suspected target samples on background reconstruction. Then, adversarial loss, target suppression constraint loss, and autoencoder loss are jointly optimized to approximate background distribution as accurately as possible. Finally, all the spectral samples are fed into the background estimation model. Experimental results on five real hyperspectral datasets show that the proposed BLTSC is superior to the other target detection methods. Its superior performance validates the importance of semisupervised learning in hyperspectral target detection.

REFERENCES

[1] X. Hadoux *et al.*, "Non-invasive in vivo hyperspectral imaging of the retina for potential biomarker use in Alzheimer's disease," *Nature Commun.*, vol. 10, no. 1, pp. 1–12, Sep. 2019.

[2] A. L. A. Gawad, Y. El-Sharkawy, H. Ayoub, A. F. El-Sherif, and M. F. Hassan, "Classification of dental diseases using hyperspectral imaging and laser induced fluorescence," *Photodiagnosis Photodyn. Therapy*, vol. 25, pp. 128–135, 2019.

[3] H. Fabelo *et al.*, "*In-vivo* hyperspectral human brain image database for brain cancer detection," *IEEE Access*, vol. 7, pp. 39 098–39 116, 2019.

[4] O. Coast *et al.*, "Predicting dark respiration rates of wheat leaves from hyperspectral reflectance," *Plant Cell Environ.*, vol. 42, no. 7, pp. 2133–2150, 2019.

[5] K. Sendin, P. J. Williams, and M. Manley, "Near infrared hyperspectral imaging in quality and safety evaluation of cereals," *Crit. Rev. Food Sci. Nutrition*, vol. 58, no. 4, pp. 575–590, 2018.

[6] F. Wang, F. Wang, Y. Zhang, J. Hu, J. Huang, and J. Xie, "Rice yield estimation using parcel-level relative spectral variables from UAV-based hyperspectral imagery," *Frontiers Plant Sci.*, vol. 10, 2019, Art. no. 453.

[7] G. A. Fricker, J. D. Ventura, J. A. Wolf, M. P. North, F. W. Davis, and J. Franklin, "A convolutional neural network classifier identifies tree species in mixed-conifer forest from hyperspectral imagery," *Remote Sens.*, vol. 11, no. 19, 2019, Art. no. 2326.

[8] M. Imangholiloo *et al.*, "Characterizing seedling stands using leaf-off and leaf-on photogrammetric point clouds and hyperspectral imagery acquired from unmanned aerial vehicle," *Forests*, vol. 10, no. 5, 2019, Art. no. 415.

[9] D. M. Varade, A. K. Maurya, and O. Dikshit, "Development of spectral indexes in hyperspectral imagery for land cover assessment," *IETE Tech. Rev.*, vol. 36, no. 5, pp. 475–483, 2019.

[10] Y. Shi, J. Lei, Y. Yin, K. Cao, Y. Li, and C.-I. Chang, "Discriminative feature learning with distance constrained stacked sparse autoencoder for hyperspectral target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1462–1466, Sep. 2019.

[11] D. Zhu, B. Du, and L. Zhang, "Target dictionary construction-based sparse representation hyperspectral target detection methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1254–1264, Apr. 2019.

[12] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.

[13] D. Zhu, B. Du, and L. Zhang, "Binary-class collaborative representation for target detection in hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1100–1104, Jul. 2019.

[14] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene River Valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 64–76, 1997.

[15] R. Zhao, Z. Shi, Z. Zou, and Z. Zhang, "Ensemble-based cascaded constrained energy minimization for hyperspectral target detection," *Remote Sens.*, vol. 11, no. 11, 2019, Art. no. 1310.

[16] S.-Y. Chen, C. Lin, C.-H. Tai, and S.-J. Chuang, "Adaptive window-based constrained energy minimization for detection of newly grown tree leaves," *Remote Sens.*, vol. 10, no. 1, 2018, Art. no. 96.

[17] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 626–637, Apr. 2013.

[18] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.

[19] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[20] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, 2015.

[21] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[22] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[23] D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[24] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[25] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognit.*, vol. 63, pp. 371–383, 2017.

[26] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[27] R. Venkatesan and S. Prabu, "Hyperspectral image features classification using deep learning recurrent neural networks," *J. Med. Syst.*, vol. 43, no. 7, 2019, Art. no. 216.

[28] B. Pan, Z. Shi, and X. Xu, "Mugnet: Deep learning for hyperspectral image classification using limited samples," *ISPRS-J. Photogrammetry Remote Sens.*, vol. 145, pp. 108–119, 2018.

[29] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[30] H. Zhang *et al.*, "StackGAN++: Realistic image synthesis with stacked generative adversarial networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1947–1962, Aug. 2018.

[31] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8789–8797.

[32] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, p. 99, Jan. 2017.

[33] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," in *Proc. ICLR*, 2016, *arXiv:1605.09782*.

[34] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," 2015, *arXiv:1511.05644*.

[35] N. Li and F. Chang, "Video anomaly detection and localization via multivariate Gaussian fully convolution adversarial autoencoder," *Neurocomputing*, vol. 369, pp. 92–105, 2019.

[36] Y. Xiao, D. Xiao, B. Hu, and C. Shi, "ANE: Network embedding via adversarial autoencoders," in *Proc. IEEE Int. Conf. Big Data Smart Comput.*, 2018, pp. 66–73.

[37] W. Xie, J. Lei, B. Liu, Y. Li, and X. Jia, "Spectral constraint adversarial autoencoders approach to feature representation in hyperspectral anomaly detection," *Neural Netw.*, vol. 119, pp. 222–234, 2019.

[38] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv: 1901.03407*.

[39] W. Xie, B. Liu, Y. Li, J. Lei, and Q. Du, "Autoencoder and adversarial-learning-based semisupervised background estimation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5416–5427, Aug. 2020.

[40] T. Jiang, Y. Li, W. Xie, and Q. Du, "Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4666–4679, Jul. 2020.

[41] E. Principi, F. Vesperini, S. Squartini, and F. Piazza, "Acoustic novelty detection with adversarial autoencoders," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2017, pp. 3324–3330.

[42] Y. Zhou and B. E. Shi, "Photorealistic facial expression synthesis by the conditional difference adversarial autoencoder," in *Proc. IEEE 7th Int. Conf. Affect. Comput. Intell. Interact.*, 2017, pp. 370–376.

[43] T. Mohiuddin and S. Joty, "Revisiting adversarial autoencoder for unsupervised word translation with cycle consistency and improved training," in *Proc. NAACL*, 2019, pp. 3857–3867.

[44] C. Ferri, J. Hernández-Orallo, and P. A. Flach, "A coherent interpretation of AUC as a measure of aggregated classification performance," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 657–664.

Xin Zhang received the B.E. degree in telecommunications engineering in 2019 from Xidian University, Xi'an, China, where she is currently working toward the M.S. degree in include hyperspectral image processing, target detection, and machine learning with the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks.

Her research interests include hyperspectral image processing, target detection, and machine learning.

Yunsong Li 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.

In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor. He is the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Services Networks, Xidian University. His research interests include image and video processing and high-performance computing.

Keyan Wang received the M.S. and Ph.D. degree in information and telecommunication engineering from Xidian University, Xi'an, China, in 2005 and 2008, respectively.

From September 2014 to September 2015, she was as a Visiting Scholar with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada. Currently, she is an Associate Professor with the School of Telecommunications Engineering, Xidian University. Her current research interests include image/video compression and processing, computer vision, and deep learning.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore, MD, USA, in 2000.

She is currently 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 SPIE-International Society for Optics and Photonics. She was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was a 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 has served as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, JOURNAL OF APPLIED REMOTE SENSING, and 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. She was the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012.

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 of Xidian University, Xi'an, China, in 2017.

Currently, she is an Associate Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. She has authored or coauthored more than 20 papers in referred journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 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.