Hyperspectral Anomaly Detection Based on Machine Learning: An Overview

Yichu Xu, *Graduate Student Member, IEEE*, Lefei Zhang ⓘ, *Senior Member, IEEE*,
Bo Du, *Senior Member, IEEE*, and Liangpei Zhang ⓘ, *Fellow, IEEE*

Abstract—**Hyperspectral anomaly detection (HAD) is an important hyperspectral image application. HAD can find pixels with anomalous spectral signatures compared with their neighbor background without any prior information. While most of the existed researches are related to statistic-based and distance-based techniques, by summarizing the background samples with certain models, and then, finding the very few outliers by various distance metrics, this review focuses on the HAD based on machine learning methods, which have witnessed remarkable progress in the recent years. In particular, these studies can generally be grouped into the traditional machine learning and deep-learning-based methods. Several representative HAD methods, including both traditional machine and deep-learning-based methods, are then conducted on four real HSIs in the experiments. Finally, conclusions regarding HAD are summarized, and prospects and future development direction are discussed.**

Index Terms—**Anomaly detection, deep learning, hyperspectral imagery, machine learning.**

I. Introduction

HYPERSPECTRAL anomaly detection (HAD), which refers to the process of finding pixels or subpixels with different spectral signatures compared with their neighbor background without any prior information, is an important task of hyperspectral image processing. As early as 2001, Prof. D. A. Landgrebe, the pioneer and leader of hyperspectral remote sensing research in IEEE GRSS, organized a special issue on analysis of hyperspectral image data, which collected very early researches on hyperspectral target detection, thus initiating the developments of the hyperspectral remote sensing target and

anomaly detection [1]. Thereafter, benefit from the scientific ideas as well as the open source software and datasets of the LARS at Purdue University, the world-wide well-known research groups, e.g., the laboratories at University of Iceland, Universidade de Lisboa, University of Extremadura, and the RS group at Wuhan University, have provided outstanding contributions in the field of hyperspectral data analysis. HAD enjoys a wide range of real-world applications, such as military reconnaissance [2], [3], camouflage identification [4], fine agriculture [5], mineral identification [6], change detection [7], and land cover classification [8]–[10]. Among these applications, HAD has elicited increasing interest because it can discover the potential targets of interest in the image scene in recent years.

However, current research faces the following three main challenges.

1) The spectrum is complex and changeable, which is usually caused by various factors, such as changes in illumination, environment, atmosphere, and time conditions.
2) The correlation between adjacent bands is strong, and the information is redundant, presenting a substantial bias in the estimation of the background statistical model.
3) The spatial resolution of the image is limited while that of hyperspectral remote sensing data is low, resulting in the widespread phenomenon of mixed pixels.

Low detection rates and high false alarms are thus involved due to the mixed pixels.

A variety of classical machine-learning-based HAD methods, including those based on feature mapping [11]–[14], representation [15]–[17], matrix decomposition [18]–[20], transformed-based [21], and GLRT-based [22] methods have been proposed in the literature and summarized in a highly recent published technical review [23]. In this review, we focus on the HAD based on machine learning methods, which have witnessed remarkable progress in recent years. In particular, the deep-learning-based HAD models, which can achieve state-of-the-art performance on various benchmarks of HAD, have been actively explored with the rapid development of deep learning techniques in recent years. A variety of deep learning methods, ranging from the unsupervised autoencoder (AE)-based method to recent promising HAD approaches using generative adversarial networks (GANs), have been applied to tackle HAD tasks.

A comprehensive overview of traditional machine learning methods to recent advances in HAD with deep learning is provided in this article. The literature contains some existing HAD surveys. Different from this study, the machine-learning-based

Yichu Xu and Lefei Zhang are with the Hubei Luojia Laboratory, Wuhan 430072, China, and also with the School of Computer Science, Wuhan University, Wuhan 430072, China (e-mail: xuyichu@whu.edu.cn; zhanglefei@whu.edu.cn).

Bo Du is with the National Engineering Research Center for Multimedia Software, School of Computer Science, Institute of Artificial Intelligence, Wuhan University, Wuhan 430072, China, and also with the Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University, Wuhan 430072, China (e-mail: dubo@whu.edu.cn).

Liangpei Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: zlp62@whu.edu.cn).

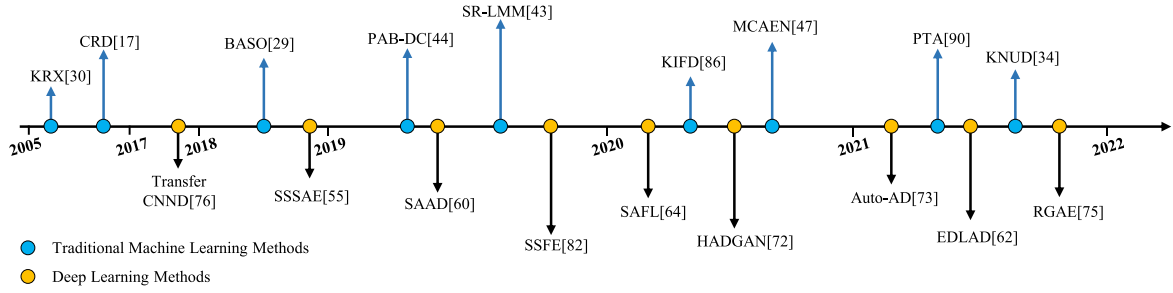Digital Object Identifier 10.1109/JSTARS.2022.3167830

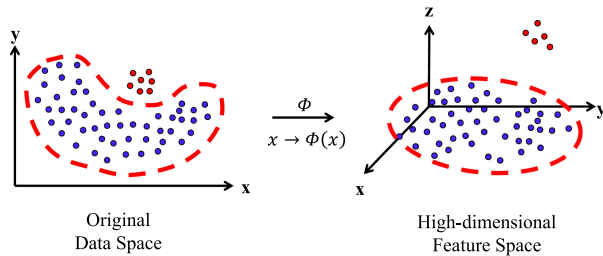Fig. 1. Milestones of hyperspectral anomaly detection.

Fig. 2. Kernel feature projection by kernel learning methods for hyperspectral anomaly detection.

methods are comprehensively reviewed, and deep learning-based HAD methods are emphasized. The main contributions of this survey are twofold.

1) A comprehensive review of HAD methods based on traditional and deep learning-based methods, which are advanced from two stages, is provided.

2) A systematic overview of past decades of traditional machine learning-based methods and recent advances of deep learning is presented hierarchically and structurally, and the advantages and limitations of each component are summarized for an effective HAD solution.

The following sections will cover various aspects of traditional and deep-learning-based methods in the HAD task. Fig. 1 shows the milestones of HAD, including traditional machine-learning-based method before 2017, and with methods after 2017 dominated by related deep networks. Section II provides a brief introduction to traditional machine-learning-based methods. Section III analyzes the main deep-learning-based structure of HAD methods. Section IV compares several representative HAD methods using real HSIs. Finally, Section V concludes this article.

II. TRADITIONAL MACHINE-LEARNING-BASED MODELS

A. Feature Mapping-Based Method

Different from the statistic-based HAD method, feature mapping is one of the machine learning techniques, and it usually maps the original data space into a new feature space that can enhance the separability of the background and target class. Fig. 2 shows the kernel feature projection by kernel learning methods for HAD.

1) Kernel Learning-Based Method: The kernel feature space has the same assumptions as those used in the RX [23] algorithm; that is, the mapped data in the feature space currently comprise two Gaussian distributions, thus modeling the two hypotheses as follows:

$$H_0 : \Phi\left(x_i\right) = n_\Phi(\text{target absent})$$

$$H_1 : \Phi\left(x_i\right) = \Phi(s) + n_\Phi(\text{target present}). \quad (1)$$

The corresponding kernel RX in the high-dimensional feature space is now represented as

$$\text{RX}\left(\Phi\left(x_i\right)\right) = \left(\Phi\left(x_i\right) - \mathbf{u}_\Phi\right)^T \Sigma_\Phi^{-1}\left(\Phi\left(x_i\right) - \mathbf{u}_\Phi\right). \quad (2)$$

The kernelization of the RX in the kernel feature space [27] considering the eigenvector decomposition and kernel methods can be obtained as follows:

$$\text{KRX}\left(x_i\right) = \left(\boldsymbol{K}_{x_i}^T - \boldsymbol{K}_u^T\right)^T \boldsymbol{K}_\Sigma^{-1}\left(\boldsymbol{K}_{x_i}^T - \boldsymbol{K}_u^T\right) \quad (3)$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$ are the samples from the hyperspectral image, and m is the number of samples.

The following algorithms published in the last decades use this idea. The kernel RX [30] provides a nonlinear version of the RX for HAD considering the nonlinear characteristic and high-order relationship of the spectral signatures. The kernel PCA [31] attempts to maximize the anomalous components and suppress the background components temporarily. The support vector data description [32] uses a single-class SVM classifier to estimate the minimum enclosing hypersphere containing the most normal instances, wherein pixels that fall outside of the hypersphere are classified as anomalies. The robust nonlinear anomaly detector [33] combines the kernelization procedure and robust iterative background distribution estimation strategy to formulate a nonlinear and robust detector, which can effectively estimate the background distribution and avoid contamination to a certain extent. The kernel-based nonlinear anomaly detection method via union dictionary [34] considers the nonlinear mixing models to demonstrate the intrinsic nonlinear characteristics of the real HSIs. A union dictionary comprising a background and anomalous atoms is also constructed by considering the local spatial correlations and global anomalies. Li *et al.* [86] proposed a iForest-based HAD method as well as its kernel version. It can fully use the global and local information in an HSI.

2) Manifold Learning-Based Method: Manifold learning methods are used for nonlinear dimensionality reduction [35],

and can discover the compact and nonlinear representation of the given high-dimensional data. Manifold learning methods can also reveal the important and intrinsic structure of the given high-dimensional data by preserving distances within the local neighborhoods of the data. Researchers usually apply manifold learning to account for the local structural information for HAD task. In the learned manifold, the samples with similar spectral signatures are clustered and those with different spectrums are far apart, which is helpful for the HAD task.

Lu *et al.* [47] proposed a manifold constrained AE network (MC-AEN)-based HAD method. The embedding manifold is first learned by the manifold learning method. Subsequently, the latent representations are learned by an AE network with the embedding manifold constraints to preserve the intrinsic structure of hyperspectral data. The anomaly scores are indicated by reconstruction errors. A manifold constrained multihead self-attention variational autoencoder (MMS-VAE) method for HAD is previously proposed [24]. The manifold learning method is utilized to first learn the embedded manifold, which is used to constrain the VAE to learn the latent feature representation and maintain the internal structure of hyperspectral data. The multihead self-attention mechanism is introduced to learn context-related information to automatically focus on abnormal areas. Finally, the respective global and local reconstruction errors of the multihead self-attention network and the latent feature space are simultaneously considered to determine whether each pixel is abnormal.

3) Orthogonal Subspace-Based Method: The subspace-based anomaly detection algorithm considers that anomaly and background pixels have a greater degree of separation in a suitable subspace. The most representative subspace method is orthogonal subspace projection (OSP). It is a versatile hyperspectral imaging technique that has shown great potential in dimensionality reduction and target detection. Orthogonal subspace-based methods project testing pixels into an orthogonal subspace of the background components where the background and anomalies can be better separated. The OSP of an abnormal pixel is a large component, while the projection of a background pixel is a small one. Thus, the component projected to the subspace is the abnormal level of the testing pixels.

The local 3-D OSP HAD method constructs the OSP from the height, width, and spectrum of hyperspectral images to realize a double utilization of spectral and spatial information [25]. Yang *et al.* [26] proposed a low-rank and sparse matrix decomposition with OSP-based background suppression and adaptive weighting for HAD. The OSP is employed to project the sparse component into the background orthogonal subspace that is estimated from the low-rank component to suppress the background interferences and highlight the anomalies. Chang *et al.* [27] presents an OSP version of go decomposition (OSP-GoDec), which implements GoDec in an iterative process by a sequence of OSPs to find desired low-rank and sparse matrices.

B. Representation-Based Method

In addition to the aforementioned methods, representation-based detectors, such as sparse, collaborative, and tensor representation-based anomaly detectors, have also received considerable attention. A background joint sparse representation detector was proposed in [15]. This detector assumes that the background pixel can be reconstructed by a linear combination of a small number of atoms in an overcomplete background dictionary. Zhao *et al.* [16] proposed a sparsity score estimation framework, which utilizes two strategies to enhance the diversity between the background and anomaly information. Different from sparse representation, collaborative representation considers that background pixels can be represented by a linear combination of surrounding pixels while anomalies cannot [45]. Li and Du proposed a collaborative representation detection (CRD) [18] method according to this assumption. After that, Hou *et al.* [87] proposed a CRD extension version, which first purifies the background in HSI. Simultaneously, saliency analysis was designed to further improve the detection performance to make full use of spatial information. Fig. 3 shows the flowchart of sparse and collaborative representation-based HAD methods.

1) Sparse Representation-Based Method: Sparse representation (SR) has been widely studied in image processing [46]. For a given dictionary, the signals are the sparsest linearly represented via atoms of the dictionary with sparsity regularization. The objective function can be explained by solving the following optimization problem:

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_0 \text{ s.t. } \mathbf{X} = \mathbf{DZ} \tag{4}$$

where $\|\cdot\|_0$ denotes the l_0 norm, which is used to evaluate the number of nonzero elements of a matrix. $\mathbf{D}$ is a dictionary, $\mathbf{Z}$ is the set of sparse coefficient vectors, and $\mathbf{X}$ is the set of test signals. The problem of (7) is an NP-hard problem. This problem can be relaxed in the following form:

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_1 \text{ s.t. } \mathbf{X} = \mathbf{DZ} \tag{5}$$

where $\|\cdot\|_1$ denotes the l_1 norm to compute the sum of the absolute value of all the nonzero entries in the matrix.

2) Collaborative Representation-Based Method: Collaborative representation (CR) assumes that collaboration among atoms improves representation accuracy [47]. For each pixel $\mathbf{x}$ of size $b \times 1$, where b is the number of bands, all the atoms from the reconstruct dictionary $\mathbf{B} \in R^{b \times s}$, where s is the number of the atoms, are used in the representation. Thus, $\mathbf{x} \approx \mathbf{B}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a weight vector, can be approximately obtained. l_2 norm is adapted to constrain the $\boldsymbol{\alpha}$, and all the atoms in $\mathbf{B}$ are assigned with similar small coefficients. The CR of pixel $\mathbf{y}$ by dictionary $\mathbf{B}$ can be formulated as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{B}\boldsymbol{\alpha}\|_2^2 + \lambda\|\boldsymbol{\alpha}\|_2^2 \tag{6}$$

where λ is a Lagrange multiplier, which controls the penalty of the norm of weight vectors.

The CRD algorithm used the collaborative representation technique for the first time in the HAD task [18]. Atoms in the reconstructed dictionary $\mathbf{B}$ contain major information of the normal pixels. Thus, the objective is to find a weight vector $\boldsymbol{\alpha}$ in which $\|\mathbf{x} - \mathbf{B}\boldsymbol{\alpha}\|_2^2$ is minimized under the constraint that $\|\boldsymbol{\alpha}\|_2^2$

Fig. 3. Flowchart of sparse and collaborative representation-based HSI anomaly detection method.
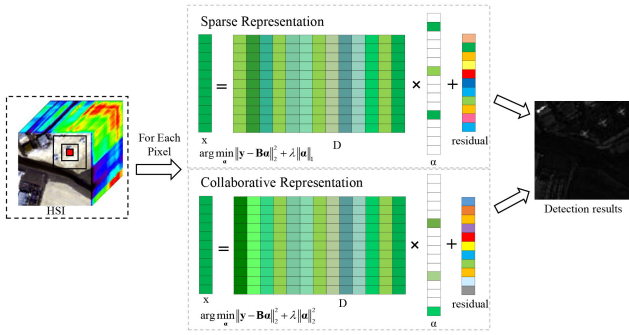

Fig. 4. Flowchart of the reconstruction-based HSI anomaly detection method.

is also minimized. Equation (6) is equivalent to

$$\arg \min_{\boldsymbol{\alpha}} \left[\boldsymbol{\alpha}^T \left(\mathbf{B}^T \mathbf{B} + \lambda I \right) \boldsymbol{\alpha} - 2 \boldsymbol{\alpha}^T \mathbf{B}^T \mathbf{x} \right]. \tag{7}$$

Taking the derivative of $\boldsymbol{\alpha}$ and setting the resultant equation to zero yields

$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{B}^T \mathbf{B} + \lambda I \right)^{-1} \mathbf{B}^T \mathbf{x}. \tag{8}$$

The residual of the pixels can be obtained as the detection score upon completion of the representation process.

$$r_1 = \| \mathbf{x} - \hat{\mathbf{x}} \|_2 = \| \mathbf{x} - \mathbf{B} \hat{\boldsymbol{\alpha}} \|_2. \tag{9}$$

C. Matrix Decomposition-Based Method

1) Low-Rank and Sparse Matrix Decomposition-Based Method: Researchers have recently provided considerable effort into extracting knowledge from backgrounds and potential anomalies [48], which provided more comprehensive information than those methods exploring only the background information. Therefore, several methods first take the low-rank and sparse matrix decomposition (LRaSMD) technique [49], [50], which decomposes the HSI matrix into low-rank, sparse, and noise matrices, and then, design a suitable detector for further processing.

Some algorithms obtain detection results by analyzing the sparsity of anomalies. Sun *et al.* [50] scored each pixel with the Euclidean distance between the corresponding sparse component vector and the mean vector of the sparse matrix. Cui *et al.* [48] scored each pixel by computing the norm of its corresponding sparse component vector. Some algorithms also design detectors based on the low rank of background. Zhang *et al.* [49] proposed the LRaSMD-based Mahalanobis distance method for HAD (LSMAD). In [51], Xu *et al.* proposed the low-rank and sparse representation (LRASR) HAD method, where a background dictionary was introduced and sparse constraints on the representation coefficient were imposed. The spectral-difference LRDL [52] is introduced to further explore the low-rank characteristic of spectral contextual information. Li *et al.* [88] proposed the LSDM-MoG detector, which based on the LSDM model in conjunction with the mixture of Gaussian to better fit the various data distribution. Furthermore,
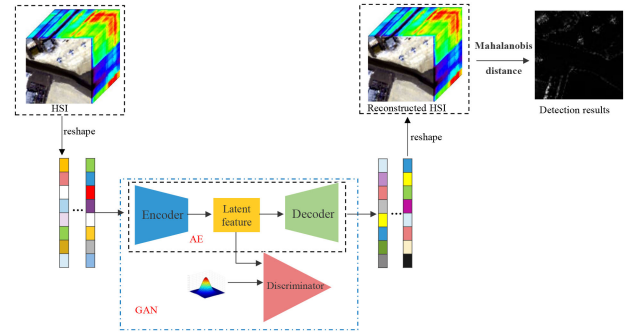
the dictionary-based low-rank decomposition algorithms were applied on the background dictionary.

2) Tensor Decomposition-Based Method: HSI is a 3-D data cube, which can be intrinsically treated as a third-order tensor. However, most of the current HAD methods, including the aforementioned methods, neglect the 3-D structure in HSI. Thus, a tensor has been employed to advance the structure for integrated consideration of all the dimensions to address this problem. Several approaches based on tensors have been proposed for HAD, thus gradually becoming a popular task.

Yang *et al.* [36] proposed an HAD method through sparse representation with tensor decomposition-based dictionary construction and adaptive weighting. Ma *et al.* [37] proposed a hyper-Laplacian regularized low-rank tensor decomposition method combing with dimensionality reduction framework which considering the spectral-spatial information at the same time. In [38], a fast and adaptive method for determining the major principal component numbers (K1, K2, and K3) along the three modes of hyperspectral data is proposed. However, this tensor is still dense in spectral dimension, thus indicating the existence of a group sparse prior in the anomaly tensor. Therefore, a prior-based tensor approximation (PTA) [90] method is proposed for HAD, which combines low-rank, sparse, and piecewise smooth with the advantages of tensor representation of HSI.

III. DEEP-LEARNING-BASED MODELS

With the development of deep learning techniques, numerous scholars have focused on hyperspectral anomaly detection tasks with deep-learning-based methods due to the capability to extract deep features. These tasks elicited several HAD studies based on deep learning. The tasks are divided into the following four categories: reconstruction-based, residual error-based, feature learning-based, and deep learning supported methods.

A. Reconstruction-Based Method

The first mainstream idea is to utilize deep learning technology to reconstruct the background component in the original hyperspectral data cube effectively. However, the anomaly part could not be reconstructed successfully temporarily. Therefore, the reconstructed hyperspectral data can be considered to obey a

Gaussian normal distribution. Finally, the Mahalanobis distance is usually applied, or some postprocessing is used to obtain the detection results. The methods of this part are divided into the following two categories according to the network architecture used in the reconstruction process: the AE-based and the GAN-based methods. Fig. 4 shows the flowchart of reconstruction-based HAD methods. The entire HSI image is fed into the reconstruction network (AE- or GAN-based), with some constraints, such as sparse and manifold constraints, to reconstruct the background components effectively.

Considering that the anomalies spectrum is significantly different from their neighbors, the HAD task is thus very concerned with the structural information of HSI data. Given that manifold learning methods can preserve distances with the local neighborhoods of the data [40], the meaningful and intrinsic structure of the given high-dimensional data can be revealed. In the learned manifold, the samples with similar spectral signatures are clustered and those with different spectrums are far apart. Researchers commonly apply this method to constrain the learning of the latent representation of AEs to create the latent representation with local structural information, which is helpful for the HAD task. As for sparse constraint, the AE requires that the output be equal to the input as much as possible, and its hidden layer must satisfy a certain sparsity. Thus, the hidden layer compresses the input, and then, decompresses it in the output layer. To ensure the sparsity of the hidden layer, a sparsity penalty term Kullback–Leibler divergence is generally added to the cost equation of the AE.

1) AE-Based Method: As a self-taught learning neural network, AE comprises an encoder and a decoder. The parameters of the network, including the weights and bias, are estimated by the backpropagation method [58]. The encoder is used to learn a mapping from the input layer x to the hidden layer y.

$$y = f(\mathbf{x}; \mathbf{W}, \mathbf{b}) \tag{10}$$

where $\mathbf{W}$ and $\mathbf{b}$ denote the weights and bias, respectively, and f denotes an activation function. The decoder is used to learn a mapping from the hidden layer y to the output layer $\tilde{x}$.

The AE aims to generate an output that approximates the input, extracting the latent features of the input. Thus, the loss function is the root-mean-square error, and the parameters can be updated using the ADAM algorithm [59]:

$$L = \|x - \tilde{x}\|_2. \tag{11}$$

The AE has been previously used to detect anomaly instances by learning anomaly scores. These processes train the parameters of encoder/decoder function by a known background training set, contributing to the effective performance of this semisupervised learning method on background replication. The reconstruction error is used as an anomaly score considering that the reconstruction error of background data will be small while those of anomaly data will be relatively large. However, the HAD research cannot provide any prior knowledge of the background training set. Moreover, the aforementioned basic AE cannot be applied to detect anomaly objects directly.

The following methods proposed in recent years are all based on AE based on the aforementioned problems and challenges.

Combining with the spatial correlation of HSI, Chang *et al.* [60] proposed a sparse AE-based AD using a dual concentric window. Zhang *et al.* [61] introduced a 3-D-convolutional variational AE-based HAD method, which can maximize the spectral-spatial information. The anomalies can be easily detected by the RX detector in the residual between the original input and the reconstructed background. Zhu *et al.* [62] proposed an encoder–decoder long short-term memory-based anomaly detector (EDLAD) for hyperspectral images. The EDLAD first utilizes a well-designed encoder–decoder LSTM to reconstruct the hyperspectral image, which tends to maintain the background and alleviate anomaly during the reconstruction process because the entire image is employed for training the network. The Mahalanobis distance is calculated to detect the probable anomalies after using the dimension reduction to further alleviate the anomaly contamination. Lei *et al.* [63] proposed a discriminative reconstruction method for HAD with spectral learning (SLDR). This method imposes a constraint on the encoder, thus forcing it to generate latent variables that follow a Gaussian distribution. The loss function of the SLDR model additionally introduces the SAD term, which constrains the model to generate a reconstruction with remarkable similarity to the input. An unsupervised spectral adversarial feature learning (SAFL) architecture, which has a powerful feature representation capability for high-dimensional HSIs, is especially designed for HAD in [64]. The morphological attribute filter is introduced to strengthen anomalies and remove the background of the fusion feature.

2) GAN-Based Method: The idea of GANs was initially proposed in [65] within a zero-sum game framework in the game theory based on the competition of two networks. Two models are available in adversarial iterative training: the generator model G and the discriminator model D. Derived from the basic framework of GANs, many variants of GANs have been proposed through changing objective function and architecture. Most existing GAN-based methods are trained on normal vectors and even normal and anomalous vectors, which are unsuccessfully applied to HAD. Therefore, their applications are limited without any normal or anomalous training samples [66]. Some scholars have also recently attempted to use the GAN to reconstruct HSI and realize HAD.

Zhong *et al.* [67] proposed a characterization of background-anomaly separability with a GAN for HAD to approximate the performance of the supervised method while eliminating the limitation of training samples. The key contribution is the proposal to constrain the background and anomaly separability explicitly by characterizing background spectral samples while avoiding anomaly reconstruction. Arisoy *et al.* [68] proposed a completely unsupervised pixel-wise anomaly detection method for hyperspectral images. Jiang *et al.* [69] introduced a discriminative semisupervised GAN with dual RX (semiDRX), which learns a discriminative reconstruction of background homogenization and anomaly saliency. Xie *et al.* [70] proposed an AE and adversarial learning-based semisupervised background estimation model, which is trained only on the background spectral samples to learn the background distribution accurately. An unsupervised background searching method is first conducted
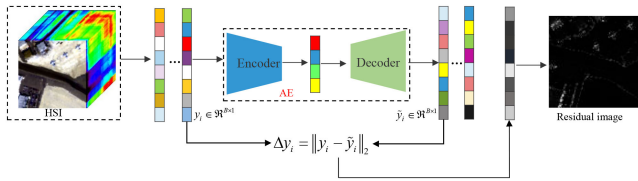
Fig. 5. Flowchart of the residual error-based HSI anomaly detection method.

on the original HSIs to search the background spectral samples. A weakly supervised discriminative learning with a spectral constrained GAN for HAD was proposed in [28]. The proposed method mainly focus on learning a discriminative end-to-end reconstruction with the background being homogenized and anomalies being salient.

B. Residual Error-Based Method

Another mainstream idea is to utilize the deep learning methods, such as AE, GAN, and its variant, to reconstruct a background successfully using the neural network, while the anomalies appear as reconstruction errors. The large errors always indicate that the corresponding pixels are anomalies, while the small errors corresponding to pixels are usually regarded as backgrounds. Fig. 5 shows the flowchart of residual error-based HAD methods. This idea mainly reconstructs background pixels successfully through the designed network. Meanwhile, reconstructing anomaly pixels with the network is difficult because anomalies are relatively smaller than the background and only occur in the image with a low probability. Thus, the reconstruction errors are used as anomaly scores. The difficulty of this processing idea lies in the development of a well-designed network to reconstruct the background pixels successfully in the entire HSI images while suppressing the anomaly pixels reconstructed temporarily. Some researchers proposed some methods based on this idea according to the aforementioned challenges and difficulties.

Zhao et al. [71] proposed a spectral-spatial HAD method via collaborative representation constraint stacked AEs. The collaborative representation constraint is imposed on the stacked AEs to extract deep nonlinear features. The CRD is then used for obtaining the detection result. Jiang et al. [72] proposed an unsupervised discriminative reconstruction constrained GAN for HAD (HADGAN). The HADGAN mainly aims to learn a discriminative background reconstruction with the suppression of anomaly targets. The initial detection image was obtained by the residual between the original and reconstruction images. An energy-based spatial and distance-based spectral joint anomaly detector is applied in the residual map to generate the final detection map. The existing AE-based methods are complicated and involve manual parameter setting and preprocessing and postprocessing procedures. Thus, an autonomous HAD network (Auto-AD) is proposed in [73]. In this network, the background is reconstructed by the fully convolutional AE network, and the anomalies appear as reconstruction errors. An adaptive weighted

loss function is designed to further suppress the anomaly reconstruction, wherein the weights of potential anomalous pixels with large reconstruction errors are reduced during training. The latent features learned from the AE occasionally fail to reflect the intrinsic structure of hyperspectral data because the locality property is disregarded during the learning process. Li et al. [74] proposed a sparse coding-inspired GAN for weakly supervised hyperspectral AD (sparseHAD), wherein a spectral mapping model is formed to reconstruct background samples with small errors. Fan et al. [75] introduced a robust graph AE detector, which is robust to noise and anomalies during training. Meanwhile, a superpixel segmentation-based graph regularization term (SuperGraph) was embedded into AE. This strategy can preserve the geometric structure and the local spatial consistency of HSI simultaneously and also effectively reduce the searching space and execution time for each pixel.

C. Feature Learning-Based Method

Another mainstream idea is to utilize pixel-wise feature learning to obtain additional discriminative features. Different from the reconstruction-based HAD methods used to acquire the same dimensional 3-D cube, this approach mainly uses the neural network to learn a discriminate feature from the original HSI data. However, the RX detector may also be used to obtain the detection results after acquiring the discriminate feature. These methods typically utilize deep learning to extract features and use the extracted features to design the subsequent detector and complete the HAD task. The difficulty of this kind of method lies in designing a neural network to extract features conducive to subsequent detection tasks. Many scholars have introduced effective solutions as follows based on existing defects and problems.

Li et al. [76] proposed a transferred deep CNN-based strategy for HAD. A reference dataset was employed to train the CNN with pixel pairs generated from labeled samples. Differences between pixel pairs are constructed by combining with its surrounding pixels for each pixel under testing. The detection results are obtained by the trained CNN with the results of similarity measurement. Zhang et al. [77] introduced a TCNNT for HAD, which is an unsupervised CNN model and employs the test and local neighboring tensors as the convolution kernel to extract deep features from the dictionary tensor. TCNNT can effectively utilize the spatial and spectral information of HSI. Furthermore, Zhang et al. [78] proposed a fractional Fourier transform and transferred CNN based on tensor (FrFTTCNNT) for HAD. The FrFTTCNNT employs tensor transformation and PCA as the preprocessing and combines FrFT and TCNNT. FrFT can effectively deal with the inherent nonstationary noise in HSI and increase the discrimination between background and targets. Zhao et al. [79] proposed bilateral-filtered GANs for HAD, which can effectively remove noise and anomalous pixels from an HSI after using dual window and adaptive parameter selection. Zhao et al. [80] introduced spectral-spatial stacked AEs based on the bilateral filter for HAD. The bilateral filter is

employed to obtain the derived anomaly and background components. The stacked AEs are utilized on the derived anomaly and background components for deep features. Ouyang *et al.* [81] introduced an unsupervised model (Line-wise SAFL) for HAD. This model utilizes LSTM in the AE network architecture, which uses spatial information to learn additional distinct latent features. A latent discriminator constrains the latent feature to follow the Gaussian distribution, which enables the RX to detect anomalies on the latent features directly. Lei *et al.* [82] proposed an approach to HAD, which depends on spectral and spatial feature extraction. A suppressed data space first extracted deep spectral features via DBN, and the RX detector is utilized to detect anomalies. Additionally, a simple differential operation is used to remove BKG and employ the guided filter (GF) to preserve the edge information and anomalies. The final detection map is achieved by combining the spectral and spatial results. Xie *et al.* [83] presented a spectral distribution-aware estimation network, which does not conduct feature extraction and anomaly detection in two separate steps but instead learns both jointly to estimate anomalies directly in an end-to-end manner without postprocessing. The unified framework can ensure that the extracted features are effective for HAD.

D. Deep Learning Supported Method

Some scholars use deep learning as an intermediate tool, which is typically utilized at endmember extraction or denoising models. The postprocessing utilizes the nondeep learning methods to obtain the final detection maps. This part mainly uses deep learning to obtain intermediate abundance or a denoise prior. Thus, the subsequent processing obtains a higher precision result than that only using the traditional methods to design a detector. This part of the method relies on the intermediate results obtained by the deep learning part. Thus, designing a good neural network is crucial in the preprocessing stage. The deep learning of this part is mainly used to deal with the part outside of an HAD task; thus, studies on this kind of idea are limited. Some scholars have proposed several methods combined with endmember extraction or based on denoising.

Song *et al.* [84] proposed an HAD method based on CNN and LRR. A CNN model is first built and trained on HSI datasets to obtain the resulting abundance maps accurately. A dictionary is then constructed by the DBSCAN to represent the background component stably. A low-rank constraint on the LRR method contributes to its suitability for abundance map data. The anomaly matrix was obtained when the convergence condition in LRR was reached. Instead of cumbersomely handcrafting a regularizer for representation coefficients, Fu *et al.* [85] proposed a plug-and-play prior for representation coefficients and constructed a new dictionary based on clustering. The deep learning technique has been applied to the image denoising area, showing promising results. An effective CNN prior is plugged into the framework to exploit the spatial correlation of representation coefficients. A modified background dictionary construction method, which carefully includes background pixels and excludes anomalous pixels from clustering results, was also proposed. An unsupervised low-rank embedded network

(LREN) was proposed in [39]. It searches the lowest rank representation based on a representative and discriminative dictionary in the deep latent space to estimate the residual efficiently. Spectral features and discriminative dictionary are learned by jointly training the autoencoders and a Gaussian mixture model. Then, an LRR-based method is performed to detect anomalies in the deep latent space.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Datasets and Setup Set

The PAVIA dataset, which was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, is a real-world dataset of the city center of Pavia in northern Italy. The spatial resolution of this image is 1.3 m per pixel. The image scene covers an area of 108×120 pixels, with 102 spectral bands in wavelengths ranging from 430 to 860 nm [29]. The main background materials are bridges and water. Some vehicles and bare soil are, respectively, found on the bridge and near the bridge pier, which is represented by 43 pixels in total and was selected as the anomalies to be detected.

The second and third datasets were collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in San Diego, CA, USA [89]. The spatial resolution is 3.5 m per pixel. The image has 224 spectral channels in wavelengths ranging from 370 to 2510 nm. A total of 189 bands remained after removing the low SNR, water absorption, and bad bands. The size of the second HSI data is 100×100 pixels, and three planes, which are represented by 58 pixels in total, were chosen as the anomalies to be detected in the scene. The size of the third HSI data is 60×60 pixels, and 22 planes covering 214 pixels were chosen as the anomalies to be detected in the scene.

The last dataset was collected by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor.[1] It is a real-world HSI downloaded from the U.S. Army Engineer Research and Development Center website. The spatial resolution is 2 m/pixel, and the spectral resolution is 10 nm. The original dataset has 210 spectral bands in the visible near-infrared range from 400 to 2500 nm. However, only 162 spectral bands remained after the water absorption, and low SNR bands (1–4, 76, 87, 101–111, 136–153, and 198–210) were removed. The scene is clustered with different land-cover types, and a scene of 80×100 pixels was selected for the experiments. Parking lot, soil, water, roads, and vehicles, which are represented by 17 pixels in total, were chosen as the anomalies to be detected.

B. Compared Methods and Parameter Settings

1) Compared Methods: Several HAD methods, including RX [11], KRX [30], KIFD [86], CRD [18], LSMAD [49], LRASR [51], PTA-HAD [90], CNND [76], RGAE [75], MCAEN [47], Auto-AD [73], and HADGAN [72], were investigated in this article. The aforementioned methods comprise traditional machine-learning-based and deep-learning-based models, which contains 12 methods in total. The 3-D and 2-D receiver operating characteristic

---

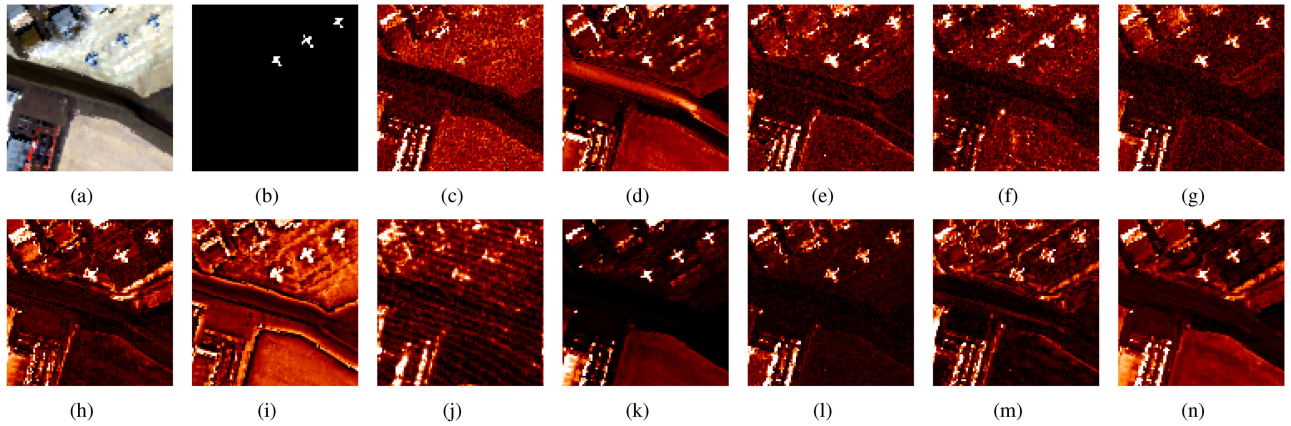[1][Online]. Available: http://www.tec.army.mil/Hypercube

Fig. 6.   AVIRIS-I dataset and color maps of the detection results. (a) Image scene. (b) Ground truth. (c) RX. (d) KRX. (e) KIFD. (f) CRD. (g) LSMAD. (h) LRASR. (i) PTA-HAD. (j) CNND. (k) RGAE. (l) MCAEN. (m) Auto-AD. (n) HADGAN.
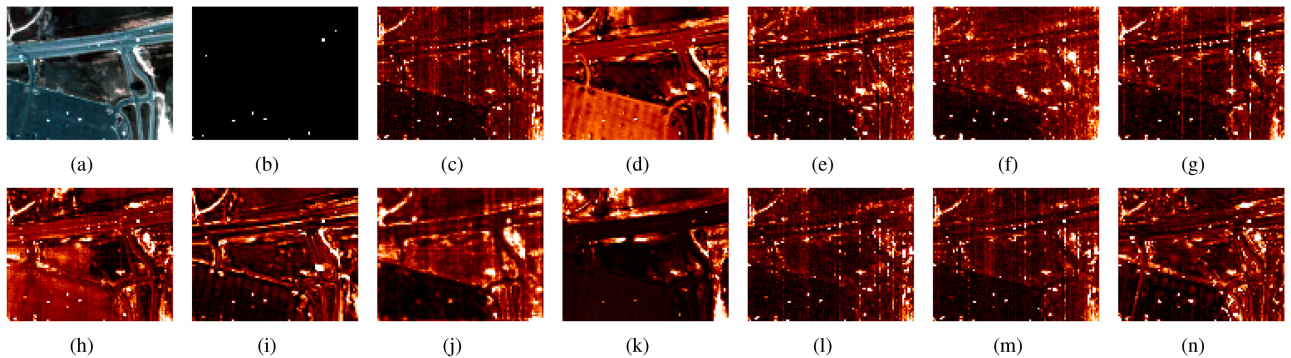

Fig. 7.   HYDICE dataset and color maps of the detection results. (a) Image scene. (b) Ground truth. (c) RX. (d) KRX. (e) KIFD. (f) CRD. (g) LSMAD. (h) LRASR. (i) PTA-HAD. (j) CNND. (k) RGAE. (l) MCAEN. (m) Auto-AD. (n) HADGAN.

(ROC) curves [11] and the area under the curve (AUC) values [12] were used as the evaluation criterions for experimental results. Furthermore, there exist two versions of the 2-D ROC curves. In one version, the probability of false alarm (P_f) and the probability of detection (P_d) were used as abscissa and ordinate, respectively. While in the other version, thresholds τ and P_f were used as abscissa and ordinate, respectively. AUC value is the area under the 2-D ROC curve, and the ideal values of $AUC(P_d, P_f)$ and $AUC(P_f, \tau)$ are 1 and 0, respectively.

2) Parameter Settings: For the CRD detector, the sizes of the two dual windows size are 15 and 17 for the all datasets. The optimal regularization parameter λ is set as 10^{-6} suggested by previous literature [18]. According to the original description of LRASR, the number of clusters K and the number of selected pixels P are set to 15 and 20, respectively, and the regularization parameters β and λ were set to 0.1 for all four experimental datasets. For the KIFD method, the default parameters are set the same as the original article. Furthermore, the number of principal components is set as 300. The parameters of PTA-HAD are set accordingly [90]. The tradeoff parameter λ, number of superpixels S, and the dimension of hidden layer n_{hid} in the RGAE method are set as optimal [75]. For the MCAEN method, the range of the dimension of embedding manifold d

and tradeoff parameter α are set following previous study [47]. The parameters of the methods Auto-AD and HADGAN are set according to [72] and [73], respectively.

C. Detection Results

Figs. 6–9 show the detection maps on the 12 HAD methods including seven traditional-based HAD methods and five deep-learning-based HAD methods for the four real HSIs datasets. For example, on the AVIRIS-I dataset, it can be seen that, although KIFD, CRD, LSMAD, and LRASR can detect all the anomaly targets in the scene, they cannot suppress the background components well. The KIFD method and the LSMAD method occur the higher false alarms in the lower left area of the image, while the CRD and LRASR methods produce the higher false alarms in the upper left area of the image. As for the PTA-HAD method, it can obtain really good performance in background suppression, the anomaly targets could also be highlighted in the scene. While in the deep learning methods, except for the CNND and Auto-AD, the other three methods have relatively good results. Especially for the RGAE method, it can obtain more convincing anomaly detection results and also have a better background suppression performance. In general, the methods based on deep
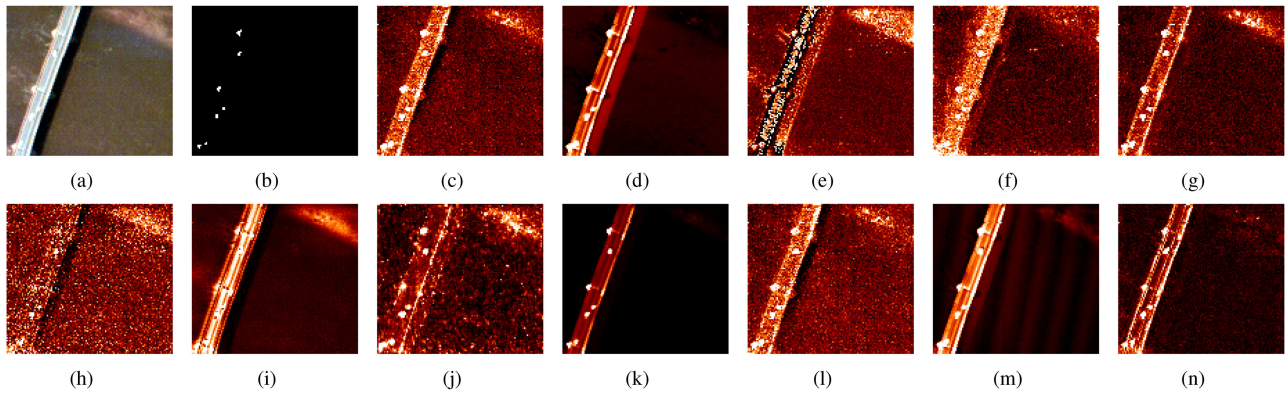
Fig. 8. PAVIA dataset and color maps of the detection results. (a) Image scene. (b) Ground truth. (c) RX. (d) KRX. (e) KIFD. (f) CRD. (g) LSMAD. (h) LRASR. (i) PTA-HAD. (j) CNND. (k) RGAE. (l) MCAEN. (m) Auto-AD. (n) HADGAN.
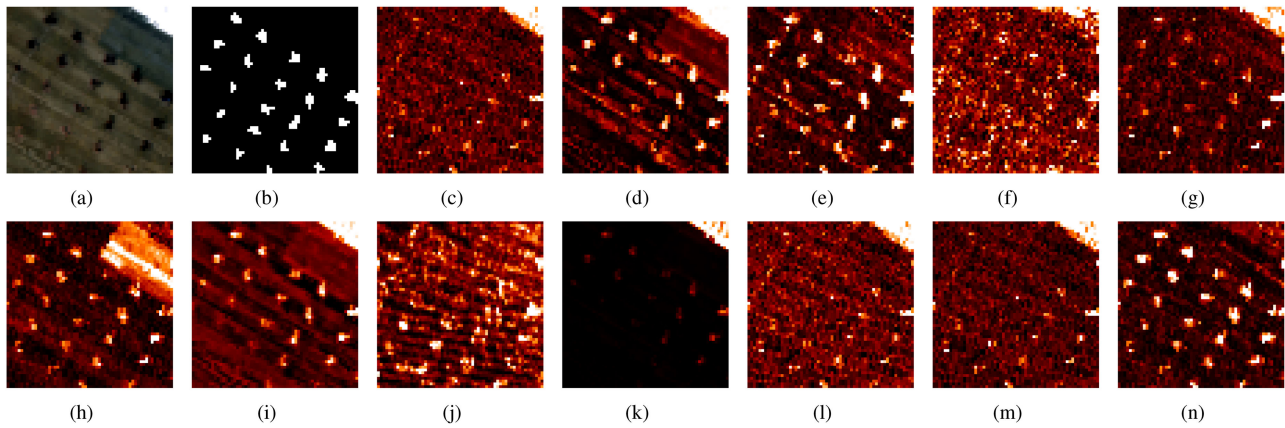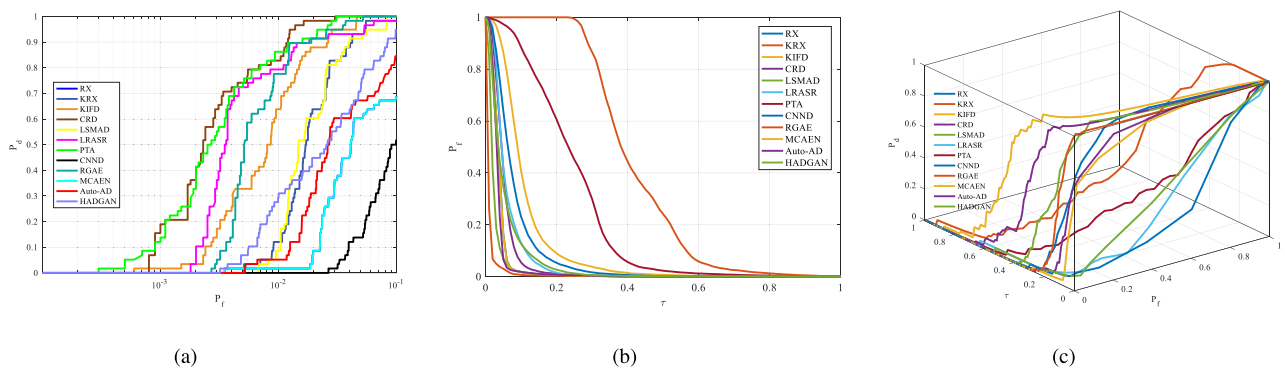

Fig. 9. AVIRIS-II dataset and color maps of the detection results. (a) Image scene. (b) Ground truth. (c) RX. (d) KRX. (e) KIFD. (f) CRD. (g) LSMAD. (h) LRASR. (i) PTA-HAD. (j) CNND. (k) RGAE. (l) MCAEN. (m) Auto-AD. (n) HADGAN.

Fig. 10. ROC curves of the 12 methods for the AVIRIS-I dataset. (a) Two-dimensional ROC curves (P_d, P_f). (b) Two-dimensional ROC curves (P_f, τ). (c) Three-dimensional ROC curves.

learning show better results in background suppression, which can effectively suppress background information such as noise in images.

Figs. 10–13 show three kinds of ROC curves corresponding to the detection maps. It can be observed that the 2-D ROC curve of (P_d, P_f), the KIFD, HADGAN, and RGAE method

is much closer to the upper left corner than other methods for the AVIRIS-II, HYDICE, and PAVIA dataset, respectively. Obviously, the ROC curves of different methods intersect with each other in Fig. 10. To deal with this issue, Table I provides two AUC indicators including AUC (P_d, P_f) and AUC (P_f, τ) of all comparing methods on four real HSI datasets for quantitative
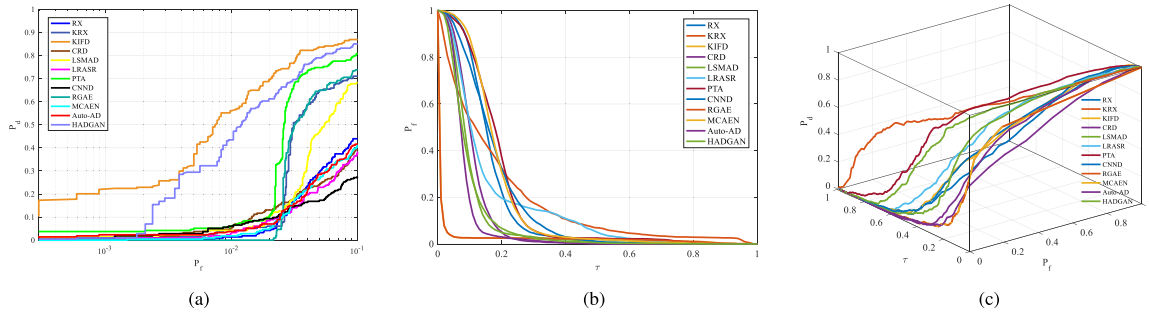
Fig. 11. ROC curves of the 12 methods for the AVIRIS-II dataset. (a) Two-dimensional ROC curves (P_d, P_f). (b) Two-dimensional ROC curves (P_f, τ). (c) 3-D ROC curves.

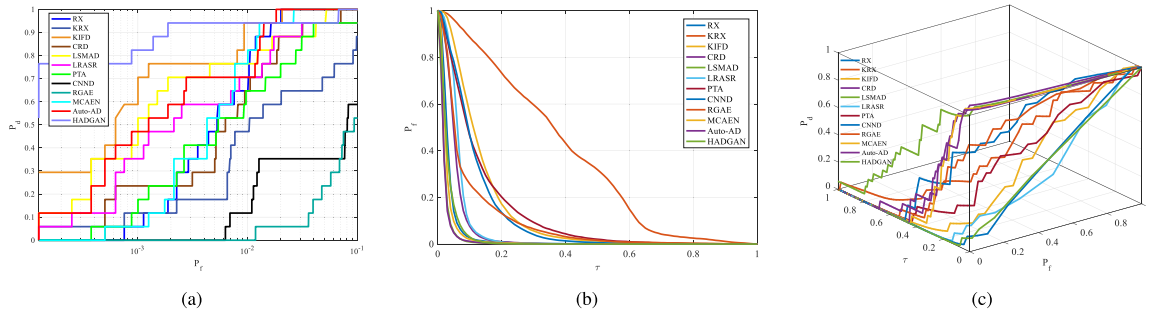
Fig. 12. ROC curves of the 12 methods for the HYDICE dataset. (a) Two-dimensional ROC curves (P_d, P_f). (b) Two-dimensional ROC curves (P_f, τ). (c) Three-dimensional ROC curves.
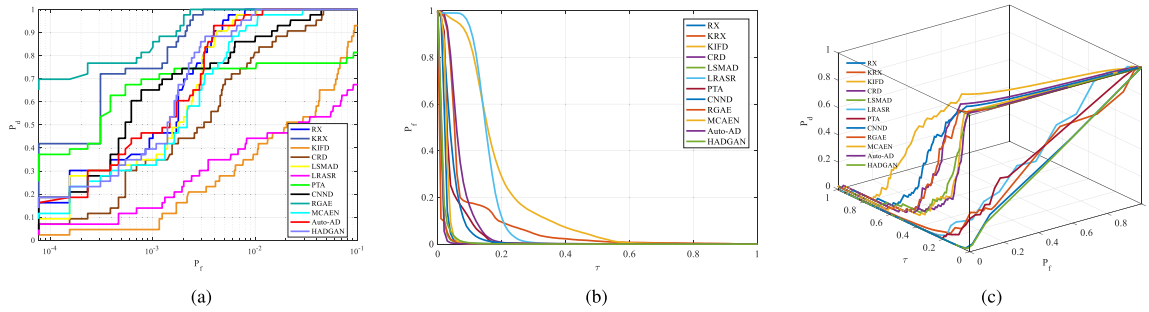

Fig. 13. ROC curves of the 12 methods for the PAVIA dataset. (a) Two-dimensional ROC curves (P_d, P_f). (b) Two-dimensional ROC curves (P_f, τ). (c) Three-dimensional ROC curves.

TABLE I
AUC VALUES $(P_d, P_f)/(P_f, \tau)$ OF THE 12 METHODS FOR THE FOUR REAL DATASETS

Detectors	AVIRIS-I	HYDICE	PAVIA	AVIRIS-II
RX[11]	0.8886/0.0381	0.9932/0.0348	0.9982/0.0235	0.7499/0.1644
KRX[30]	0.9759/0.4161	0.9313/0.3697	0.9994/0.0854	0.8774/0.1802
KIFD[86]	0.9890/0.1099	0.9929/0.1353	0.9424/0.1876	**0.9386**/0.0276
CRD[18]	0.9955/0.0521	0.9878/0.0554	0.9925/0.0684	0.7616/0.1082
LSMAD[49]	0.9775/0.0247	0.9920/**0.0204**	0.9979/0.0118	0.8581/0.0998
LRASR[51]	0.9876/0.0602	0.9755/0.0657	0.8646/0.1537	0.8359/0.1580
PTA-HAD[90]	**0.9961**/0.2377	0.9825/0.1321	0.9547/0.0445	0.9037/0.1869
CNND[76]	0.8849/0.0804	0.8517/0.1157	0.9957/0.0462	0.7820/0.1744
RGAE[75]	0.9911/**0.0113**	0.8264/0.0961	**0.9996/0.0082**	0.8805/**0.0253**
MCAEN[47]	0.9386/0.0379	0.9934/0.0287	0.9967/0.0289	0.7248/0.1815
Auto-AD[73]	0.9165/0.0345	0.9952/0.0213	0.9980/0.0117	0.7035/0.0809
HADGAN[72]	0.9646/0.0536	**0.9958**/0.0319	0.9970/0.0153	0.9299/0.0954

evaluation. In addition, the best results of AUC (P_d, P_f) and AUC (P_f, τ) are highlighted in boldface. The higher the AUC (P_d, P_f) value, the better the detection result. While the smaller the AUC value (P_f, τ), the better the suppressing effect on the background. The AUC values show similar conclusions with the visualization results. The PTA, KIFD, HADGAN, and RGAE methods obtain the highest AUC (P_d, P_f) on four HSIs, respectively. It is consistent with the visualization maps, the aforementioned corresponding methods have the best results in background suppression and highlighting anomalies. As for AUC (P_f, τ), the RGAE method shows good competitiveness. It achieves the best results in the three datasets. While the LSMAD method shows the best background suppression performance in the HYDICE dataset. Overall, the RGAE method offers a good performance in the background suppression, which is consistent with the visualization maps and AUC (P_f, τ).

In general, for both of the AVIRIS datasets, the matrix decomposition-based and kernel learning-based methods achieve the highest AUC (P_d, P_f) values among the all compared methods. While for the HYDICE and PAVIA datasets, the deep-learning-based methods obtain better results compared to traditional machine learning methods. The RGAE method can achieve very good results in terms of a background suppression effect on different datasets according to the AUC (P_f, τ) values, although the method only has the highest AUC (P_d, P_f) value on one set of datasets. Therefore, the RGAE method is most beneficial for background suppression among all methods.

V. CONCLUSION

This article presents a comprehensive overview of HAD techniques based on machine learning. In particular, both of the traditional machine-learning-based models and recently emerged deep-learning-based models are introduced in the parallel taxonomies, respectively. Furthermore, experiments on four real HSIs with 12 compared methods are conducted. Finally, in this section, we discuss the prospects and future directions. With the rapid and exciting development of deep learning techniques in the computer-vision-related tasks, the HAD task has also clearly moved from the traditional machine learning to deep learning era. As reviewed in this article, many successful attempts of deep-learning-based methods for HAD have been reported. However, the following aspects should be addressed in the future study.

1) Both of the traditional machine-learning and deep-learning-based models have many parameters to set, which is hardly to be tuned in practice due to the unavailable of evaluation data. Therefore, it is necessary to design the parameter-robust and parameter-adaptive machine learning models for HAD.

2) The current machine-learning-based HAD algorithms often trained on single or very few hyperspectral images, thus the overfitting phenomenon may imperceptibly happen. To relieve such issue, some advanced machine learning techniques, e.g., regularization, model compression, and network architecture search, are expected to be investigated.

3) Since the HAD is a full unsupervised task, it is beneficial to employ the recent deep generative models (e.g., GAN) to bring large volume of synthesis data to support training. In the future study, it is more important to further embed the physical factors (e.g., the linear mixture model) into the data generation and make the whole learning architecture to be explainable.

4) In the deep learning society, the regularly updated backbone networks (usually for feature extraction) push forward various of computer vision tasks to better performance. Consequently, it is also reasonable to rapidly introduce and improve these latest backbone networks to deal with the HAD task.

5) Recent researches show there exist some potential risks that the deep learning algorithms may get attacked by specific deception algorithms, which are crucial for some HAD applications. Therefore, it is extremely important to design the HAD algorithms that are robust and not vulnerable to external attacks.

REFERENCES

[1] D. A. Landgrebe, S. B. Serpico, M. M. Crawford, and V. Singhroy, "Introduction to the special issue on analysis of hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1343–1345, Jul. 2001.

[2] 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.

[3] M. Shimoni, R. Haelterman, and C. Perneel, "Hyperspectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[4] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[5] P. Goel, S. Prasher, R. Patel, J. Landry, R. Bonnell, and A. Viau, "Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn," *Comput. Electron. Agricult.*, vol. 39, no. 2, pp. 67–93, May 2003.

[6] K. Tan, F. Wu, Q. Du, P. Du, and Y. Chen, "A parallel Gaussian–Bernoulli restricted Boltzmann machine for mining area classification with hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 2, pp. 627–636, Feb. 2019.

[7] L. Wang, L. Wang, Q. Wang, and P. M. Atkinson, "SSA-SiamNet: Spectral-spatial-wise attention-based siamese network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, Jul. 2022, doi: 10.1109/TGRS.2021.3095899.

[8] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[9] X. X. Zhu et al., "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[10] G. Cheng, X. Xie, J. Han, L. Guo, and G. -S. Xia, "Remote sensing image scene classification meets deep learning: Challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, Jun. 2020.

[11] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[12] M. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 374–387, Feb. 2005.

[13] N. Billor, A. S. Hadi, and P. F. Velleman, "BACON: Blocked adaptive computationally efficient outlier nominators," *Comput. Statist. Data Anal.*, vol. 34, no. 3, pp. 279–298, Sep. 2000.

[14] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.

[15] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.

[16] R. Zhao, B. Du, and L. Zhang, "Hyperspectral anomaly detection via a sparsity score estimation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3208–3222, Jun. 2017.

[17] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[18] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. Neural Inf. Process. Syst.*, Dec. 2009, pp. 2080–2088.

[19] E. Candes, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. ACM.*, vol. 58, no. 3, pp. 1–37, 2009.

[20] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, May 2018.

[21] R. Tao, X. Zhao, W. Li, H. -C. Li, and Q. Du, "Hyperspectral anomaly detection by fractional Fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.

[22] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, "Multipixel anomaly detection with unknown patterns for hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1–11, Apr. 2021, doi: 10.1109/TNNLS.2021.3071026.

[23] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, pp. 2–28, Sep. 2021, doi: 10.1109/MGRS.2021.3105440.

[24] H. Jiang, "A manifold constrained multi-head self-attention variational autoencoder method for hyperspectral anomaly detection," in *Proc. Int. Conf. Electron. Inf. Technol. Smart Agriculture*, 2021, pp. 11–17.

[25] X. Zhang and G. Wen, "A hyperspectral imagery anomaly detection algorithm based on local three-dimensional orthogonal subspace projection," in *Proc. Image Signal Process. Remote Sens.*, 2015, pp. 1–8.

[26] Y. Yang, J. Zhang, S. Song, C. Zhang,, and D. Liu, "Low-rank and sparse matrix decomposition with orthogonal subspace projection-based background suppression for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1378–1382, Aug. 2020.

[27] C. -I. Chang, H. Cao, S. Chen, X. Shang, C. Yu,, and M. Song, "Orthogonal subspace projection-based go-decomposition approach to finding low-rank and sparsity matrices for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2403–2429, Mar. 2021.

[28] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, "Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst*, pp. 1–14, May 2021, doi: 10.1109/TNNLS.2021.3082158.

[29] S. Chang, B. Du, and L. Zhang, "BASO: A background–anomaly component projection and separation optimized filter for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3747–3761, Jul. 2018.

[30] H. Kwon and N. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[31] B. Scholkopf, A. Smola, and K.-R. Muller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.*, 1997, pp. 583–588.

[32] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[33] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[34] Y. Gao, J. Gu, T. Cheng, and B. Wang, "Kernel-based nonlinear anomaly detection via union dictionary for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2021, doi: 10.1109/TGRS.2021.3093591.

[35] P. Yan, W. Zhang, B. Turkbey, P. L. Choyke, and X. Li, "Global structure constrained local shape prior estimation for medical image segmentation," *Comput. Vis. Image Understanding*, vol. 117, no. 9, pp. 1017–1026, 2013.

[36] Y. Yang, S. Song, D. Liu, J. C.-W. Chan, J. Li, and J. Zhang, "Hyperspectral anomaly detection through sparse representation with tensor

[37] X. Ma, X. Zhang, N. Huyan, X. Tang, B. Hou, and L. Jiao, "Hyper-Laplacian regularized low-rank tensor decomposition for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 6380–6383.

decomposition-based dictionary construction and adaptive weighting," *IEEE Access.*, vol. 8, pp. 72121–72137, 2020.

[38] X. Zhang and G. Wen, "A fast and adaptive method for determining K_1, K_2, and K_3 in the tensor decomposition-based anomaly detection algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 15, no. 1, pp. 3–7, Jan. 2018.

[39] K. Jiang, W. Xie, J. Lei, T. Jiang, and Y. Li, "LREN: Low-rank embedded network for sample-free hyperspectral anomaly detection," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 4139–4146.

[40] G. Mishne, U. Shaham, A. Cloninger, and I. Cohen, "Diffusion nets," *Appl. Comput. Harmon. Anal.*, vol. 47, no. 2, pp. 259–285, 2017.

[41] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.

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

[43] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.

[44] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.

[45] H. Su, Z. Wu, A. Zhu, and Q. Du, "Low rank and collaborative representation for hyperspectral anomaly detection via robust dictionary construction," *ISPRS J. Photogrammetry Remote Sens.*, vol. 169, pp. 195–211, Nov. 2020.

[46] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[47] X. Lu, W. Zhang, and J. Huang, "Exploiting embedding manifold of autoencoders for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1527–1537, Mar. 2020.

[48] M. D. Farrell and R. M. Mersereau, "On the impact of covariance contamination for adaptive detection in hyperspectral imaging," *IEEE Signal Process. Lett.*, vol. 1, no. 9, pp. 649–652, Sep. 2005.

[49] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[50] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, pp. 1–18, May 2014.

[51] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.

[52] X. Zhang, X. Ma, N. Huyan, J. Gu, X. Tang, and L. Jiao, "Spectral-difference low-rank representation learning for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10364–10377, Dec. 2021.

[53] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.

[54] C. Zhao, L. Xueyuan, and Z. Haifeng, "Hyperspectral anomaly detection based on stacked denoising autoencoder," *J. Appl. Remote Sens.*, vol. 11, no. 4, 2017, Art. no. 042605.

[55] C. Zhao and L. Zhang, "Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection," *Infrared Phys. Technol.*, vol. 92, pp. 166–176, 2018.

[56] Z. Chen, B. Yang, and B. Wang, "A preprocessing method for hyperspectral target detection based on tensor principal component analysis," *Remote Sens.*, vol. 10, no. 7, Jun. 2018, Art. no. 1033.

[57] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022, doi: 10.1109/TNNLS.2020.3038659.

[58] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[59] H. Fan, Y. Chen, Y. Guo, H. Zhang, and G. Kuang, "Hyperspectral image restoration using low-rank tensor recovery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4589–4604, Oct. 2017.

[60] S. Chang, B. Du, and L. Zhang, "A sparse autoencoder based hyperspectral anomaly detection algorithm using residual of reconstruction error," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 5488–5491.

[61] J. Zhang, Y. Xu, T. Zhan, Z. Wu, and Z. Wei, "Anomaly detection in hyperspectral image using 3D-convolutional variational autoencoder," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2512–2515, doi: 10.1109/IGARSS47720.2021.9554184.

[62] D. Zhu, B. Du, and L. Zhang, "EDLAD: An encoder-decoder long short-term memory network-based anomaly detector for hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4412–4415, doi: 10.1109/IGARSS47720.2021.9553145.

[63] J. Lei, S. Fang, W. Xie, Y. Li, and C. I. Chang, "Discriminative reconstruction for hyperspectral anomaly detection with spectral learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7406–7417, Oct. 2020.

[64] W. Xie, B. Liu, Y. Li, J. Lei, C. Chang, and G. He, "Spectral adversarial feature learning for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2352–2365, Apr. 2020.

[65] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[66] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020.

[67] J. Zhong, W. Xie, Y. Li, J. Lei, and Q. Du, "Characterization of background-anomaly separability with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6017–6028, Jul. 2021.

[68] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, "Unsupervised pixel-wise hyperspectral anomaly detection via autoencoding adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2021.

[69] T. Jiang, W. Xie, Y. Li, and Q. Du, "Discriminative semi-supervised generative adversarial network for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2420–242.

[70] 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.

[71] C. Zhao, C. Li, S. Feng, and W. Li, "Spectral-spatial anomaly detection via collaborative representation constraint stacked autoencoders for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jan. 2021, Art. no. 21408939, doi: 10.1109/LGRS.2021.3050308.

[72] 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, Jun. 2020.

[73] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Mar. 2022.

[74] Y. Li, T. Jiang, W. Xie, J. Lei, and Q. Du, "Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, Aug. 2022, doi: 10.1109/TGRS.2021.3102048.

[75] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Jul. 2022, doi: 10.1109/TGRS.2021.3097097.

[76] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.

[77] L. Zhang and B. Cheng, "Transferred CNN based on tensor for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2115–2119, Dec. 2020.

[78] L. Zhang and B. Cheng, "Fractional Fourier transform and transferred CNN based on tensor for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3072249.

[79] C. Zhao, C. Li, S. Feng, and N. Su, "Hyperspectral anomaly detection using bilateral-filtered generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4408–4411, doi: 10.1109/IGARSS47720.2021.9553233.

[80] C. Zhao, C. Li, S. Feng, and N. Su, "Spectral-spatial stacked autoencoders based on the bilateral filter for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2209–2212.

[81] T. Ouyang, J. Wang, X. Zhao, and S. Wu, "LSTM-adversarial autoencoder for spectral feature learning in hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 2162–2165, doi: 10.1109/IGARSS47720.2021.9553371.

[82] J. Lei, W. Xie, J. Yang, Y. Li, and C. Chang, "Spectral–spatial feature extraction for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8131–8143, Oct. 2019.

[83] W. Xie, S. Fan, J. Qu, X. Wu, Y. Lu, and Q. Du, "Spectral distribution-aware estimation network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Jun. 2021, doi: 10.1109/TGRS.2021.3089711.

[84] S. Song, H. Zhou, Y. Yang, and J. Song, "Hyperspectral anomaly detection via convolutional neural network and low rank with density-based clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3637–3649, Sep. 2019.

[85] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.

[86] S. Li, K. Zhang, P. Duan, and X. Kang, "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 319–329, Jan. 2020.

[87] Z. Hou, W. Li, R. Tao, P. Ma, and W. Shi, "Collaborative representation with background purification and saliency weight for hyperspectral anomaly detection," *Sci. China Inf. Sci.*, vol. 65, no. 1, pp. 1–12, 2022.

[88] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2021.

[89] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.

[90] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.

Yichu Xu (Graduate Student Member, IEEE) received the B.S. degree in geographic information system from Lanzhou University, Lanzhou, China, in 2018. He is currently working toward the Ph.D. degree in computer science with the School of Computer, Wuhan University, Wuhan, China.

His research interests include machine learning and hyperspectral anomaly detection.
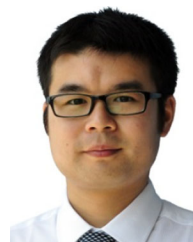
Lefei Zhang (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

He was a Big Data Institute Visitor with the Department of Statistical Science, University College London, London, U.K., and a Hong Kong Scholar with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. He is currently a Professor with the School of Computer Science, Wuhan Universit, and also with Hubei Luojia Laboratory, Wuhan. His research interests include pattern recognition, image processing, and remote sensing.

Dr. Zhang serves as an Associate Editor for *Pattern Recognition* and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Bo Du (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, in 2010.

He is currently a Professor with the School of Computer Science and the Institute of Artificial Intelligence, Wuhan University, where he is also the Director with the National Engineering Research Center for Multimedia Software. He has more than 80 research articles published in the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), IEEE TRANSACTIONS ON CYBERNETICS (TCYB), IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS), IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS), and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL), and 13 of them are ESI hot papers or highly cited papers. His main research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du regularly serves as a Senior PC Member of International Joint Conference on Artificial Intelligence (IJCAI) and Association for the Advancement of Artificial Intelligence. He served as the Area Chair for International Conference on Pattern Recognition (ICPR). He won the Highly Cited Researcher 2019 by the Web of Science Group. He was the recipient of the IJCAI Distinguished Paper Prize, the IEEE Data Fusion Contest Champion, and the IEEE Workshop on Hyperspectral Image and Signal Processing Best Paper Award in 2018. He serves as an Associate Editor for *Neural Networks*, *Pattern Recognition*, and *Neurocomputing*. He serves as a reviewer for 20 Science Citation Index (SCI) magazines, including IEEE TPAMI, TCYB, TGRS, TIP, JSTARS, and GRSL.

Liangpei Zhang (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China in State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He was a Principal scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the Remote Sensing Program in China. He has authored and coauthored more than 700 research papers and five books. He is the Institute for Scientific Information (ISI) highly cited author. He is the holder of 30 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the Institution of Engineering and Technology (IET). He was the recipient of the 2010 Best Paper Boeing award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS), the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE), the IEEE GRSS David Landgrebe Award 2020, and the IEEE GRSS Transactions Prize Paper Award 2020. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) student paper contest in recent years. He serves as an Associate Editor or Editor of more than ten international journals. He is the founding Chair of IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He is currently serving as an Associate Editor for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.