Weakly Supervised Discriminative Learning With Spectral Constrained Generative Adversarial Network for Hyperspectral Anomaly Detection

Tao Jiang, Weiying Xie, Yunsong Li, and Jie Lei

Abstract—Anomaly detection (AD) using hyperspectral images (HSIs) is of great interest for deep space exploration and Earth observations. This article proposes a weakly supervised discriminative learning with a spectral constrained generative adversarial network (GAN) for hyperspectral anomaly detection (HAD), called weaklyAD. It can enhance the discrimination between anomaly and background with background homogenization and anomaly saliency in cases where anomalous samples are limited and sensitive to the background. A novel probability-based category thresholding is first proposed to label coarse samples in preparation for weakly supervised learning. Subsequently, a discriminative reconstruction model is learned by the proposed network in a weakly supervised fashion. The proposed network has an end-to-end architecture, which not only includes an encoder, a decoder, a latent layer discriminator, and a spectral discriminator competitively but also contains a novel Kullback–Leibler (KL) divergence-based orthogonal projection divergence (OPD) spectral constraint. Finally, the well-learned network is used to reconstruct HSIs captured by the same sensor. Our work paves a new weakly supervised way for HAD, which intends to match the performance of supervised methods without the prerequisite of manually labeled data. Assessments and generalization experiments over real HSIs demonstrate the unique promise of such a proposed approach.

Index Terms—Anomaly detection (AD), category thresholding, generative adversarial network (GAN), hyperspectral images (HSIs), spectral constraint (SC), weakly supervised learning (WSL).

I. INTRODUCTION

HYPERSPECTRAL image (HSI) appears in three dimensions: one dimension of which provides continuous spectral reflectance vectors containing essential attributes of different materials, and the other two dimensions show the spatial information of materials [1]. As an advanced means of Earth observation and deep space exploration, hyperspectral imaging has received significant attention in application areas, such as environmental monitoring, resource exploration, fine agriculture, land classification, and management [2]. Based on such applications, various data analysis methods are derived, including land-cover classification [3], target detection [4], and anomaly detection (AD) [5]. Especially, AD, known as an essential task in these methods, requires no prior knowledge of targets, aiming to locate anomaly instances that differ dramatically spatially or spectrally from plentiful background categories. Another key in the relationship between anomaly and background is the occurrence of anomalies is much lower than background [6]. Thus, anomalies can be detected concerning a background model or an anomaly-background separation model. In practice, there are many challenges in hyperspectral anomaly detection (HAD), including unavailable prior information, complex background with various classes, unbalanced samples, and inaccurate labels. To address these problems, traditional and deep learning-based methods are two major techniques for existing findings of studies.

According to the literature, traditional mainstream HAD algorithms can be further roughly subdivided into two categories: background modeling and background removal. The most famous background modeling algorithm is the Reed–Xiaoli (RX) detector [7]. Depending on the assumption that the background obeys the Gaussian distribution, the RX detector estimates the mean and variance of samples in the entire image to build the background model. Subsequently, numerous variants of the RX detector were proposed [8], [9]. For instance, compared to the global RX, the local RX (LRX) detector [8] models the background in a local window. Not content with contamination of the background statistics by anomaly pixels of the abovementioned detector, representation-based background modeling algorithms have been proposed in recent years. For example, the collaborative representation-based detector (CRD) [10] assumes that a pixel can be represented by a linear approximation of

Manuscript received 3 April 2020; revised 4 November 2020; accepted 17 May 2021. Date of publication 31 May 2021; date of current version 28 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62071360, Grant 61801359, and Grant 61671345, Grant 91538101, Grant 61502367 and Grant 61701360; in part by the Young Talent fund of University Association for Science and Technology in Shaanxi of China under Grant 20190103; in part by the Special Financial Grant from the China Postdoctoral Science Foundation under Grant 2019T102087; in part by the 111 Project under Grant B08038; in part by the Fundamental Research Funds for the Central Universities under Grant JBI80104; in part by the National Natural Science Foundation of China under Grant 2019Q153, Grant 2016JQ0623, and Grant 2016JQ0618; in part by the General Financial Grant from the China Postdoctoral Science Foundation under Grant 2017M620440; in part by the Youngtsve Rive Scholar Bonus Schemes under Grant CJT160102; in part by the Ten Thousand Talent Program; in part by the Innovation Fund of Xidian University under Grant 20109205456; and in part by the Fundamental Research Funds for the Central Universities. (Corresponding authors: Weiying Xie; Yunsong Li.)

Tao Jiang, Weiying Xie, Yunsong Li, and Jie Lei are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi’an 710071, China (e-mail: tjiang_2@stu.xidian.edu.cn; wyxie@xidian.edu.cn; ysl@mail.xidian.edu.cn).

Qian Du is with the Department of Electronic and Computer Engineering, Mississippi State University, Starkville, MS 39759 USA (e-mail: du@ece.msstate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TNNLS.2021.3082158.

Digital Object Identifier 10.1109/TNNLS.2021.3082158
its neighborhood pixels if it belongs to the background. The background of the CRD is built by making the above decision on all the pixels in the HSI. SRD [11] is a sparse version of CRD. The latest CRD variant, SWCRD [12], takes further account of the importance of different bands. Besides, the low-rank representation detector [low-rank and sparse matrix decomposition-based Mahalanobis distance for hyperspectral anomaly detection (LSMAD)] [13] establishes the background model by fully exploring the low-rank property of the background after removing sparse anomalies. Recently, Li et al. [5] proposed a detector (LSDM_MoG) based on low rank and sparse matrix decomposition in conjunction with a mixture noise model, which accurately describes noise characteristics. Considering the local geometrical structure and spatial relationships, the graph and total variation regularized low rank representation-based detector (GTVLRR) [14] method induces the graph regularization and total variation to the low-rank representation. The emerging typical algorithms for background removal are attribute and edge-preserving filters-based detector (AED) [15] and structure tensor and guided filter-based detector (STGD) [16], both of which remove the background mainly by the combination of attribute filtering and difference operation. More recently, other interesting studies have emerged. Fractional Fourier entropy-based detector (FrFE) [17] innovatively introduced the theory of fractional Fourier entropy into hyperspectral images (HSIs). Subpixel-pixel-superpixel guided fusion-based detector (SPSGF) [18] fully studied the advantages of feature fusion at different pixel levels for AD. Subspace selection-based discriminative forest method (SSDF) [19] borrowed from the concept of isolation and intended to avoid background contamination. In summary, modeling the background or the relationship between background and anomaly is crucial for HAD. However, with complex scenes and applications, background of HSIs becomes complex, which limits the detection performance based on traditional methods.

Deep learning-based methods can capture nonlinear and deep features, achieve complex function approximation through layer-by-layer combination, and have become increasingly popular in AD for various applications [20], [21]. Generally, depending on the availability of labels, deep learning-based AD models can be roughly grouped into three categories: supervised, semisupervised, and unsupervised [22]. A supervised learning model can achieve superior detection performance theoretically with labeled samples, whereas in the domain of hyperspectral detection, there are some problems, such as unstable atmospheric environment, inaccurate pixel-level labeling, and unavailable prior knowledge [23]. Such problems make unsupervised and semisupervised AD methods more practical. Some works on semisupervised learning have been carried out, such as [24] and [25], but a common flaw is that they cannot obtain the expected training samples purely from the background, which casts a cloud over a semisupervised learning method. Thus, unsupervised methods have become a dominant research topic due to not requiring label information. Especially, some generative models, such as autoencoder (AE) [26] and generative adversarial network (GAN) [27], have emerged as building blocks of an unsupervised AD model in HSI. For example, adversarial AEs are used to extract in-depth features in latent layer in [28]. Although similar feature extraction operations can reduce the dimension, they cannot always reflect the intrinsic structure of HSI due to the absence of locality property [29]. Thus, instead of extracting features in the middle latent layer, the end-to-end unsupervised AD is widely concerned. For instance, by imposing the learned embedding manifold constraints on the latent layer and combining all reconstruction errors of AE, manifold constrained auto-encoder network-based detector (MC-AEN) [29] can fully exploit the intrinsic structure of HSI. SDLR [30] uses a spectral angular distance to measure the reconstruction error of adversarial AE. Unlike using the AE model, HADGAN [31] is the first attempt to use the GAN for background modeling in HAD. Li et al. [32] used a transferred convolutional neural network (CNN) to capture the difference between pixel pairs, while neural network and dictionary-based low-rank detector (NNDLR) [33] employs a CNN to extract the abundance maps. The above-unsupervised detection methods can map the distribution of input high-dimensional data without any prior information [34]; however, the lack of prior knowledge still limits the detection performance with the risk of high false alarm and low detection accuracy. Therefore, it is urgent to explore novel learning methods to offer a better performance gain for AD.

To approach the detection performance of the supervised method while releasing the limitation of sample size, we put forward a weakly supervised detection method, i.e., weakly supervised discriminative learning technique with spectral constrained GAN (SCDGAN) (weaklyAD), for the first time in the field of HAD. In general, supervised learning has certain labels of both background and anomaly data instances, and semisupervised learning has pure background labels, while unsupervised learning needs no labels [22]. Speaking of the weakly supervised learning (WSL), it depends on the fact that the initially given labels are not always ground truth, which are different from the afore-mentioned deterministic labels [35]. As for the cornerstone theory of the proposed weaklyAD method in HAD, it mainly includes two assumptions: 1) the occurrence probability of abnormal samples is much lower than that of background samples and 2) background samples are subject to the multivariate Gaussian distribution. The Gaussian hypothesis originated from the RX algorithm [7], the recognized ancestor of the HAD algorithm. To verify the accuracy of the Gaussian hypothesis, background samples of some bands in the HSI are input into the Kolmogorov–Smirnov (K-S) test to determine whether the bands obey the Gaussian distribution. Fig. 1 shows that bands approximately following the Gaussian distribution can achieve the ideal detection
performance by the RX detector, while the performance of bands not following the Gaussian distribution is far from ideal. It shows that the closer the background obeys a Gaussian distribution, the better it performs in the HAD, which proves the rationality of Gaussian assumption in HAD.

Considering that the availability and accuracy of HSI’s labels are limited by the influence of the atmospheric environment and imperfect equipment, we introduce the concept of WSL to match the performance of supervised learning. In preparation for weak supervision, first, a salient category searching method is proposed to search for coarse labels of samples, i.e., inaccurate labels required for the beginning of WSL. In practice, the complex and diverse backgrounds make it easy for anomalies to get lost in the background, which inspires the idea of constructing a reconstruction network with an enhanced anomaly-background discrimination degree to be proposed. Thus, considering the characteristic of high-dimensionality, limited, and unbalanced hyperspectral data samples, the second step in the weaklyAD method is to learn a reconstructed model $R(\cdot)$ through a weakly supervised discriminant learning network based on the GAN with spectral distance constraint. The final step is to feed all the spectral samples into the well-learned model, resulting in a novel reconstruction with background homogenization and anomaly saliency.

Comparing other existing AD algorithms in HSI, the main contributions of our weaklyAD method lie in the following.

1) For the first time, the concept of WSL is introduced into HAD, which not only breaks the dilemma of limited detection performance without prior information but also opens up more flexible ways to study HAD.
2) To satisfy the concept of WSL, a novel salient category searching stage is designed to search for coarse sample labels. Especially, a novel probability-based category thresholding (PCT) is proposed, which can divide the category with the most occurrence probability into the background category, realizing the coarse classification of category label.
3) By examining the relative relationship between background and anomaly in HSIs, an innovative end-to-end discriminative reconstruction with background homogenization and anomaly-background discriminability enhancement is modeled by SCGANs in a weakly supervised way. Moreover, several designed and applicable GANs are fully discussed and validated both experimentally and theoretically.
4) A new orthogonal projection divergence (OPD) spectral distance constraint is imposed to improve the discrimination capacity between the anomaly and the background, in conjunction with the basic GAN structure. Besides, a Kullback–Leibler (KL) divergence term ($D_{\text{KL}}$) is used to maximize the OPD distance in the full loss function.
5) The generalization of the real application of deep neural networks in HAD is discussed for the first time. For the network learned on an HSI, the common deep learning-based HAD methods can only detect input HSI, while our weaklyAD method can detect several other HSIs captured from the same sensor, which makes the proposed method naturally offer great feasibility for real application.

The remainder of this article is organized as follows. Section II briefly reviews the basic theories related to GAN and AE. The proposed weaklyAD is detailed in Section III. Section IV shows the assessments and generalization experiments. Conclusions are drawn in Section V.

II. RELATED WORK

This section is to briefly describe the deep architectures employed in this article. This includes GAN and AE, both of which are commonly used in different applications of AD.

A. Generative Adversarial Network

A GAN [36] consists of two adversarial subnetworks, which compete with each other in a zero-sum game, called generator and discriminator. The task of the generator is to map input data $z \sim p_z(z)$ to a given real data $x \sim p_{\text{data}}(x)$ as much as possible. While the discriminator works to identify the real sample data $x$ and discard the generate sample. In other words, the generator tries to trick the discriminator with the generated samples, but the discriminator does not fail for it until the generator and the discriminator are at equilibrium, that is, the generator learns the distribution of the given dataset. The zero-sum game they participate in, also known as the min–max game, has the following cost function:

$$\min_G \max_B \mathbb{E}_{x \sim p_{\text{data}}(x)} \left[ \log D(x) \right] + \mathbb{E}_{z \sim p_z(z)} \left[ \log(1 - D(G(z))) \right].$$

The above structure enables GAN to learn exact data distribution resulting in generating new data samples with expected variations, which makes GAN earn the most approval in diverse tasks, particularly AD [20], [37]. Besides, when modeling complex high-dimensional data distributions, GAN opens new possibilities for HSI processing. To reach the abovementioned equilibrium of the generator and discriminator, different objective functions can be designed according to actual requirements and the most common one is to minimize the Jensen–Shannon (JS) divergence. Thus, by changing the objective function and the interior structure of the generator and discriminator, GAN has spawned many variations. In this article, we choose GAN as the backbone of our solution and then select the proper interior structure for GAN as follows.

B. Auto-Encoder

An AE [26] is a feedforward neural network architecture with two components (i.e., an encoder and a decoder), which maps an input image $x \in \mathbb{R}^n$ to a middle latent layer $z \in \mathbb{R}^m$ $(m \ll n)$ by encoder function [i.e., $f(\cdot)$] and then maps $z$ back to the output layer $x' \in \mathbb{R}^n$ by decoder function (i.e., $g(\cdot)$) has a symmetric structure with respect to the encoder). By minimizing the reconstruction error $L(x, x') \in \mathbb{R}^n$, AE can effectively represent data within multiple latent layers. The middle latent layer of AE can linearly or nonlinearly compress the redundancies while preserving essential information of the input. Thus, the AE was originally introduced for nonlinear dimensionality reduction and feature extraction [38]. Meanwhile, due to the fact that AE has strong modeling and
representing capability for the data distribution, which has also emerged as a suitable approach to AD in recent years [39]. Several variants of AE architectures [40], [41] produce promising results in AD, including HAD [42]. Especially, in this article, we construct an expected AE structure relying on the nature of hyperspectral data for the purpose of discriminative learning.

### III. Proposed Method

This work proposes a weakly supervised approach for HAD and learns a discriminative end-to-end reconstruction with background homogenization and anomaly saliency by adversarially training the proposed SCGAN architecture in a weakly supervised manner. Fig. 2 shows a high-level overview of the proposed weaklyAD approach, which is composed of three major components: salient category searching, weakly supervised discriminative learning, and AD on reconstructed HSI. Mathematically, we define and formulate our problem as follows.

The HSI with \( L \) dimensions is denoted as \( Y = \{ y_1, y_2, \ldots, y_{M \times N} \} \), where \( y_i \in \mathbb{R}^{L \times 1} \) is the \( i \)th spectral vector and \( i = (i_1, i_2) \), \( i_1 = 1, \ldots, M \), \( i_2 = 1, \ldots, N \). \((i_1, i_2)\) represents the position coordinates of the \( i \)th spectral vector in the spatial domain, which means that \( i = N \cdot (i_1 - 1) + i_2 \). Thus, the HSI can also be set as \( Y = \{ y_i \in \mathbb{R}^{L \times 1} \mid i = (i_1, i_2) = N \cdot (i_1 - 1) + i_2 \mid i_1 = 1, i_2 = 1, \ldots, N \} \). In this article, an input HSI \( Y \) is split into anomaly samples set \( A \) and background samples set \( B \), i.e., \( Y = [A, B] \), \( y_i \subset A \cup B \). Note that \( A \cup B = Y \) and \( A \cap B = \emptyset \).

Based on the notations defined above, a reconstruction model \( \mathcal{R}(\cdot) \) can be established as

\[
\hat{y}_i = \mathcal{R}(y_i; A, B | A \cup B = Y, A \cap B = \emptyset)
\]  

(2)

where \( \hat{y}_i \in \mathbb{R}^{L \times 1} \), \( \hat{y}_i \subset \hat{Y} \in \mathbb{R}^{M \times N \times L} \), and \( \hat{Y} \) refers to \( Y \)'s estimate. The learning objective of the model \( \mathcal{R}(\cdot) \) is to capture the discriminative feature between anomaly and background such that the reconstructed \( \hat{Y} \) is more discriminative than the original \( Y \). Next, we introduce how to design and learn \( \mathcal{R}(\cdot) \) in the weaklyAD framework.

#### A. Salient Category Searching

This salient category searching step is to prepare for WSL and intends to predict coarse labels of the given input. According to the predicted coarse labels, this process searches for two salient category sets, including a coarse anomaly samples set where pixels have a higher probability of belonging to anomalies and a coarse background samples set where pixels have a higher probability of belonging to the background. In other words, the searching problem can be regarded as an initial anomaly-background separation problem. In particular, considering the premise that HAD does not have prior knowledge and anomalies occur with a lower probability than the background, we design salient category searching in an unsupervised manner, which can be subdivided into three major components: salient category extraction, coordinate index, and coarse samples set construction, as shown in Fig. 2.

Fig. 3 shows the details of the salient category extraction step. Without prior knowledge, we first apply an unsupervised clustering method called density-based spatial clustering of applications with noise (DBSCAN) [43] to obtain a category probability map (i.e., clustering results map). Due to its ability to discover clusters with different sizes and shapes and its low requirements on specifying cluster numbers in the presence of noise and outlier, DBSCAN has been applied to HAD [25].
In the clustering process, the DBSCAN scans each pixel’s neighbors within Eps (i.e., the minimum distance between adjacent points) distance. Once the number of neighbors of the core pixel exceeds MinPts (the minimum number of points), a cluster is formed. Subsequently, the neighbors within Eps distance of the core pixel are assembled iteratively. Here, the Euclidean distance is employed for the DBSCAN clustering. Given an input HSI $Y$, under the condition of $(\text{Eps}, \text{MinPts})$, the category probability map $P = \{p_{i,j}\}_{i=1}^{M} \in \mathbb{R}^{M \times N \times 1}$ can be obtained, as shown in Fig. 3.

However, for different HSIs, the final number of categories obtained by clustering is quite different and may be as large as 50. To address this problem, a novel PCT method is designed based on the fact that the probability of background samples appearing in the HAD is much higher than that of abnormal samples. The PCT step starts with counting the number of samples for each category by category statistical bar. Since the probability of the background appearing is high, the number of samples for the background category should be the largest. According to the experimental test, for an HSI used for HAD, the number of category samples obtained from the category statistical bar with category “1” is always far higher than that of all other categories. For example, from the category statistics bar drawn in Fig. 3, it can be found that the sample number of category “1” is 9903, which is far outnumbered by 97 for all other categories. In this case, the objective is to get a coarse labels division, not for fine classification. Thus, samples belonging to the background are with category label 0, while other samples can be roughly classified as anomalies with category label 1 as

$$S = \{s_i = 0, \ p_i \in "1" \\
= 1, \ p_i \notin "1"\} \quad (3)$$

where $S = \{s_i| i = (i_1, i_2) \}_{i=1}^{M \times N}$ refers to the category saliency map and is regarded as the category labels set of $Y$. $(i_1, i_2)$ represents the position coordinates of the $i$th pixel point in the 2-D image space, which means that $i = N \cdot (i_1 - 1) + i_2$. There are only the values 0 and 1 in $S$.

According to the coordinate index of the category saliency map $S$, the coarse anomaly samples set $A$ and the coarse background samples set $B$ are searched by

$$A = \{y_i|s_i = (i_1, i_2) = 1\} \quad (4)$$

$$B = \{y_i|s_i = (i_1, i_2) = 0\} \quad (5)$$

The coarse anomaly samples set $A = \{a_i \in \mathbb{R}^{L \times 1}|i = i_1 \}_{i=1}^{C_A}$ and the coarse samples set $B = \{b_i \in \mathbb{R}^{L \times 1}|i = i_1 \}_{i=1}^{C_B}$ are finally constructed, where $C_A$ and $C_B$ represent the number of samples in sets $A$ and $B$, respectively, and $C_A + C_B = M \times N$. Meanwhile, for the convenience of subsequent training, coarse anomaly samples are squeezed into an anomaly sample vector as

$$a = \frac{1}{C_A} \sum_{i=1}^{C_A} a_i \quad (6)$$

where $a \in \mathbb{R}^{L \times 1}$ means the average of coarse anomaly samples. The coarse anomaly samples $A$ (especially, vector $a$) and the coarse background samples $B$ are used to learn model $\mathcal{R}(\cdot)$. For coarse anomaly or background samples, some errors can be tolerated, i.e., the predicted coarse labels in $S$ are not always ground truth. Consequently, the model $\mathcal{R}(\cdot)$ is learned on the aforementioned coarse anomaly and background samples in the form of WSL.

### B. Weakly Supervised Discriminative Learning

In this section, three novel GANs are constructed to dig and enhance the discriminative features between anomaly and background. The abovementioned coarse background and anomaly samples are fed into these proposed GANs to learn model $\mathcal{R}(\cdot)$ from the original HSI space to the reconstructed HSI space in a weakly supervised way. A detailed description of the network structure, learning process, and reconstruction process is given in the following.

1) **Network Architecture:** Considering the strong generation capacity of GAN and its applicability to HSIs [31], GAN is chosen as the basic network structure for the proposed weaklyAD method. When designing the GAN structure, the network is desired to be able to map the input data back to the original input space again for a new representation (that is, reconstruction). Thus, for reconstruction and overcoming the imbalance between the generator and discriminator of GAN, it is motivated to build GAN with AE, as in “OCGAN” [44].

There are three basic structures shown in Fig. 4, including weaklyAD_GAN_Dz, weaklyAD_GAN_Ds, and weaklyAD_DGAN. However, in our method, the network not only consists of an encoder, a decoder, a latent layer discriminator, or a spectral discriminator but also includes a novel KL divergence-based OPD spectral constraint (SC) that measures the spectral distance between the coarse background samples and coarse anomaly samples, which makes basic structure reborn as weaklyAD_SCGAN_Dz, weaklyAD_SCGAN_Ds, and weaklyAD_DSCGAN. The network architecture in Fig. 4 shows that the three basic network structures are all configured with AE. The AE plays a role in reconstructing, which consists of an encoder (E) and a decoder (D). First, the learning of $\mathcal{R}(\cdot)$ on coarse samples set can be decomposed into encoding and decoding processes

$$Z = f(WB + \beta) \quad (7)$$

$$\hat{B} = f(WZ + \beta) \quad (8)$$
where the encoder $E$ maps input $B$ to latent feature representation $Z$ by an identity linear function $f(\cdot)$ with learnable parameters: the weight matrix $W$ and the bias vector $b$. Then, the decoder $D_e$ maps $Z$ back to the original input layer $B$ by a tanh nonlinear function $f(\cdot)$ with the weight matrix $W$ and the bias vector $b$. Furthermore, our objective is to learn the discriminative features between anomaly and background. In addition, we add that the spectral distance constraint (i.e., the OPD [46] except for HAD) is imposed on the reconstruction network to maximize the distance between coarse anomaly samples and other background samples in the input space. The larger the value of OPD, the more discriminative the two spectral vectors. The spectral distance between sets $A$ and $B$ is defined as

$$\text{OPD}(a, b_i) = (\alpha^{T}P_{b_i}a + b_i^{T}P_{a}b_i)^{1/2} \quad (9)$$

where $P_{i} = I - k(k^{T}k)^{-1}k^{T}$ for $k = a, b$, and $I$ is an identity matrix. $P_{i}$ means the orthogonal subspace of $k$.

Another major force in our network structure is GAN-based adversarial learning, which is realized by adding discriminators. Considering the original assumption of HAD that the background of HSI obeyes the multivariate Gaussian distribution, a latent discriminator $D_z$ is added to join the zero-sum game with $E$ for matching the prior distribution $N(0, I)$ of the latent feature space, which facilitates the network to learn and to generate more homogenized background information, as shown in Fig. 4(a). However, such learning often leads to blurring of the reconstructed images [47], which harms the detection performance. Thus, we attach a spectral discriminator $D_s$ to match the output of the decoder in the original input space, as shown in Fig. 4(b). This allows $D_s$ to help $D_e$ competitively reconstruct more reliable samples with less blurriness and more details [48]. It is worth noting that in Fig. 4(c), the spectral discriminator is also included in the reconstructed layer, which enables structure weaklyAD_DGAN to absorb the advantages of both two discriminators and learns the characteristics of higher dimension (i.e., original input space) and low dimension (i.e., latent feature space). Therefore, in the following, we focus on the structure of Fig. 4(c) with SC, i.e., weaklyAD_DSCGAN, which consists of two GANs, each of which trained by competing while collaborating to understand the underlying discriminative features.

2) Learning Process: The weaklyAD_DSCGAN architecture is composed of the encoder $E$, the decoder $D_e$, the latent discriminator $D_z$, and a spectral discriminator $D_s$, resulting in having two adversarial and one reconstruction losses. Based on the weaklyAD_DSCGAN architecture, the learning process of model $\mathcal{R}(\cdot)$ is to optimize the learnable parameters $\theta = (W, \beta, \hat{W}, \hat{\beta})$ with the following three losses.

The reconstruction loss of AE is to minimize the deviation between the decoded images and the original input image, which generally uses mean squared error (MSE) as

$$L_R = \|b - \hat{b}_i\|_2 \quad (10)$$

where $b_i \in \mathbb{R}^{L \times 1}$ is an input sample from the coarse background samples set $B$.

The spectral distance constraint intends to achieve discriminative learning by maximizing the OPD distance between each spectral vector in $B$ and the anomalous vector $a$. For this purpose, the reconstruction loss is reformulated as

$$L_R = a_1\|b - \hat{b}_i\|_2 - a_2D_{KL}(1.0||\text{OPD}(a, b_i)) \quad (11)$$

where $a_1$ and $a_2$ control the influence of corresponding terms on $L_R$ and are set to 10 empirically. Besides, a KL divergence term ($D_{KL}$) [49] is added to maximize OPD distance. By minimizing this loss, $E$ and $D_e$ are encouraged to better reconstruct the expected distribution.

The latent space adversarial loss for matching the distribution of the latent feature space with the prior distribution $N(0, I)$ is formed as

$$L_{D_z} = \mathbb{E}[\log D_z(N(0, I))] + \mathbb{E}[\log (1 - D_z(E(b_i)))] \quad (12)$$

By minimizing $E$ and maximizing $D_z$ on $L_{D_z}$, $E$ can learn more homogeneous background information.

The spectral space adversarial loss is employed to match the distribution of the decoded images and the known input data distribution

$$L_{D_s} = \mathbb{E}[\log D_s(b_i)] + \mathbb{E}[\log (1 - D_s(\hat{b}_i))] \quad (13)$$

where $AE$ reconstructs its input $b_i$, generates $\hat{b}_i$, and tries to fool $D_s$, while $D_s$ takes responsibility for identifying $AE$'s tricks. $AE$ intends to minimize $L_{D_s}$ against $D_s$ that intends to maximize it.

The final objective of the model $\mathcal{R}(\cdot)$ gives

$$L_f = L_R + L_{D_z} + L_{D_s} \quad (14)$$

The model’s parameters are optimized by minimizing the loss function. When the encoder, decoder, and two discriminators are jointly learned for 3000 epochs, the parameters $\theta = (W, \beta, \hat{W}, \hat{\beta})$ of model $\mathcal{R}(\cdot)$ are fixed and used to reconstruct the original HSI.

C. AD on Reconstructed HSI

1) Reconstructing Process: With the parameters $\theta = (W, \beta, \hat{W}, \hat{\beta})$ being well-optimized, an input HSI $Y = [A, B] = \{y_i \in \mathbb{R}^{L \times 1}\}_{i=1}^{M \times N}$ can be end-to-end reconstructed by the well-learned model $\mathcal{R}(\cdot)$ into a new discriminative reconstruction HSI $\hat{Y} = \{\hat{y}_i \in \mathbb{R}^{L \times 1}\}_{i=1}^{M \times N}$ as

$$\hat{y}_i = \mathcal{R}(y_i; \theta) \quad (15)$$

By the above reconstruction process, the model $\mathcal{R}(\cdot)$ facilitates the reconstructed $\hat{Y}$ more discriminative than the original $Y$ when identifying anomalies from the background. Moreover, Figs. 5 and 6 show some visual reconstructed results, spectrally or spatially, to promote the understanding of the reconstructed HSI. The locations and spectral variation of seven representative anomaly pixels and seven representative background pixels are shown in Fig. 5. It is worth noting that after the reconstruction, in addition to maintaining the spectral curves of anomaly pixels well, the spectral vectors from the background also tend to be more uniform (i.e., homogenization). Furthermore, although the reconstructed background
spectra are homogenized, the trend at the typical peaks and troughs remains approximately the same as the original background, meaning that the reconstruction does not violate the physical properties of the HSI. Meanwhile, according to the 2-D visual results in Fig. 6, discrimination between the background and anomaly of the reconstructed image is significantly larger than that of the original image. This means that, without violating the physical characteristics of the HSI, the nonlinear model $\mathcal{R}(\cdot)$ based on the weaklyAD_DSCGAN architecture not only generates the background with uniform distribution but also enhances anomaly-background separation. Thus, using the reconstructed output $\hat{Y}$ of the model $\mathcal{R}(\cdot)$ is much easier to detect anomalies than using the original $Y$.

2) Anomaly Detection: Finally, the classical RX detector is adopted on $\hat{Y}$ to detect anomalies as

$$D = (\hat{y}_i - \mu)^T \Gamma^{-1} (\hat{y}_i - \mu)$$

(16)

where $\mu \in \mathbb{R}^{L \times 1}$ and $\Gamma \in \mathbb{R}^{L \times L}$ represent the mean vector and covariance matrix of the matrix $\hat{Y}$, respectively. Since there is more discriminative information in $\hat{Y}$ than $Y$, the RX detector can more effectively detect anomalies with resistance to background interference.

IV. EXPERIMENTS

In this section, extensive experiments are conducted to evaluate and verify the effectiveness and generalization of the proposed weaklyAD method. It is compared qualitatively and quantitatively with the most advanced and typical HAD methods.

A. Experimental Setup

1) Data Description: There are four real HSIs captured over different scenes by two different sensors. In these images, some of the anomalies appear as points, some have structural information, and most of the anomalies appear in the form of different scales, which are listed as follows.

1) Gulfport Dataset: The first dataset was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor from Gulfport, MS, USA, in 2010. After removing SNR and water-absorption bands, the Gulfport image is of size $100 \times 100$ with 191 spectral bands in spectral coverage ranging from 400 to 2500 nm. The spatial resolution is 3.4 m per pixel. The three airplanes in different scales are anomalous targets of interest, which consists of 60 pixels and accounts for 0.6% of the image.

2) Pavia Dataset: The second dataset is a Reflective Optics System Imaging Spectrometer (ROSIS-03) image, which was acquired at the location of Pavia city center in northern Italy. The image scene covers an area of $150 \times 150$ pixels, with 102 spectral bands in wavelengths ranging from 430 to 860 nm. There are some vehicles on the bridge, which are selected as anomalies. These targets are composed of 63 pixels, accounting for 0.28% of the image. The main background classes are bridge, water, and bare soil.

3) Los Angeles-1 (LA-1) Dataset: The third dataset was acquired by the AVIRIS sensor over the area of Los Angeles city. The image scene covers an area of $100 \times 100$ pixels, with 205 spectral bands varying from 430 to 860 nm, with a 7.5-m spatial resolution. There are some houses considered as anomalies, which, in total, consists of 232 pixels and accounts for 2.32% of the whole image.

4) San Diego Dataset: The fourth dataset is widely used HSIs, which was collected by the AVIRIS sensor over the San Diego airport area, CA, USA. This image contains $100 \times 100$ pixels, with 189 spectral bands in wavelengths ranging from 400 to 2500 nm. We consider three airplanes as the objects to be detected in this dataset. This dataset is an airport scene in which the main background types are hangars, parking aprons, and exposed soil. The ground truth of this dataset includes 134 anomalous pixels and accounts for 1.34% of the image.

2) Evaluation Metrics: We employ the three most commonly used AD evaluation indexes to explore the detection performance of the proposed method and its comparison methods quantitatively. First of them, derived from boxplot, is the separability range, which represents the detector’s ability to extract anomalies from background [50]. A better detector offers more obvious separation between anomalies and background. The second index is the receiver operating characteristic (ROC) curve [51], which can be plotted by the true positive rate (TPR, i.e., $P_T$) and the false positive rate.
(FPR, i.e., $P_f$) at various thresholds ($r$) based on the ground truth. It can be found in the ROC curve that the method with high detection probability is superior to the method with low detection probability under the same false alarm condition. The third index is area under ROC curve (AUC) [52], which calculates the area of an ROC curve. The closer the AUC of $(P_d, P_f)$ value is to 1, the better the detection performance. Conversely, the closer the AUC of $(P_f, r)$ value is to 0, the lower the probability of false detection.

3) Comparative Methods: Five frequently cited and state-of-the-art AD methods are compared, including a deep learning-based method spectral adversarial feature learning for anomaly detection (SAFL) [28] and four traditional-based methods, RX [7], CRD [10], FrFE [17], and LSMAD [13] methods. For fine comparison, we reproduced the SAFL network as recommended in [28] and replaced the detector with the RX detector, termed SAFL_RX.

4) Implementation Details: All learning and reconstructing processes based on the network model are carried out on a GeForce RTX 2080 graphics card with running Python 3.6.0, TensorFlow-GPU 1.10.0, and CUDA 10.0. Meanwhile, all the detection processes are implemented in the MATLAB R2017a environment. We optimize our network by using a stochastic Adam [53] optimizer with an initial learning rate of 0.0001. The batch size is the same as the number of input spatial pixels, and the epoch is 3000. The encoder $E$ and the decoder $D_e$ both adopt the two fully connected layers followed by the leaky ReLU activation. The size of the latent feature space between the encoder and the decoder is limited to 20 units. The structure of latent discriminator $D_z$ is similar to the decoder $D_e$, but the identity linear function is replaced with the sigmoid nonlinear function as output activation function, and the structure of spectral discriminator $D_s$ is the same as that of the decoder $D_e$.

5) Parameter Settings: In the weaklyAD method, there are two parameters: the radius value $Eps$ and the value MinPts. The AUC scores of $(P_d, P_f)$ on salient category maps are used to evaluate the effects of the parameters $Eps$ and MinPts for the proposed method. To analyze the sensitivity of parameter $Eps$, MinPts is fixed as 1. With $Ept$ varying from 0.01 to 0.25, the optimal $Ept$ values for different datasets are shown in Fig. 7(a). For the Gulfport dataset, it achieves the highest AUC score of $(P_d, P_f)$ when $Eps$ is 0.13. The AUC score of $(P_d, P_f)$ over the Pavia dataset achieves the best point with $Eps$ of 0.15. As for the Los Angeles-1 and San Diego datasets, the optimal $Ept$ values should be $Ept = 0.07$ and $Ept = 0.22$, respectively. Fig. 7(b) shows the AUC scores of $(P_d, P_f)$ of saliency category extraction results when the MinPts change from 1 to 25. Note that the best $Eps$ value obtained from Fig. 7(a) is selected to help analyze the sensitivity of the MinPts parameters. For the Los Angeles-1 and San Diego datasets, changes to MinPts have little impact on the AUC scores of $(P_d, P_f)$. When MinPts is 1, the Pavia dataset achieves the best AUC scores. As for the Gulfport dataset, the highest AUC value is obtained by MinPts = 7. However, the experimental evaluation shows that there is only a 0.0001 difference between the final detection result of weaklyAD method when MinPts is 7 and 1. MinPts that equals 1 is acceptable for the Gulfport dataset. For simplicity and less human intervention, MinPts = 1, which provides stable and acceptable performance for all datasets tested in our experiments.

For the comparative methods, according to the corresponding AUC scores of $(P_d, P_f)$, we select the optimal parameters for each dataset, which are listed in Table I. As for the RX and FrFE algorithms, there is no parameter dependence on them. The network structure parameters of SAFL_RX algorithm are as recommended in [28]. The CRD method is sensitive to the size of the inner and outer windows (i.e., $w_{in}$ and $w_{out}$). For the LSMAD method, there are two important parameters: the rank $r$ of the background matrix and the cardinality $k$ of the sparse matrix, which are empirically fixed to 3 and 0.003, respectively, on four datasets.

B. Effectiveness Assessment

To assess the effectiveness of the network of the weaklyAD in background homogenization and anomaly extraction, we analyze the anomaly-background separability via plotting the box plots in Fig. 8. The “RX” column means the anomaly-background separability result obtained by inputting the original HSI directly to the RX detector. The column “weaklyAD” refers to the separable result obtained by feeding the reconstructed HSI into the RX detector, where the reconstructed HSI is generated by feeding the original HSI into the proposed network. Each detector column (“RX” or “weaklyAD”) corresponds to two boxes, in which the red box represents the distribution range of detection values of the anomaly category and the green box refers to the distribution range of detection values of the background category. The lines in the middle of the box are averages. The position and compactness of the boxes reflect the trend of the pixel distribution of

![Fig. 7. Effects of the parameters (a) Eps and (b) MinPts over the AUC scores of $(P_d, P_f)$ on each dataset.](image-url)
the background and anomaly. Thus, the anomaly-background separability can be determined by the relative distance of the red and green boxes. In other words, the distance between the lower bound of the red box and the upper bound of the green box indicates anomaly-background separability. Meanwhile, the homogenization of the background can be expressed by the compactness of the green box.

By observing Fig. 8, separability distances between red boxes and green boxes of weaklyAD on four experimental HSIs are quite larger than that of the RX, which means that the network model of weaklyAD makes anomaly and background more discriminative which means that the network model of weaklyAD makes anomaly and background more discriminative. It can also be found that each green box of weaklyAD is squeezed to a greater degree than RX, which means that detection values of background pixels of weaklyAD are more similar. By using the same detector, background pixels of the reconstructed HSI are more likely to be judged as the same substance by the detector than background pixels of the original HSI, so weaklyAD has more similar detection values. For the above results, the most reasonable inference is that weaklyAD has a more homogeneous background, which is also consistent with the conclusions in Figs. 5 and 6.

Conversely, on different datasets, the red boxes of weaklyAD are longer than the red box of RX, indicating that the distribution of anomaly of weaklyAD is more diverse. As shown in Fig. 5(b), the spectral curves of anomalies at different locations are inherently different to a certain extent, and these differences are respected. Therefore, the more dispersed detection values of anomalies in weaklyAD have no negative effect on our experiment. Combined with the observations in Figs. 5, 6, and 8, the evaluation results of the separability range index are consistent with the reconstruction results obtained in Section III Part C, which verifies the reliability and effectiveness of our method.

C. Detection Quality Assessment for Different Networks

To explore the influence of different network structures on detection performance, we assess the detection quality of different network models mentioned in Section III-B by using the AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\). Specifically, we set up six comparative structures. The first structure is our weaklyAD framework based on GAN with the latent discriminator \(D_z\) and the spectral discriminator \(D_s\) under the spectral distance constraint (named weaklyAD_SCDGAN, also as weaklyAD).

For the second structure, only the spectral discriminator \(D_s\) is removed compared with weaklyAD. The latent discriminator \(D_z\) and spectral distance constraint are employed for constructing GAN (named weaklyAD_SCGAN_Dz). Compared with the second structure, the third structure uses the spectral discriminator \(D_s\) instead of latent discriminator \(D_z\) to complete GAN (named weaklyAD_SCGAN_Ds). The above three network structures are all constrained by spectral distance. The next three structures (i.e., weaklyAD_DGAN, weaklyAD_GAN_Dz, and weaklyAD_GAN_Ds) are the spectral distance-free versions of the above three structures (i.e., weaklyAD_SCDGAN, weaklyAD_SCGAN_Dz, and weaklyAD_SCGAN_Ds), respectively.

Table II shows the evaluation AUC scores of the different network-based methods on different datasets. The best scores are highlighted in bold in Table II. By analyzing the AUC scores over four real datasets, the following observations can be found.

In terms of the detection accuracy, first, the AUC scores of \((P_d, P_f)\) for weaklyAD_SCDGAN have reached the top compared with the other structures, which means that...
weaklyAD_SCDGAN achieves the best detection performance. Second, the performance of the three structures with spectral distance constraint (i.e., weaklyAD_SCDGAN, weaklyAD_SCGAN_Dz, and weaklyAD_SCGAN_Ds) is superior to that of the corresponding version without spectral distance constraint (i.e., weaklyAD_DGAN, weaklyAD_GAN_Dz, and weaklyAD_GAN_Ds) respectively, which fully reflects the reliability and importance of the application of spectral distance constraint. Third, the detection accuracy of weaklyAD_SCGAN_Dz is higher than that of weaklyAD_SCGAN_Ds in most cases, so is the version without spectral distance constraint. This means that Dz contributes more than Ds. As far as the false alarm rate, the weaklyAD_GAN_Ds structure can obtain the minimum average of false alarm rate. Most importantly, under the condition of fully ensuring the detection accuracy, the false alarm rate obtained by structure weaklyAD_SCDGAN is relatively low and acceptable. In general, the structure of weaklyAD_SCDGAN is the optimal one that meets the expectation of AD.

D. Detection Quality Assessment for Different Methods

According to the parameter settings in Table I, all methods achieve an optimal performance. The visual detection maps over four datasets are shown in Fig. 9, and the corresponding visual ROC curves and quantized AUC values are shown in Fig. 10 and Table III, respectively.

First, by observing the visual comparisons in Fig. 9, we notice that the weaklyAD algorithm can well highlight anomaly objects with different sizes and shapes and suppressing the backgrounds with rich categories. Other algorithms are more likely to miss the anomalies or be disturbed by the background. In other words, the weaklyAD algorithm always finds more complete, more evident, and less background-contaminated anomalies than other comparison algorithms. Concretely, in Fig. 9(a), the weaklyAD method can accurately detect the largest airplane and obtain the position of the other two small airplanes. The LSMAD algorithm has the same detection capability as the weaklyAD algorithm, but it is more vulnerable to background intrusion. The RX, CRD, and FrFE methods can only recognize the locations of one largest airplane, while the shape of the airplane is blurry and the rest of the airplanes are completely undetected. Although the SAFL_RX method can also detect the airplane, the contrast between the anomaly and the background in its detection
results is not enough, which makes it difficult to see the airplane visually. Taking Fig. 9(b), the weaklyAD method obtains the most complete extraction of small anomalies. None of the four methods, RX, FrFE, LSMAD, and SAFL_RX, can detect all small anomalies. The CRD method can find anomalies, but almost all the detected anomalies are mixed in the backgrounds and noises, as in Fig. 9(c). For complex scenarios with many anomaly objects with different sizes and shapes, Fig. 9(c) shows that compared with weaklyAD method, the RX, CRD, FrFE, and SAFL_RX methods miss anomalies to different degrees, and the LSMAD method is more likely to be confused by the background. In the scenario of the anomalies with more structural information, as shown in Fig. 9(d), our method still does a good job of detecting anomaly objects, even their edges. By contrast, the RX and CRD methods find low-quality anomalies mixed with more background, and the SAFL_RX method again misses anomalies. As discussed above, the RX and FrFE methods are robust and have good detection effects in simple scenarios, but their discriminability between background and anomalies is poor. The CRD method is susceptible to window size. Representation-based methods are often susceptible to noise and background due to their linear or nonlinear representation, such as CRD, while the LSMAD method can alleviate background and noise interference relatively well due to its separation of background, anomaly, and noise. Although the SAFL_RX method is almost free of background interference, it cannot detect most of the anomalies.

Meanwhile, in the ROC curves of \((P_d, P_f)\) of Fig. 10, the weaklyAD method always shows higher TPRs as FPRs vary from 0 to 0.1. Furthermore, Table III calculates the AUC scores of \((P_d, P_f)\) and \((P_f, \tau)\) for comparative methods over four datasets. It can be seen that the weaklyAD has maximal AUC of \((P_d, P_f)\), which means that it is capable of achieving the best performance. In terms of false alarm rate, it can achieve a relatively low false alarm rate on the premise of ensuring the highest detection accuracy. Considering that HAD focuses on whether anomalies can be detected, the AUC scores fully meet the requirements of HAD. Note that the best AUC scores in Table III are consistent with the visual detection maps in Fig. 9 and ROC curves in Fig. 10, and thus, it can conclude that the weaklyAD method exhibits competitive performances compared with other advanced methods for HAD.

E. Generalization of Networks

In this section, we further discuss how the proposed architecture has the potential for generalization, that is, to generically reconstruct different datasets by a well-learned network. As far as we know, all known deep learning-based HAD methods are trained and tested on a single dataset without exploring the generalization of the method [31], which means that the generalization experiment in this article is of great significance. To be more convincing, we conduct experiments across six datasets from two groups: group I contains three original datasets, namely Gulfport, Los Angeles-1 (LA-1), and San Diego. The AUC scores in Table III are consistent with the visual detection maps in Fig. 9 and ROC curves in Fig. 10, and thus, it can conclude that the weaklyAD method exhibits competitive performances compared with other advanced methods for HAD.

Besides, the average computing time of each comparison method on all the tested images is recorded in seconds in Table IV. It is obvious that the weaklyAD method is far more efficient than CRD, FrFE, and LSMAD methods except for the RX and SAFL_RX methods. In comparison to the RX and SAFL_RX methods, our method is more time-consuming, but it is acceptable for HAD. In essence, the learning of a neural network greatly improves the accuracy of detection, but at the cost of time. Fortunately, with the use of more efficient computational resources in the future, time consumption will continue to decrease. The above analysis again exhibits the competitiveness of the proposed detector in HAD.

TABLE IV

<table>
<thead>
<tr>
<th>Datasets</th>
<th>WeaklyAD</th>
<th>SAFL_RX</th>
<th>RX</th>
<th>CRD</th>
<th>FrFE</th>
<th>LSMAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gulfport</td>
<td>0.4845</td>
<td>0.3827</td>
<td>0.2380</td>
<td>0.2380</td>
<td>0.2380</td>
<td>0.2380</td>
</tr>
<tr>
<td>Pavia</td>
<td>0.3146</td>
<td>0.2280</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2140</td>
</tr>
<tr>
<td>LA-1</td>
<td>0.3594</td>
<td>0.2617</td>
<td>0.2417</td>
<td>0.2417</td>
<td>0.2417</td>
<td>0.2417</td>
</tr>
<tr>
<td>San Diego</td>
<td>0.3146</td>
<td>0.2280</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2140</td>
<td>0.2140</td>
</tr>
<tr>
<td>Average</td>
<td>0.3884</td>
<td>0.2617</td>
<td>0.2417</td>
<td>0.2417</td>
<td>0.2417</td>
<td>0.2417</td>
</tr>
</tbody>
</table>

E. Generalization of Networks

In this section, we further discuss how the proposed architecture has the potential for generalization, that is, to generically reconstruct different datasets by a well-learned network. As far as we know, all known deep learning-based HAD methods are trained and tested on a single dataset without exploring the generalization of the method [31], which means that the generalization experiment in this article is of great significance. To be more convincing, we conduct experiments across six datasets from two groups: group I contains three original datasets, namely Gulfport, Los Angeles-1 (LA-1), and
San Diego, and group II includes three additional datasets, namely Texas Coast-1 (TC-1), Texas Coast-2 (TC-2), and Los Angeles-2 (LA-2), respectively. It is important to note that all the datasets for this section need to come from the same sensor, which is the reason we abandoned the Pavia dataset captured by the ROSIS-03 sensor and choose six datasets with different scenes from the AVIRIS sensor. Another point to note is that the number of bands in the reconstructed dataset should be the same as the number of bands in the learned dataset. The specific process is shown in Fig. 11. According to Table V, the San Diego dataset has the least number of bands after useless bands being discarded, so we first choose the San Diego dataset to learn the optimal parameters $\theta = (W, \beta, \hat{W}, \hat{\beta})$ in the learning process. Subsequently, for the reconstructing step, all the above six datasets are reconstructed by inputting the datasets into the network with the well-learned parameters $\theta = (W, \beta, \hat{W}, \hat{\beta})$. Finally, the RX detector is employed to obtain the detection results. The bands selected for the reconstruction of the six datasets are listed in Table V. The result maps of AD over six datasets and the detection maps of the comparison methods on datasets of group II are shown in Fig. 11. Comparing the AUC scores in Tables III and V for datasets in group I, the following observations are made. First, the AUC scores of $(P_d, P_f)$ in Table V are only slightly lower than those in Table III for the proposed weaklyAD method, and for other comparison algorithms, the AUC scores of $(P_d, P_f)$ in Table V are also better than those over the corresponding dataset in Table III. Second, false alarm rates do not vary much. Meanwhile, the comparison maps of visual detection in Figs. 9 and 11 are consistent with the above observations. Subsequently, according to Table VI, except dataset Texas Coast-1, the generalization experiment-based detection accuracy of the new additional datasets is higher than that of other comparison datasets. Thus, the conclusion can be drawn that the satisfactory results obtained by feeding several HSIs, which captured by the same sensor into the network learned by a certain HSI verify that our network has generalization property. In essence, the generalization performance of the proposed weaklyAD method opens a broad prospect for deep learning-based HAD in practical engineering applications.

V. CONCLUSION

In this article, we propose a novel weaklyAD method with the SCDGAN for HAD. The main attention of our method is to learn a discriminative end-to-end reconstruction with the background being homogenized and anomalies being salient. To this end, several aspects have been taken into consideration. First of all, based on the observation that anomalies have low occurrence probability, a novel category saliency searching is adopted to generate coarse anomaly and coarse background samples in preparation for WSL. Second, the proposed SCDGAN is learned to construct the reconstruction model over the coarse samples in a weakly supervised way. Moreover, the SC and the adversarial learning of the SCDGAN structure maximize the discriminative learning ability and

![Image of a diagram showing the process of network generalization and detection maps.

Table V: Details and AUC Scores of Network Generalization on Different Datasets

<table>
<thead>
<tr>
<th>Function</th>
<th>Datasets</th>
<th>Sizes</th>
<th>Selected bands $(P_d, P_f)$</th>
<th>$P_d$, $P_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>San Diego</td>
<td>100 $\times$ 100 $\times$ 189</td>
<td>1-189</td>
<td>0.99138, 0.01570</td>
</tr>
<tr>
<td></td>
<td>Gulfport</td>
<td>100 $\times$ 100 $\times$ 191</td>
<td>3-191</td>
<td>0.98254, 0.01838</td>
</tr>
<tr>
<td></td>
<td>LA-1</td>
<td>100 $\times$ 100 $\times$ 205</td>
<td>6-194</td>
<td>0.97907, 0.01484</td>
</tr>
<tr>
<td>Reconstructing</td>
<td>TC-1</td>
<td>100 $\times$ 100 $\times$ 204</td>
<td>6-194</td>
<td>0.99292, 0.01650</td>
</tr>
<tr>
<td></td>
<td>TC-2</td>
<td>100 $\times$ 100 $\times$ 207</td>
<td>6-194</td>
<td>0.99825, 0.01026</td>
</tr>
<tr>
<td></td>
<td>LA-2</td>
<td>100 $\times$ 100 $\times$ 205</td>
<td>6-194</td>
<td>0.99238, 0.00812</td>
</tr>
</tbody>
</table>

Table VI: Evaluation AUC Scores of Network Generalization of the Comparative Methods on Datasets in Group II

<table>
<thead>
<tr>
<th>Datasets</th>
<th>The AUC scores of $(P_d, P_f)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>WeaklyAD</td>
<td>SFAL RX</td>
</tr>
<tr>
<td>TC-1</td>
<td>0.99292, 0.98271, 0.99064, 0.99390, 0.98975, 0.97944</td>
</tr>
<tr>
<td>TC-2</td>
<td>0.99825, 0.99342, 0.99463, 0.95475, 0.99540, 0.99759</td>
</tr>
<tr>
<td>LA-2</td>
<td>0.99258, 0.98559, 0.98170, 0.96234, 0.97984, 0.97744</td>
</tr>
</tbody>
</table>
make the distribution of the background tend to be uniform and normalized. Third, a discriminative end-to-end reconstruction with the background being homogenized and anomalies being salient can be obtained by feeding all spectral samples into the reconstruction model. By doing so, anomaly-background separation can be further enhanced even when the anomalous samples are limited and similar to the background.

Furthermore, the effectiveness experiment ensures the reliability and realizability of the proposed ideas and methods. Experiments of detection quality assessment show that the weaklyAD method outperforms many state-of-the-art detection methods on four publicly available datasets. The empirical findings in the generalization experiment provide an insight into the application capability of the proposed method to real engineering tasks. The superior performance of our method exhibits the value of WSL in hyperspectral tasks. Thus, our work paves a new way of studying HAD.

REFERENCES

Jiang et al. WEAKLY SUPERVISED DISCRIMINATIVE LEARNING WITH SPECTRAL CONSTRAINED GAN 6517


Tao Jiang received the B.E. degree in telecommunications engineering from Xidian University, Xi’an, China, in 2018, where she is currently pursuing the Ph.D. degree with the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center.

Her research interests include anomaly detection, hyperspectral image, and deep learning.

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 currently 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 ON GEOSCIENCE 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.

In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor and the Director of the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center. 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 at the Department of Computer Science, University of California at Los Angeles, Los Angeles, CA, USA, from 2014 to 2015. He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University. He is also a member of the State Key Laboratory of Integrated Services Networks, Image Coding and Processing Center, Xidian University. 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 currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is also a fellow of the SPIE—International Society for Optics and Photonics. 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 has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS), the Journal of Applied Remote Sensing, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE J-STARS. 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.