Dynamic Negative Sampling Autoencoder for Hyperspectral Anomaly Detection

Jingxuan Wang, Jinqiu Sun, Yong Xia, *Member, IEEE*, and Yanning Zhang, *Senior Member, IEEE*

***Abstract*—Hyperspectral anomaly detection (HAD) aims at detecting the anomalies without any prerequisite information, which gains lots of attention in recent years. Most of existing detectors locate the anomalies by eliminating the background. The background is usually reconstructed by utilizing only the global and local homogenous attribute, no matter the matrix decomposition-based or the deep learning-based methods. In this article, a dynamic negative sampling autoencoder is proposed for the hyperspectral anomaly detection (DNA-HAD). Some pixels are randomly selected and altered as negative samples. Both the rest original pixels and the altered negative samples are sent into the network. An adaptive adjusted loss function is designed to suppress the reconstruction error for the original pixels, and to enlarge the error for the negative samples. Meanwhile, skip connection is designed to ensure features at both shallow levels and deep levels being utilized for the reconstructing process. In this way, by importing some negative samples in the reconstructing process, the proposed DNA-HAD is not only robust to reconstruct the background but also sensitive to detect the anomalies. Experiments on six hyperspectral imagery which are captured by different sensors have demonstrated the effectiveness of the proposed method.**

***Index Terms*—Anomaly detection, autoencoder (AE), hyperspectral image, negative sampling.**

I. Introduction

HYPERSPECTRAL imagery (HSI) with hundreds of spectral bands can intuitively be represented as a 3-D data cube, including two spatial dimensions and one spectral dimension [1]. The spectral dimension of the HSI usually ranges from the visible to the near-infrared wavelength by a step of less than 10 nm. The rich and detailed spectral information is the key for accurate identification of the subtle difference between different objects [2], [3].

With this spectral discriminative property, HSI has been widely utilized in many applications, including target detection [4], [5], classification [6], change detection [7], and the others [8], [9]. Hyperspectral target detection can be viewed as a binary classification problem which determines each pixel in the spatial domain as background or target. According to whether signature information of the target is known, the object detection can be classified into the target detection with prior information [10], [11] and the anomaly detection without prior information [12], [13]. A weakly supervised low rank representation is designed to convert the deep learning-based anomaly detection into a convex low rankness representation optimization problem [14]. The weakly supervised back estimation model is more flexible than supervised manner, and more robust than the unsupervised manner. To deal with the challenges in separating the background and the anomaly, a class saliency map extraction algorithm is utilized to obtain pseudobackground and anomaly samples for adversarial training [15]. In addition, a discriminator is induced to make the encoded representation resemble Gaussian distribution.

In reality, it is often difficult or even impossible to access the spectral information of the targets. In this way, the unsupervised hyperspectral anomaly detection (HAD) attracts much attention of researchers for its practicability. As to the anomalies without any prerequisite information, they have not been exactly defined, and usually refers to the pixels those are obviously different from the spectrum of the surrounding background and appear with a small probability. Specifically, according to [16], there are four features suggested to characterize anomalies, as follows.

1) There is no prior information about neither the backgrounds nor the anomalies.
2) The anomalies occur with a low probability.
3) The anomalies exhibit insignificance in spectral statistics.
4) The anomalies are with small size compared to the backgrounds.

In this way, a lot of works have been proposed according to these four features, which can be roughly divided into two categories (RGAE) [17].

The first category is the distance-based detectors. This kind of methods usually suppose that, the distribution of the background is simulated via some typical distribution, such as the Gaussian distribution, the mixture of Gaussian distribution or some others. Every pixel in the spatial domain is measured to evaluate its derivation from the background. The derivation denotes the tendency of it to be the anomaly. The classical global Reed Xiaoli detector (Global-RX) is the typical representor of this kind of detectors [18]. They assume the whole background follows a Gaussian distribution, which is consistent to the aforementioned third feature of the anomalies. Once the covariance matrix and

Manuscript received 22 August 2022; revised 20 October 2022; accepted 3 November 2022. Date of publication 8 November 2022; date of current version 21 November 2022. This work was supported by the National Natural Science Foundation of China under Grant 61901384, Grant 61971356, and Grant U19B2037. *(Corresponding authors: Jingxuan Wang; Yanning Zhang.)*

Jingxuan Wang, Yong Xia, and Yanning Zhang are with the School of Computer Science, Northwestern Polytechnical University, Xi’an 710072, China, also with the National Engineering Laboratory for Integrated Aerospace-Ground-Ocean Big Data Application Technology, Northwestern Polytechnical University, Xi’an 710072, China, and also with the Shaanxi Provincial Key Laboratory of Speech and Image Information Processing, Northwestern Polytechnical University, Xi’an 710072, China (e-mail: jxwang0130@163.com; yxia@nwpu.edu.cn; ynzhang@nwpu.edu.cn).

Jinqiu Sun is with the School of Astronautics, Northwestern Polytechnical University, Xi’an 710072, China (e-mail: sunjinqiu@nwpu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3220514

mean vector of the scene are calculated, the probability density function of the background can be estimated. In this way, all the pixels are measured by their derivations from the distribution of the background, and the pixels far from the background distribution are supposed to be anomalies.

Inspired by the RX detector, many other methods have been proposed to make some improvements over the RX detector. Some works aim at formulating a more accurate distribution for the background, such as the local RX detector [19], the weighted RX detector [20], the subspace RX detector [21], kernel RX [22], and some other detectors [23], [24]. Meanwhile, some works mainly focus on enhancing the spatial discrimination ability of the input HSI. They preprocess the original HSI to highlight the anomalies, and further utilize the RX detector to make a final detection. Considering the superiority of the fractional Fourier transform in handling the nonstationary noise, it is employed to obtain features in an intermediate domain. The discrimination between the anomalies and backgrounds in the fractional domain is enhanced, and the RX detector is applied to locate the anomalies (FrFT-RX) [25]. Considering the noise during the imagery process, local gradient profiles of the probable anomalies have been transformed, and to enhance the spatial information of the HSI. The enhanced HSI is finally detected by the RX detector (LGG-RX) [26]. In ref. [27], the anomalies are detected with the background following a gaussian distribution. Two adaptive detectors are proposed based on the generalized likelihood ratio test (GLRT) design procedure and ad hoc modification of it. Demonstration has shown that the one step GLRT is equivalent to the two-step GLRT (2S-GLRT), but the 2S-GLRT consumes a less computational complexity.

The second category is the reconstruction-based method. Considering that the original HSI can be viewed as a combination of the backgrounds and the anomalies, many methods detect the anomalies by reconstructing the background. For this kind of methods, accurate reconstructing the background is important to locate the anomalies. Both the matrix decomposition-based detectors and most of the deep learning-based methods belong to this category.

For the matrix decomposition-based detectors, the most common tools for reconstructing the background including the low rankness and the sparse decomposition [28]. For the HSIs, highly correlated spatial and spectral information makes the low-rankness of the HSIs [29]. The anomalies are sparse due to the aforementioned second and the fourth feature of them, which can be formulated as a low-rank and sparse decomposition (LRaSMD) problem [30]. The classical GoDec algorithm is usually adopted to solve for the optimal variables [31]. Inspired by the LRaSMD, many other detectors have been proposed by importing some other more effective constraints, such as the graph and total variation constraints to suppress the noises and to preserve the local structure of the HSIs (GTVLRR) [32]. A spatial constraint was incorporated to smooth the coefficients, which are based on single or multiple local windows and the low-rank representation sum-to-one model (SLW_LRRSTO/MLW_LRRSTO) [33]. Meanwhile, to avoid the large computational complexity caused by the redundant dictionary atoms, the initial dictionary was constructed via a random selection method. By formulating the sparse component as a mixture of Gaussian noises (LSDM-MoG) in the reconstructing process, the anomalies can be detected more accurately [34].

In contrast, collaborative representation detector (CRD) assumes that collaboration between dictionary atoms is more important than competition in the case of small samples, and it allows all atoms in a dictionary to participate in linear representation to obtain better performance [35]. CRD has rapidly attracted a lot of attention because it is simple and efficient. However, its dictionary construction is also implemented via a dual window, which does not take account for the pollution of abnormal pixels. An improved outlier removal anomaly detector based on CRD (CRDBPSW) has been proposed that incorporates a background purification framework to automatic remove the outlier. It avoids the heterogeneous pixels being involved in the background reconstruction process, and achieves a more appealing detection performance [36].

In addition, because of the 3-D characteristics of the HSIs, the three-order tensors are also utilized to reconstruct the background HSI [37]. The tensor decomposition-based detector (TenB) is applied to eliminate the background, and highlights the anomalies. To be specific, the TenB abandons the first principal component in the three dimensions, a process to eliminate the background [38]. Considering that the mislocating anomalies usually happen for the pixels with comparative large but not the largest responses, the tensor completion-based detector sets a comparatively strict threshold for the background [39]. In this way, most of the selected backgrounds are correctly selected, and can be utilized to reconstruct the entire background and to detect the anomalies. Considering the group sparse prior in the HSI, a prior-based tensor approximation detector is proposed, which combines the low-rankness, sparsity and piecewise smoothness with the advantages of the tensor representation [40].

Motivated by the development of the deep learning in image processing, it has also been extensively applied to the HSI processing. When it comes to the HAD with no prior information about the anomalies, the autoencoders (AEs) [41], [42] and generative adversarial networks (GANs) are widely employed [43]. The basic idea is that the trained AEs and GANs act as the feature extractors, which can accurately preserve the main feature of the input HSIs [44]. In this way, the background can be accurately reconstructed via the network, while the anomalies cannot be. Residual between the original pixel and the reconstructed pixel can directly be regarded as the abnormal level. A robust graph AE detector is proposed by embedding a super-pixel segmentation-based graph regularization term in the AE, which preserves the geometric structure in reconstructing the background [17]. To avoid the manual parameter setting in the training process, an autonomous HAD (autoAD) method is proposed by utilizing a fully convolutional AE with adaptive-adjusted loss function [45]. The potential anomalies have been allocated smaller weight during the training process. For most of the GAN based detectors, the AEs usually act as generators; discriminators are added to evaluate the difference between the real pixel and the generated pixel [46].

In this article, a dynamic negative-sampling AE is proposed for hyperspectral anomaly detection (DNA-HAD). There are three modules involved in the network, including generation of the negative samples, feature extraction via the AE, and the adaptive-adjusted loss back propagation. To be specific, some pixels in the scene are first randomly selected and altered as negative samples via a fixed ratio. The HSIs with already known negative samples are sent into the AE with skip connections. Skip connection in the network is designed to utilize features at both shallow levels and deep levels for the reconstructing. In addition, the adaptive-adjusted loss function is formulated to suppress the reconstruction error for the background pixels, and to enhance the discrimination ability for the altered negative samples. In this way, the network is more sensitive to the abnormal anomaly pixels, and can achieve a more precise detection accuracy.

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

1) Dynamic negative-sampling is used to improve the sensitivity of the proposed detector for the anomalies, where fixed percentage of pixels are randomly selected and altered as negative samples. These negative samples are supposed to be highlighted by the AEs, which can be treated as some manually made anomalies.
2) To enhance the robustness of the network in dealing with the negative samples, an adaptive-adjusted loss function is designed to suppress the negative samples and to highlight the positive samples. Meanwhile, skip connections are utilized to make a combination between the features at both deep and shallow levels. In this way, the network is more sensitive to the pixels which are severally different from the background, and can achieve a more precise detection accuracy.
3) All the parameters are fixed to reconstruct the background. Detection map is also automatically generated by separating the background from the original HSI. No preprocessing or postprocessing is involved in the reconstruction process, making the detection process flexible.

The rest of this article is organized as follows. Detailed description of the proposed method is in Section II. Experimental setup and data analysis have been elaborated in Section III. Finally, Section IV concludes the article.

II. Proposed Method

In this article, an autonomous dynamic negative-sampling AE is proposed for hyperspectral anomaly detection. The flowchart of the DNA-HAD is provided in Fig. 1, in which the left branches and the right branches of the network denote the encoder and the decoder, respectively. Some pixels are first randomly selected from the original scene and altered as the negative samples which follow the gaussian distribution. The altered HSI is sent into an AE with skip connections. In this way, the entire background can be reconstructed. The negative samples are suppressed via the adaptive-adjusted loss function, making the network more sensitive to the spectrally different pixels. The final detection map is achieved by eliminating the background from the original HSI.

There are three main modules in the proposed DNA-HAD, including generation of the negative samples, reconstruction of the background via AE with skip connections, and the adaptive-adjusted loss function to enhance the sensitivity of the network for the spectrally different pixels. Details of these three modules are described as follows.

A. Dynamic Generation of the Negative Samples

Different from the target detection, there is no prior information during the anomaly detection. The anomalies refer to the pixels those are spectrally different from the background. In this way, some pixels in the scene have been randomly selected and altered via a Gaussian distribution. These altered pixels can be viewed as some already known man-made "anomalies," and act as the negative samples for the training process, making the network more sensitive to the spectrally different pixels. Suppose the original HSI as $H \in R^{w \times h \times b}$, the ratio of the selection is r. The number of the altered pixels can be calculated as $N_a = floor(w * h * r)$. N_a locations have been randomly selected from the whole scene. Reflectance at the N_a locations are altered via a Gaussian distribution. The generation of the input $\hat{H}$ can be denoted as

$$\hat{H} = f_{rand}(H). \tag{1}$$

Suppose all the locations except the N_a altered ones come into the set S, the pixels of $\hat{H}$ and H which belong to the S set are the same.

B. Background Reconstruction via AE With Skip Connections

The HSI can be viewed as a combination of the background and the anomalies. In this way, accurate reconstruction of the background is crucial for the detection [47]. In this article, the HSIs with randomly altered negative samples have been sent into the AE with skip connections. It should be noted that the network mainly contains a chain of convolutional layers and symmetric deconvolutional layers, which is demonstrated in Fig. 1. The left branch denotes the encoder, and contains four "M" modules. There are three main submodules in each M module, which are 1) the convolutional layers; 2) the leaky rectified linear unit (leaky Relu) layer to enhance the nonlinearity; and 3) the Residual block (ResB). The ResB can preserve the original information and make the network deeper [48]. The M modules can be viewed as the feature extractor, which aims at preserving the main components of the scene and eliminating the abnormal information. The right branch represents the decoder, and contains four corresponding "DM" modules. There are also three submodules in each DM module, including the deconvolutional layer, the batch normalization layer, and the Relu layer.

The details of the convolutional layers in the four M modules are listed in Table I, whose kernel sizes are 3, 4, 7, and 3. It is noted that there are two convolutional layers in each ResB of the M module. Their input channel and output channel are the same as the output channel of its previous convolutional layer. The kernel sizes are fixed at 3×3.Corresponding for the deconvolutional layers in the four DM modules are listed in Table II.

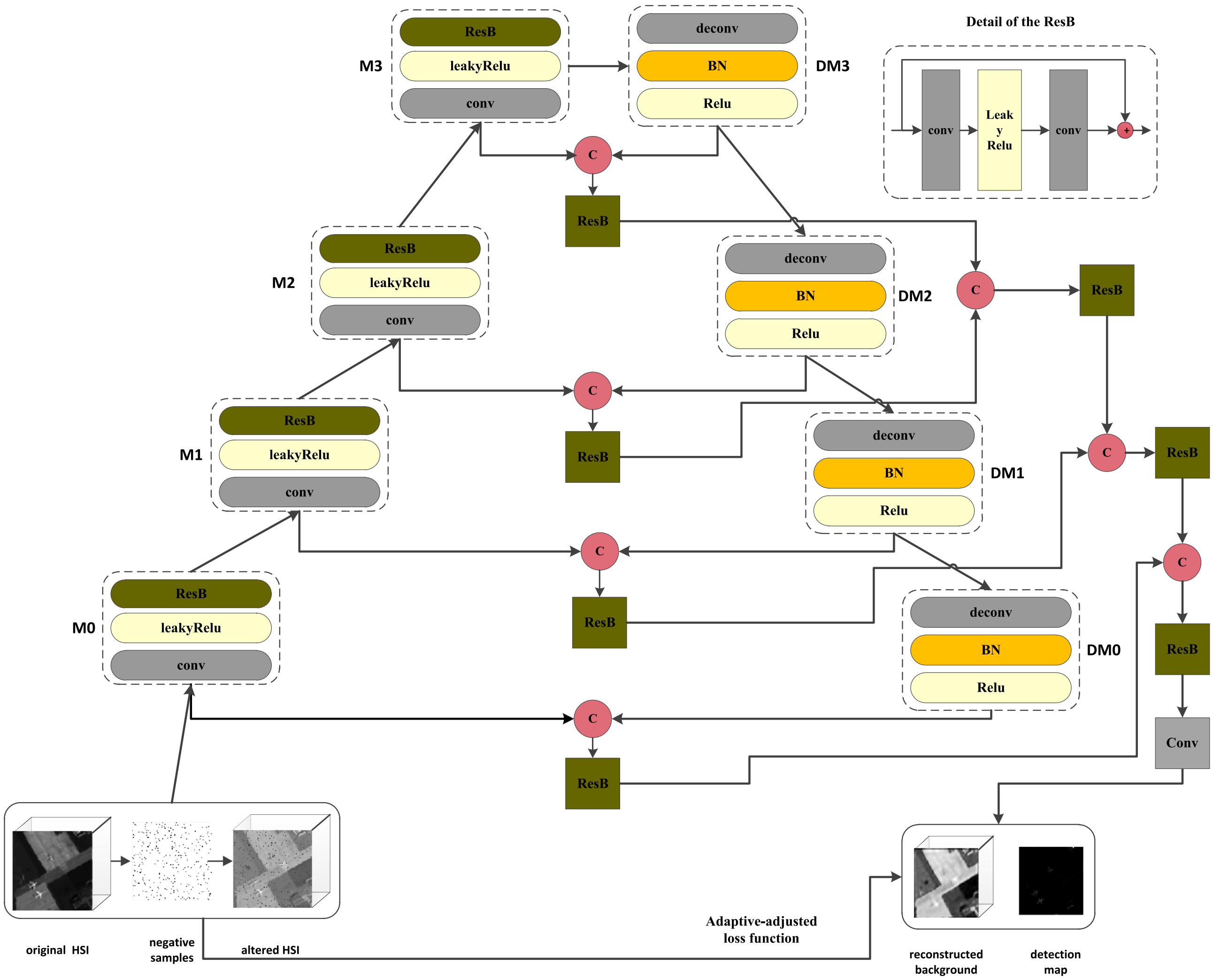


Fig. 1. Network architecture of the proposed DNA-HAD.

TABLE I
DETAILS OF THE FOUR M MODULES

	Input channel	Output channel	Kernel size	stride
M0	b	100	3	1
M1	100	64	4	2
M2	64	32	7	5
M3	32	16	3	1

TABLE II
DETAILS OF THE FOUR DM MODULES

	Input channel	Output channel	Kernel size	stride
DM3	16	32	3	1
DM2	32	64	3	1
DM1	64	100	7	5
DM0	100	b	4	2

Combination between the convolution layers and deconvolution layers was first proved to be efficient for semantic segmentation, such as the classical Unet for medical image segmentation [49]. In the proposed method, the M module which mainly consists of the convolutional operations, and the DM module which mainly consists of deconvolutional operation are concatenated. The concatenated features are sent into the ResB for better information extraction. There are two main advantages of this skip connections between the M module and the DM module. First, as the network goes deeper, there are more image details lost, making the deconvolution more difficult. Connections between the convolution layers at the shallow level and the deconvolution layers at the deep level ensure that more information is utilized for recovering the scene. Second, skip connections also provide a way for the gradient being back-propagated to the bottom layers. This operation makes the training process easier.

It is noted that no pooling layers are utilized in the network. This is because of that pooling layers usually focus on extracting the low-level features of the input by discarding useful image details. This is different from the aim of proposed method in reconstructing the background. During reconstructing the background, it is of great significance to eliminate the low-level noises while keeping the image details. Skip connections are connected symmetrically between convolutional layers and deconvolutional layers, to make a

TABLE III
OPTIMAL PARAMETERS OF VARIOUS DETECTORS IN DIFFERENT DATASETS

Methods	FrFT-RX	GTVLRR [M,P,λ,β,γ,k,σ]	LSDM-MoG [r,K]	2S-GLRT [out,in]	RGAE [λ, s, n_hid]
SanDiego	0.8	[15,20,0.5,0.2,0.05,10,1]	[10,2]	[25,13]	[10^{-2},150,160]
HYDICE	/	[20,20,0.5,0.2,0.05,10,1]	[20,7]	[5,3]	[10^{-3},150,160]
airport-1	/	[15,20,0.5,0.2,0.05,10,1]	[10,4]	[11,9]	[10^{-2},150,160]
airport-2	/	[20,20,0.5,0.2,0.05,10,1]	[50,6]	[19,15]	[10^{-3},150,120]
beach-4	0.8	[15,20,0.5,0.2,0.05,10,1]	[10,5]	[21,5]	[10^{-2},50,20]
urban-2	/	[15,20,0.5,0.2,0.05,10,1]	[10,2]	[25,3]	[10^{-4},50,80]

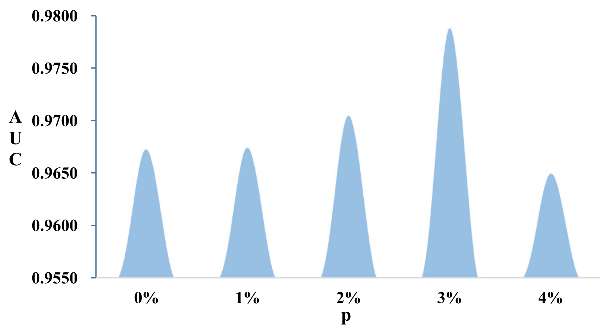


Fig. 2. Different percentages of the negative samples and the corresponding average AUCs.

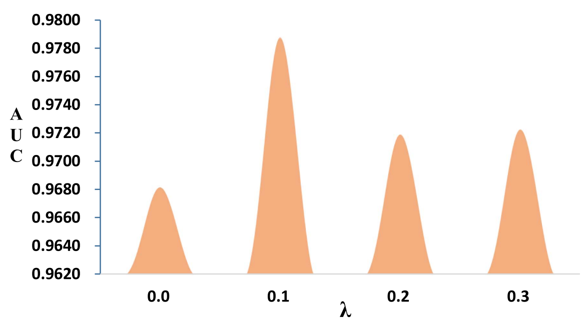


Fig. 3. Different λ and the corresponding average AUCs.

combination between the low-level features and the high-level features.

The final concatenation between the input of the M0 and the output of the DM0 is further sent into a M module, which is shown in right bottom of the Fig. 1. This operation makes a further extraction of the input HSI and the deep extracted features, and obtains an image with the same size of the input HSI. The obtained output HSI is referred as the reconstructed background HSI.

C. Adaptive-Adjusted Loss Function

During the training process, the reconstruction error of the original pixels in the scene is supposed to decrease. On the contrary, the reconstruction error of the dynamic negative samples is expected to increase. Reconstruction errors of the original pixels and negative samples are expected to follow different trends. In the generation process of the negative samples, two variables have been employed to mark the backgrounds and the altered samples, respectively. These two matrices are utilized to control the samples involved in the training process. Suppose the matrix utilized for marking the background as W_1, the matrix utilized for marking the altered negative samples can be represented as W_0. The adding result between W_1 and W_0 is a matrix of unit elements. In this way, the reconstruction error of the original positive pixels can be expressed as

$$loss1 = ||W_1.*\hat{H} - W_1.*\tilde{H}||_2. \tag{2}$$

$\tilde{H}$ is the output of the network, and $loss1$ is the error for the original positive pixels. The reconstruction error of the altered samples can be expressed as

$$loss2 = ||W_2.*\hat{H} - W_2.*\tilde{H}||_2. \tag{3}$$

$loss2$ is the error for the negative-sampling pixels. During the training process, the $loss1$ is supposed to decrease, and the $loss2$ is expected to increase. To make a unification of both the $loss1$ and $loss2$, the final loss function for the network is defined as

$$loss = loss1 - lambda * loss2. \tag{4}$$

$lambda$ is a positive variable which controls the $loss2$ to follow an increasing tendency in the training process. Once every iteration finished, there is a background being reconstructed. The corresponding loss is calculated via the adaptive-adjusted loss function eq. 4. The loss is then fed backward, and makes an update of all the parameters in the network via the typical Adam optimizer.

III. EXPERIMENTAL SETUP AND DATA ANALYSIS

To validate the effectiveness of the proposed DNA-HAD, experiments have been conducted on six real HSIs with different spatial resolutions. These HSIs are captured by three different sensors. Meanwhile, both the experimental setup and the detection results are also introduced in the section. All the experiments are conducted on an Intel Core i5-8400 CPU with 16 GM of RAM, Geforce 1060.

A. Datasets

This article employs six HSIs which are captured by three different sensors. The detailed descriptions of these datasets are shown as follows.

SanDiego: This dataset was collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor, which partly describes the different areas of the San Diego airport, CA, USA. The spatial resolution of this image is 3.5 meters per pixel. There are 224 spectral bands in the wavelength ranging from 370 to 2510 nm. Once removing the bands that are corresponding to

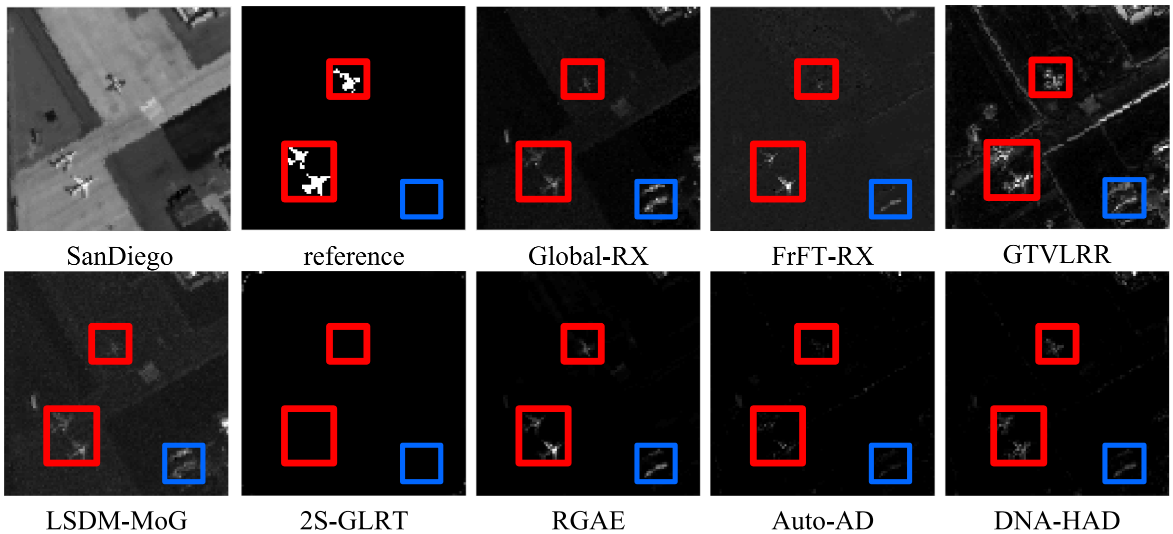


Fig. 4. Visual detection maps of the SanDiego achieved by different detectors.

TABLE IV
ABLATION EXPERIMENTS OF THE MODULES IN THE DNA-HAD

Skip Connections	Negative sampling	Adaptive-adjusted loss function	Average AUC
×	×	×	0.9525
✓	×	×	0.9627
✓	✓	×	0.9682
✓	✓	✓	0.9788

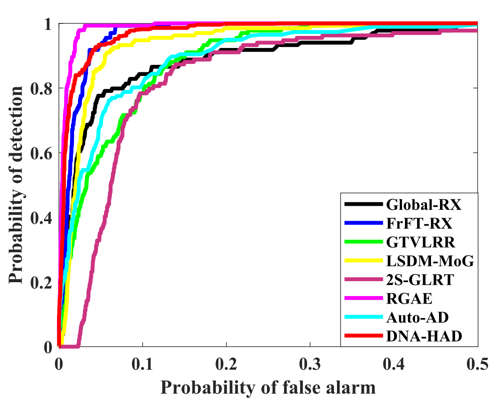


Fig. 5. ROC curves of the SanDiego achieved by different detectors.

the water absorption region, low-signal noise ratio (SNR), and bad bands, there are 189 bands are remaining. The spatial size is 100×100. Three aircrafts are marked as the anomalies, which consist of 134 pixels, respectively.

HYDICE: This dataset was collected by the hyperspectral digital imagery collection experiment (HYDICE) sensor, which describes a suburban residential area of Michigan, USA. The background landcovers types include the soil, parking lot, water, and road. Ten man-made vehicles are marked as anomalies with a total number of 17 pixels. This dataset is also collected ranging from the visible 400 nm to the near-infrared 2500 nm. 162 spectral bands are retained for the detection. The spatial size of this dataset is 80×100, whose spatial resolution is 1.56 meters per pixel.

Airport-1 and Airport-2: These two datasets were also captured by the AVIRIS sensor on 11/9/2011. Spatial resolution of these two HSIs is 7.1 meters per pixel. Airplanes in the scene are marked as anomalies, whose number is 144 pixels and 87 pixels. Both HSIs are captured manually extracted from large images downloaded from the AVIRIS website. Regions with a spatial size of 100×100 are utilized for the experiment [50].

Beach-4: This dataset was collected by the reflective optics system imaging spectrometer (ROSIS) sensor, which mainly depicts a bridge in the Pavia Center. The background includes the bridge, river, bare soil, and buildings. The vehicles on the bridge are marked as anomalies, including 68 pixels. There are 150×150 pixels in the spatial domain, whose spatial resolution is 1.3 meters per pixel.

Urban-2: This HSI captured by the AVIRIS sensor on 29 August 2010. It depicts the Texas coast, in which the buildings are marked as the anomalies. The spatial resolution is 17.2 meters per pixel, which is the coarsest among all the experimental HSIs.

B. Experimental Setup

In order to validate the performance of proposed method, the most widely used metrics receiver operating characteristics (ROC) curves and the area under the curve (AUC) are employed as the quantitatively metrics [51]. Meanwhile, the qualitative detection maps via different detectors are exhibited to visually distinguish the detection results.

To make a comprehensive evaluation of the proposed method, there are seven competitors employed for the comparison, including three statistical-based methods, two representation-based detectors, and two deep learning-based detector. The three statical-based methods include the Global-RX, the FrFE-RX, and the 2S-GLRT. The representation-based methods are the GTVLRR and the LSDM-MoG. The deep-learning based methods are the RGAE and the Auto-AD. For Global-RX, there are no

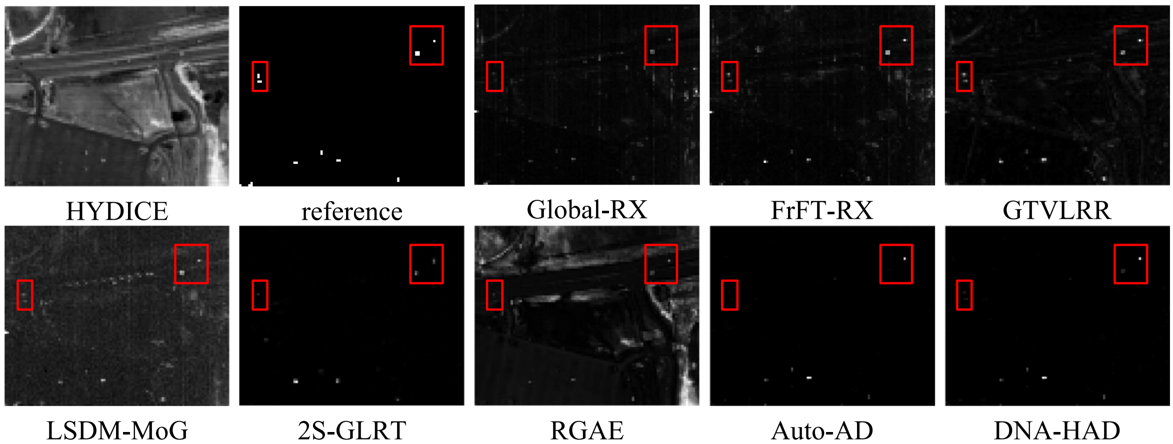


Fig. 6. Visual detection maps of the HYDICE achieved by different detectors.

TABLE V
AUC Values of the Six HSIs Achieved by Different Detectors

	Global-RX	FrFT-RX	GTVLRR	LSDM-MoG	2S-GLRT	RGAE	Auto-AD	DNA-HAD
SanDiego	0.9403	0.9818	0.9413	0.9667	0.9074	**0.9922**	0.9440	0.9850
HYDICE	0.9857	0.9930	0.9930	0.9916	0.9829	0.7809	0.9975	**0.9997**
airport-1	0.8221	0.8741	0.8998	0.8487	0.9320	0.6406	0.8808	**0.9311**
airport-2	0.8404	0.9604	0.8928	0.8619	**0.9873**	0.7484	0.8623	0.9672
beach-4	0.9531	0.9501	0.9813	0.9566	0.9868	0.9199	0.9905	**0.9931**
urban-2	0.9945	0.9952	0.8888	0.9907	0.9210	**0.9994**	0.9514	0.9964
AVERAGE	0.9227	0.9591	0.9328	0.9360	0.9529	0.8469	0.9378	**0.9788**

The optimal and sub-optimal values have been highlighted by the bold and underline format.

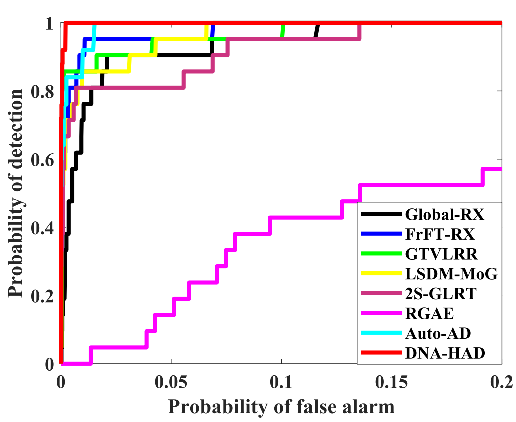


Fig. 7. ROC curves of the HYDICE achieved by different detectors.

parameters to be adjusted. For the FrFT-RX, the optimal order is autodetermined, except for the SanDiego and the beach-4. The autodetermined orders correspond to dissatisfactory detection accuracies for these two HSIs. Hence, we have iterated the optimal order from 0.1 to 1.5 by a step of 0.1 to find the optimal orders for these two HSIs. For the GTVLRR, there are several parameters to be determined. According to the description in the corresponding paper, the λ and β in the GTVLRR model are empirically set as 0.5 and 0.2. the tradeoff parameter is fixed as 0.05. The number of nearest neighbors is 10, and the scalar parameter is set as 1. 20 pixels are selected from each cluster to ensure sufficient diversity of the spectral signature. The number of clusters M is iterated from 10 to 20 by a step of 5 to find the optimal detection accuracy. For the LSDM-MoG, there are two variables to be determined, including the initial rank and the initial number of mixture Gaussian noise. According to the descriptions in the paper, experiments have been conducted by iterating the initial rank from 10 to 100 with an interval 10, and the initial number of mixture Gaussian noise K ranging from 1 to 10. For the 2S-GLRT, there is a double concentric sliding window employed. There is a outer window and an inner window whose size need to determined. According to the paper, the outer window size varies from 5 to 25 by a step of 2, and the inner size ranges from 3 to 15. For the RGAE, it embeds a super pixel segmentation-based graph regularization term into the AE to preserve the geometric structure and the local spatial consistency. There are three parameters to be adjusted, including tradeoff parameter λ, the number of super pixels, and the dimension of hidden layer. As described in the paper, the λ is ranging from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. The number of super pixels is set to $\{50, 100, 150, 300, 500\}$. The dimension of the hidden layer is set to $\{20, 40, 60, 80, 100, 120, 140, 160\}$. Iterations have been done to find the optimal AUCs. For the Auto-AD, it is an autonomous HAD detector, and all the parameters are empirically set and fixed. There are no parameters to be adjusted.

Corresponding optimal parameters for different detectors have been listed in Table III.

C. Parameter Setting

Experiments have been conducted to evaluate the parameter sensitivity of percentage of the negative-sampling. It is noted

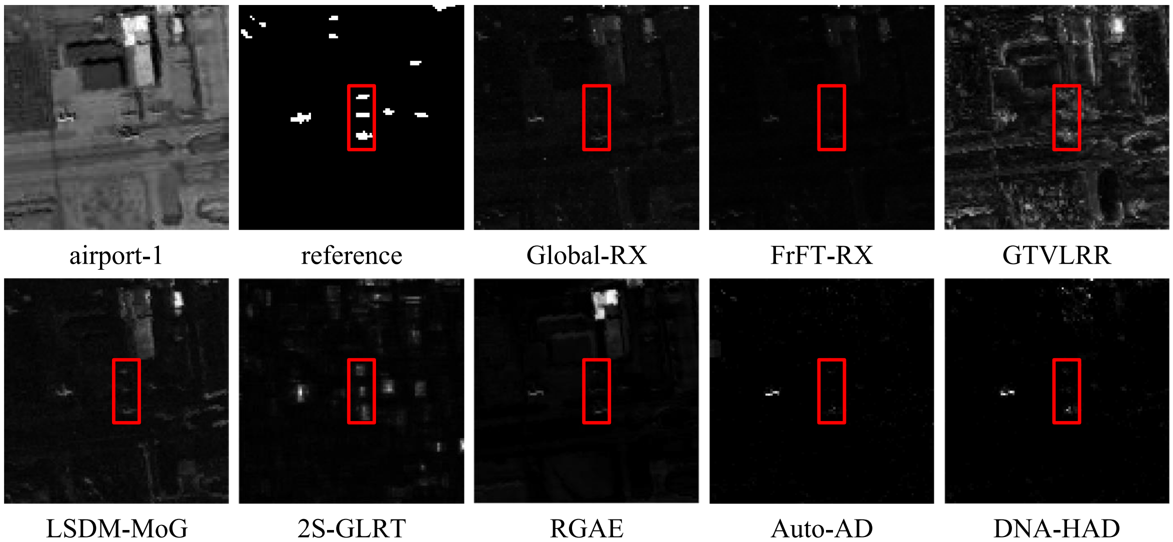


Fig. 8. Visual detection maps of the airport-1 achieved by different detectors.

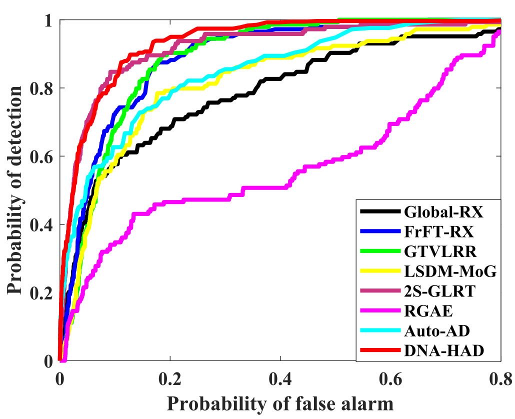


Fig. 9. ROC curves of the airport-1 achieved by different detectors.

that the anomalies usually refer to the pixels which occupy for a small amount of the total pixels, such as the 1.34% of the SanDiego, and the percentage of the negative sampling is ranged from 1% to 5% by a step of 1%. Once the optimal percentage is determined, the weighting parameter between the background error and the negative-sampling error is to be determined, which ranges from 0.1 to 0.5 by a step of 0.1.

Impact of the percentage on the detection accuracy has been plotted in Fig. 2. The x axis denotes the percentage of the negative sampling. The y axis represents the average AUC of the 6 test HSIs. Seen from the average AUC value, it gradually increases as the percentage grows at the first stage, and becomes the optimal when p $=$ 3%. In this way, the percentage is fixed as 3%.

Once the percentage p is determined, the weighting parameter λ is ranged to find the optimal one. Corresponding values have been listed in Fig. 3. It is seen that the DNA-HAD with the adaptive-adjusted loss function always outperforms the DNA-HAD whose λ is 0, which demonstrates the effectiveness of the loss function. Meanwhile, the proposed method achieves the optimal performance when λ is 0.1. In this way, λ is fixed as 0.1 in the experiments.

The other parameters, including the maximum iteration number, the negative slope of the leaky Relu, and the eps of the BN layer are set as 1000, 0.2, and 0.8, respectively.

D. Ablation Experiment

Experiments have been conducted to illustrate the effectiveness of each module in the proposed method, including dynamic sampling of the negative samples, designation of the loss function, and skip connections between the encoder and the decoder. The corresponding average AUC of the six HSIs have been listed in Table IV. Experiments have been first conducted on the network without skipping connections and the negative-sampling module. Because of there is no negative-sampling, the adaptive-adjusted loss function is the same as the normal loss function. In this way, the DNA-HAD degenerates into a traditional AE, just with four M modules as the encoder and another four DM modules as the decoder. The corresponding average AUC is listed in the first row in Table IV.

The second row depicts the performance of the network in which skip connections are made between the M modules of the encoder and the DM modules of the decoder. But there is no negative-sampling involved, and no adaptive-adjusted loss function is set. Comparison between the first two rows has demonstrated the effectiveness of the skip connection. Concatenation between the lower feature at the encoder and the deeper feature at the decoder is benefit for reconstructing the background.

The third row represents the situation that negative sampling is involved, but the loss function just focuses on minimizing the error between the input and the reconstructed HSI. The weighting factor λ is set as 0, a loss function which neglects the effects of the negative samples in the reconstruction process. The fourth row corresponds to the performance of the proposed DNA-HAD.

Comparison between the second row and the third row has demonstrated the effectiveness of the negative-sampling. It is

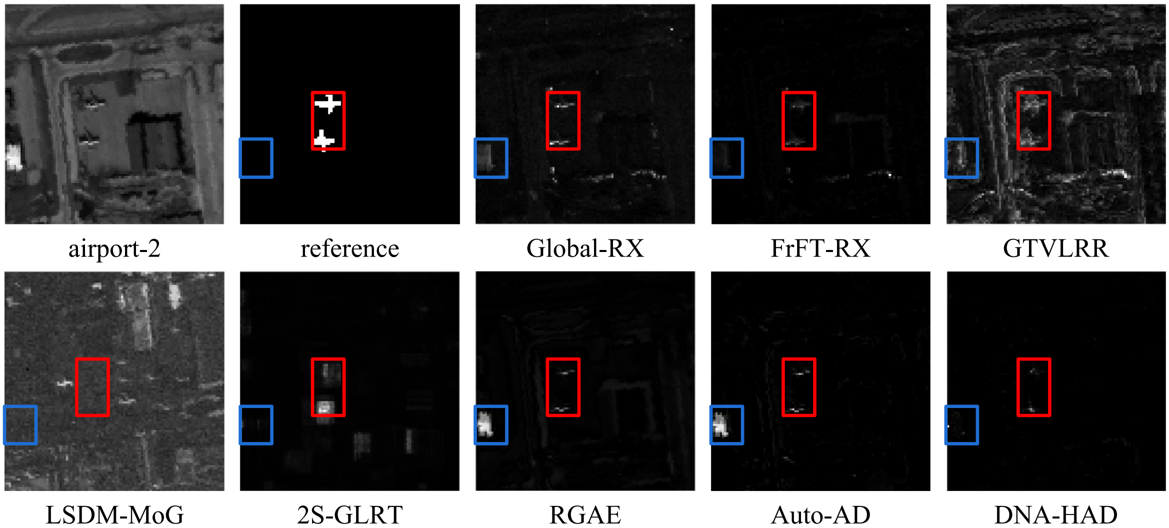


Fig. 10. Visual detection maps of the airport-2 achieved by different detectors.

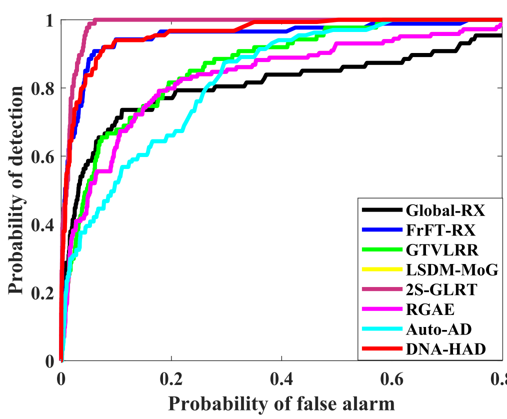


Fig. 11. ROC curves of the airport-2 achieved by different detectors.

noted that with the fixed percentage of negative samples, the network is more robust to the abnormal pixels in recovering the background, making a detection accuracy enhancement. Meanwhile, with the adaptive-adjusted loss function in the DNA-HAD, which corresponds to the fourth row in Table IV, the difference between the input negative samples and the reconstructed ones is further enhanced. In this way, the detection accuracy is further improved.

E. Data Analysis

Figs. 4, 6, 8, 10, 12, and 14 exhibit the visual detection maps achieved by different detectors. The first two columns represent the 100th band of the input HSI and the referenced detection map. Red rectangles denote the regions that contain the anomalies. Blue rectangles represent the regions those are backgrounds.

For the detection maps of SanDiego in Fig. 4, it is noted that the Global-RX, the GTVLRR, and the LSDM-MoG exhibit the comparatively high responses for the right bottom background regions, which will induce a high alarm rate. As to the anomalous regions, the proposed DNA-HAD exhibits high responses to both red rectangle regions. Even the FrFT exhibits high responses for the left red region, the proposed DNA-HAD outperforms it for the top-right anomalous region. Both the 2S-GLRT and the Auto-AD exhibit the low responses to the whole scene. Corresponding AUCs have been listed in Table V, it is noted that the proposed method achieves the second-optimal one, whose value is slightly smaller than the by 0.0072. All the optimal and suboptimal values have been highlighted by the bold and underline format. Fig. 5 plots the ROC curves of these detectors. It is noted that the false alarm rate ranges from 0 to 0.5. Given a false alarm rate larger than 0.5, most of detectors have detected all the anomalies. The ROC curve of the proposed DNA-HAD locates at the left-top region at the most cases, except for the RGAE. It means the proposed method achieves the suboptimal detection accuracy at most of the false alarm rate, which further validates the superiority of the DNA-HAD.

As to the detection maps of HYDICE in Fig. 6, the ten vehicles are randomly distributed in the scene, especially the ones in the red rectangles. It can be seen that the Global-RX, FrFT-RX, GTVLRR, LSDM-MoG, and RGAE all exhibit the high false alarm rates, which cannot suppress the background effectively. Given the small false alarm rate, such as values smaller than 0.01, the 2S-GLRT outperforms the Global-RX. Both the Auto-AD and the DNA-HAD compress the background efficiently, but the DNA-HAD outperforms the Auto-AD in detecting the anomalies, such as both the red rectangles. Corresponding ROC curves have been demonstrated in Fig. 7. Compared with the SanDiego, all the detectors have detected all the anomalies with a lower false alarm rate. In this way, the x axis of Fig. 7 is focused from the range of 0 to 0.2. The curve of the DNA-HAD still locates at the left-top region, which outperforms all the rest detectors. Corresponding AUC values have been listed in Table V, and the proposed DNA-HAD achieves the optimal one.

When it comes to the airport-1, anomalies of this HSI are more than those of the SanDiego and HYDICE. Accurate detection of the anomalies in this HSI is more difficult than detection of

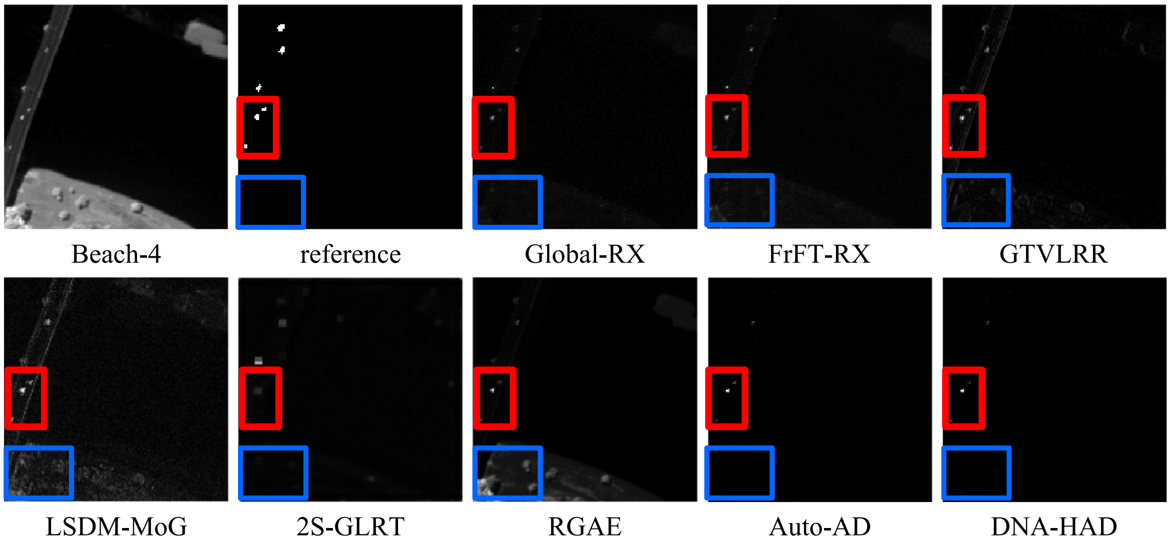


Fig. 12. Visual detection maps of the beach-4 achieved by different detectors.

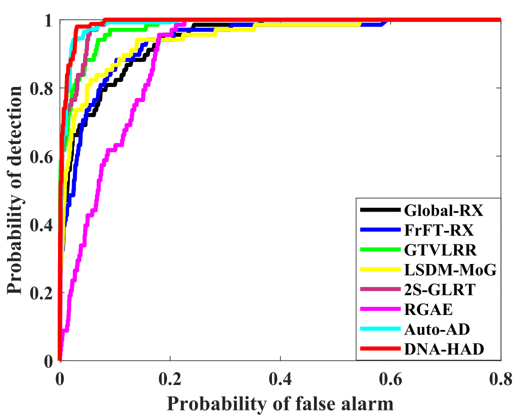


Fig. 13. ROC curves of the beach-4 achieved by different detectors.

the others. Detection maps of this HSI are shown in Fig. 8. The GTVLRR exhibits a high false alarm rate for the whole scene. The RGAE mistakenly detects the top region as the anomalies, and the anomalies in the red region have been missed. The 2S-GLRT detects the anomalies in the red region accurately, but some neighboring backgrounds are also mistakenly detetced. The LSDM-MoG exhibits a high false alarm rate for the top-right region. As to the red rectangle, the proposed detector achieves the best detection accuracy. Corresponding AUC curves have been plotted in Fig. 9. The false alarm rate is nearly 0.8 when the detection accuracy reaches 1. It is noted that given a detection accuracy, the RGAE achieves the largest false alarm rate in most cases, which is consistent to the low AUC in Table V. The proposed method is still the most next to the left-top region of the ROC curves.

As to the airport-2 in Fig. 10, the GTVLRR and the LSDM-MoG still exhibit the high false alarm rates. The Global-RX, the Auto-AD, and especially the RGAE achieve a comparatively high response for the left blue rectangle. The FrFT-RX exhibits low response for the blue rectangle, but it mistakenly regards some other regions as the probable anomalies. The 2S-GLRT shows a high detection accuracy for the red rectangle region, and a good suppression ability for the blue rectangle region. The ROC curves are plotted in Fig. 11. It is noted that the FrFT-RX is with a smaller false alarm rate when the detection rate is from 0.85 to 0.9. However, when the detection rate is above 0.95, the proposed DNA-HAD outperforms the FrFT-RX and the other detectors, besides the 2S-GLRT. This is caused by the rectangle region which locates at the right-bottom of the red anomalous region. Objective AUC values are also listed in the fourth row of Table V.

The beach-4 was captured by the ROSIS sensor, whose spatial resolution is 1.3 meters per pixel. The detection maps are shown in Fig. 12. It is seen that the FrFT-RX, the GTVLRR, the LSDM-MoG and the RGAE are mistakenly sensitive to the left-bottom regions. For the rest detectors, the proposed DNA-HAD exhibits the strongest responses for the red anomalous region. Corresponding ROC curves are plotted in Fig. 13. It is noticed that the DNA-HAD achieves the largest detection rate among the others in most cases, when given a fixed false alarm rate. The objective evaluation AUC values have been listed in the fifth row of Table V.

When it comes to the urban-2 HSI which depicts the Texas coast, the buildings in the scene are marked as the anomalies. The detection maps are shown in Fig. 14. It is shown that the right up region is with small responses in most of the detection maps. For the red rectangle region, both the Global-RX and the DNA-HAD achieve the acceptable responses. However, for the right blue region, both the Global-RX and RGAE still exhibit comparatively high responses. The ROC curves have been plotted in Fig. 15. It is seen that when the alarm rate is set between 0.03 and 0.05, the FrFT-RX slightly outperforms the proposed method. But if the false alarm rate is very strict, such as smaller than 0.015, the proposed method will achieve a superiority over all the other detectors, except for the RGAE. Even the RGAE is robust to SanDiego and the urban-2, its performance for the other HSIs are really unstable. The corresponding overall AUC values have been listed in the last but two row of Table V. It is

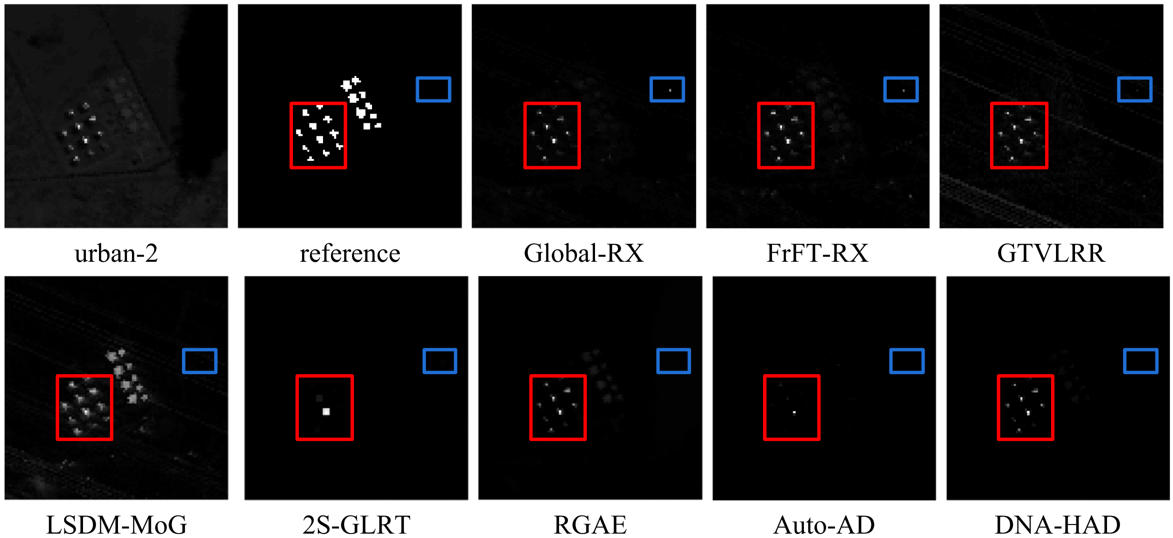


Fig. 14. Visual detection maps of the urban-2 achieved by different detectors.

TABLE VI
EXECUTION TIME OF VARIOUS ANOMALY DETECTORS USING DIFFERENT DATA SETS (UNIT: SECONDS)

	Global-RX	FrFT-RX	GTVLRR	LSDM-MoG	2S-GLRT	RGAE	Auto-AD	DNA-HAD
SanDiego	0.16	18.28	370.21	760.44	4233.75	45087.13	40.93	132.44
HYDICE	0.10	6.42	293.11	519.34	2763.37	40499.70	35.48	113.17
airport-1	0.15	9.15	376.02	369.11	7778.84	35767.54	42.95	130.45
airport-2	0.15	8.42	496.53	1004.83	5250.72	46156.36	61.31	130.77
beach-4	0.14	30.14	686.10	1265.80	2745.99	57234.91	39.95	123.31
urban-2	0.14	8.70	329.64	368.08	5221.90	38533.93	40.28	123.35
AVERAGE	0.14	13.52	425.27	714.60	4665.76	43879.93	43.48	125.58

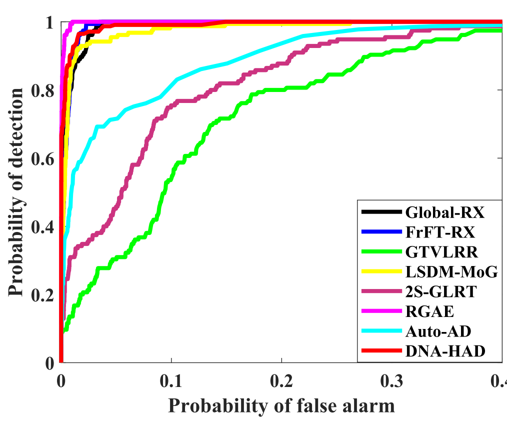


Fig. 15. ROC curves of the urban-2 achieved by different detectors.

observed that the proposed method is still the second-optimal detector for the urban-2 whose spatial resolution is more than 15 meters per pixel, which is just 0.003 smaller than the optimal one.

Both the visual detection maps and the quantified AUC values have demonstrated the effectiveness of the proposed method. By importing the negative sampling strategy and the skip connection between the encoders and the decoders, the background information can be well reconstructed. In this way, it is noted from the detection maps achieved by the proposed DNA-HAD, the false alarm rates are comparatively low. The detection rate achieved by the DNA-HAD is satisfactory.

Table VI provides the computational costs of various detectors. All the experiments have been conducted on the same platform. It is noted that both the Global-RX and the FrFT-RX requires little computational cost. For the Global-RX, the main cost is caused by calculating the covariance matrix and the mean. For the FrFT-RX, there is a preprocessing of autodetermining the optimal fractional order. It is noted that the automatically determined orders for the SanDiego and the beach-4 make the dissatisfactory detection accuracy. Iterations have been done from 0.1 to 1.5 by a step of 0.1, making the comparative high computational cost of this detector.

For GTVLRR, LSDM-MoG, 2S-GLRT, and RGAE, there are several parameters to be determined during the detection process. In this way, the computational costs of these four detectors are comparatively high.

For both the Auto-AD and the proposed DNA-HAD, all the involved hyperparameters have been empirically set before the detection process. In this way, the detection process is faster than the detectors which involved iterations. It is noted that the proposed DNA-HAD consumes much more time than the Auto-AD. It is mainly caused by two reasons. First, the architecture of the DNA-HAD is much deeper than that of the Auto-AD. Second, fixed percentage of pixels have been selected from the whole pixels in the preprocess, which also consumes some costs.

IV. Conclusion

In this article, a novel hyperspectral anomaly detection method via dynamic negative sampling is proposed. Different from the existing methods which reconstruct the background by mainly utilizing the global and local homogeneous attribute of the original HSI. Dynamic negative sampling and adaptive-adjusted loss function are utilized to enhance the sensitivity of the network for the anomalies. Meanwhile, skip connections between the encoder and the decoder makes an insurance for the combination between the lower feature and the deeper feature, which make the reconstruction process more accurate. Experiments on six HSIs which were captured by different sensors and are with different resolutions have demonstrated the effectiveness of the proposed method.

References

[1] P. K. Sethy, C. Pandey, Y. K. Sahu, and S. K. Behera, "Hyperspectral imagery applications for precision agriculture-a systemic survey," *Multimedia Tools Appl.*, vol. 81, pp. 3005–3038, 2022.

[2] I. Racetin and A. Krtalić, "Systematic review of anomaly detection in hyperspectral remote sensing applications," *Appl. Sci.*, vol. 11, no. 11, 2021, Art. no. 4878.

[3] J. Zhang, R. Su, Q. Fu, W. Ren, F. Heide, and Y. Nie, "A survey on computational spectral reconstruction methods from RGB to hyperspectral imaging," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 11905.

[4] X. Wang et al., "A double dictionary-based nonlinear representation model for hyperspectral subpixel target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5524516.

[5] R. A. Borsoi, T. Imbiriba, J. C. M. Bermudez, and C. Richard, "Fast unmixing and change detection in multitemporal hyperspectral data," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 975–988, 2021.

[6] Z. Gong, P. Zhong, and W. Hu, "Statistical loss and analysis for deep learning in hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 322–333, Jan. 2021.

[7] Z. Hou, W. Li, L. Li, R. Tao, and Q. Du, "Hyperspectral change detection based on multiple morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5507312.

[8] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, "Model-guided deep hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 5754–5768, 2021.

[9] L. Zhuang and M. K. Ng, "Fasthymix: Fast and parameter-free hyperspectral image mixed noise removal," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2021.3112577.

[10] C. Zhao, M. Wang, S. Feng, and N. Su, "Hyperspectral target detection method based on nonlocal self-similarity and rank-1 tensor," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500815.

[11] D. B. Gillis, "An underwater target detection framework for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1798–1810, 2020.

[12] 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.

[13] S. Yuan, L. Shi, B. Yao, F. Li, and Y. Du, "A hyperspectral anomaly detection algorithm using sub-features grouping and binary accumulation," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6007505.

[14] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3889–3900, Aug. 2021.

[15] 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.

[16] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.

[17] 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, 2022, Art. no. 5511314.

[18] 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.

[19] J. M. Molero, E. M. Garzon, I. Garcia, and A. Plaza, "Analysis and optimizations of global and local versions of the rx algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.

[20] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[21] 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.

[22] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.

[23] X. Yang, X. Huang, M. Zhu, S. Xu, and Y. Liu, "Ensemble and random RX with multiple features anomaly detector for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6009505.

[24] J. A. P. Hidalgo, A. Pérez-Suay, F. Nar, and G. Camps-Valls, "Efficient nonlinear RX anomaly detectors," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 231–235, Feb. 2020.

[25] 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.

[26] J. Hu, Y. Zhang, M. Zhao, P. Li, and Y. Li, "Transformation of local gradient profiles for hyperspectral anomaly detection," *J. Image Graph.*, vol. 26, no. 8, pp. 1847–1859, 2021.

[27] 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.*, vol. 33, no. 10, pp. 5557–5567, Oct. 2022.

[28] 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.

[29] H. Ning, 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.

[30] 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, 2014, Art. no. 083641.

[31] T. Zhou and D. Tao, "Godec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 33–40.

[32] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.

[33] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1578.

[34] 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.

[35] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[36] 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.s*, vol. 65, no. 1, 2022, Art. no. 112305.

[37] Z. Hou, W. Li, R. Tao, and Q. Du, "Three-order tucker decomposition and reconstruction detector for unsupervised hyperspectral change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6194–6205, 2021.

[38] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5801–5820, Oct. 2016.

[39] J. Wang, Y. Xia, and Y. Zhang, "Anomaly detection of hyperspectral image via tensor completion," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1099–1103, Jun. 2021.

[40] 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.

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

[42] Y. Liu, W. Xie, Y. Li, Z. Li, and Q. Du, "Dual-frequency autoencoder for anomaly detection in transformed hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523613.

[43] K. Jiang, W. Xie, Y. Li, J. Lei, G. He, and Q. Du, "Semisupervised spectral learning with generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 5224–5236, Jul. 2020.

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

[45] 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, 2022, Art. no. 5503314.

[46] K. Kayabol, E. B. Aytekin, S. Arisoy, and E. E. Kuruoglu, "Skewed t-distribution for hyperspectral anomaly detection based on autoencoder," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 5510705.

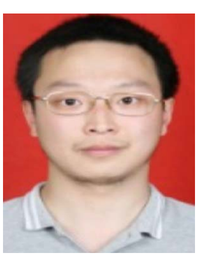
[47] Z. Wu, W. Zhu, J. Chanussot, Y. Xu, and S. Osher, "Hyperspectral anomaly detection via global and local joint modeling of background," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3858–3869, Jul. 2019.

[48] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.

[49] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support, : 4th Int. Workshop, 8th Int. Workshop, Held Conjunction*, 2018, vol. 11045, pp. 3–11.

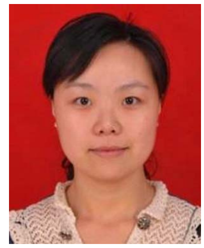
[50] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.

[51] J. M. Lobo, A. Jimenez-Valverde, and R. Real, "Auc: A misleading measure of the performance of predictive distribution models," *Glob. Ecol. Biogeogr.*, vol. 17, no. 2, pp. 145–151, 2008.



Jingxuan Wang received the B.S. degree in IC design and integrated system and the M.S. degree in integrated circuit engineering from Xidian University, Xi'an, China, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree in electronics and information with the School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, China.

His research interests include neural networks, tensor decompositions and applications, hyperspectral image processing, and parallel computing.



Jinqiu Sun received the B.S., M.S., and Ph.D. degrees in guidance navigation and control from Northwestern Polytechnical University, Xi'an, China, in 1999, 2004, and 2005, respectively.

She is currently a Professor with the School of Astronomy, Northwestern Polytechnical University. Her research work focuses on signal and image processing, computer vision, and pattern recognition.

Yong Xia (Member, IEEE) received the B.E., M.E., and Ph.D. degrees in computer science and technology from the School of Computer Science and Engineering, Northwestern Polytechnical University (NPU), Xi'an, China, in 2001, 2004, and 2007, respectively.

He is currently a Professor with the School of Computer Science and Engineering, NPU. His research interests include medical image analysis, computer-aided diagnosis, pattern recognition, machine learning, and data mining.

Yanning Zhang (Senior Member, IEEE) received the B.S. degree in electronic engineering from the Dalian University of Technology, Dalian, China, in 1988, the M.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 1993, and the Ph.D. degree from the School of Marine Engineering, Northwestern Polytechnical University, Xi'an, in 1996.

She is currently a Professor with the School of Computer Science, Northwestern Polytechnical University. She is also a Cheung Kong Professor with the Ministry of Education, Beijing, China. She has authored or coauthored more than 200 articles. She is also the Organization Chair of the ACCV2009 and the Publicity Chair of the ICME2012. Her current research interests include remote sensing image analysis, computer vision, and pattern recognition.

Dr. Zhang is an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.