Tensor Decomposition-Inspired Convolutional Autoencoders for Hyperspectral Anomaly Detection

Bangyong Sun, Zhe Zhao, Di Liu, Xiaomei Gao, and Tao Yu

Abstract—Anomaly detection from hyperspectral images (HSI) is an important task in the remote sensing domain. Considering the three-order characteristics of HSI, many tensor decomposition based hyperspectral anomaly detection (HAD) models have been proposed and drawn much attention during the past decades. However, as most tensor decomposition based detectors are directly performed on the original HSI, the detection accuracy is usually limited due to the high-dimension and noise corruption of the HSI. Benefiting from the good capacity of autoencoders (AE) for feature extraction, in this article, an enhanced tensor decomposition-inspired convolutional AE for HAD is proposed to address those problems, named TDNet. Within the proposed TDNet, the traditional canonical-polyadic (CP) tensor decomposition model is innovatively alternated by a deep neural network (DNN), and the DNN tensor decomposition model performs more stably and robustly for noise. Specifically, a potential abnormal pixels remove strategy is first built to obtain the background training sets. Then, a DNN tensor decomposition-inspired convolutional AE is used to recover the original background information, which consists of an encoder, a low-rank tensor decomposition network, and a decoder. Finally, the residual errors between input HSI and recovered background are used for anomaly detection. Extensive experiments demonstrate the superiority of the TDNet in terms of both AUC values and ROC curves.

Index Terms—Anomaly detection, hyperspectral image (HSI), tensor decomposition network.

I. INTRODUCTION

HYPERSPECTRAL images (HSI) include hundreds of spectral bands with much more information than RGB images, and are usually used for change detection [1], [2], unmixing [3], [4], classification [5], [6], and anomaly detection [7]–[9]. Among these tasks, hyperspectral anomaly detection (HAD) has attracted much interest due to its significance in military surveillance and mineral exploration [10]. Generally, anomalies are different from their surrounding pixels in spatial domain or spectral domain, and the goal of HAD is to separate the abnormal elements from the background.

In the past decades, researchers have proposed a large number of methods for HAD. Those methods can be divided into traditional methods and deep-learning based methods. The traditional methods mainly consist of statistical modeling based methods, representation modeling based methods, matrix or tensor decomposition modeling based methods and other methods [11]. The statistical modeling-based methods use statistical variables to estimate background and detect anomalies, such as Reed-Xiaoli detector (RX) [12] and local RX detector (LRX) [13]. But the assumption of background distribution is not reasonable due to the complex image scene in the real world [14]. The representation modeling-based methods include sparse representation [15] and collaborative representation [16]. These methods assume that background pixels can be represented by a set of bases extracted from input HSI, while anomalies cannot.

Within the matrix decomposition modeling based methods, the background generally has low-rank priority due to the redundancy of HSI, and anomalies are sparsely distributed in the background. Xu et al. [17] used dictionary learning and low-rank representation for HAD. Qu et al. [14] first proposed an abundance and dictionary-based low-rank decomposition method for HAD. Afterward, the tensor decomposition technique is introduced for HAD, which can explore the spatial-spectral features of the HSI [10], [18]–[20]. However, the matrix or tensor decomposition methods may not be stable due to the high-dimensional and noise corruption of the HSI. For other methods, such as support vector data description (SVDD) [21], SVDD assumes that the background of HSI is enveloped by a minimum enclosing hypersphere, and anomalies fall outside this hypersphere. In [22], Xie et al. used the structure tensor and guided filter (STGF) for HAD. In STGF, the informational bands are selected by structure tensor, and then the background is removed through attribute filter. Moreover, the selected first band as the guided filter to rectificat the final detection map.

Recently, various deep learning based HAD methods have been proposed due to the high representation capability of deep neural network (DNN). The deep-learning based HAD methods mainly consist of convolutional neural network (CNN)
based methods, autoencoder (AE) network based methods, and generative adversarial network (GAN) based methods. To the best of our knowledge, Li et al. [23] first introduced CNN into the HAD domain. The authors use transferred learning and reference data to measure the similarity of the test pixel, and then for anomaly detection. To further utilize the spatial and spectral information of HSI, Zhang et al. proposed TCNN detector [24]. The TCNN employs tensor theory and CNN into a joint unsupervised framework, and anomalies are detected by the difference between test block and local neighboring tensor. Nevertheless, the performance of CNN based methods is limited due to the lack of labeled training sets. The AE-based HAD methods are the main component of deep-learning based HAD [25], which generally use reconstruction errors for anomaly detection. A sparse AE for HAD was proposed in [26], and the anomalies are determined under the two cases of their network. Wang et al. [27] designed an adaptive weighted loss to train the skip connected AE, and the reconstruction errors are used for HAD. Lu et al. [7] introduced the manifold learning skill into the AE architectural, and the local structure of the latent representation can be preserved. Concurrent to this method, a graph AE was proposed in [28].

Besides, for better modeling the background or constraining the latent representation, Xie et al. broadened the prospects of HAD by introducing the GAN into this domain [8], [29]–[32]. An unsupervised discriminative reconstruction constrained GAN for HAD is proposed in [30], which is named HADGAN. To learn the distribution of background, the encodings are constrained by Gaussian distribution. Meanwhile, Xie et al. [32] proposed an adversarial AE (AAE) for HAD. The features are learned by AAE, and the anomalies are detected by Mahalanobis distance.

Considering the three-order characteristics of HSI and the strong feature extraction capability of AE, a tensor decomposition inspired convolutional AE is proposed for HAD in this article (TDNet). The TDNet tries to recover the background of the input HSI by using a limited number of training sets, and then anomalies are represented as the residual errors between input HSI and the recovered background. Specifically, the TDNet mainly consists of three steps. First, a potential abnormal pixels remove strategy is built to obtain the background training sets. Then, those training sets are fed into the DNN based decomposition network to recover the background of the input HSI. Note that the traditional CP decomposition is alternated by DNN in this article, which performs more stably and robust for noise. Finally, the residual errors between recovered background and input HSI are employed to detect anomalies.

The rest of this article is organized as follows: Section II reviews some related works. Section III presents the proposed TDNet in detail. The parameter analysis and detection results are shown in Section IV. Finally, Section V concludes this article.

II. RELATED WORK

A. Statistical Based Methods

As a classic statistical based method, let $x$ represents the pixel vector, RX detector [12] is defined as follows:

$$D(x) = (x - \mu)^T \Gamma^{-1} (x - \mu)$$

(1)

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i, \quad \Gamma = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)(x_i - \mu)^T$$

(2)

where $\mu$ is the mean vector of the samples, $\Gamma$ denotes the covariance matrix of the image, and $N$ represents the total number of pixels. The RX detector can detect anomalies effectively when the background is close to the multivariate normal distribution, but it can hardly find anomalies in the complex image scene. There are two main improvements for RX detector. One improvement is to enhance the adaptive selection of the background [11], e.g., weighted RX detector (WRX). WRX improves anomaly detection accuracy by giving smaller weights to potential anomalous pixels and larger weights to background pixels [33]. For the other improvement, some methods remove anomaly contamination and get a relatively pure background [34]. In [34], the detector eliminates the influence of potential abnormal pixels on the final detection result through multiple iterations, and stops iteration when two adjacent results are continuous.

B. Representation Based Methods

The CRD detector [16] assumes that the background pixel can be approximately represented by its neighborhoods, while anomalies cannot. The objective function is

$$\arg \min_{\alpha} ||y - X_s \alpha||_2^2 + \lambda ||\alpha||^2_2$$

(3)

where $y$ represents the pixel under test, and $X_s$ denotes the pixels inside in the outer window while outside in the inner window. When the representation process is finished, the anomalies can be determined from the residual image

$$r_1 = ||y - \hat{y}||_2 = ||y - X_s \hat{\alpha}||_2$$

(4)

$\hat{y}$ is the represented pixel vector, $r_1$ is an anomaly pixel if it is larger than a predefined threshold. The main drawback of the CRD method is that it ignores the correlations of all the pixels, and there is no general rule for choosing appropriate window sizes [35]. To improve the computational speed of the CRD algorithm, Ma et al. reduce the iteration time of CRD by constructing the correlation matrix between adjacent pixels [36]. Wang et al. improved the performance of CRD algorithm by giving different weights to different spectral bands [37]. Moreover, Zhao et al. [19] imposed the collaborative representation constraint into the stacked AE, thus making the extracted features easier for CRD detection.
C. Matrix Decomposition Based Methods

The matrix decomposition based method utilizes the low-rank and sparse priors of the hyperspectral data for anomaly detection. Given a hyperspectral data $\mathbf{X}$, LRASR method [17] is given by

$$\mathbf{X} = \mathbf{DS} + \mathbf{E}$$  \hspace{1cm} (5)

where $\mathbf{D}$ is the background dictionary, which is obtained by k-means algorithm. $\mathbf{S}$ represents the coefficients of $\mathbf{D}$, and $\mathbf{E}$ denotes the anomalous part. (5) can be written as the following energy minimization problem:

$$\min_{\mathbf{S}, \mathbf{E}} \|\mathbf{S}\|_* + \beta \|\mathbf{S}\|_1 + \lambda \|\mathbf{E}\|_{2,1} \quad \text{s.t.} \quad \mathbf{X} = \mathbf{DS} + \mathbf{E}$$  \hspace{1cm} (6)

where $\|\cdot\|_*$ denotes the matrix nuclear norm, $\|\cdot\|_1$ and $\|\cdot\|_{2,1}$ represent the $\ell_1$ norm and $\ell_{2,1}$ norm, respectively. $\beta$ and $\lambda$ are the parameters to control rankness and sparsity, and the linearized alternating direction method with adaptive penalty (LADMAP) [38] is adopted to solve (6). Actually, those parameters are difficult to determine due to the lack of priori knowledge of anomalies or background [27]. To reduce the redundancy of adjacent bands, a spectral difference low-rank representation learning framework was proposed in [39]. The residuals of adjacent bands are added to the background dictionary, thus improving the accuracy of the original LRASR. Feng et al. [40] used both the sparse part and the low-rank part for anomaly detection, which achieved considerable detection results. Afterward, Zhang et al. used Tucker decomposition method for HAD. Wang et al. [20] recovered the background of the HSI from a preprocessing data, then anomalies are detected by the residual errors.

D. Autoencoder Based Methods

As a typical unsupervised DNN [41], AE is popularly used for anomaly detection, which is encoded by an encoder and a decoder. The encoder aims to learn a low dimensional latent representation from input data while the decoder tries to reconstruct the input from the latent representation. Assume $\mathbf{x} \in \mathbb{R}^D$ is the input data with $D$ dimension, AE first encodes it to a latent representation $\mathbf{z} \in \mathbb{R}^d$ via a nonlinear transform function $f : \mathbf{z} = f(\mathbf{x})$. The decoder is used to reconstruct the input data $\tilde{\mathbf{x}}$ from latent representation $\mathbf{z} : \tilde{\mathbf{x}} = g(\mathbf{z})$. The general loss function of AE is mean square error (MSE), which is defined by

$$L = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}^{(i)} - \tilde{\mathbf{x}}^{(i)}\|^2$$  \hspace{1cm} (7)

where $N$ is the number of training samples, and the parameters are updated by the back-propagation method [42]. Compared with original data, the latent representation is more abstractive and informational. In [43], the autoencoder is first used for HAD. The input image is encoded by a sparse encoder and then decoded, and the final anomaly score of each pixel is determined by the reconstruction error. To effectively extract the spatial-spectral features of hyperspectral images, Sun et al. combined low-rank representation and 3-D convolutional autoencoder for HAD [44]. In which, the 3-D convolutional autoencoder extracts features, and then uses low-rank representation to obtain the initial results on the encoding features. The final abnormal score is obtained by fusing the initial result and reconstruction error of autoencoder. Zhang et al. designed a 3-D variational autoencoder for HAD, and the reconstruction errors are regarded as abnormal score [45]. Arisoy et al. [46] proposed a HAD algorithm based on GANs. The algorithm uses a GAN to generate a synthetic background image that is as close as possible to the original background image. Then, the background information in the original image is removed by the synthesized image. Finally, the detection result is obtained by RX detector on the residual image.

Besides that, the most similar works to TDNet are [47]–[49], but these works are mainly designed for HSI recovery tasks. Saragadam et al. [47] used deep generative networks for low-rank decomposition, and the method is robust to a wide range of distributions. Luo et al. [48] utilized a multilayer neural network to learn the nonlinear nature of multidimensional images, which is more effective than the linear transforms. Zhang et al. [49] constructed the Kroneker bases by performing a maxpooling operation on the input features, and then the degraded HSI is recovered by a Conv layer with the spatial size of $3 \times 3$. Unlike the above mentioned works, this article focuses on the HAD task. The proposed method first extracts the low-dimensional embedding of the revised HSI using an encoder, then performs a low-rank tensor decomposition on the feature space, and finally recovers the background information of the original HSI by a decoder.

III. PROPOSED METHOD

In this section, the proposed TDNet is presented in detail. As shown in Fig. 1, the TDNet mainly consists of three parts: potential abnormal pixels remove, deep CP decomposition for background recovery, and anomaly detection. In summary, the proposed TDNet is based on autoencoder framework. In particular, we use neural network to simulate the traditional CP decomposition model on the latent representations, so as to improve the accuracy of anomaly detection. Different from the traditional iterative optimization update of tensor decomposition, we update parameters by using back propagation through end-to-end training, so that the network can adapt to different datasets.

A. Potential Abnormal Pixels Remove

The potential abnormal pixels remove part is used to construct background training sets for deep CP decomposition network. There are many strategies that can achieve this goal, such as clustering based method [8], attribute filter based method [50], and naive detector based method [20]. In this article, RX detector is used to remove potential abnormal pixels due to its effectiveness and simplicity. Given a hyperspectral dataset $\mathbf{X} \in \mathbb{R}^{W \times H \times B}$, here $W$, $H$, $B$ denote the width, height, and spectral bands of the HSI, respectively, then the detection result of RX detector can be denoted as $\mathbf{M} \in \mathbb{R}^{W \times H}$. Obviously, the larger the values in $\mathbf{M}$, the higher the probabilities of being abnormal pixels. Thus, the pixels $W \times H \times \eta$ with the smallest response values in the
M are set as 1, and the others are set as 0, here \( \eta (\eta \in (0, 1]) \) denotes the rate of background training set. Finally, the rectified \( \mathbf{M} \) is used to obtain the background training set \( \mathbf{X}_B \in \mathbb{R}^{W \times H \times B} \).

### B. Deep CP Decomposition for Background Recovery

The deep CP decomposition network (CPNet) aims to learn a low rank tensor from the background training sets, and then to recover the background information of the original HSI. As shown in Fig. 1, the CPNet consists of an encoder, a low-rank tensor generate network (LTGN), and a decoder. The encoder is used to reduce the dimension of the input training sets, then, LTGN generates the low-rank tensor from the encodings [49], [51]. Fig. 2(b) shows the structure of LTGN, which is composed of rank-1 tensor generate block (RTGB) and residual learning. After the low-rank tensor is obtained, the decoder is used to recover the background information.

The CP decomposition is a higher order of image singular value decomposition, which can describe the discriminative information of different components of the image. For a three-order tensor \( \mathbf{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), it can be decomposed into a linear combination of \( N \) rank-1 tensors

\[
\mathbf{X} \approx \sum_{n=1}^{N} \alpha_n \mathbf{a}_n \odot \mathbf{b}_n \odot \mathbf{c}_n \tag{8}
\]

where \( \alpha_n \) is the weight parameter of \( n \)th rank-1 tensor, \( \mathbf{a}_n \in \mathbb{R}^{I_1} \), \( \mathbf{b}_n \in \mathbb{R}^{I_2} \), \( \mathbf{c}_n \in \mathbb{R}^{I_3} \) are the Kronecker bases in the three directions, and \( \odot \) denotes the Kroneker product. Generally, (8) can be transformed as an optimization problem, and the alternating least squares (ALS) algorithm is usually used to solve it. However, due to the high-dimension and noise corruption of the HSI, the solutions to this problem may not be stable. Fortunately, with the powerful representation capacity of DNN, Chen et al. [51] proposed a tensor low-rank reconstruction network for semantic segmentation, and Zhang et al. [49] exploited the tensor low-rank prior for HSI reconstruction. Those works provide a new perspective to solve the problem of tensor decomposition, and it is more flexible and effective.

In this article, a CP-inspired convolutional AE is proposed for HAD. As shown in Fig. 2(a), RTGB is designed to generate rank-1 tensors. With the image features fed to RTGB, the global average pooling (GAP)-1×1 convolution (Conv)-Sigmoid is employed to generate Kroneker bases in the channel, height, and width directions. Then, those bases are applied to generate rank-1 tensor by Kronecker product.

As shown in Fig. 2(b), the LTGN is used to generate a low rank tensor. First, the encodings of training sets act as the input of LTGN, and the first rank-1 tensor \( O_1 \) is generated by RTGB. Then, the residual part between the input of RTGB and the generated rank-1 tensor is used to generate the second rank-1 tensor \( O_2 \). It can obtain \( N \) rank-1 tensors by repeating \( N \) times, and each of them can be regarded as the different frequencies of the encoding features [49]. Finally, those rank-1 tensors are aggregated into a low-rank tensor along channel directions, and a 3×3 Conv is used to reduce the dimension. Note that, the skip connection is used to avoid overfitting and enhance the discriminative of feature representation [49]. The above processing can be expressed as

\[
\mathbf{X}_{\text{low}} = \mathbf{X}_{\text{encoding}} \odot \text{Conv} (\text{Concat} (O_1, O_2, \ldots, O_N)) \tag{9}
\]

where \( \odot \) is the element-wise product, and \( \mathbf{X}_{\text{low}} \in \mathbb{R}^{W \times H \times S} \), \( \mathbf{X}_{\text{encoding}} \in \mathbb{R}^{W \times H \times S} \) are the obtained low-rank tensor and the...
C. Anomaly Detection

1) Training Loss: The MSE is usually used for training the AE network, but they may suffer from blurry reconstructions [52]. In this article, inspired by [53], the structure similarity (SSIM) is added into the loss function to alleviate the blurry problem. Given the input image $x$ and the reconstructed image $\hat{x}$, the SSIM can be defined as

$$SSIM(x, \hat{x}) = \frac{(2\mu_x\mu_{\hat{x}} + c_1)(2\sigma_x\sigma_{\hat{x}} + c_2)}{\mu_x^2 + \mu_{\hat{x}}^2 + c_1(\sigma_x^2 + \sigma_{\hat{x}}^2 + c_2)}.$$  

In (10), $\mu_x, \mu_{\hat{x}}$ denote the mean of input image $x$ and reconstructed image $\hat{x}$, respectively, while $\sigma_x$ and $\sigma_{\hat{x}}$ represent their variances. As described in [53], $c_1$ and $c_2$ are set as 0.01, 0.03, respectively. Obviously, the larger the SSIM value is, the higher the similarity between $x$ and $\hat{x}$ will be. Thus, the negative SSIM as a penalty term is defined [53]

$$Loss_{ssim}(X, \hat{X}_B) = -SSIM(X, \hat{X}_B)$$  

where $X \in \mathbb{R}^{W \times H \times B}$ is the input HSI and $\hat{X}_B \in \mathbb{R}^{W \times H \times B}$ represents the reconstructed background. So, the total loss function in this article is defined as

$$Loss = \|X - \hat{X}_B\|_2^2 + \lambda \cdot Loss_{ssim}(X, \hat{X}_B).$$  

The first term is MSE, and $\lambda$ as a parameter to balance the SSIM term. The effect of $\lambda$ is discussed in Section IV-C.

2) Abnormal Score: During the test phase, the original HSI $X \in \mathbb{R}^{W \times H \times B}$ is directly fed into the network. The abnormal score is calculated as

$$Map = \|X - \hat{X}_B\|_2^2$$  

where $Map$ denotes the final detection result.

IV. EXPERIMENTS

In this section, we first introduce the datasets and comparing methods used in the experiments. Then, we verify the effectiveness of TDNet on four real hyperspectral datasets. Finally, the ablation studies are performed to evaluate the necessity of each component in the framework.
SUN et al.: TENSOR DECOMPOSITION-INSPIRED CONVOLUTIONAL AES FOR HAD

Fig. 3. Image scenes and ground truth.

Fig. 4. Parameter analysis. (a) Number of hidden nodes. (b) Number of tensor rank. (c) Weight parameter $\lambda$. (d) Rate of training set.
B. Experimental Setup

1) Network Architecture and Hyperparameters Setting: In this article, the encoder and decoder consist of three layers, and each layer is composed of Conv-LReLU operation. The spatial kernel size of Conv is $1 \times 1$, and the filter number of encoder and decoder is set as $E[L, 128, 64, S], D[S, 64, 128, L]$, respectively. Here, $S$ is the hidden nodes of the encoder. Considering the tradeoff between detection accuracy and computation time, $S$ is set as 24. Furthermore, the rate of training set is 0.5, the rank number is 15, and the weight parameter $\lambda$ is 0.5. The effect of those parameters for the final detection accuracy is discussed in Section IV-C.

2) Compared Methods and Evaluation Metric: The receiver operating characteristic (ROC) curve and AUC value are generally used for HAD [56], seven mostly used detectors are employed for comparison. The RX [12] and LRX [13] detectors are statistical based methods, and the CRD [16] detector is based on collaborative representation theory. The main parameters of LRX and CRD detector are the window size. The optimal window size of LRX is set as (13, 25), (9, 31), (17, 19), (5, 25) for AVIRIS-I, El Segundo, Texas Coast, and Los Angeles datasets, respectively. For CRD detector, the window size is (19, 23), (13, 31), (23, 35), and (7, 13) for the four datasets, respectively. The LRASR [17] is based
on low-rank matrix decomposition skill. The DAEAD [57] is based on AE architecture, and the TDAD [10] method is a tensor decomposition based detector. The KIFD [54] is a state-of-the-art detection method. The DAEAD method and TDNet are implemented by the tensorflow framework, and other methods are implemented by MATLAB 2018b.

C. Parameter Analysis

As described in Fig. 4, the main parameters of TDNet contain the rate of training set $\eta$, the number of hidden nodes $S$, the number of rank tensor $N$, and the weight parameter $\lambda$ in the loss function. Fig. 4(a) reflects that the Texas Coast dataset is stable in different hidden nodes, and the other datasets achieve best detection accuracy when $S$ is set to 24. Limited by this premise, Fig. 4(b)–(d) shows the effect of $N$, $\lambda$, and $\eta$, respectively. $N$ controls the redundancy of the background training set, when $N$ is set to 15, the accuracy of the four datasets is stable, and three of them obtained the highest AUC values. The $\lambda$ balances the effect of SSIM term in the loss function. For a small value of $\lambda$, the reconstructed background may lose structural information, while for a very large $\lambda$, the detection results are decreased. For a larger $\eta$, the anomalies are mistakenly regarded as one part of the background training set, and it will decrease the final accuracy.
D. Detection Performance

The detection maps, ROC curves, and AUC values are shown in Figs. 5–9, and Table II, respectively. For AVIRIS-I dataset, the proposed TDNet obtains the second-highest AUC values. The RX, LRX, and TDAD detector cannot detect anomalies well, since they get poor AUC values. Compared with LRASR and KIFD detector, the intensity of the anomalies is weak for TDNet, but TDNet can suppress the background well. To further illustrate the effectiveness of the TDNet, the ROC curves of seven detectors are shown in Fig. 9(a). The curve of TDNet and LRASR is higher than other methods, which demonstrates the effectiveness of the TDNet.

The size of Segundo is larger than other datasets, and the image scene is more complex. From Fig. 6, we can see that RX, LRX, CRD, and TDAD detector failed to detect anomalies, and the LRASR and DAEAD detector just find a few parts of targets. The detection maps of KIFD and TDNet are closer to GT than others, and the shapes of anomalies are preserved. Compared with KIFD, the TDNet has a lower false alarm rate, and obtains the highest AUC values, 0.9943. Fig. 9(b) shows the ROC curve of the detectors on Segundo dataset, while the curves of the compared methods are all surrounded by TDNet, which indicates the superiority of the TDNet.

As shown in Fig. 7, for Texas Coast dataset, RX, DAEAD, TDAD, and TDNet achieve better visual inspection. LRASR and KIFD detector is influenced by stripe noise, so they get lower AUC values. From the detection map of TDNet, we can see clearly that TDNet can not only find the locations of abnormal targets, but also keep the shape of the anomalies. The ROC curves in Fig. 9(c) further validate the effectiveness of the TDNet.

The anomalies of Los Angeles dataset mainly consist of buildings, and the image scene is relatively simple. All of the seven detectors obtain satisfactory detection accuracy, and TDNet gets the highest AUC values. The ROC curve in Fig. 9(d) clearly and intuitively shows the performance among these detectors, and the other detectors are surrounded by TDNet.

Overall, from ROC curves, AUC values and detection maps, the proposed TDNet not only gets better visual inspection, but also obtains the highest AUC values. Meanwhile, the ROC curves of other detectors are surrounded by TDNet, which can verify the effectiveness of our proposed TDNet.
E. Ablation Study

In order to analyze the effect of SSIM term and CP decomposition network in the proposed framework, an ablation study is shown in the last three columns of Table II. AE just uses AE network for anomaly detection by calculating the reconstruction errors, and TDNet(M) uses the MSE as the loss function. Compared with the original AE, the CP decomposition network can improve the detection accuracy by 0.0106, 0.1160, 0.0043, 0.0058 for AVIRIS-I, Segundo, Texas Coast, and Los Angeles dataset, respectively, which indicates the effectiveness of the CP decomposition network. When the SSIM term is added into the loss function, the AUC values of the four datasets can be further improved. As mentioned above, SSIM term can preserve the structure of the original dataset.

V. CONCLUSION

In this article, a CP-inspired convolutional AE is proposed for HAD (TDNet). In particular, the traditional CP decompositional model is replaced with DNN, which makes the model more stable and robust for noise. To alleviate the blurry reconstruction problem, the SIMM loss is added into the loss function. The comparative experiments on four datasets demonstrate the effectiveness of the TDNet, especially in large abnormal targets. Furthermore, the ablation studies verified the superiority of TDNet. In fact, TDNet also has some weaknesses. For instance, the detection ability is limited on the small target. In the future, we will analyze the performance of TDNet on other object detection tasks.

REFERENCES

[1] X. Zheng, X. Chen, X. Lu, and B. Sun, “Unsupervised change detection by cross-resolution difference learning,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, Mar. 2022.

[2] Q. Guo, J. Zhang, C. Zhong, and Y. Zhang, “Change detection for hyperspectral images via convolutional sparse analysis and temporal spectral unmixing,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 4417–4426, Apr. 2021.

[3] K. T. Shahid and I. D. Schizas, “Unsupervised hyperspectral unmixing via nonlinear autoencoders,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, May 2022.

[4] F. Li, S. Zhang, B. Liang, C. Deng, and S. Wang, “Hyperspectral sparse unmixing with spectral-spatial low-rank constraint,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 6119–6130, Jun. 2021.

[5] X. Zheng, Y. Yuan, and X. Lu, “A deep scene representation for aerial scene classification,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 7, pp. 4799–4809, Jul. 2019.

[6] Z. Zhao, D. Hu, H. Wang, and X. Yu, “Center attention network for hyperspectral image classification,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 3415–3425, Mar. 2021.

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

[8] 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.

[9] L. Zhang and B. Cheng, “A stacked autoencoders-based adaptive subspace model for hyperspectral anomaly detection,” Infrared Phys. Technol., vol. 96, pp. 52–60, 2019.

[10] 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.

[11] 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.

[12] 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.

[13] S. Matteoli, T. Verraci, M. Diani, and G. Corsini, “A locally adaptive background density estimator. An evolution for RX-based anomaly detectors,” IEEE Geosci. Remote Sens. Lett., vol. 11, no. 1, pp. 323–327, Jan. 2014.

[14] Y. Qu et al., “Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 8, pp. 4391–4405, Aug. 2018.

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

[16] 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.

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

[18] C. Zhao and L. Zhang, “Spectral-spatial stacked autoencoders based on low-rank and sparse matrix decomposition for hyperspectral anomaly detection,” Infrared Phys. Technol., vol. 92, pp. 106–176, 2018.

[19] C. Zhao, C. Li, S. Feng, and W. Li, “Spectral-spatial anomaly detection via collaborative representation constraint stacked autoencoders for hyperspectral images,” IEEE Geosci. Remote Sens. Lett., vol. 19, pp. 1–5, Jan. 2022.

[20] J. Wang, Y. Xia, and Y. Zhang, “Anomaly detection of hyperspectral image via tensor completion,” IEEE Trans. Geosci. Remote Sens., vol. 18, no. 6, pp. 1099–1103, Jun. 2022.

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

[22] W. Xie, T. Jiang, Y. Li, X. Jia, and J. Lei, “Structure tensor and guided filtering-based algorithm for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 7, pp. 4218–4230, Jul. 2019.

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

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

[25] W. Xie et al., “Unsupervised spectral mapping and feature selection for hyperspectral anomaly detection,” Neural Netw., vol. 132, pp. 144–154, 2020.

[26] 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.

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

[28] G. Pan, 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, Jul. 2022, Art. no. 5511314.

[29] 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.

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

[31] 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.

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

[33] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, “Weighted-RDX and linear filter-based RDX: 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.
[34] Y. Taitano, B. Geier, and K. Bauer, “A locally adaptable iterative RX detector,” J. Adv. Signal Process., vol. 10, no. 1, 2010, Art. no. 341908.

[35] T. Cheng and B. Wang, “Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 2, pp. 1472–1486, Feb. 2021.

[36] N. Ma, Y. Peng, and S. Wang, “A fast recursive collaboration representation anomaly detector for hyperspectral image,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 4, pp. 588–592, Apr. 2019.

[37] R. Wang, H. Hu, F. He, F. Nie, S. Cui, and Z. Ming, “Self-weighted collaborative representation for hyperspectral anomaly detection,” Signal Process., vol. 177, 2020, Art. no. 107118.

[38] Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” in Proc. Adv. Neural Inf. Process. Syst., 2011, pp. 612–620.

[39] X. Zhang, X. Ma, H. Ning, J. Gu, X. Tang, and L. Jiao, “Spectral-difference low-rank representation learning for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 59, no. 12, pp. 10364–10377, Dec. 2021.

[40] S. Feng, S. Tang, C. Zhao, and Y. Cui, “A hyperspectral anomaly detection method based on low-rank and sparse decomposition with density peak guided collaborative representation,” IEEE Trans. Geosci. Remote Sens., vol. 60, Feb. 2022, Art. no. 5501513.

[41] B. Schölkopf, J. Platt, and T. Hofmann, “Greedy layer-wise training of deep networks,” in Proc. Adv. Neural Inf. Process. Syst., 2007, pp. 153–160.

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

[43] E. Bati, A. Çalışkan, A. Koz, and A. A. Alatan, “Hyperspectral anomaly detection method based on auto-encoder,” in Proc. SPIE Image Signal Process. Remote Sens., 2015, pp. 220–226.

[44] S. Bangyong, Z. Zhe, H. Bingliang, and Y. Tao, “Hyperspectral anomaly detection based on 3D convolutional autoencoder and low rank representation,” Acta Photonica Sinica, vol. 30, no. 4, 2021, Art. no. 254.

[45] J. Zhang, Y. Xu, T. Zhan, Z. Wu, and Z. Wei, “Anomaly detection in hyperspectral image using 3D-convolutional variational autoencoder,” in Proc. IEEE Int. Geosci. Remote Sens. Symp., 2021, pp. 2512–2515.

[46] S. Arisoy, N. M. Nasrabadi, and K. Kayabol, “GAN-based hyperspectral anomaly detection,” in Proc. 28th Eur. Signal Process. Conf., 2021, pp. 1891–1895.

[47] V. Saragadam, R. Balestrieri, A. Veeraraghavan, and R. G. Baraniuk, “DeepTensor: Low-rank tensor decomposition with deep network priors,” 2022, arXiv:2204.03145.

[48] Y.-S. Luo, X.-L. Zhao, T.-X. Jiang, Y. Chang, M. K. Ng, and C. Li, “Self-supervised nonlinear transform-based tensor nuclear norm for multi-dimensional image recovery,” IEEE Trans. Image Process., vol. 31, pp. 3793–3808, May 2022.

[49] S. Zhang, L. Wang, L. Zhang, and H. Huang, “Learning tensor low-rank prior for hyperspectral image reconstruction,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 12001–12010.

[50] 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.

[51] W. Chen et al., “Tensor low-rank reconstruction for semantic segmentation,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 52–69.

[52] Y. Ma, B. Wei, P. Feng, P. He, X. Guo, and G. Wang, “Low-dose CT image denoising using a generative adversarial network with a hybrid loss function for noise learning,” IEEE Access, vol. 8, pp. 67519–67529, 2020.

[53] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, “Improving unsupervised defect segmentation by applying structural similarity to autoencoders,” in Proc. VISAPP (5: VISAPP), 2019.

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

[55] 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.

[56] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, “Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms,” IEEE Signal Process. Mag., vol. 31, no. 1, pp. 24–33, Jan. 2014.

[57] A. Taghipour and H. Ghassemian, “Unsupervised hyperspectral target detection using spectral residual of deep autoencoders,” in Proc. 4th Int. Conf. Pattern Recognit. Image Anal., 2019, pp. 52–57.