Hyperspectral Anomaly Change Detection Based on Autoencoder

Meiqi Hu, *Student Member, IEEE*, Chen Wu ⬤, *Member, IEEE*, Liangpei Zhang ⬤, *Fellow, IEEE*, and Bo Du ⬤, *Senior Member, IEEE*

Abstract—With the hyperspectral imaging technology, hyperspectral data provides abundant spectral information and plays a more important role in the geological survey, vegetation analysis, and military reconnaissance. Different from normal change detection, hyperspectral anomaly change detection (HACD) helps to find those small but important anomaly changes between multitemporal hyperspectral images (HSI). In previous works, most classical methods use linear regression to establish the mapping relationship between two HSIs and then detect the anomalies from the residual image. However, the real spectral differences between multi-temporal HSIs are likely to be quite complex and of nonlinearity, leading to the limited performance of these linear predictors. In this article, we propose an original HACD algorithm based on autoencoder (ACDA) to give a nonlinear solution. The proposed ACDA can construct an effective predictor model when facing complex imaging conditions. In the ACDA model, two siamese autoencoder networks are deployed to construct two predictors from two directions. The predictor is used to model the spectral variation of the background to obtain the predicted image under another imaging condition. Then the mean square error between the predictive image and corresponding expected image is computed to obtain the loss map, where the spectral differences of the unchanged pixels are highly suppressed and anomaly changes are highlighted. Ultimately, we take the minimum of the two loss maps of two directions as the final anomaly change intensity map. The experiments results on public "Viareggio 2013" datasets demonstrate the efficiency and superiority over traditional methods.

Index Terms—Anomaly change detection, autoencoder (AE), feature extraction, hyperspectral image (HSI).

I. Introduction

CHANGE detection for remote sensing images refers to acquiring the different information of landscapes in the same location by observing it at different times [1]–[3]. With high spectral resolution, the hyperspectral image could distinguish various objects more accurately [4], [5]. Nowadays, HSI change detection has been extensively applied in land use and land cover change analysis [6], resource exploration [7], vegetation change analysis [8], and damage assessment [9].

Compared to the conventional hyperspectral change detection, hyperspectral anomaly change detection (HACD) [10]–[13] focuses on finding those small and rare changes, which have different distribution with the background changes. The background refers to all ground objects that present at the same location in both HSIs. And the anomalous changes may arise from insertion, disappearance, or movement of small size objects (generally man-made). Besides, the camouflage, concealment, and deception of little stationary objects also bring about anomaly changes. HACD serves as a reminder of these easily overlooked changes for decision makers [14] and has been widely applied in airborne defense and surveillance, agricultural surveying, and illicit crop identification [15]–[17].

A straightforward procedure for HACD is to detect anomaly from the difference images of multitemporal HSIs. However, the spectral variations resulting from diverse imaging conditions in the difference image can result in plenty of pseudo changes, since the solar height angle, illumination, and atmosphere condition may have altered tremendously. In theory, these pseudo changes could be suppressed if the multitemporal images were acquired under the same imaging condition and the unchanged landscapes would show quite similar spectral features. Therefore, anomaly change detection could be done by establishing the mapping relationship from one image to another in order to get a predictive image and then comparing the spectral differences. The idea of predicting the image under the imaging condition of another image is concluded as a predictor model [18]. Chronochrome (CC) [19] is such a classical predictor that models the spectral differences of background by the least square linear regression. CC is the first global linear predictor that obtains a predicted image and detects the anomaly changes from the residual image. Since misregistration error quite influences the performance of CC, the emergence of covariance equalization (CE) [20] solves this problem. CE functions as whitening [21] and is assumed to be able to unify the distribution of two HSIs. Concretely, the two classical methods for HACD, CC, and CE employ the statistical features of the two images to establish the relationship mapping from one imaging space to another. But both of them obey the linear space-invariant assumption, which refers to the affine transformation containing no target changes for global linear predictors. Carlotto [22] combined clustering with

Reed-Xiaoli (RX) [23], [24] to detect man-made changes, which first segmented the reference image into several clusters and then used RX to detect anomaly changes within each cluster. RX aims at finding out those pixels that deviate from the main distribution. And the clustering provides additional freedom to adapt to the space-variant background. Eismann *et al.* [11], [18] proposed a new method that divided the background into several classes and mapped one by one to get a prediction. The approach really obtains better detection effect with the segmented linear prediction. But the number of clusters is hard to determine and need to be selected by trial-and-error in the experiment. It is worth noting that most previous predictor models are all linear models.

To sum up, the traditional methods use the statistical features of images to construct linear predictors, which suffer from the inability of representing the complex relationship of the imaging conditions. Specifically, there are abundant bands of hyperspectral images, which cover a wealth of surface features. Thus a predictor model for HACD requires strong capability of feature extraction. And a nonlinear predictor based on deep learning may give a solution.

Deep learning is composed of multiple processing layers, which are similar to neurons of the human brain, and is able to learn multiple levels of representation of data [25], [26]. And deep learning has gained remarkable performance in image processing of remote sensing [27], such as classification [28], [29], target detection [30], [31], and change detection [32], [33]. Among deep learning architectures, autoencoder (AE) has shown powerful feature extraction ability in [34], which makes it possible to establish a nonlinear complex mapping relationship. The under-complete autoencoder (hereafter referred to as AE) is characterized by the design of the bottleneck, which limits the dimension of information transmitted after the input layer and helps to extract the crucial features.

Therefore, inspired by the ability of the feature extraction of AE, we proposed a method that employs AE as a nonlinear predictor for HACD called hyperspectral anomaly change detection based on autoencoder (ACDA). In the proposed method, we utilize two AE networks to get two predictive images, respectively. The special structure "bottleneck" of AE is able to extract the essential features at a lower dimension and reconstruct the input. Using predetected unchanged pixels as training samples, the input is the spectral vector of one HSI and the desired output is corresponding to one of another HSI. These two spectral vectors both belonging to the background own consistent essential features but differ in the spectral values. The spectral differences between them result from the different imaging conditions. Thus, the AE network transforms the HSI into another imaging condition to get a predictive image. Then the loss maps will be calculated by computing the mean square error (MSE) between the predictive images and corresponding expected images. Finally, we take the minimum of the two loss maps as the anomaly intensity map. In addition, the ability of feature extraction and nonlinearity of AE enable the model to deal with the problem of space-variant background. The intention of AE predictor is to minimize the spectral differences of multitemporal HSIs and highlight the anomaly changes. The main contributions of this article are concluded as follows.

1) A nonlinear predictor method based on AE denoted as ACDA is proposed for HACD, which gains greater detection performance against other state-of-the-art approaches. And the network structure of the model is compact and simple, which is easy to implement and less time-consuming.

2) By utilizing the bottleneck structure of the AE, the proposed method is effective in extracting the intrinsic information of the high dimensional spectral vector, which is critical to construct the predictive relationship of different imaging conditions.

3) In practice, one of the major issues encountered by HACD lies in the variant background space, which causes the violent and unbalanced spectral variance between two backgrounds of multitemporal HSIs. Combined the feature extraction and nonlinear mapping, the proposed ACDA obtains good results in this case.

The rest of this article is organized as follows. Section II gives a representation of the proposed method ACDA. Then we implement the algorithm and two experiments results on real-world datasets are presented in Section III. And Section IV concludes this article.

II. METHODOLOGY

A. Autoencoder

AE is an unsupervised deep learning network that is able to learn features from the unlabeled data [34]. AE aims at replicating the output from the input and learning the representation of the input and has broad application, such as dimension reduction [35], image classification [36], and hyperspectral unmixing [37]. Generally, the architecture of AE consists of an encoder and a decoder, where the encoder is used to extract the features of the input, and the decoder is designed to decode the feature and reconstruct the input. Concretely, for a single-layer AE network, the encoder is composed of an input layer as well as a hidden layer, and the decoder is made up of the hidden layer and an output layer. And joint hidden layer between the encoder and decoder is also called the code layer. The structure of AE is generally symmetrical.

There are two significant characteristics of the AE. 1) The neural unit number of the input layer is equal to the one in the output layer. 2) The size of the hidden layer is smaller than the input layer. As for a high-dimensional input, the encoder first transforms it into a low-dimensional code, and then the decoder recovers the data from the code. Such a special structure is referred to as a bottleneck, which is vital to learn feature representation in an unsupervised manner.

For an input sample $z \in \mathbb{R}^P$, the output of the encoder can be written as

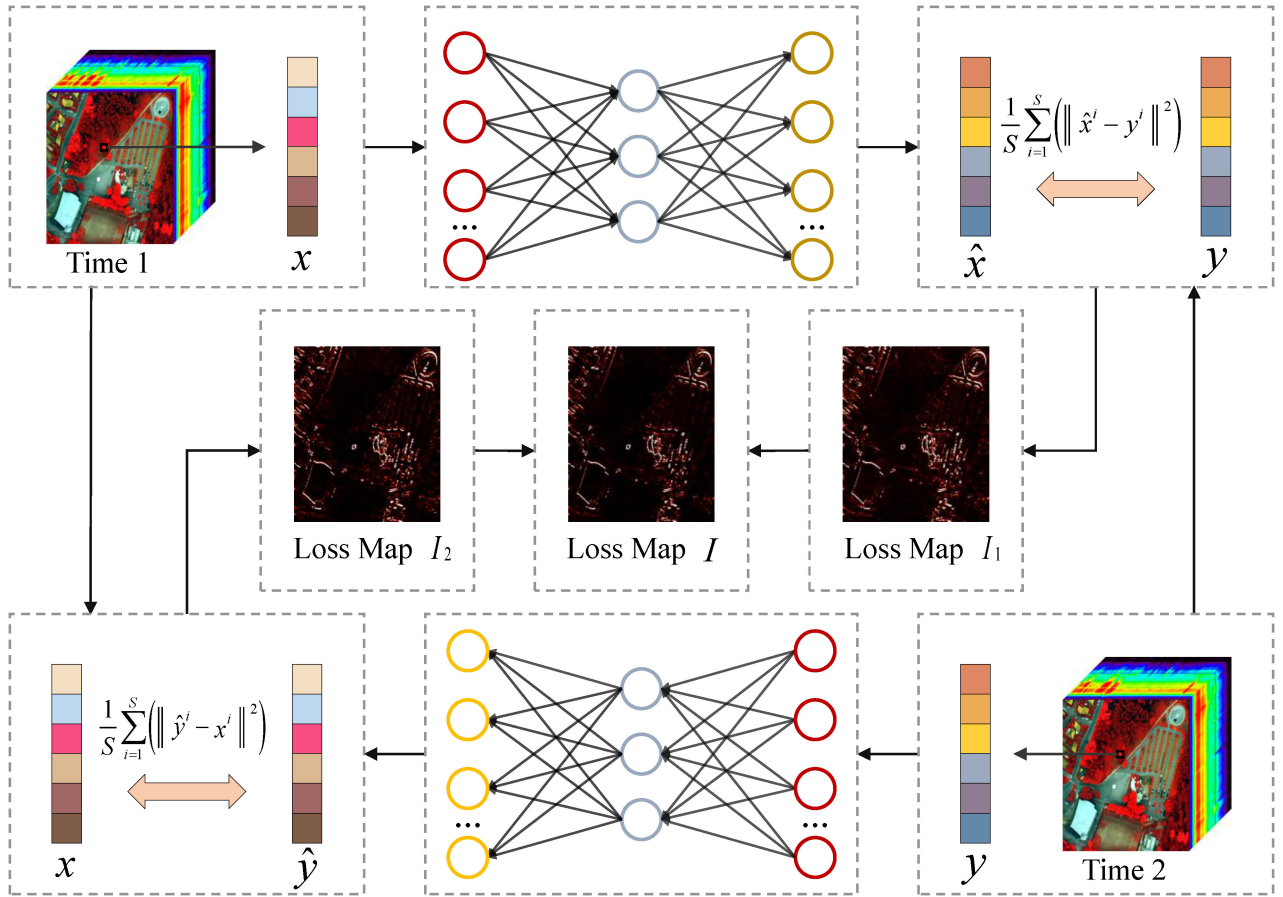$$E(z) = g\left(w^{(e)}z + b^{(e)}\right) \tag{1}$$

Fig. 1. Architecture of the proposed ACDA. The top part of the network trains a predictor from HSI of Time 1 to HSI of Time 2, and the bottom part of the network trains a predictor from the HSI of Time 2 to Time 1. Then the predicted vector of Time 2, $\hat{x}$ and the original spectral vector of Time 2, y is used to compute the loss map I_1. So is the loss map I_2. After two loss maps are acquired from two directions, the minimum of the two loss maps are computed as the final loss map I.

where $w^{(e)} \in \mathbb{R}^{K \times P}$ is the weight matrix with K features, $b^{(e)} \in \mathbb{R}^{K \times 1}$ is the bias vector, and g refers to the activation function, individually. Then the hidden code is put into the decoder to get a reconstructed result $\hat{z}$

$$\hat{z} = g\left(w^{(d)} E(z) + b^{(d)}\right) \qquad (2)$$

where $w^{(d)} \in \mathbb{R}^{P \times K}$ is the weight matrix with P features and $b^{(d)} \in \mathbb{R}^{P \times 1}$ is the bias vector, separately. And θ refers to all parameters $\{w^{(e)}, b^{(e)}, w^{(d)}, b^{(d)}\}$ that need to be trained in the AE network. Given the training sets z^i, $i = 1, 2, \ldots, N$, the θ can be iteratively updated by minimizing the reconstruction error which adopts the MSE as the cost function, as the following equation depicts:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} \left(\left\|\hat{z}^i - z^i\right\|^2\right). \qquad (3)$$

The AE tries to learn an approximation that the output is as similar as possible to the input. Under the constraint that the unit number of the hidden layer is less than the one of the input layer, the code layer is forced to learn a compressed feature representation of the input information. Furthermore, the hidden representation extracts the essential information, which can reconstruct the input.

B. HACD Based on AE

As mentioned above, AE is capable of extracting the essence from the input data due to the bottleneck. Generally, when detecting the anomaly changes from multitemporal hyperspectral images, we suppose that the ground objects of background do not change, but the spectral features may alter because of the different imaging conditions. Therefore, taken the spectral vector of Time 1 as the training input and the corresponding vector of Time 2 as the training output, the spectral variation of the pair pixel vectors can be fitted by the AE network, since the two spectral vectors of the unchanged pixel have the same essential information. AE functions as a predictor that establishes the mapping relationship between two imaging conditions. And anomaly changes could be detected from the residual image where the spectral differences of the background between the predicted image and the expected image are suppressed.

The overview of the proposed ACDA is shown in Fig. 1. The inputs of ACDA are pairwise spectral vectors of hyperspectral imagery. Then ACDA could be roughly divided into two parts:

predictor module and postprocessing module. In the predictor module, two systematic AE, whose layers are all fully connected (FC) layers, are used to get two predictive images. In Fig. 1, the red nodes represent the nodes of the input layer, the orange nodes are used to represent the output layer, and the blue nodes are the nodes of hidden layers. After acquiring two predicted images, we choose the MSE between the predictive image and the expected image as the loss map. And in the postprocessing module, we use the minimum between the two loss maps as the final anomaly change intensity map.

1) AE Predictor Model: Generally, AE is used to reproduce the output from the input. In this article, AE is deployed as a predictor by changing the loss function, where the spectral vector of one HSI is taken as input and the corresponding vector of another multitemporal HSI is served as the desired output. We choose the training samples from the predetection results of unsupervised slow feature analysis (USFA) [38], [39], which has showed outstanding comprehensive results in [40]. USFA tries to extract the invariant features and detects the anomalies from transformed difference features.

Mathematically, the AE predictor model is defined as follows. Let us denote $X \in \mathbb{R}^{M \times Q}$ as a HSI acquired at Time 1 and $Y \in \mathbb{R}^{M \times Q}$ as another HSI acquired at Time 2, individually, where M and Q refer to the number of pixels and spectral channels, separately. Let us also denote $x = [x_1, x_2, \ldots, x_Q]^T$ as a training pixel vector of X, $y = [y_1, y_2, \ldots, y_Q]^T$ as the corresponding pixel vector of Y.

Supposing there are n hidden layers and h_i denotes the number of nodes of ith hidden layer. Given x fed into the AE model, the output of the first hidden layer can be written as

$$f_1^1(x) = g\left(w_1^1 x + b_1^1\right) \qquad (4)$$

where $w_1^1 \in \mathbb{R}^{h_1 \times P}$ and $b_1^1 \in \mathbb{R}^{h_1 \times 1}$ are weight matrix and bias vector, g refers to the activation function. The superscript of f_1^1 means the first predictor model mapping from X to Y and subscript of f_1^1 means the first hidden layer. The same are true for w_1^1 and b_1^1. The output of the last hidden layer can be written as

$$f_n^1(x) = g\left(w_n^1 f_{n-1}^1(x) + b_n^1\right) \qquad (5)$$

where $w_n^1 \in \mathbb{R}^{h_n \times P}$ and $b_n^1 \in \mathbb{R}^{h_n \times 1}$ are weight matrix and bias vector, separately. The output of the model can be formulated as

$$\hat{x} = f(\theta_1, x) = g\left(w_{n+1}^1 f_n^1(x) + b_{n+1}^1\right) \qquad (6)$$

where $w_{n+1}^1 \in \mathbb{R}^{Q \times h_n}$ and $b_{n+1}^1 \in \mathbb{R}^{Q \times 1}$ are weight matrix and bias vector, respectively. And θ_1 refers to all parameters $\{w_1^1, b_1^1, \ldots, w_n^1, b_n^1, w_{n+1}^1, b_{n+1}^1\}$ that need be trained in the first predictor model. Given the training sets $x^i, i = 1, 2, \ldots, S$ and corresponding desired output $y^i, i = 1, 2, \ldots, S$, the loss function of the model can be described as

$$L(\theta_1) = \frac{1}{S} \sum_{i=1}^{S} \left(\|\hat{x}^i - y^i\|^2\right) + \lambda_1 \sum_{j=1}^{n+1} \|w_j^1\|^2 \qquad (7)$$

where the cost function is the MSE, w_j^1 refers to the weight parameter of different layers, and λ_1 is the regularization coefficient of the first predictor model.

After the predictor has been trained, all pixel vectors of the image X are fed into the predictor to get a predictive image $\hat{X}$. The predictor is not a prophet that predicts what will happen on Time 2, but rather a transformer that maps the landscape from the imaging condition of Time 1 to the imaging condition of Time 2. Consequently, the imaging condition of the predicted image $\hat{X}$ is the same as the one of expected image Y. Furthermore, the spectral differences of the background between $\hat{X}$ and Y are highly suppressed. While the location with a new appearance or disappearance will get high-intensity value. We use the MSE between the predicted $\hat{X}$ and the expected Y as the anomaly change intensity as follows:

$$I_1 = \text{MSE}\left(\hat{X}, Y\right). \qquad (8)$$

Considering the change direction of two images, we can get another AE predictor model with predicted image $\hat{Y}$ when feeding the pixel spectral vectors of Y. The loss function of the training model from Y to X be described as

$$L(\theta_2) = \frac{1}{S} \sum_{i=1}^{S} \left(\|\hat{y}^i - x^i\|^2\right) + \lambda_2 \sum_{j=1}^{n+1} \|w_j^2\|^2 \qquad (9)$$

where θ_2 is the training parameter $\{w_1^2, b_1^2, \ldots, w_n^2, b_n^2, w_{n+1}^2, b_{n+1}^2\}$ of the second predictor. w_j^2 refers to the weight parameter of different layers and λ_2 is the regularization coefficient. The loss map I_2 can be described as the MSE of the spectral vector between $\hat{Y}$ and X as follows:

$$I_2 = \text{MSE}\left(\hat{Y}, X\right). \qquad (10)$$

2) Postprocessing: After two loss maps are obtained, we adopt the minimum operation on the two loss maps I_1 and I_2 in order to get better detection performance with less noise. Since the anomaly change detector adopts MSE, both loss maps cover bidirectional changes, such as the appearance and disappearance of small objects. The final anomaly change intensity map I can be expressed as follows:

$$I = \min(I_1, I_2). \qquad (11)$$

The reasons for using min operation can be summarized as follows.

1) If a pixel belongs to anomalous change, the anomaly intensity value in both loss maps is large, and then the smaller one is also a big value. As a result, the anomalous information is preserved after min operation.

2) If a pixel does not have an anomalous change, both of the intensity values in the two loss maps are small, the minimum of those two makes it less possible to be an anomalous one.

3) If one of the two intensity values is largely resulting from the training error or local environment, the minimum operation can cut down the possibility of being anomaly change, leading to less noise of the final result. Thus, the

---

Algorithm 1: Process of Training and Generating Anomaly Change Intensity Map for ACDA

Input:

Hyperspectral images X and Y;

Output:

The anomaly change intensity map I;

1:    Employ USFA pre-detection to generate training samples x^i, $i = 1, 2, \ldots, S$ and y^i, $i = 1, 2, \ldots, S$;

2:    Initialize the AE network's parameters $\{\theta_1, \theta_2\}$;

3:    **while** $epoch < max\_epochs$ **do**

4:    Calculate the predictive spectral vector $\hat{x}$ and $\hat{y}$;

5:    Calculate the loss function value $L(\theta_1)$ and $L(\theta_2)$;

6:    Back propagation to update the parameters with gradient descent algorithm;

7:    $epoch++$;

8:    **end while**

9:    Calculate the predicted image $\hat{X}$ and $\hat{Y}$;

10:   Calculate the loss map $I_1 = \text{MSE}(\hat{X}, Y)$ and $I_2 = \text{MSE}(\hat{Y}, X)$;

11:   Compute the final anomaly intensity map $I = \min(I_1, I_2)$;

12:   **return** I;

---

min operation can absorb and amplify the advantage of the two loss maps.

3) Scheme of Proposed ACDA: The detailed implementation of ACDA is depicted in Algorithm 1.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In order to illustrate the effectiveness of the proposed method, abundant experiments on two benchmark hyperspectral datasets for HACD have been conducted. In this section, the description of the two real hyperspectral datasets and the experiments are first introduced in detail. Then the anomaly change detection performance and analyses are represented, which are followed by the parameter analyses about the hidden units and the selection strategies of training samples. Finally, we present the time cost analysis of four involved deep learning based algorithms.

A. Hyperspectral Datasets

The benchmark datasets, "Viareggio 2013" datasets [41], include three hyperspectral images acquired on May 8–9, 2013, in Viareggio, Italy, by an airborne hyperspectral sensor SIM.GA. The sensor collects spectral information ranging from 400 to 1000 nm at a spectral resolution of 1.2 nm approximately. Besides, the ground spatial resolution is 0.6 m. As shown in Fig. 2, (a) is D1F12H1 that acquired on May 8, 2013, 14.18 (Greenwich Mean Time, GMT), (b) is D1F12H2 that acquired on May 8, 2013, 14.33 (GMT), and (c) is D2F22H2 that acquired on May 9, 2013, 12.64 (GMT). Since D1F12H1 and D1F12H2 are acquired on a cloudy day and the radiation comes from the scattered light of the sun, the imaging conditions of D1F12H1
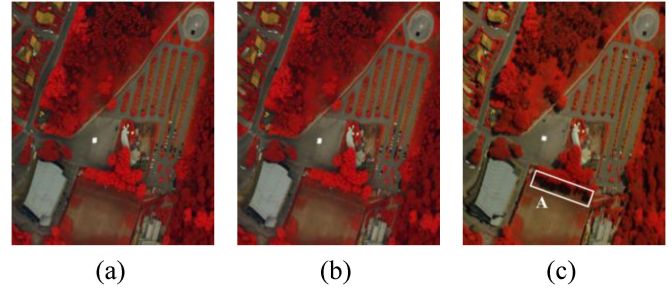

Fig. 2. Viareggio datasets used for experiments. (a) D1F12H1. (b) D1F12H2. (c) D2F22H2.

and D1F12H2 are closely similar. While D2F22H2 is obtained on a clear sunny day, it is obvious that there is a shadow in region A, which indicates that D2F22H2 is under a space-variant condition. In this article, two groups of experiments are carried out, where D1F12H1 and D1F12H2 make up EX-1, while EX-2 consists of D1F12H1 and D2F22H2. These public datasets are available with five different preprocessing levels. We choose preprocessed data with destriping, noise-whitening, and spectrally binning. Concretely, the data was processed at first by multilinear regression to eliminate residual striping noise. And noise-whitening was used to normalize the noise variance. A spectral binning was also employed to reduce random noise by averaging the four consecutive spectral channels. All images have been coregistered available on these public datasets. For each HSI, there are 127 spectral bands with a scene size of 450×375.

It should be noted that all anomaly changes including appearance and disappearance are taken into consideration for the ground truth map, for these two types are all anomaly changes.

B. Experimental Setting

In our ACDA model, AE predictors are composed of FC layers. There are three hidden layers with ReLU [25] as activation function and all weight along with bias matrices are initialized by he-normal way [42]. Using Adam [43] as an optimizer, the ACDA is implemented by PyTorch. Besides, adding L2 regularization to loss function is good for avoiding over-fitting. For each experiment, the regularization coefficient is selected as 0.001 by trial from the range [10e-8, 10e-1]. The training epochs are set as 200 in all experiments, and the learning rate and batch size are 0.001 and 256, respectively. As for the number of different hidden units, we design an AE network with three hidden layers, which are defined as h_1, h_2, and h_3, respectively. The size of h_3 is the same as the size of h_1. We denote ACDA-h_1-h_2 as ACDA model with the unit size of h_1 in the first hidden layer and h_2 in the second hidden layer. In the experiments, we adopt ACDA-60-40 model and the detailed analysis about the influence of the different number of hidden units on algorithm performance is discussed in Section III-D.

Training samples are selected from predetected results of USFA instead of manual selection. For each experiment, 10 000 samples, which are 6% approximately of the total number of

the pixels, are chosen from the predetected results at random. Specifically, the number of bands whose 1/eigenvalue is greater than 1 is automatically selected for the anomaly detection of USFA. Then the anomaly change intensity result is sent to the K-means classifier with 3 clusters and the cluster samples with the smallest cluster center are assigned as the desired training set. Owning to the random initialization of weight matrices and bias vectors, we take the average of ten independently repeated results as the final anomaly intensity map. And the assessment is based on the average result.

Ten comparable algorithms are conducted on the datasets mentioned above, including difference reed-xiaoli algorithm (Diff-RX) [24], straight anomalous change detector (SACD), simple difference anomalous change detector (SDACD), simple difference hyperbolic anomalous change detector (SDHACD) [44], CC [19], CE [20], three FC [45] neural networks and USFA. RX is known for anomaly detection, and the diff-RX algorithm detects anomalies directly from the difference of two images to achieve the purpose of detecting anomaly changes. Based on the assumption that images are subject to multivariate Gaussian model, SACD, SDACD, and SDHACD detect anomaly changes from the likelihood ratio between probability density functions of HSIs. CC and CE are two canonical predictor methods based on the statistical features of the images. As for the three FC neural networks, they are all predictor models for unusual change detection, which uses no labeled data and a three-layer quick prop-trained neural network to build a nonlinear predictor used for multispectral and panchromatic data sets. We refer to this algorithm here for HACD. The training settings and cost function adopted by the FC networks are the same as that of ACDA, but the number of hidden layer units is greater than or equal to the number of input layer units. Concretely, each network has three hidden layers and refers to FC-h_1-h_2, with the first hidden layer having the same number of cells as the third. The three networks, FC-127-127, FC-127-200, and FC-200-200, are specifically designed to be compared with bottleneck structure in this article. And USFA is the predetection algorithm used in this article. Various types of predictor models are used for comparison, including CC, CE, FC-127-127, FC-127-200, and FC-200-200. Note that the total framework as well as the min operation suggested in this article is also used for these predictor models to produce the final results for them. Besides, all the methods based on deep learning are repeated ten times and the average of the anomaly change intensity maps is taken as the final anomaly change map as well as the evaluation.

Moreover, receiver operating characteristic (ROC) [46] curve as well as area under curve (AUC) are adopted for quantitative assessments. ROC is widely used for no threshold to test the performance of anomalous change detection. The horizontal axis of the ROC curve is the false alarm rate and the vertical axis is the detection rate. In detail, the closer the ROC is to the upper left corner, the better the performance of the method is. With the same false alarm rate, the method with a higher detection rate gains superior detection results. And the AUC is the area enclosed by the ROC curve, horizontal axis and vertical axis. The bigger the AUC is, the greater the algorithm is.

C. Results and Analysis

In this section, the detection results of the experiments are first analyzed. Then the analysis of the impact of the min operation on the final results is presented.

Figs. 3 and 4 show the anomaly change intensity maps of D1F12H1-D1F12H2 and D1F12H1-D2F22H2 by diff-RX, SACD, SDACD, SDHCD, CC, CE, FC-127-127, FC-127-200, FC-200-200, USFA, and ACDA-60-40, respectively. There are all 2% linearly stretched for visual comparison. For the D1F12H1-D2F22H2 dataset, there are 849 pixels of anomaly changes showed in Fig. 3(l). And Fig. 4(l) is the ground truth map of D1F12H1-D2F22H2, with 1778 pixels of anomalous changes. The brighter region is more likely to be the anomaly change. As shown in Fig. 3, most of the bright regions in the result of SACD, SDHACD, FC-200-200, and ACDA-60-40 correspond to the anomalous changes of the ground truth map for the D1F12H1-D1F12H2 dataset. And there are lots of noises on results of diff-RX, SACD, SDACD, CC, and CE. By contrast, FC-127-127, FC-127-200, FC-200-200, USFA, and ACDA-60-40 are effective in suppressing the background, where the methods based on deep learning work better. For the D1F12H1-D2F22H2 dataset, it is obvious that SACD, CE, USFA, and ACDA-60-40 detect most of the anomaly changes presented in Fig. 4. Yet there are lots of noises and pseudo changes in the result of diff-RX, SACD, SDACD, CC, and CE, which are consistent with the results of D1F12H1-D1F12H2. Since D2F22H2 is acquired on a sunny day and there are large distinct shadows of the vegetation on the image, it is a challenge to deal with huge gaps between the imaging conditions. Except for several algorithms seriously affected by noise mentioned above, e.g., SACD, the analyses about the performance of the other algorithms on the shadow problem are summarized as follows. It is apparent that there are high anomaly intensity values in the shadow area for the results of SDHACD and USFA. As for SDHACD, it belongs to the space invariant assumption and supposes that the spectral vector is sampled from the multivariate Gaussian model. But D1F12H1-D2F22H2 dataset is under a variant background space and the dataset might not obey the distribution perfectly. And USFA is designed to suppress the spectral differences of the slowly varying pixels to highlight the changed ones. But the performance of USFA, in this case, is not satisfying, which may lie in the large differences of the shadow area hard to be compressed. Whereas the ACDA-60-40, FC-127-127, FC-127-200, and FC-200-200 suppress the big spectral differences greatly, of which ACDA-60-40 detects more anomaly changes. It is indicated that ACDA-60-40 shows good performance on shadow suppression and anomaly change detection. The powerful nonlinearity of neural networks promotes the effect of the predictor methods based on deep learning. Furthermore, compared to the other FC networks, the special structure bottleneck enables AE to extract the essential features of the input and is critical to establish a valid mapping relationship between the two backgrounds.

Fig. 5 represents the ROC performance of the eleven techniques for two data sets and the details are zoomed in. It can be observed that the ROC of ACDA-60-40 in black is closest to the top left. Concretely, when the false alarm rate is under
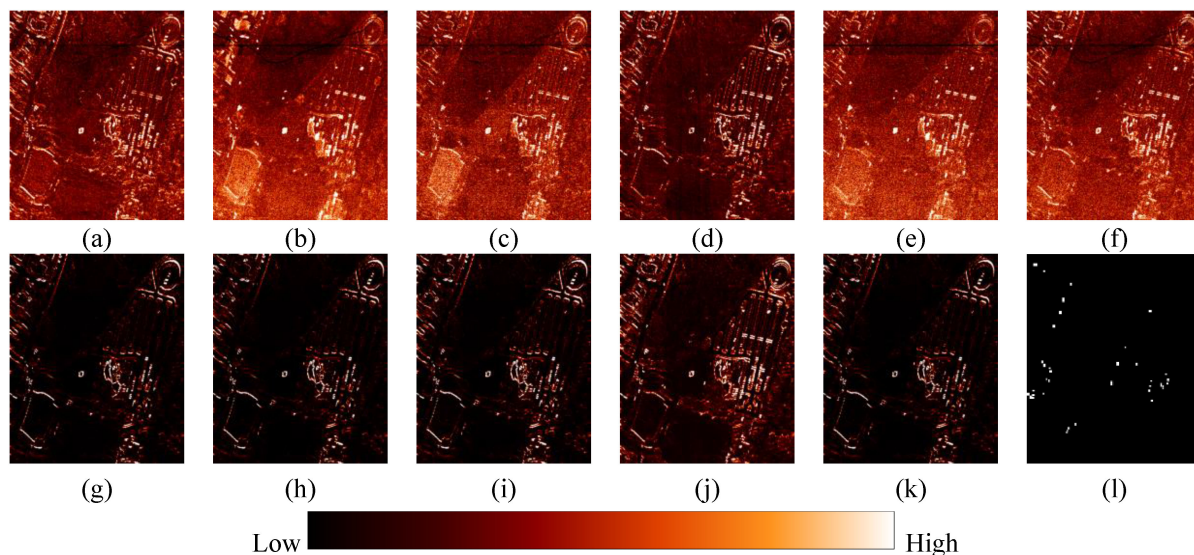
Fig. 3.   Anomaly change intensity map of D1F12H1-D1F12H2. (a) Diff-RX. (b) SACD. (c) SDACD. (d) SDHACD. (e) CC. (f) CE. (g) FC-127-127. (h) FC-127-200. (i) FC-200-200. (j) USFA. (k) ACDA-60-40. (l) Ground truth map.
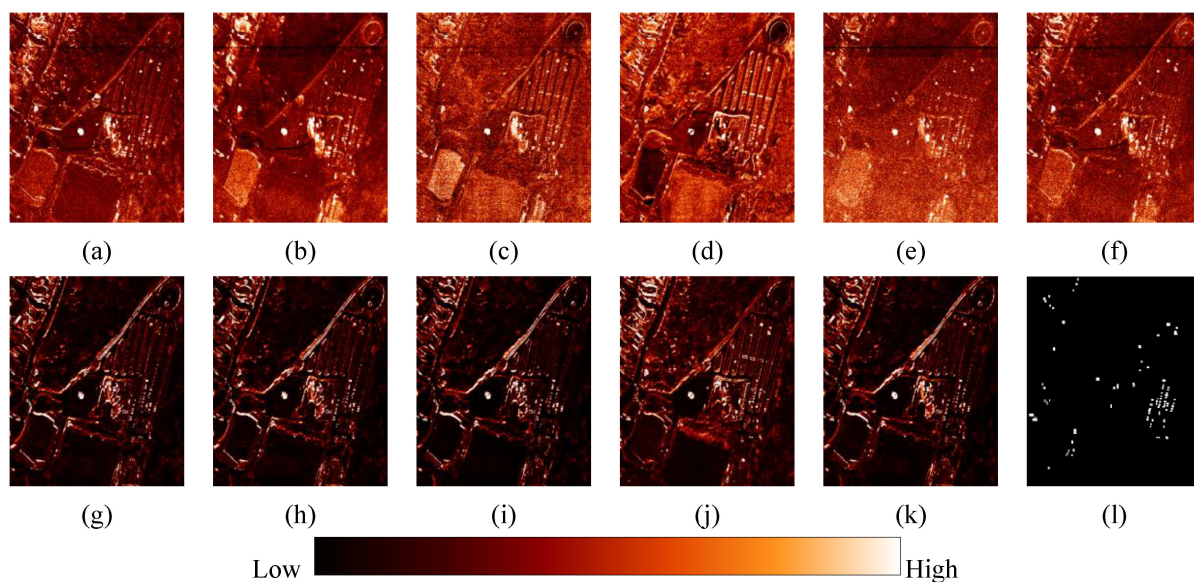

Fig. 4.   Anomaly change intensity map of D1F12H1-D2F22H2. (a) Diff-RX. (b) SACD. (c) SDACD. (d) SDHACD. (e) CC. (f) CE. (g) FC-127-127. (h) FC-127-200. (i) FC-200-200. (j) USFA. (k) ACDA-60-40. (l) Ground truth map.

the low-level range, CE, diff-RX, and USFA all show high detection rates from the left enlarged view of Fig. 5(a) and (b). And ACDA-60-40 obtains the best detection rate from the right enlarged view under a high false alarm rate range. Fig. 6 presents the quantitative evaluation of two experiments. The bigger AUC value refers to the greater effect. And the maximum is highlighted in bold in figures. For D1F12H1-D1F12H2, the proposed method has the largest AUC values of 0.8221. The performance of USFA ranks second with 0.8112, followed by FC-127-127 with 0.8019, FC-200-200 with 0.7996, and diff-RX with 0.7984. And other approaches based on deep learning also perform well, especially compared with SACD and CC. In addition, CC works

worst at an AUC value of 0.7412, which may be attributed to the impact of misregistration. In the case of D1F12H1-D2F22H2, the proposed method gains the largest AUC values again with 0.8451. FC-200-200, FC-127-200, and CE also acquire satisfying results with 0.8387, 0.8365, and 0.8359. Besides, the AUC values of SDACD and SDHACD are extremely low at 0.6778 and 0.6708, respectively, whereas SACD gains a nice AUC value with 0.8252. These three methods are all under the multivariate Gaussian model, but the performances are quite different. The reason may be that the SACD adopts a joint vector observation model while SDACD and SDHACD both employ the difference vector observation model, since the spectral differences between
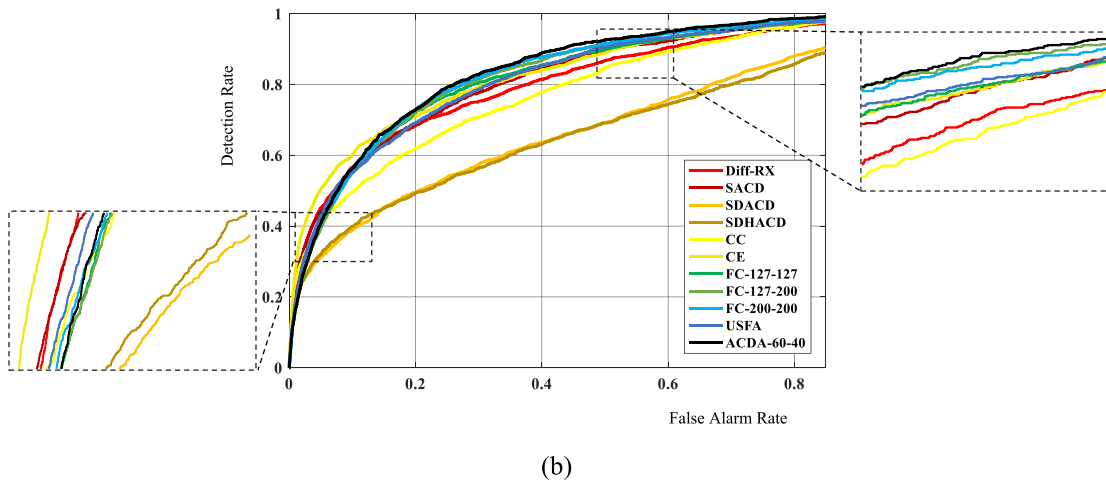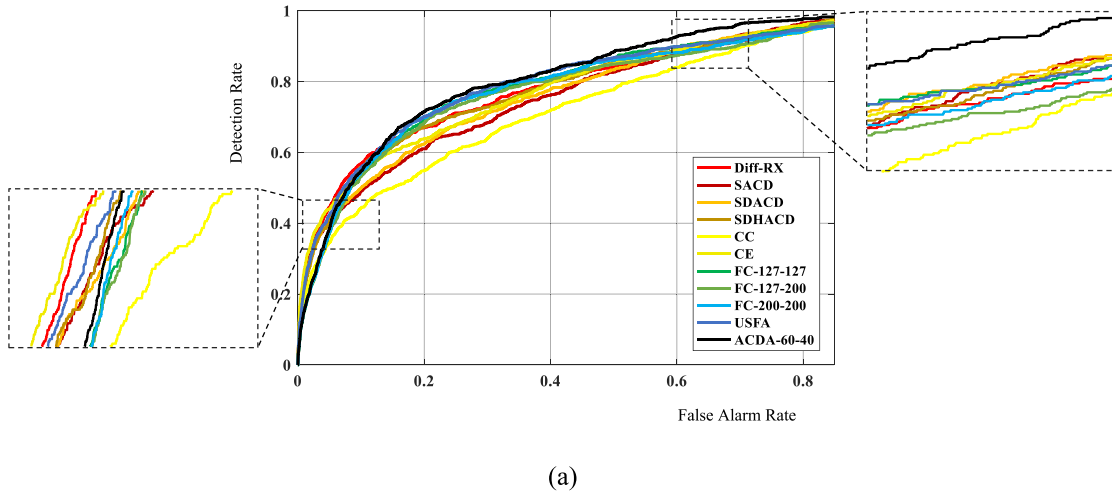
Fig. 5.   ROC performance for two data sets. (a) D1F12H1-D1F12H2. (b) D1F12H1-D2F22H2.
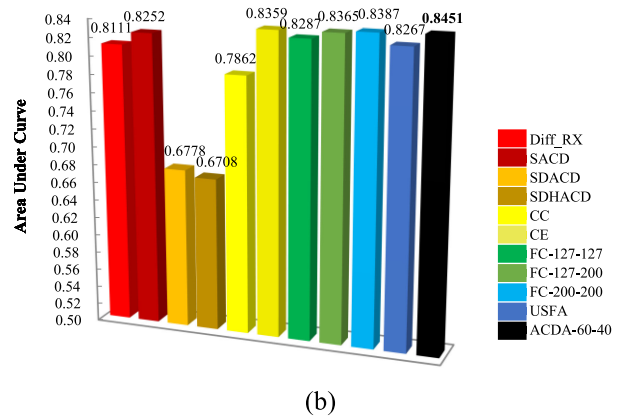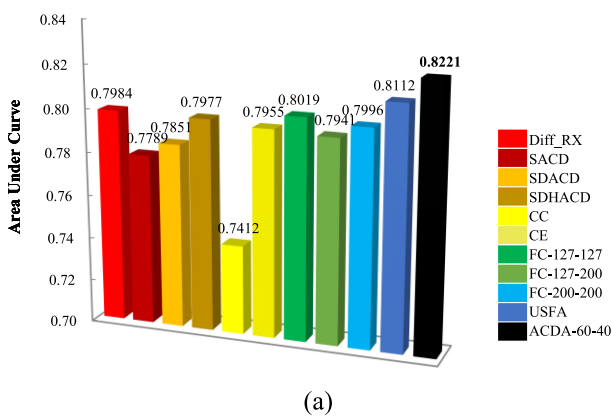

Fig. 6.   AUC performance for two data sets. (a) D1F12H1-D1F12H2. (b) D1F12H1-D2F22H2.
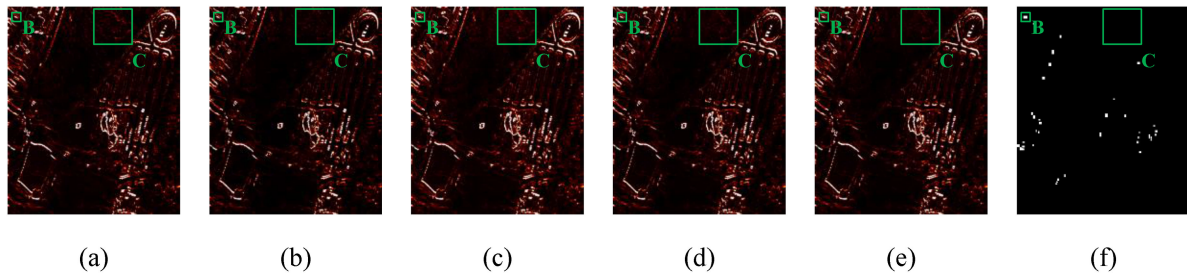
Fig. 7. Influence of min, max, and average operation on the final loss maps about the D1F12H1-D1F12H2. (a) Loss map I_1 of predictor $X \rightarrow Y$. (b) Loss map I_2 of predictor $Y \rightarrow X$. (c) Loss map of max operation. (d) Loss map of average operation. (e) Loss map I of min operation. (f) Ground truth map.
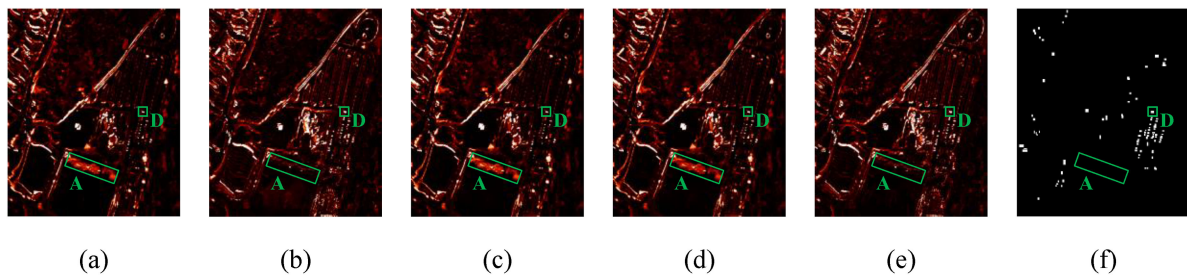

Fig. 8. Influence of min, max, and average operation on the final loss maps about the D1F12H1-D2F22H2. (a) Loss map I_1 of predictor $X \rightarrow Y$. (b) Loss map I_2 of predictor $Y \rightarrow X$. (c) Loss map of max operation. (d) Loss map of average operation. (e) Loss map I. of min operation. (f) Ground truth map.

the two images are quite large. The quantitative assessments of two experiments manifest that the proposed ACDA has the best performance on anomaly change detection and background suppression.

And there is another trick need to be discussed as follows. For each experiment, there are two predictors trained to reap two loss maps. The final result comes from the min operation of the two loss maps as above mentioned. Fig. 7 displays the influence of min, max, and average operation on the final loss maps about the first experiment. As Fig. 7 shows, it can be known from the ground truth map that region B has anomaly changes, which is also bright in the loss map I_1 and I_2. And the min operation still reserves the bright anomalous information. The max operation and average operation also perform well in keeping the anomalous change information. For region C, there is no anomalous change in fact. But there are still some twinkling spots detected in region C of loss map I_1. And region C in loss map I_2 seems darker. The min operation takes the minimum of the two anomaly intensity values as the final results, leading to smaller anomaly values and reducing the probability of being anomaly changes and less noise. Since the anomalous intensity of max operation and average operation is larger than the one of min operation, the min operation does a better job of suppressing the background and noise. Fig. 8 shows another example of D1F12H1-D2F22H2. Region A is the shadow area and does not have anomaly changes. But there are large spectral differences between the two HSIs. As Fig. 8 shows, region A in loss map I_1 is bright and the whole shadow area is likely to be detected as anomalous change. While region A in loss map I_2 has low anomaly intensity value and is less possible to be anomalous change. So the minimum result of the two loss maps yields a low possibility of being the anomalous change. In contrast, the max and average operations make region A maintain a high probability of being anomalous. And when a region such as region D has high values in both two loss maps, the result of the min operation can still preserve the high-intensity information. The max and min operations have a good performance on holding the anomalous change information too. Moreover, Figs. 9 and 10 give a quantitative evaluation to test the effectiveness of min operation. As Fig. 9(a) shows, the five curves are closely similar during the stage of low false alarm rate, while the ROC curve of the final loss map I in red obtains the best performance at a higher false alarm rate. For EX-2: D1F12H1-D2F22H2, it is obvious that the ROC curve of final loss map I in red overtops the other in Fig. 9(b). Fig. 10 makes a more intuitive comparison, where the AUC performance of min operation outstands the performance of the other operations on two experiments.

All in all, when a pixel has anomaly change, in fact, the anomaly intensity values detected by two predictors are both high. The minimum of them is still a large value that still retains anomaly change information. When a pixel does not contain anomaly change actually, the anomaly intensity values detected by two predictors are both low. The minimum of them becomes lower that makes the pixel less likely to be anomalous change. Another alternative max operation is proved to be less effective than min operation by trial. Taking region A, for example, the maximum of the anomaly value on two loss maps is the one on loss map I_1, which makes it high likely to be anomaly change. The min operation reduces the noise on the final intensity map but at the expense of some anomaly changes, which brings about a low detection rate under a low false alarm rate range.
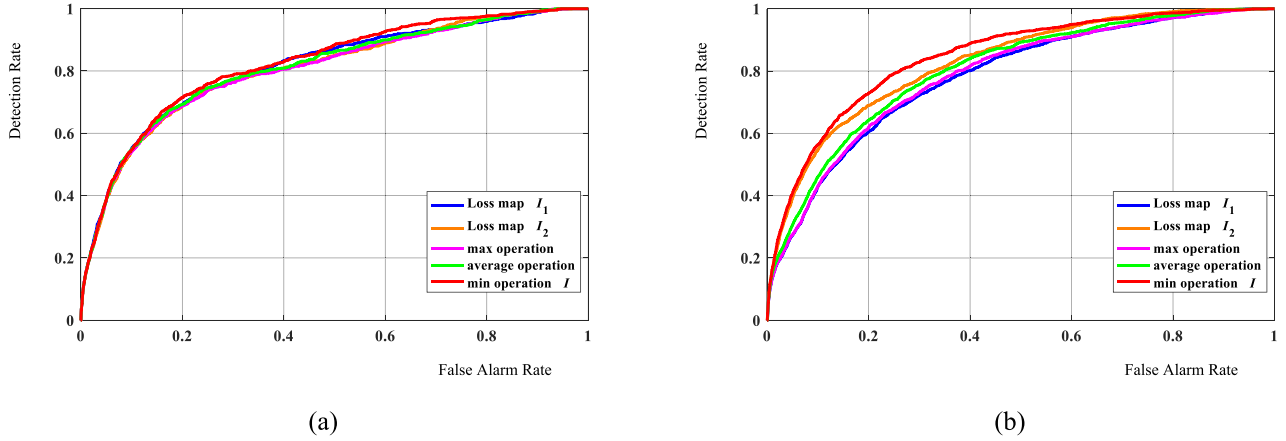
Fig. 9.   ROC performance about the influence of min, max, and average operation on two data sets. (a) D1F12H1-D1F12H2. (b) D1F12H1-D2F22H2.

TABLE I
AUC COMPARISON UNDER DIFFERENT PARAMETER h_1 AND h_2 FOR D1F12H1-D1F12H2

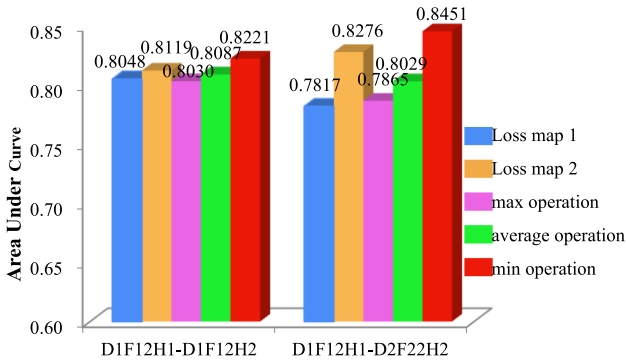| h_2	h_1				
	120	100	80	60	40
100	0.8021	——	——	——	——
80	0.8225	0.8133	——	——	——
60	0.8035	0.7959	0.8089	——	——
40	0.8102	0.7940	0.8112	0.8221	——
20	0.8048	0.7981	0.8004	0.8157	0.8159

Fig. 10.   Comparison of AUC performance between the influence of min, max, and average operation on two data sets. (a) D1F12H1-D1F12H2. (b) D1F12H1-D2F22H2.

D. Parameter Analysis

This section discusses the influence of the units of hidden layers and the selection of training samples on the experiment performance.

The number of the units of hidden layers is a group of crucial parameters of ACDA. Considering that deeper networks aggravate the time cost and the possibility of over-fitting, we design an ACDA model based on AE with three hidden layers. Since the configuration of a neural network can be various and there are no accurate means to choose the optimum number of the units of hidden layers, we set the range of h_1 as [120 100 80 60 40] and h_2 as [100 80 60 40 20] to tune the parameters.

The size of the input layer is equivalent to the channel of the HSI, which is equal to 127. For the special structure bottleneck of AE mode, h_1 starts from 120 and h_2 starts from 100. Tables I and II are the evaluation of AUC results on D1F12H1-D1F12H2 and D1F12H1-D2F22H2. According to Tables I and II, the AUC decreases when the number of h_1 is very large. This phenomenon can be explained as follows. When the number of h_1 and h_2 are both large, the code feature with high dimensions cannot represent the essence of the input. When the number of h_1 is large and h_2 is small, the gaps between these two dimensions lead to a crack in the process of feature transformation and flow, thus losing part of features. In addition, when the number of h_1 is too small, the AUC turns low, which can be attributed to the large dimensional difference between the two connecting layers. Thus, we select $h_1 = 60$ and $h_2 = 40$ as the optimal result for both experiments.

Fig. 11 represents the impact of three different ways of training sample selection on the effectiveness of experiments. Random strategy and USFA refer to choose samples from the whole images and the predetection results of USFA at random, respectively. And the ground truth strategy refers to select the examples from the ground truth map. As Fig. 11 shows, the results of both experiments show the same pattern, which can be summarized as two points. First, the results trained by USFA and ground truth are very similar, which demonstrates the effectiveness of the USFA. The reason for the subtle difference between the performance of USFA and ground truth strategy can be concluded as follows. Note that most of the unchanged pixels

TABLE II
AUC COMPARISON UNDER DIFFERENT PARAMETER h_1 AND h_2 FOR D1F12H1-D2F22H2

h_2	h_1				
	120	100	80	60	40
100	0.8402	——	——	——	——
80	0.8374	0.8443	——	——	——
60	0.8334	0.8399	0.8379	——	——
40	0.8417	0.8364	0.8358	0.8451	——
20	0.8442	0.8451	0.8428	0.8451	0.8359

TABLE III
RUNNING TIME OF DIFFERENT DEEP LEARNING APPROACHES

Dataset	FC-127-127	FC-127-200	FC-200-200	ACDA-60-40
D1F12H1-D1F12H2	157.92 ± 5.21(s)	159.15 ± 4.20(s)	159.59 ± 2.98(s)	**152.29 ± 4.43(s)**
D1F12H1-D2F22H2	160.42 ± 4.81(s)	159.22 ± 5.77(s)	153.56 ± 4.07(s)	**137.23 ± 4.66(s)**

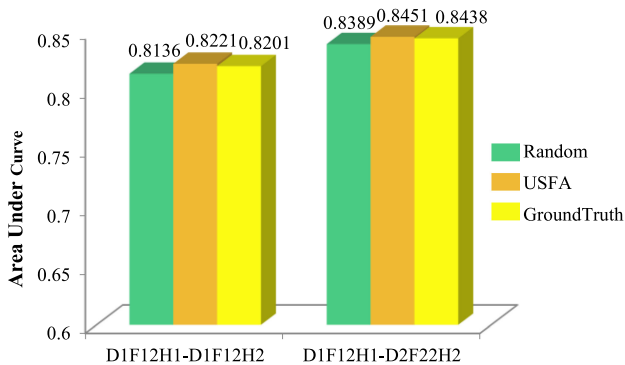

Fig. 11. Influence of different training sample selection strategies on the AUC performance of two data sets.

based on neural networks involved in this article, which are FC-127-127, FC-127-200, FC-200-200, and ACDA-60-40.

Table III displays the run time analysis of them. And the CPU used in this article is Intel Xeon E3-1220 3.00-GHz processor with RAM of 16 GB. And the GPU adopts an NVIDIA RTX 2080 TI graphic card. According to Table III, the proposed method runs the fastest in both experiments owing to the smaller size of network. The time consumption of the other approaches is a little longer but within an acceptable range.

IV. CONCLUSION

In this article, we point out that the classical linear predictors have a limited effect in dealing with the problem of variant space. And a nonlinear predictive model based on AE has been proposed for HACD, which maps a hyperspectral image to the imaging condition of another and obtain a predictive image. The bottleneck structure of AE is capable of extracting the essence feature from the input. Rather than replicating the output from the input, we design the training out as the corresponding spectral vector of another multitemporal HIS to construct a predictor model. And we suggest using a min operation between the loss maps, which is a nonparameter method and helps to reduce the noise of the final result. In the experiments, the performances on public hyperspectral data sets demonstrate that the proposed ACDA outperforms other state-of-the-art techniques, especially compared with another five classical and deep learning based predictor approaches. On the whole, the discussion can be summarized as follows.

1) The designed bottleneck structure is proved to perform better than the normal FC network, which uses large size of unit at the hidden layer.
2) The ACDA model performs well under the imaging condition of space variability, which is a challenging problem of HACD.
3) As for how to build an appropriate architecture for practical application, the results of parameter analysis indicate that the gaps of the number of units between adjacent hidden layers should be uniformly decreasing for the encoder, and uniformly increasing for the decoder.

detected by the USFA are background objects that are slightly or moderately affected by external imaging conditions. Thus, the training samples of the USFA strategy are representative of the main spectral difference caused by the imaging conditions. While the samples of the ground truth strategy are possible to be mixed and messy. Consequently, the performance of the predictor trained by USFA strategy might be a little better than the performance of the predictor trained by the ground truth strategy. Second, there is a slight difference between the AUC performance of random selection and USFA selection. This is because there are a small number of anomaly changes in the ground truth map. And the total number of samples used for training only occupies around 6% of the total number of pixels. Though the probability is very small that the pixels of anomalous change are selected, it cannot be ruled out. And once pixels of anomaly change are fed into the model, the mapping relationship between different imaging conditions would be contaminated, thus impairing the effect of the predictive model.

E. Run Time Cost Analysis

The time cost of the method is significant for timely detection in practical application, e.g., emergency response (quickly find available roads after a hurricane). There are four algorithms

4) The ACDA runs the fastest among all deep learning based methods.

Since the ACDA mainly focuses on the spectral features of multitemporal HSIs, we will put more attention on the combination of spectral and spatial features in further study for complex space variant background changes.

REFERENCES

[1] S. Ashbindu, "Review article digital change detection techniques using remotely-sensed data," *Int. J. Remote Sens.*, vol. 10, no. 6, pp. 989–1003, 06-01 1989.

[2] D. Lu, E. Moran, E. Brondizio, and P. Mausel, "Change detection techniques," *Int. J. remote Sens.*, vol. 25, no. 12, pp. 2365–2407, 2004.

[3] P. C. C. Author, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, "Review ArticleDigital change detection methods in ecosystem monitoring: A review," *Int. J. Remote Sens.*, vol. 25, no. 9, pp. 1565–1596, 2004.

[4] J. Peng, Y. Zhou, W. Sun, Q. Du, and L. Xia, "Self-paced nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1501–1515, Feb. 2021.

[5] W. Sun, J. Peng, G. Yang, and Q. Du, "Fast and latent low-rank subspace clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3906–3915, Jun. 2020.

[6] G. Xian, C. Homer, and J. Fry, "Updating the 2001 National Land Cover Database land cover classification to 2006 by using Landsat imagery change detection methods," *Remote Sens. Environ.*, vol. 113, pp. 1133–1147, 2009.

[7] F. A. KRUSE, "Identification and mapping of minerals in drill core using hyperspectral image analysis of infrared reflectance spectra," *Int. J. Remote Sens.*, vol. 17, no. 9, pp. 1623–1632, 1996.

[8] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman, "Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1246–1259, Jun. 2003.

[9] A. Koltunov and S. L. Ustin, "Early fire detection using non-linear multitemporal prediction of thermal imagery," *Remote Sens. Environ., Interdiscipl. J.*, vol. 110, pp. 18–28, 2007.

[10] J. Theiler and S. Perkins, "Proposed framework for anomalous change detection," 2006.

[11] M. T. Eismann, J. Meola, A. D. Stocker, S. G. Beaven, and A. P. Schaum, "Airborne hyperspectral detection of small changes," *Appl. Opt.*, vol. 47, no. 28, pp. F27–F45, 2008.

[12] S. Resta, N. Acito, M. Diani, G. Corsini, T. Opsahl, and T. V. Haavardsholm, "Detection of small changes in airborne hyperspectral imagery: Experimental results over urban areas," in *Proc. 6th Int. Workshop Anal. Multi-temporal Remote Sens. Images*, 2011, pp. 5–8.

[13] N. Acito, M. Diani, A. Rossi, and S. U. de Ceglie, "Hyperspectral anomalous change detection in the presence of non-stationary atmospheric/illumination conditions," 2014, pp. 92440Z–92440Z–10.

[14] J. Theiler and B. Wohlberg, "Local coregistration adjustment for anomalous change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 8, pp. 3107–3116, Aug. 2012.

[15] M. Shimoni, R. Heremans, and C. Perneel, "Detection of small changes in complex urban and industrial scenes using imaging spectroscopy," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 4757–4760.

[16] J. Theiler, "Quantitative comparison of quadratic covariance-based anomalous change detectors," *Appl. Opt.*, vol. 47, no. 28, pp. F12–F26, 2008.

[17] J. Meola, M. T. Eismann, R. L. Moses, and J. N. Ash, "Application of model-based change detection to airborne VNIR/SWIR hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3693–3706, Oct. 2012.

[18] M. T. Eismann, J. Meola, and R. C. Hardie, "Hyperspectral change detection in the presenceof diurnal and seasonal variations," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 237–249, Jan. 2008.

[19] S. A. Schaum A, "Long-interval chronochrome target detection," in *Proc. Int. Symp. Spectral Sens. Res.*, 1998, pp. 1760–1770.

[20] A. Schaum and A. Stocker, "Hyperspectral change detection and supervised matched filtering based on covariance equalization," *Proc. SPIE*, vol. 5425, pp. 77–90, 2004.

[21] R. Mayer, F. Bucholtz, and D. Scribner, "Object detection by using "whitening/dewhitening" to transform target signatures in multitemporal hyperspectral and multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 5, pp. 1136–1142, May 2003.

[22] M. J. 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.

[23] J. Y. Chen and I. S. Reed, "Detection algorithm of optical targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 46–59, Jan. 1987.

[24] 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.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[27] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state-of-the-art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[28] S. Liu, Q. Shi, and L. Zhang, "Few-Shot Hyperspectral Image Classification With Unknown Classes Using Multitask Deep Learning," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–18, Sep. 2020.

[29] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 645–657, Feb. 2017.

[30] Q. Shi *et al.*, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1430–1434, Aug. 2020.

[31] 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.*, 2019, pp. 5488–5491.

[32] B. Du, L. Ru, C. Wu, and L. Zhang, "Unsupervised deep slow feature analysis for change detection in multi-temporal remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9976–9992, Dec. 2019.

[33] H. Chen, C. Wu, B. Du, L. Zhang, and L. Wang, "Change detection in multisource VHR images via deep siamese convolutional multiple-layers recurrent neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2848–2864, Apr. 2020.

[34] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[35] Y. Wang, H. Yao, and S. Zhao, "Auto-encoder based dimensionality reduction," *Neurocomputing*, vol. 184, pp. 232–242, 2016.

[36] X. Ma, H. Wang, and J. Geng, "Spectral–spatial classification of hyperspectral image based on deep auto-encoder," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4073–4085, Sep. 2016.

[37] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[38] C. Wu, B. Du, and L. Zhang, "Slow feature analysis for change detection in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2858–2874, May 2014.

[39] C. Wu, L. Zhang, and B. Du, "Hyperspectral anomaly change detection with slow feature analysis," *Neurocomputing*, vol. 151, pp. 175–187, 2015.

[40] C. Wu, Y. Lin, B. Du, and L. Zhang, "A study for hyperspectral anomaly change detection on 'Viareggio 2013 Trial' dataset," in *Proc. 10th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.

[41] N. Acito, S. Matteoli, A. Rossi, M. Diani, and G. Corsini, "Hyperspectral Airborne 'Viareggio 2013 Trial' data collection for detection algorithm assessment," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2365–2376, Jun. 2016.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.

[43] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *Proc. Int. Conf. Learn. Representations*, 2014.

[44] N. Acito, M. Diani, G. Corsini, and S. Resta, "Introductory view of anomalous change detection in hyperspectral images within a theoretical Gaussian framework," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 32, no. 7, pp. 2–27, Jul. 2017.

[45] C. Clifton, "Change detection in overhead imagery using neural networks," *Appl. Intell.*, vol. 18, pp. 215–234, 2003.

[46] J. A. Hanley and B. J. McNeil, "A method of comparing the areas under receiver operating characteristic curves derived from the same cases," *Radiology*, vol. 148, no. 3, pp. 839–843, 1983.

Meiqi Hu (Student Member, IEEE) received the B.S. degree in surveying and mapping engineering from the School of Geoscience and Info-Physics, Central South University, Changsha, China, in 2019. She is currently working toward the M.S. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote sensing, Wuhan University, Wuhan, China.

Her research interests include deep learning and multitemporal remote sensing image change detection.
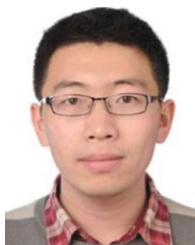
Chen Wu (Member, IEEE) received the B.S. degree in surveying and mapping engineering from Southeast University, Nanjing, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from State Key Laboratory of Information Engineering in Surveying, Mapping and Remote sensing, Wuhan University, Wuhan, China, in 2015.

He is currently an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China. His research interests include multitemporal remote sensing image change detection and analysis in multispectral and hyperspectral images.

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 Chair Professor with 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 or 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 Institute of Electrical and Electronic Engineers (IEEE) and the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS) and 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams received 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 is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He also serves as an Associate Editor or Editor for more than ten international journals. He serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Bo Du (Senior Member, IEEE) received the Ph.D. degree in photogrammetry and remote sensing from 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 Institute of Artificial Intelligence, Wuhan University. He is also the Director of National Engineering Research Center for Multimedia Software, Wuhan University. He has authored or coauthored more than 80 research papers published in 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), etc. Thirteen of them are ESI hot papers or highly cited papers. His major research interests include pattern recognition, hyperspectral image processing, and signal processing.

Dr. Du serves as an Associate Editor for neural networks, pattern recognition and neurocomputing. He also serves as a Reviewer of 20 Science Citation Index (SCI) magazines including IEEE TPAMI, TCYB, TGRS, TIP, JSTARS, and GRSL. He was the recipient of the Highly Cited Researcher 2019 by the Web of Science Group, the IJCAI (International Joint Conferences on Artificial Intelligence) Distinguished Paper Prize, IEEE Data Fusion Contest Champion, and IEEE Workshop on Hyperspectral Image and Signal Processing Best Paper Award, in 2018. He regularly serves as a Senior PC member of IJCAI and AAAI. He served as an Area Chair for ICPR.