

Tucker Decomposition-Based Network Compression for Anomaly Detection With Large-Scale Hyperspectral Images

Yulei Wang¹, Member, IEEE, Hongzhou Wang, Enyu Zhao¹, Meiping Song¹, and Chunhui Zhao

Abstract—Deep learning methodologies have demonstrated considerable effectiveness in hyperspectral anomaly detection (HAD). However, the practicality of deep learning-based HAD in real-world applications is impeded by challenges arising from limited labeled data, large-scale hyperspectral images (HSIs), and constrained computational resources. In light of these challenges, this article introduces a convolutional neural network (CNN)-based HAD model through the incorporation of Tucker decomposition, named TD-CNN. Drawing inspiration from transfer learning, the proposed model initially constructs pixel sample pairs from known labeled HSIs in the source domain, feeding them into the designed CNN to train the network learning spectral feature differences to obtain a CNN containing knowledge from the source domain. Subsequently, to prevent the need for network retraining caused by structural changes and to reduce model parameters for improving detecting timeliness, a general network compression scheme based on Tucker decomposition is applied to the CNN, where the convolutional layers of the above CNN undergo Tucker tensor decomposition to compress the network and alleviate parameter redundancy. Finally, spectral features realignment is used to recover the detection accuracy loss caused by Tucker tensor decomposition. In addition, a dual-windows structure is implemented during the detection phase, incorporating spatial information to the aforementioned spectral-level learning model, facilitating spectral-spatial collaborative HAD. Experimental evaluations using three real hyperspectral datasets and artificially expanded datasets demonstrate that, in comparison with state-of-the-art methods, the proposed TD-CNN method exhibits effectiveness and superiority in terms of both time cost and detection accuracy, where the notable advantages in terms of time cost become more pronounced with an increasing number of pixels.

Index Terms—Anomaly detection, convolutional neural network (CNN), hyperspectral images (HSIs), transfer learning, tucker decomposition.

Manuscript received 14 February 2024; revised 19 April 2024; accepted 15 May 2024. Date of publication 23 May 2024; date of current version 5 June 2024. This work was supported in part by the National Nature Science Foundation of China under Grant 42271355 and Grant 61801075, in part by the Natural Science Foundation of Liaoning Province under Grant 2022-MS-160, in part by China Postdoctoral Science Foundation under Grant 2020M670723, and in part by the Fundamental Research Funds for the Central Universities. (*Corresponding author: Enyu Zhao*)

Yulei Wang, Hongzhou Wang, Enyu Zhao, and Meiping Song are with the Information Science and Technology College, Dalian Maritime University, Dalian 116026, China (e-mail: wangyulei@dltu.edu.cn; whz1579@dltu.edu.cn; zhaoenyu@dltu.edu.cn; smping@dltu.edu.cn).

Chunhui Zhao is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: zhaochunhui@hrbeu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3404607

I. INTRODUCTION

HYPERSPECTRAL imaging sensors have the capability to capture surface information across hundreds of continuous bands, facilitating its precise spectral identification in many applications, yielding an increasingly important role in the military and civilian fields [1], [2], [3]. Hyperspectral images (HSIs) are typically conceptualized as a three-dimensional (3-D) data cube, integrating two spatial dimensions with a spectral dimension. The rich spectral information embedded within the HSI is pivotal in identifying subtle differences among different objects, making HSIs widely used across various applications, such as target detection [4], [5], anomaly detection [6], [7], [8], classification [9], band selection [10], [11], etc. Hyperspectral target detection, in particular, is often approached as a binary classification problem, determining the probability of each pixel as either background or target by the prior target spectrum of interest. However, the intricate imaging environments coupled with low spatial resolution [12] often make it difficult or even impossible to acquire desired spectral information for the target of interest. As a result, hyperspectral anomaly detection (HAD) without any prior information has received significant research attention in real applications.

HAD primarily focuses on separating anomalous targets from the observed background data, based on the spectral difference between anomalous targets and other pixels. Until now, the definition of anomalous targets remains ambiguous and is characterized as pixels significantly distinct from the surrounding background spectrum, along with a low probability of occurrence. Anomaly detection is used to locate unusual objects with particularly small proportions and great differences from their neighbors in spectral characteristics. Depending on the application scenario, anomalous targets may appear as different entities, such as infected trees in woods, rare minerals in geological applications, and man-made objects (vehicles, aircraft, and tanks) in defense applications.

Initial HAD methodologies predominantly relied on traditional statistical approaches, with the assumption that the background distribution of HSI could be adequately modeled using standard statistical distributions, with the Gaussian distribution being particularly prevalent. Anomalies were thus identified as those pixels exhibiting significant statistical divergence from the established Gaussian model of the background. The most representative method is the Reed-Xiaoli (RX) method [13] proposed

in 1990, assuming that the background obeys a multivariate normal distribution and the Mahalanobis distance is used to estimate the anomalous level of each pixel. Extensive experiments have shown that RX performs satisfactorily in HSIs with simple backgrounds, but not in complex scenes. In view of this, various improved methods emerged, such as the kernel-RX [14], [15], regularized RX [16], and local-RX [17]. These approaches, while effective in certain contexts, often face limitations in more complex real-world hyperspectral scenes where the background distribution might not conform neatly to simple Gaussian or similar statistical models [18].

In response to the limitations inherent in the background distribution hypothesis, representation-based approaches have been gradually proposed. These methods seek to reconstruct background distribution accurately by considering HSI as a combination of background and anomalies. To avoid making any inaccurate assumptions about the HSIs, the collaborative representation-based detector (CRD) [19] detects anomalies based on the concept that the background pixel can be linearly represented by its adjacent pixels, whereas the anomalous pixel cannot. The CRD has garnered significant attention due to its simplicity and high efficiency. However, the anomaly detection results of CRD may be significantly influenced by the size of the dual window in different datasets. The concept of low-rank representation (LRR) [20] involves computing the lowest rank representation of all pixels in the dataset, which is then employed to detect anomalies by subtracting the reconstructed background from the original image. While effective in capturing global pixel relationships, it overlooks the local structure of the pixel coefficient, which is crucial for accurate representation, potentially affecting the accuracy of anomaly detection results. Given the fact that the background in HSI is widely distributed while the target rarely appears, further refinement is achieved through the low-rank [21] and sparse constraints. Representative approaches have been proposed such as low-rank and sparse representation (LRASR) [22] based HAD, low-rank matrix decomposition based on Mahalanobis distance for anomaly detector (LSMAD) [23], and so on [24]. Numerous studies have demonstrated that the low-rank constraint is essential for HAD, allowing the model to effectively distinguish background and anomaly under a delicately designed dictionary.

The advancement of deep learning methodologies in image processing has facilitated their integration into HSI processing, [25], [26], [27]. In terms of HAD, convolutional neural networks (CNNs), auto-encoders (AEs), and generative adversarial networks (GANs) have been extensively studied. CNNs have demonstrated proficiency in extracting intricate hierarchical features from HSIs, where the subsequent convolution layer can extract data features from the preceding convolution layer through convolution kernel operation, thereby completing the extracting of high-level feature information from the underlying features [28], [29]. CNN-based anomaly detectors are usually designed using the differences between pixel spectra, such as the transferred CNN anomaly detector [30], which detects anomalous targets by learning the spectral differences of a known labeled classification dataset. For the unsupervised HAD methods, the dominant behavior is to achieve anomaly

detection by reconstructing backgrounds without anomalies. So, the latest CNN-based method provides a new solution for HAD, which designs a blind spot strategy to train a background reconstruction network to detect anomalies [31]. The AE-based anomaly detectors consider that the main background distribution in HSI can be more easily reconstructed by a well-designed AE network than that of anomalies. Therefore, the key goal is to design an AE network that almost exclusively reconstructs the background. On this basis, the RGAE [32] imposes the graph regularization on the latent layer of AE to consider the relationship between pixels. Besides, to avoid manual parameter setting during training, an automatic HAD (Auto-AD) method [33], based on a fully convolved AE with an adaptive adjustment loss function, is proposed. Xie et al. [34] adopt the Markov chain Monte Carlo and Gaussian mixture model so that the network adapts to complex scenes. To learn intrinsic low-dimensional attributes of HSIs, Wang et al. [35] build a dense residual self-attention module and a low-rank loss function to constrain the generation of the latent features. In terms of avoiding to extract features mixed with anomalous information, a guided map is generated to restrain the participation of anomalies in GAED [36].

The deep neural networks, leveraging backpropagation algorithms, exhibit the ability to extract intricate features, yielding promising results in addressing HAD challenges. Despite the advancements in deep learning methodologies for HAD compared with tradition methods, these methodologies are still in the preliminary exploration stage, and the following challenging issues still need to be further explored.

- 1) To converge to a good local minimum of the loss function, the over-parameterization of the networks used in training, brings a certain degree of network parameter redundancy, resulting in increased time costs of detection. The resource consumption and the timeliness of detection of the model are critical for the performance of HAD. Therefore, when designing the depth model for optimal performance, it is important to consider its complexity, including the number of model parameters and computational requirements.
- 2) The capability of deep learning-based HAD to handle large spatial domain images remains a challenge. The current approaches are primarily designed for smaller spatial sizes, whereas practical applications often involve large-scale images. Within the constraints of limited computational resources, the images with large spatial domains may cause a dramatic increase in the number of parameters and computation for these methods, which does not meet the anomaly detection requirements for detection timeliness.
- 3) Both the AE and GAN-based methods detect anomalies by reconstructing the background, but experiments show that most of them inevitably reconstruct anomalies at the same time, which is especially obvious in the scenarios with the large target. Furthermore, these methods still lack the use of spatial information, which is an important component of HSIs.

In response to the aforementioned challenges, this article presents a general network compression scheme for the CNNs.

For large-scale images, this scheme can significantly reduce the number of parameters and improve the computational efficiency, resulting in improving detection timeliness. To assess the effectiveness of this scheme, a novel HAD method with Tucker decomposition is proposed, named TD-CNND, which builds upon the existing CNN-based method. This method avoids the issue of reconstructing anomalies during background reconstruction, which is a common problem with existing methods. HAD is regarded as a binary classification problem by the TD-CNND method, with learning the differences between pixel spectra to detect anomalies. Specifically, inspired by transfer learning, the proposed TD-CNND extends pixel pairs by leveraging label information from HSI in the source domain, thereby constructing sufficient training data. These pixel pairs are then fed into the CNN to learn spectral feature differences. After training, a CNN containing knowledge of spectral feature differences in the source domain is obtained. Subsequently, to prevent the need for network retraining caused by structural changes and to reduce model parameters for improving detecting timeliness, a general network compression scheme based on Tucker decomposition is applied to the CNN obtained from transfer learning. Tucker tensor decomposition is employed to decompose the convolutional layer into two 1×1 convolutional layers and a core convolutional layer, where achieving the network compression and reducing the parameter redundancy in the network. After that, spectral features re-alignment (SFRA) is used to recover the detection accuracy loss caused by decomposition. Finally, anomaly detection in the target domain is performed by combining spectral and spatial information in a dual-window structure. The main contributions of this article are outlined as follows.

- 1) *Innovative Network Compression Scheme:* A general network compression scheme that applies Tucker tensor decomposition to CNNs. This strategy significantly reduces the network parameters and prevents the need for retraining due to structural changes. A novel component, SFRA, is specifically designed to counteract the loss in detection accuracy that might accompany the compression process. This ensures that the model maintains high performance in anomaly detection tasks even after substantial parameter reduction.
- 2) *Enhanced Model Scalability for Large-Scale Applications:* Unlike existing deep learning-based methods that struggle with scalability for large field views, the proposed model can render the CNN-based methods more suitable for anomaly detection tasks across large field of view in real-world scenarios, without a dramatic increase in parameters and computation cost. This is particularly critical in real-world scenarios where a timely and efficient process is essential. This approach not only enhances the efficiency and scalability of CNNs but also preserves their performance, addressing a significant challenge in the practical deployment of HAD systems.
- 3) *Broad Applicability of the Compression Scheme:* The proposed network compression scheme has potential applications beyond HAD. It is suitable for any application involving 1-D CNNs, particularly those dealing with large datasets that include extensive spectral and spatial

information. This versatility extends to various tasks in hyperspectral imaging, such as target detection and classification, thereby demonstrating the wide-ranging impact of our approach within and potentially outside the field of hyperspectral imaging.

The rest of this article is organized as follows. Section II provides a comprehensive description of the proposed TD-CNND, and Section III presents experimental studies and analysis to validate the proposed approach, finally, conclusions are drawn in Section IV.

II. PROPOSED METHOD

This section shows the details of the proposed HAD method using Tucker decomposition (TD-CNND). The method comprises three stages: source domain transfer learning, Tucker decomposition, and SFRA, and spectral-spatial anomaly detection in the target domain, as illustrated in Fig. 1.

A. Source Domain Transfer Learning

In recent years, the field of remote sensing processing has witnessed the ascendancy of deep learning, culminating in the advent of methodologies predicated on deep learning-based HAD. Given the inherent absence of a priori information amenable to direct utilization for anomalies, training deep neural networks in a supervised learning paradigm for anomaly detection may become unfeasible. However, transfer learning, which has the ability in applying labeled knowledge from a source domain dataset to the unlabeled task within the target domain, is able to handle this problem. A transfer learning-based method named CNND for hyperspectral detection was proposed at the beginning, wherein knowledge from a source domain dataset with known labels is effectively transferred to the task of anomaly detection in the target domain. The TD-CNND proposed in this section represents a notable advancement over the CNND, particularly in the context of time efficiency for HAD within a large field of view. This section mainly describes the process of source domain transfer learning for the TD-CNND method in detail.

1) *Training Data Preparation:* The preparation of the training data involves expanding the training dataset through the label information from a source domain hyperspectral classification image. This process involves categorizing the difference between spectral pixel pairs of the same category as the “similarity” class (labeled as 0), and that between spectral pixel pairs of different categories as the “dissimilarity” class (labeled as 1). Consequently, a sufficient number of sample pairs from both similarity and dissimilarity classes are obtained to form the training dataset for the 1-D-CNN. This strategy facilitates the network in discerning differences between spectra.

2) *One-Dimensional CNN:* Following the acquisition of the training dataset for transfer learning, a 1-D-CNN is employed to learn the spectral feature differences. The network is a feed-forward network, which encompasses 1-D convolutional layers, fully connected layers, and activation function layers. This design enables the incorporation of locally relevant information through inter-neuronal connections across adjacent layers. As

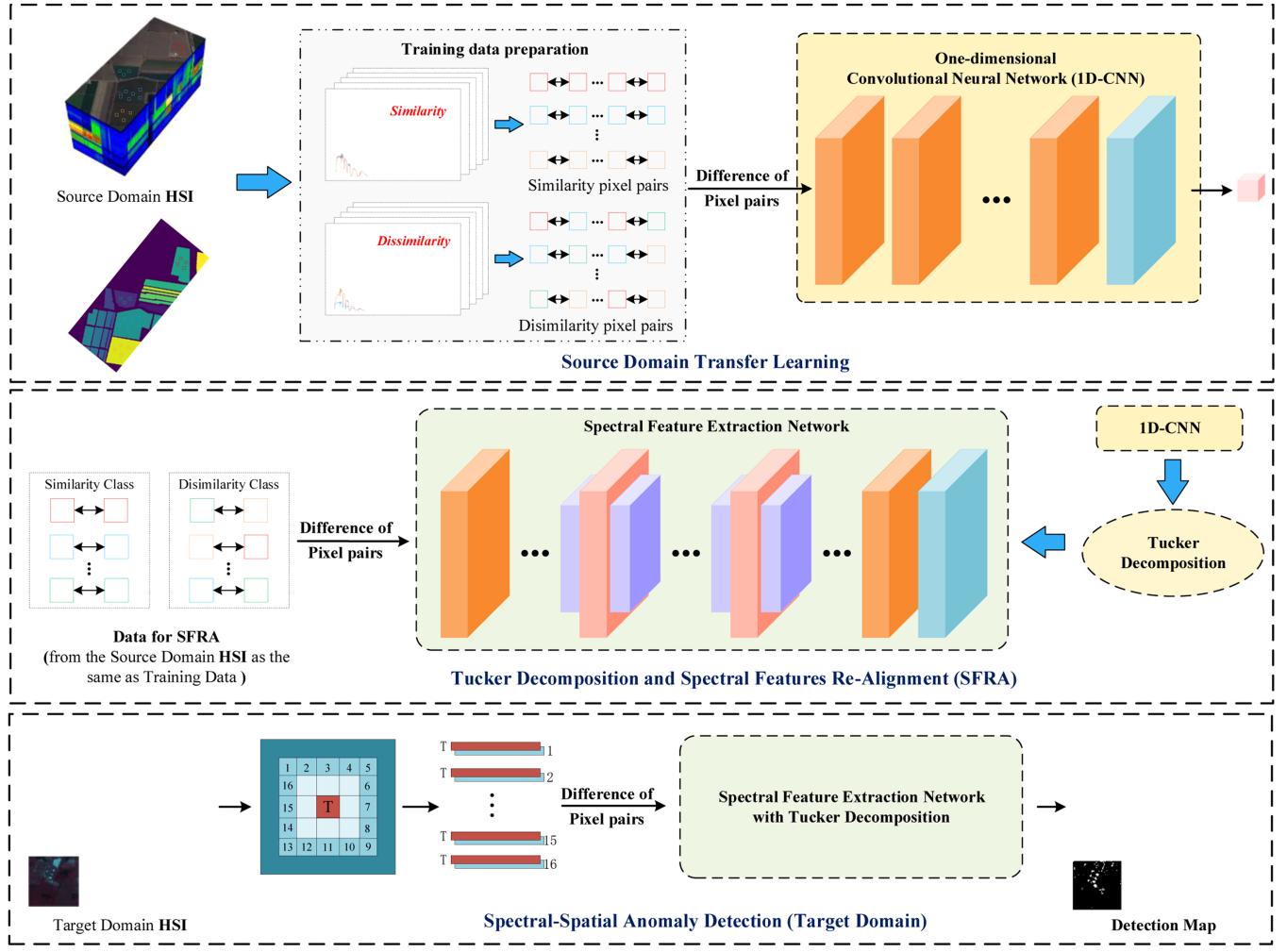


Fig. 1. Flowchart of the proposed TD-CNND method.

shown in Fig. 2, the main network configuration includes 17 learnable 1-D convolutional layers. In the designed framework, the 1-D-convolutional layers, namely C1, C2, C3, C5, C6, C8, C9, C11, C12, C14, and C15, are configured with a stride of 1, while the 1-D-convolutional layers C4, C7, C10, C13, and C16 employ a stride of 2. The convolution kernel size for the aforementioned 16 convolutional layers is set to 1×3 . The final learnable 1-D-convolutional layer is the Pool layer, distinguishing itself from the average pooling layer employed in the original paper by incorporating a 1×1 1-D-convolutional layer, thereby contributing to the preservation of spectral information within the HSIs. Noteworthy optimizations to various components of the CNN, such as augmenting the number of convolutional kernels to increase the difficulty of convolution, and modifying the pooling layer to prevent the loss of detailed information, have been implemented rigorously to substantiate the efficacy of the proposed TD-CNND method.

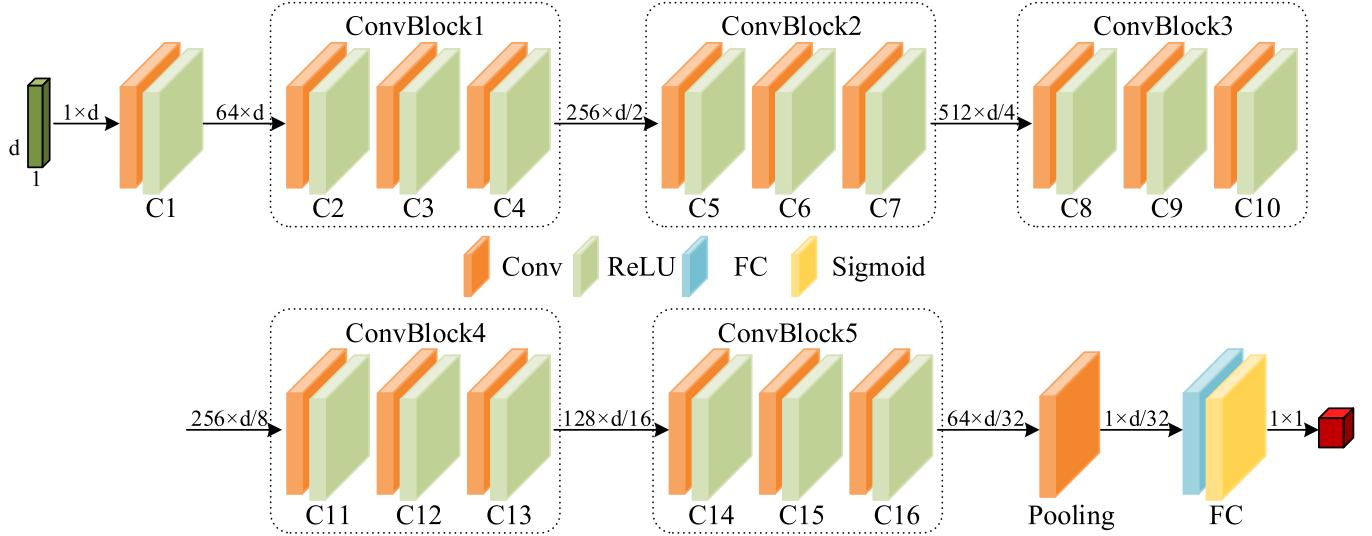
Differing from both the AE and GAN-based methods indirectly detecting anomalies by reconstructing background, the HAD is regarded as a binary classification problem by the TD-CNND method, with learning the differences between pixel spectra to detect anomalies. This effectively avoids that most of

the above-mentioned methods inevitably reconstruct anomalies at the same time, which is especially obvious in the scenarios with large targets. Consequently, using the sigmoid activation function serves as the final layer to produce an output representation in terms of scores (or labels), deriving the probability that a given pixel belongs to an anomaly. The optimization function employed for the training of the 1-D-CNN in transfer learning is the binary cross entropy, expressed as follows:

$$\text{Loss}_{\text{BCE}} = -\frac{1}{B} \sum_{i=1}^B [y_i \cdot \log f_i + (1 - y_i) \cdot \log (1 - f_i)]. \quad (1)$$

B. Tucker Decomposition and Spectral Feature Realignment

The integration of deep learning-based methods has indeed substantially advanced the extraction of intricate features for HAD, resulting in notable improvements in detection performance. However, this advancement has come at the expense of increased hardware requirements and heightened computational complexity. In contrast to traditional anomaly detection methods, deep learning-based HAD methods leverage a multitude of network parameters to guide the trained network toward a



C1,C2,C3,C5,C6,C8,C9,C11,C12,C14,C15,Pooling: Conv(1×3),stride=1 C4,C7,C10,C13,C16: Conv(1×3),stride=2

Fig. 2. One-dimensional CNN for TD-CNNND.

favorable local minimum of the optimization function. However, these methods would introduce redundancy in network parameters, consequently increasing computational complexity. Though efficacious for HAD in scenarios characterized by HSIs with smaller field of views and relatively diminutive spatial dimensions, these methods may prove less apt for practical applications that entail a substantial volume of large-scale HSIs. Consequently, the imperative arises to design efficient deep learning-based HAD methods that mitigate time costs while concurrently enhancing or sustaining detection accuracy.

1) Network Compression Scheme: To prevent the need for network retraining caused by dramatic structural changes and to reduce model parameters for improving detecting timeliness, this article addresses this challenge by introducing a general network compression scheme for CNNs, incorporating Tucker tensor decomposition, aiming to alleviate the network parameter redundancy inherent in 1-D-CNN due to the over-parameterization. Subsequently, SFRA is specifically designed in the scheme to ensure the alignment of the network parameter distribution after decomposition with the source-domain migration knowledge, thereby reinstating anomaly detection accuracy. It is worth noting that this general network compression scheme has well transferability, which can be applied directly to any 1-D convolutional layers-based CNNs. This compression scheme proposed is primarily based on the technology of Tucker tensor decomposition/Tucker decomposition [37], [38], [39].

Tucker Decomposition is a mathematical technology within multilinear algebra applied to decompose higher order tensors, which represent multidimensional arrays of data, into a set of core tensors and factor matrices. This process is instrumental in extracting latent information and mitigating the complexity of high-dimensional data representations. Tucker decomposition serves as a higher-order extension of the singular value decomposition commonly applied to matrices. In this technique,

mode-n matrices within the original tensor are analyzed and combined with a core tensor, as shown in Fig. 3. The core tensors capture the essential information, while the factor matrices preserve the relationships between mode-n matrices. Consequently, Tucker decomposition is valuable for effectively mitigating the over-parameterization and reducing the computational demands associated with CNNs.

The genera network compression scheme proposed is shown in Fig. 4. Distinct from 2-D convolutional operations on the spatial dimensions of images and 1-D-convolution extracts hidden features along the spectral dimension of HSIs. Distinct from 2-D-convolutional operations on the spatial dimensions of images, 1-D-convolution extracts hidden features along the spectral dimension of HSIs. Consequently, Tucker decomposition is applied to the convolution kernel tensor of the 1-D-convolution layer, and the 1-D-convolution operation can be described as follows: it maps an input tensor χ of size $N \times S$ into an output tensor γ of size $N' \times T$ by using the following linear mapping:

$$\gamma_{n'_i,t} = \sum_{i=1}^D \sum_{s=1}^S \kappa_{i,s,t} \chi_{n_i,s} \quad (2)$$

where κ is a convolution kernel tensor of size $D \times S \times T$. And Fig. 4 shows the flowchart of the network compression scheme proposed. It utilizes the technology of Tucker decomposition to decompose each 1-D-convolutional layer into a novel structure, consisting of two 1×1 convolutional layers and a core convolutional layer. This process aims to construct the lightweight spectral feature extraction network to eliminate redundant parameter information within the network. The process can be described mathematically as follows.

With Tucker model, the kernel tensor κ is decomposed according to the rank (R_1 , R_2 , R_3) into the following

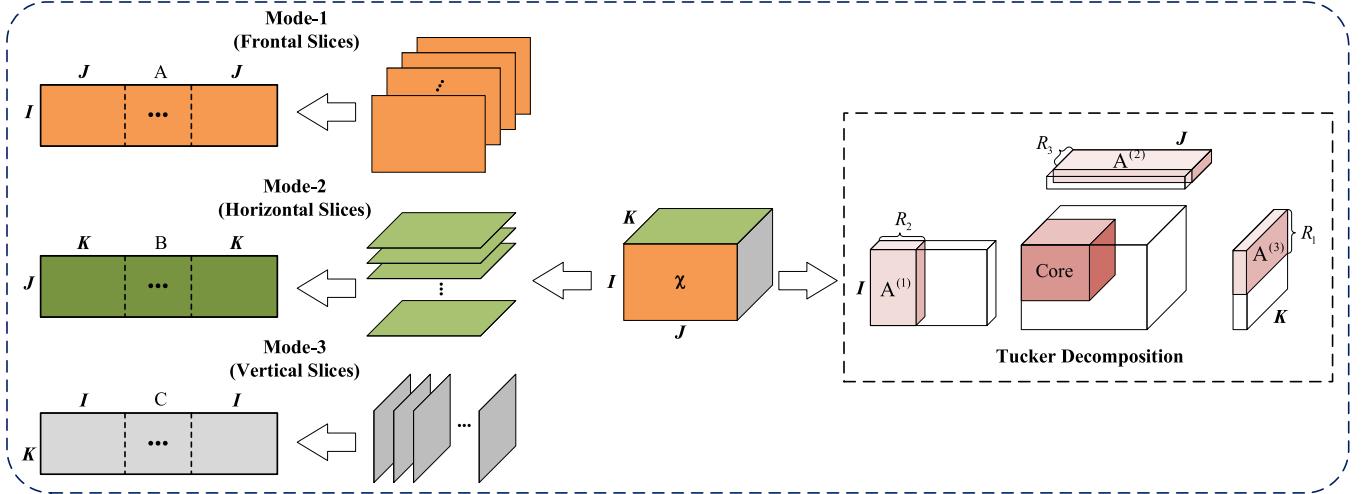


Fig. 3. **Left:** (a) mode-1, (b) mode-2, and (c) mode-3 matrix of the 3-way tensor χ . They are constructed by concatenation of frontal, horizontal, and vertical slices, respectively. **Right:** illustration of 3-way tensor χ tucker decomposition. It is decomposed into the core tensor of size $R_1 \times R_2 \times R_3$ and the factor matrices $A^{(1)}$, $A^{(2)}$, and $A^{(3)}$. R_1 , R_2 , and R_3 is the rank of different dimensions.

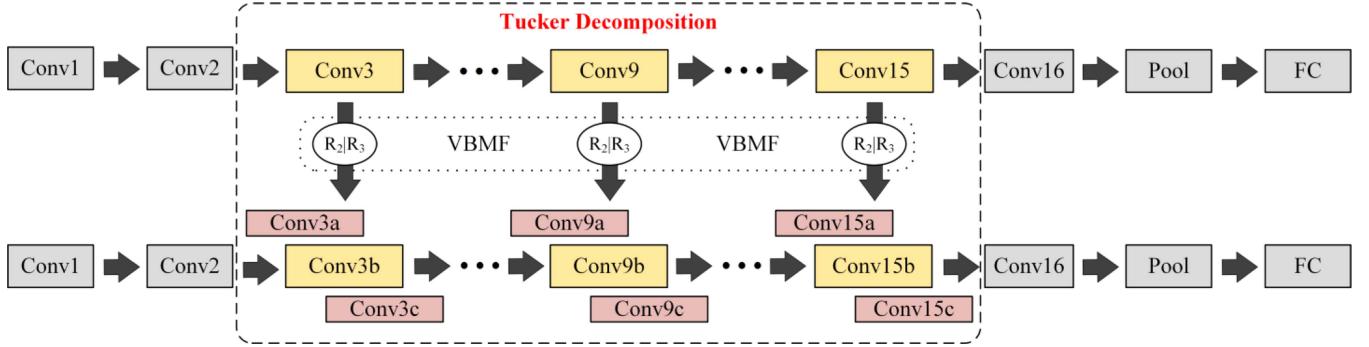


Fig. 4. Flowchart of the designed scheme of network compression.

form:

$$\kappa_{i,s,t} = \sum_{r_1=1}^{R_1} \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} C'_{r_1,r_2,r_3} U_{i,r_1}^{(1)} U_{s,r_2}^{(2)} U_{t,r_3}^{(3)} \quad (3)$$

where C' is a core tensor of size $R_1 \times R_2 \times R_3$, $U^{(1)}$, $U^{(2)}$ and $U^{(3)}$ are factor matrices of sizes $D \times R_1$, $S \times R_2$ and $T \times R_3$, respectively.

In fact, it is crucial to underscore that not every dimension necessitates the application of Tucker decomposition. Within the network compression *scheme* delineated in this article, a deliberate decision has been made to abstain from decomposing the dimension associated with the size of the convolution kernel tensor κ , primarily due to its typically diminutive scale (commonly denoted as D , and typically equal to 3 or 5). Within this specific variant of Tucker decomposition, referred to as Tucker-2 decomposition [40], the convolution kernel tensor κ is decomposed as

$$\kappa_{i,s,t} = \sum_{r_2=1}^{R_2} \sum_{r_3=1}^{R_3} C_{i,r_2,r_3} U_{s,r_2}^{(2)} U_{t,r_3}^{(3)} \quad (4)$$

where C is a core tensor of size $D \times R_2 \times R_3$.

Through substituting the factorization (4) of the kernel tensor κ into the mathematical expression (2) for 1-D-convolution and summing the rearranged groups, three continuous expressions for the approximate evaluation of 1-D-convolution are derived

$$Z_{n,r_2} = \sum_{s=1}^S U_{s,r_2}^{(2)} \chi_{n,s} \quad (5)$$

$$Z'_{n',r_3} = \sum_{i=1}^D \sum_{r_3=1}^{R_3} C_{i,r_2,r_3} Z_{n_i,r_3} \quad (6)$$

$$\gamma_{n',t} = \sum_{r_3=1}^{R_3} U_{t,r_3}^{(3)} Z'_{n',r_3} \quad (7)$$

where Z and Z' are intermediate tensors of sizes $N \times R_3$ and $N' \times R_3$, respectively.

Based on the aforementioned formula, it can be observed that the computing of Z' involves a 1-D-convolution with the core tensor κ as the convolution kernel. And the computing of Z and γ correspond to 1×1 convolution that essentially performs pixel-wise linear recombination of input maps. Note that there

is no nonlinear ReLU function in the computations of (5), (6) and (6), (7). The computational process entails the generation of a smaller intermediate output tensor in (6), subsequently restored to its original size in (7). The integration of Tucker decomposition inherently combines two compression techniques, seamlessly incorporating both forms of network compression within the process of Tucker-2 decomposition.

2) Rank Selection With Global Analytic VBMF: The determination of the rank (R_2, R_3) for the convolution kernel tensor κ serves as a crucial hyperparameter within the network compression scheme proposed, controlling the delicate balance between performance improvement and accuracy preservation. Instead of engaging in a continuous debugging approach through traditional parameter tuning to ascertain the optimal rank (R_2, R_3), this article adopts a data-driven, one-shot decision-making process, facilitated by empirical Bayes with automatic relevance determination prior. Variable Bayesian matrix factorization (VBMF) [41] proves advantageous due to its capacity to automatically finding the noise variance and rank while offering theoretical conditions for perfect rank recovery [42]. Consequently, this paper determines the rank R_2 and R_3 by applying global analytic VBMF on mode-2 matrices (size of $S \times TD$) and mode-3 matrices (size of $T \times SD$) of kernel tensor κ , respectively. This strategic approach contributes to the robust and principled selection of the rank, thereby enhancing the efficiency of the network compression scheme.

3) Spectral Features Realignment: While the parameters of the network post-Tucker decomposition undergo a substantial reduction in comparison to those of the 1-D CNN without Tucker decomposition, it is inevitable that the parameter distribution of the new network may not align well with the transfer knowledge from the source domain, leading to the degradation of the detection accuracy in the target domain. To address this issue, the process of SFRA is specifically designed in this scheme. Random data, which is also from the source domain dataset, is utilized to realign drift features to establish the lightweight spectral feature extraction network, thus restoring the detection accuracy degraded by the network decomposition.

An important detail should be but not explicitly mentioned in previous sections is that the network compression scheme presented in this article does not decompose all convolutional kernel tensors within the learnable 1-D convolutional layers. Instead, Tucker-2 decomposition is selectively applied solely to specific internal convolutional layers, namely C3 to C15, within the 1-D-CNN architecture. This deliberate selection avoids the risk of over-compression across the entire network, a circumstance that could lead to a substantial decline in detection accuracy due to the loss of numerous network parameters, potentially causing irrecoverable during the process of SFRA. The determination of the quantity and specific convolutional layers subjected to Tucker decomposition necessitates a nuanced consideration tailored to different HSI datasets and tasks. In the proposed TD-CNND model, the choice of convolutional layers for Tucker decomposition is based on empirical insights and validation experiments, reflecting a context-specific approach that may not be universally representative across datasets and tasks.

C. Spectral-Spatial Anomaly Detection

In contrast to hyperspectral target detection tasks benefiting from prior information of known target spectra, HAD faces the challenge of lacking such informative priors. The goal of HAD is to identify pixels within HSIs that exhibit significant differences from their surroundings, commonly referred to as anomalies. To effectively detect these anomalies, the method employed in this article involves comparing spectral differences between the under-tested pixel and its neighboring pixels acquired through a dual-window, where the utilization of dual-window leverages spatial information for accurate detection. As depicted in Fig. 1, the under-tested pixel, denoted as T , is the central pixel of the dual window. The dual window employed in this article consists of an inner window of size $W_{\text{in}} \times W_{\text{in}}$ and an outer window of size $W_{\text{out}} \times W_{\text{out}}$. For each central pixel to be detected, surrounding pixels are extracted, forming pairs to be assessed for anomaly detection. These pairs are then fed into the spectral feature extraction network, generating corresponding feature vectors. These vectors are then averaged to obtain the detection result.

Finally, the detection result is compared against a predefined threshold η . Pixels larger than the threshold η are considered anomalies, while those smaller than η are considered normal pixels. This spectral-spatial anomaly detection methodology effectively integrates both spectral and spatial information to enhance the precision of anomaly detection in HSIs.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, a comprehensive set of experiments is conducted on three real hyperspectral datasets to validate the effectiveness of the proposed TD-CNND method in terms of both time cost and detection accuracy.

In Section III-A, we briefly describe the real hyperspectral datasets used. Section III-B shows the evaluation criteria for HAD. Section III-C gives the relevant experimental setups for the proposed method and the comparison methods. It also introduces the corresponding hardware environments. In Section III-D, the detection results and analysis of TD-CNND and the comparison methods on the real datasets are given. It proves the effectiveness and superiority of TD-CNND method. Section III-E briefly analyses the effect of hyperparameter—the dual windows size ($W_{\text{in}}, W_{\text{out}}$) on detection accuracy. Section III-F provides a more definite complexity analysis of the proposed method and the deep learning-based comparison methods. In Section III-G, we perform ablation experiments by manually extending the real datasets to validate the time-cost advantage of TD-CNND on large spatial domain datasets.

A. Hyperspectral Datasets

This section provides the detailed information about the real hyperspectral datasets used for detection in the target domain, along with a labeled hyperspectral dataset of the same sensor used for transfer learning in the source domain.

Salinas Dataset: The Salinas dataset, acquired by AVIRIS sensors over the Salinas Valley in California, USA, has 224

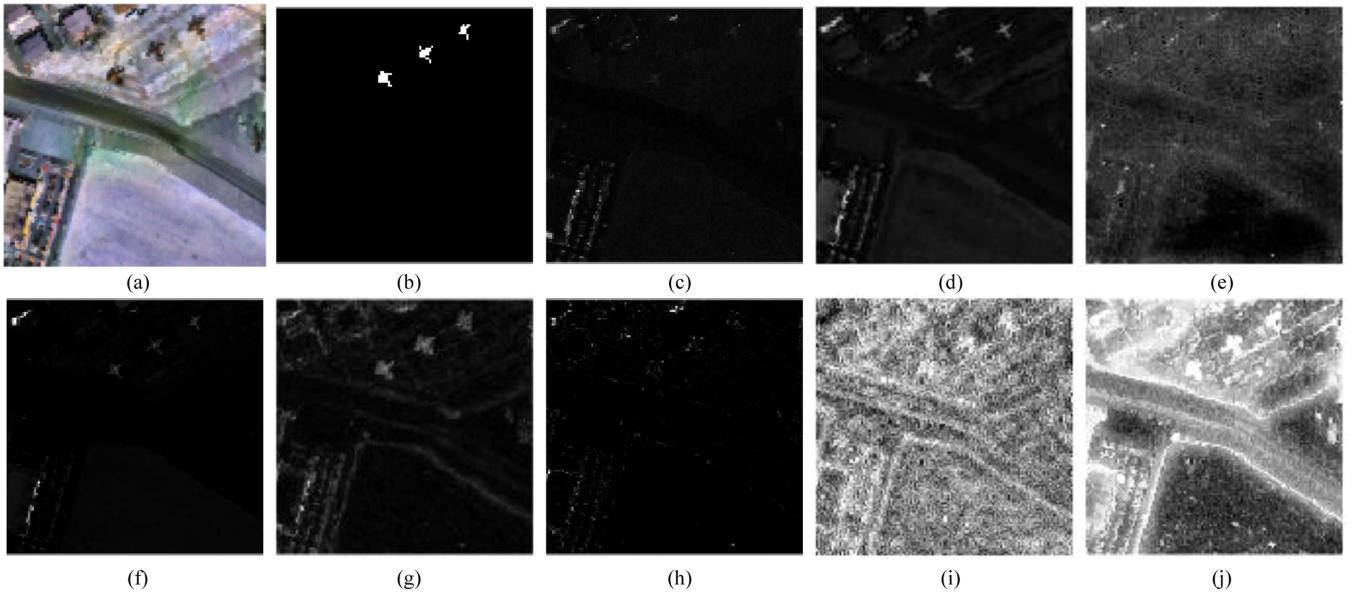


Fig. 5. Anomaly detection maps for SanDiego dataset. (a) Pseudo-color image. (b) Ground truth. (c) RX. (d) LRASR. (e) CRD. (f) RGAE. (g) GNLTR. (h) Auto-AD. (i) CNND. (j) TD-CNN.

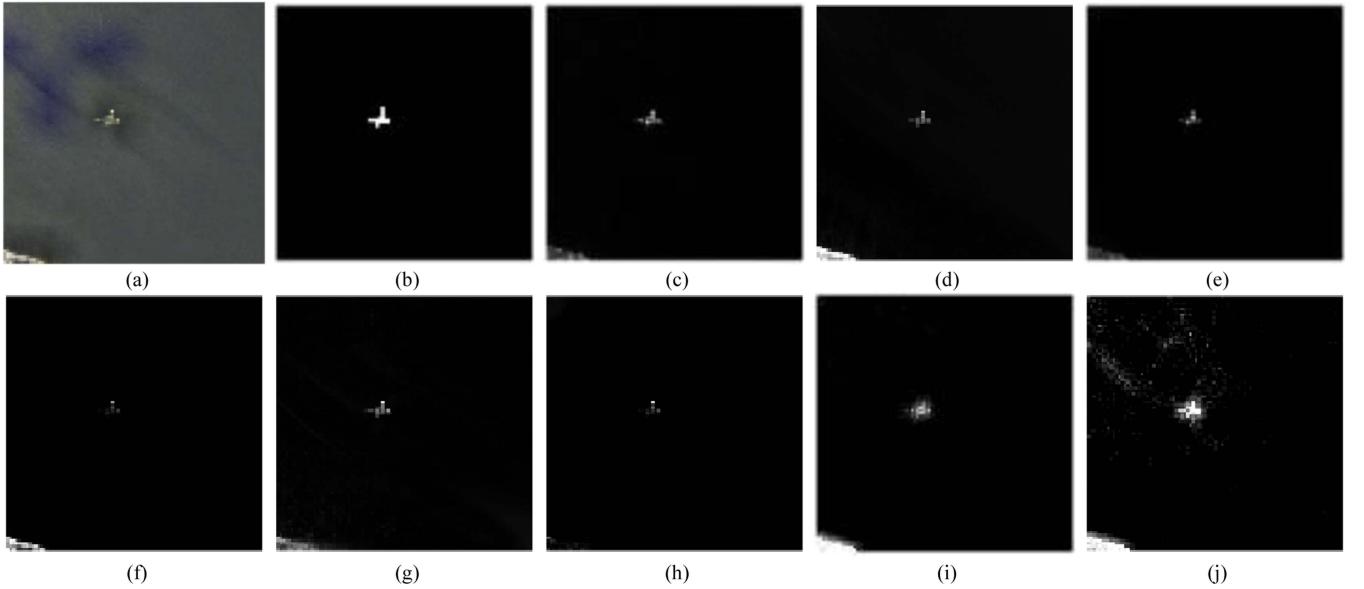


Fig. 6. Anomaly detection maps for beach dataset. (a) Pseudo-color image. (b) Ground truth. (c) RX. (d) LRASR. (e) CRD. (f) RGAE. (g) GNLTR. (h) Auto-AD. (i) CNND. (j) TD-CNN.

bands with a spatial resolution of 3.7 m. The original image size is 512×217 , featuring 16 categories of label samples, including vegetables, bare soil, vineyards, etc., each containing hundreds of label samples. Since labeled information is provided, this dataset is used the source domain transfer learning.

SanDiego Dataset: The San Diego dataset, collected by AVIRIS at the San Diego Airport area, California, USA, exhibits a spatial resolution of 3.5 m and image of size 120×120 , with a total of 224 bands, spectral resolution of 10 nm and a wavelength range of 370–2510 nm. After removing low SNR and water absorption bands, a total of 189 bands are retained for detection. The pseudo-color image and the corresponding ground truth are

shown in Fig. 5(a) and (b). The aircraft (58 in total) in a scene is treated as an anomaly.

Beach Dataset: The Beach dataset is captured by the AVIRIS sensor on Cat Island, with a spatial resolution of 17.2 m. After preprocessing such as removing the noise bands, an image with the size of $90 \times 90 \times 188$ is obtained in this experiment. The pseudo-color image and the corresponding ground truth are shown in Fig. 6(a) and (b), including 19 anomaly points.

Urban Dataset: The Urban dataset is captured by AVIRIS sensors off the coast of Texas in the United States. The whole image has a total of 100×100 pixels, leaving 204 bands after removing the low signal-to-noise ratio bands. The pseudo-color

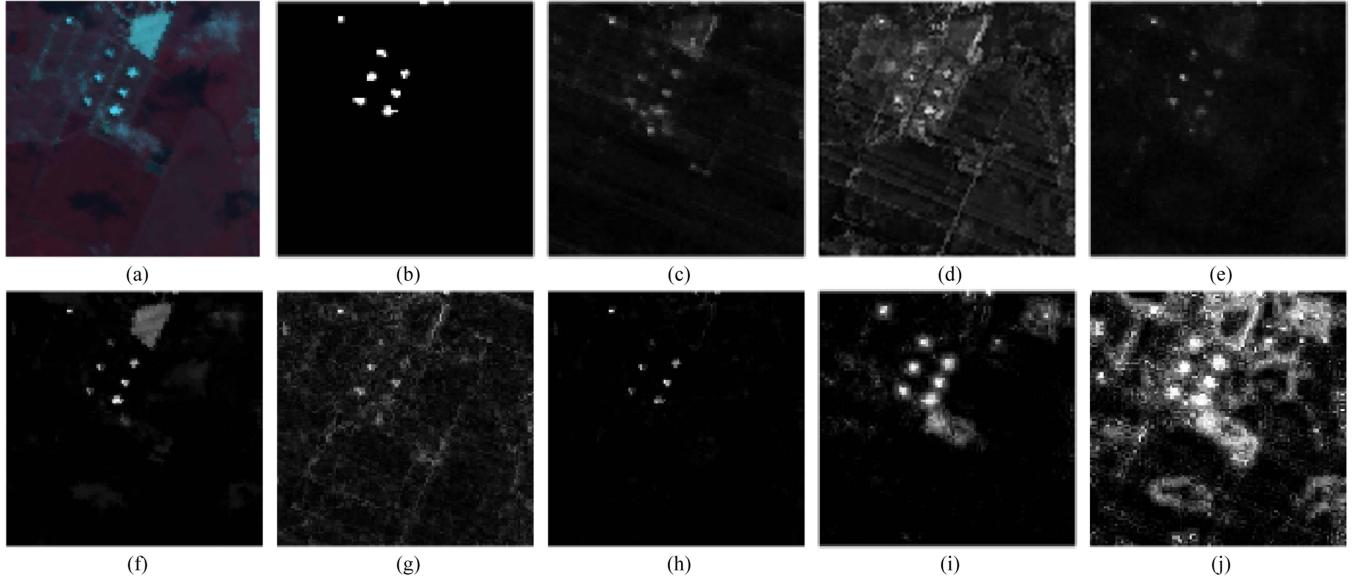


Fig. 7. Anomaly detection maps for Urban dataset. (a) Pseudo-color image. (b) Ground truth. (c) RX. (d) LRASR. (e) CRD. (f) RGAE. (g) GNLTR. (h) Auto-AD. (i) CNND. (j) TD-CNNND.

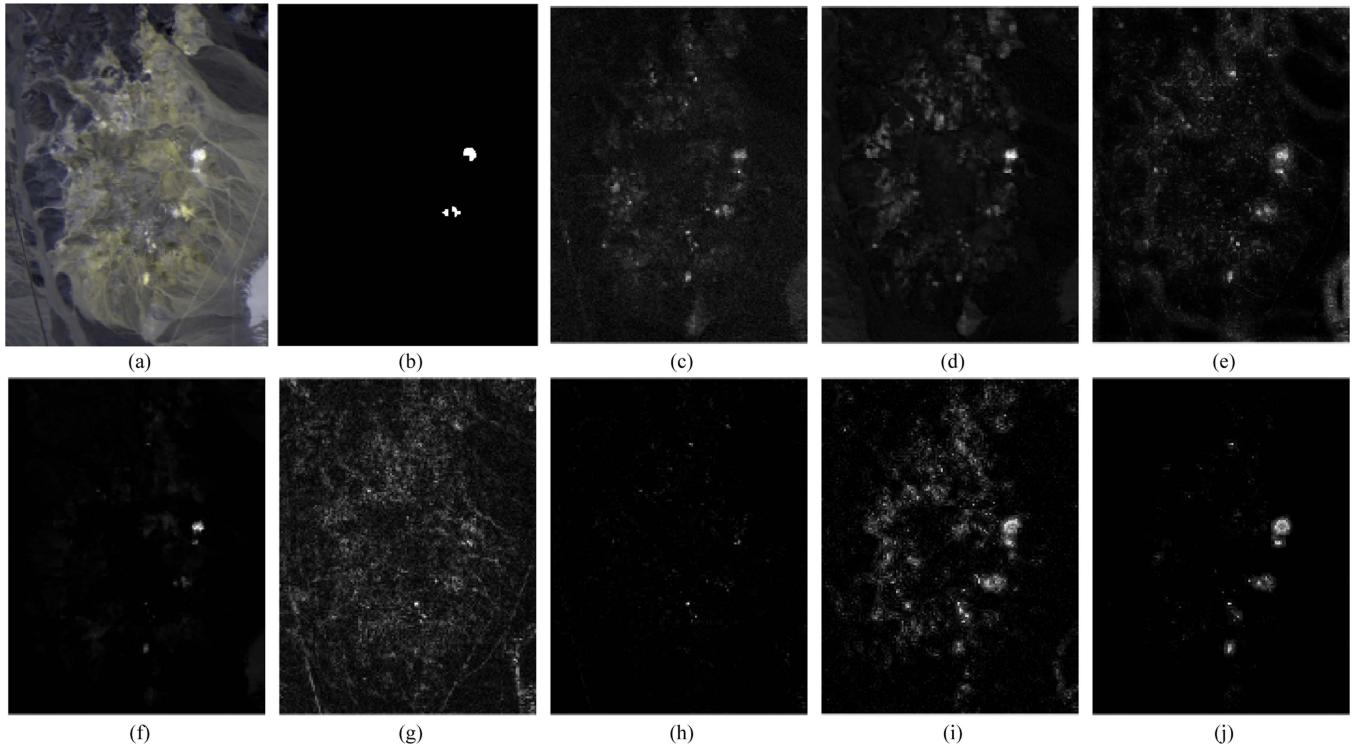


Fig. 8. Anomaly detection maps for Cuprite dataset. (a) Pseudo-color image. (b) Ground truth. (c) RX. (d) LRASR. (e) CRD. (f) RGAE. (g) GNLTR. (h) Auto-AD. (i) CNND. (j) TD-CNNND.

image and corresponding ground truth are shown in Fig. 7(a) and (b), including 64 anomaly points.

Cuprite Dataset: The Cuprite dataset was obtained by the AVIRIS sensor, in the Cuprite mining district of Nevada in 1997. There are about 14 kinds of minerals in this image, including buddingtonite, Na-Montmorillonite, Nontronite

(Fe clay), Kaolinite, etc. We use a 250×191 pixel subset of this image to conduct our experiment. After removing the low SNR and water absorption bands, 188 bands are left to conduct our experiment. The pseudo-color image and corresponding ground truth are shown in Fig. 8(a) and (b).

B. Evaluation Criteria

To evaluate the performance of the proposed method in comparison with the state-of-the-art methods, quantitative analysis is performed using the receiver operating characteristic curve (ROC) and its area under the curve (AUC) [43], [44]. The ROC curve has been widely used as an evaluation tool for the target detection in HSIs. The ROC curve obtains different detection probability P_D and false alarm probability P_F by changing the threshold value τ . Detection probability P_D and false alarm probability P_F can be calculated by the following equation:

$$P_D(\tau) = \frac{n_{D,\tau}}{n_{D,\tau} + n_{FN,\tau}} \quad (8)$$

$$P_F(\tau) = \frac{n_{F,\tau}}{n_{F,\tau} + n_{TN,\tau}} \quad (9)$$

where $n_{D,\tau}$, $n_{FN,\tau}$, $n_{F,\tau}$, and $n_{TN,\tau}$ represent the number of correctly detected target pixels, the number of pixels that are targets but not detected as targets, the number of background pixels that are detected as target pixels, and the number of correctly detected background pixels below the threshold, respectively.

Due to the interaction between the detection probability P_D and the false alarm probability P_F , the ROC curve (P_D , P_F) with a higher AUC value does not necessarily mean that the detector has a good background suppression ability. Therefore, in order to evaluate the detector performance more accurately, this article uses 3-D ROC curve [43] as the evaluation standard, and three 2-D ROC curves (P_D , P_F), (P_D , τ), and (P_F , τ) are used to evaluate the detector's effectiveness, detection ability, and background suppression ability, respectively.

The AUC is the value of area under the ROC curve, used to quantitatively evaluate the performance of the detector. For the 2-D ROC curve (P_D , P_F), AUC (P_D , P_F) value between 0.5 and 1 indicates that the detector is effective, with closer values to 1 signifying better performance. AUC (P_D , τ) is the AUC of the 2-D ROC curve (P_D , τ), quantitatively representing the target detection capability of the detector, with the larger values indicating stronger detection ability. While AUC (P_F , τ) value is the AUC of the 2-D ROC curve (P_F , τ), measuring the ability of the background suppression, with smaller values indicating better suppression of the background. Besides, a new quantitative detection index designed in [43] takes the three AUC values as a whole to measure the total performance, named AUC_{OD} , with a range of [-1, 2], which is defined as

$$AUC_{OD} = AUC_{(P_D, P_F)} + AUC_{(P_D, \tau)} - AUC_{(P_F, \tau)}. \quad (10)$$

C. Experimental Setup

This section mainly introduces the parameter setting and the experimental platform used by TD-CNND and the comparison methods.

The experimental environment of both the proposed TD-CNND method and the CNND method includes an AMD Ryzen Threadripper 3990X 64-core CPU and Quadro RTX 8000 48GB GPU. The implementation is carried out using Python 3.8.0, PyTorch 1.12, and MATLAB R2023a. Traditional methods and

the machine learning-based methods use an Intel Core i5-9300h 8-core CPU, and are implemented with MATLAB R2023a.

The proposed TD-CNND method involves three phases: training data preparation and CNN training, Tucker decomposition and SFRA, and spectral-spatial anomaly detection. For training data preparation, TD-CNND utilizes source domain hyperspectral data (Salinas) with labeled information to expand pixel pairs for learning spectral similarities and differences. In the phase of training, Adam optimizer is employed to optimize the CNN, with a learning rate of 0.001. The network is trained 50 epochs for both San Diego and Cuprite datasets, and 30 epochs for both Beach and Urban datasets, with a batch size of 256 for all datasets. During Tucker decomposition, all the 1×3 1-D convolution layers (except C1, C2, and C16) in the CNN are decomposed, with the ranks of Tucker decomposition determined by VBMF. The decomposed model is restored with the same optimizer and learning rate. To align spectral transfer features deflected by the decomposition process, SFRA module is employed with a small amount of data from the same dataset as the source domain, where "a small amount" is specifically defined as approximately 15% of the whole training set. For spectral-spatial anomaly detection, the inner and outer dual windows with sizes of (3, 15), (3, 5), (3, 5), and (5, 9) are used for the San Diego, Beach, Urban, and Cuprite dataset.

To evaluate the performance of the proposed TD-CNND, seven state-of-the-art methods are used for comparison, including traditional methods of RX, representation-based methods such as LRR, LRASR, CRD, GNLTR [45], as well as deep-learning-based methods such as RGAE, Auto-AD, and CNND. Among these methods, RX and CRD are regarded as the most classical methods for the HAD task, while the other methods have also gained widespread attention in recent literature and have shown competitive performance. And we empirically tune the parameters of all compared methods. Specifically, RX does not require any human-set parameters. According to the original paper of LRASR, the number of clusters and selection pixels of these datasets are set to 15 and 20, respectively, and the regularization parameters β and λ are set to 0.1 and 1, respectively. For CRD, the regularization parameter λ is set to 1e-6, then varies w_{out} and w_{in} from 5 to 25 and from 3 to 15, respectively. Finally, the optimal detection performance is achieved using specific window sizes (w_{out} , w_{in}). For the GNLTR model, four variables are supposed to be optimized to gain satisfactory detection results. The first two hyper-parameters θ and α are selected in {0.0005, 0.001, 0.01, 0.03, 0.05, 0.07, 0.1}. The other two are in connection with the nonconvex surrogate selection, which is chosen as capped L_1 . According to the original paper of RGAE and sizes of anomalies of datasets used, the tradeoff parameter λ , the number of super-pixels S, and the hidden layers dimension n_hid are set to 0.01, 150, and 100, respectively. In Auto-AD, the training process stops when the average variation of the loss falls below a certain threshold σ , which is varied from 1.0×10^{-5} to 1.5×10^{-5} to achieve the optimal detection performance. For the transfer learning-based CNND method, in order to ensure consistency of experiments, the settings of the training phase are the same as those of the proposed TD-CNND method.

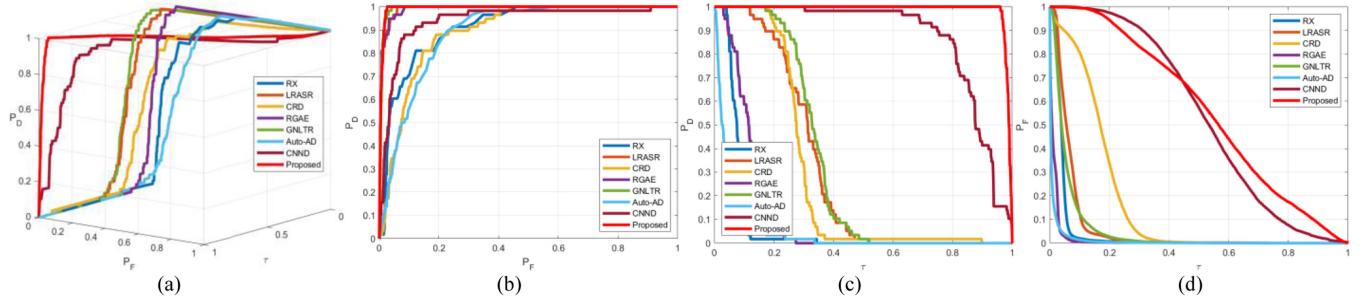


Fig. 9. ROC curves for SanDiego dataset. (a) 3-D ROC curve. (b) 2-D ROC curve of (P_D, P_F) . (c) 2-D ROC curve of (P_D, τ) . (d) 2-D ROC curve of (P_F, τ) .

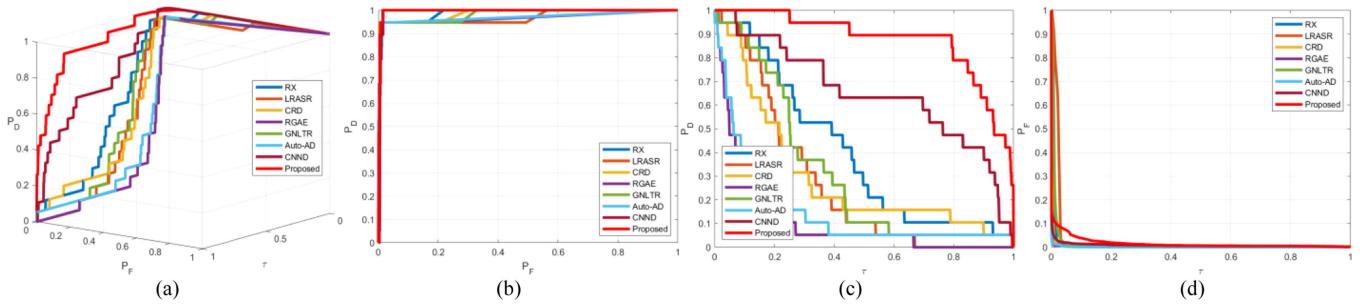


Fig. 10. ROC curves for Beach dataset. (a) 3-D ROC curve. (b) 2-D ROC curve of (P_D, P_F) . (c) 2-D ROC curve of (P_D, τ) . (d) 2-D ROC curve of (P_F, τ) .

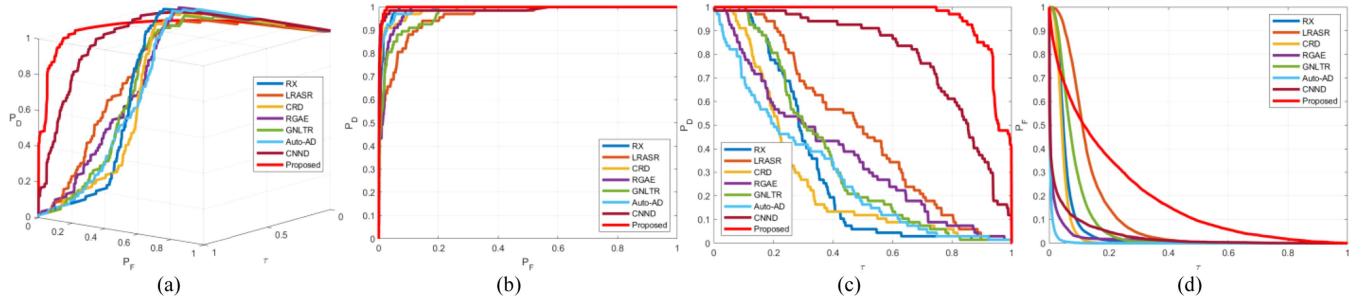


Fig. 11. ROC curves for Urban dataset. (a) 3-D ROC curve. (b) 2-D ROC curve of (P_D, P_F) . (c) 2-D ROC curve of (P_D, τ) . (d) 2-D ROC curve of (P_F, τ) .

D. Results and Analysis by Different Detectors

This section presents a comprehensive evaluation of the proposed TD-CNN method through a series of experiments, comparing it with several state-of-the-art methods. As mentioned before, the comparison methods include traditional methods of RX, representation-based methods such as LRASR, CRD, and GNLTR, as well as deep-learning-based methods such as RGAE, Auto-AD, and CNND. The assessment includes both visual analysis and quantitative metrics. The visual comparisons are shown in Figs. 5–8 by four HSIs, with pseudo-color images in (a), ground-truth in (b), and detection maps by different detectors in (c)–(j), respectively.

To overcome the subjectivity inherent in visual analysis, 3-D ROC curves and three corresponding 2-D ROC curves are employed for quantitative evaluation. The areas under the curve (AUC) for each 2-D ROC curve are provided, offering a nuanced understanding of method performance. The 3-D ROC curves, illustrating comprehensive method performance, consistently position the TD-CNN method's lines above other methods [see

Figs. 9–12(a)]. In addition, the AUC_{OD} values of the proposed TD-CNN surpass those of other methods across all datasets [see Tables I–IV], affirming its robust and effective detection capabilities. The 2-D ROC curve (P_D, P_F) focusing on anomaly detection ability [see Figs. 9–12(b)] showcases the proposed TD-CNN's superiority for all datasets. These results align with the $AUC(P_D, P_F)$ values presented in Tables I–IV. Furthermore, the preservation ability of anomalies is evaluated through 2-D ROC curves of (P_D, τ) [see Figs. 9–12(c)], demonstrating that TD-CNN consistently outperforms other methods across three datasets, corroborated by $AUC(P_D, \tau)$ values in Tables I–III. For Cuprite dataset, it performs close to the optimal result. The background suppression ability, as assessed by 2-D ROC curves of (P_F, τ) [see Figs. 9–12(d)], reveals relative shortcomings in the TD-CNN method, reflected in higher values of $AUC(P_F, \tau)$ compared to other methods. However, the detection maps show that while the other methods have excellent background suppression ability, they also significantly suppress some anomalous pixels, making it difficult to clearly observe anomalies for some datasets.

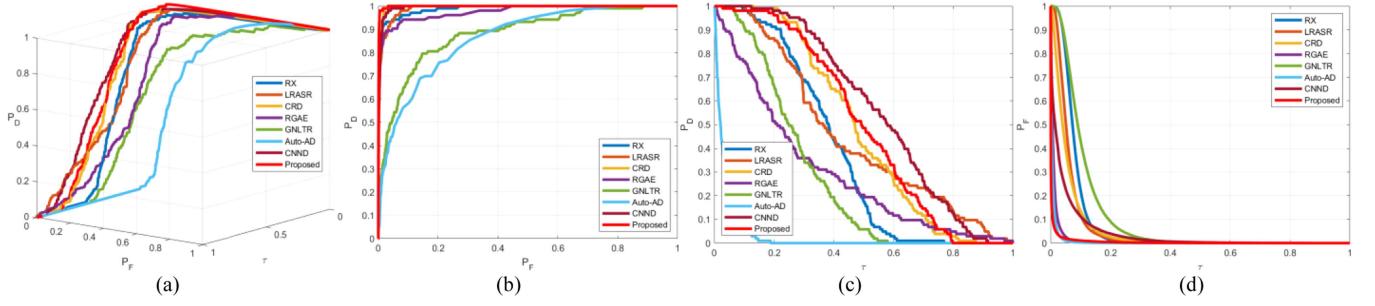


Fig. 12. ROC curves for Cuprite dataset. (a) 3-D ROC curve. (b) 2-D ROC curve of (P_D, P_F) . (c) 2-D ROC curve of (P_D, τ) . (d) 2-D ROC curve of (P_F, τ) .

TABLE I
ACCURACY COMPARISON OF DIFFERENT METHODS FOR SANDIEGO DATASET

Method	RX	LRASR	CRD	RGAE	GNLTR	Auto-AD	CNND	Proposed
AUC (P_D, P_F)	0.9118	0.9891	0.8932	0.9889	<u>0.9919</u>	0.8892	0.9389	0.9933
AUC (P_D, τ)	0.0790	0.3048	0.2853	0.1159	0.3306	0.0451	<u>0.8697</u>	0.9868
AUC (P_F, τ)	0.0406	0.0659	0.1667	<u>0.0155</u>	0.0566	0.0103	0.5370	0.5603
AUC _{OD}	0.9502	1.2280	1.0118	1.0894	1.2659	0.9240	<u>1.2716</u>	1.4198

The optimal performance are marked with bold, while the sub-optimal performance is marked with underline.

TABLE II
ACCURACY COMPARISON OF DIFFERENT METHODS FOR BEACH DATASET

Method	RX	LRASR	CRD	RGAE	GNLTR	Auto-AD	CNND	Proposed
AUC (P_D, P_F)	0.9880	0.9671	0.9838	0.9646	0.9798	0.9690	<u>0.9935</u>	0.9964
AUC (P_D, τ)	0.3990	0.2618	0.2935	1.1075	0.3109	0.1496	<u>0.6500</u>	0.8743
AUC (P_F, τ)	0.0111	0.0260	<u>0.0036</u>	0.0049	0.0125	0.0012	0.0092	0.0165
AUC _{OD}	1.3758	1.2029	1.2737	1.0672	1.2783	1.1174	<u>1.6342</u>	1.8542

The optimal performance are marked with bold, while the sub-optimal performance is marked with underline.

TABLE III
ACCURACY COMPARISON OF DIFFERENT METHODS FOR URBAN DATASET

Method	RX	LRASR	CRD	RGAE	GNLTR	Auto-AD	CNND	Proposed
AUC (P_D, P_F)	0.9906	0.9551	0.9871	0.9822	0.9687	<u>0.9910</u>	0.9896	0.9966
AUC (P_D, τ)	0.3112	0.4863	0.2718	0.3711	0.3601	0.2848	<u>0.7951</u>	0.9534
AUC (P_F, τ)	0.0555	0.1308	0.0458	<u>0.0178</u>	0.0815	0.0050	0.0329	0.1923
AUC _{OD}	1.2463	1.3105	1.2131	1.3355	1.2474	1.2708	<u>1.7518</u>	1.7577

The optimal performance are marked with bold, while the sub-optimal performance is marked with underline.

TABLE IV
ACCURACY COMPARISON OF DIFFERENT METHODS FOR CUPRITE DATASET

Method	RX	LRASR	CRD	RGAE	GNLTR	Auto-AD	CNND	Proposed
AUC (P_D, P_F)	0.9886	0.9904	<u>0.9962</u>	0.9769	0.8813	0.8638	0.9445	0.9983
AUC (P_D, τ)	0.3748	0.4500	0.4811	0.2901	0.2706	0.0325	0.5669	<u>0.4965</u>
AUC (P_F, τ)	0.0803	0.0615	0.0557	0.0107	0.1079	<u>0.0067</u>	0.0408	0.0055
AUC _{OD}	1.2831	1.3789	1.4216	1.2563	1.0440	0.8896	<u>1.5207</u>	1.4894

The optimal performance are marked with bold, while the sub-optimal performance is marked with underline.

Quantitative evaluations unveil nuanced insights. For the SanDiego dataset, TD-CNND shows the best comprehensive detection ability (AUC_{OD}), while exhibiting poor background suppression with bad AUC (P_F, τ) . Though LRASR and RGAE exhibit the most outstanding anomaly detection ability with higher AUC (P_D, P_F) , their highlighting ability of anomalies is deficient, as represented by AUC (P_D, τ) . The traditional RX detector struggles in complex scenes due to its reliance solely on Mahalanobis distance, ignoring the global information. For the Beach and Urban datasets, TD-CNND surpasses other methods in comprehensive detection ability, anomaly detection

ability, and anomaly highlighting ability, though exhibits poor performance in background suppression. The inherent limitations of traditional and representation-based methods become apparent in complex scenes, where they struggle to resist the influence of interfering information such as shadows and edges, resulting in the decline and instability of detection results. The CNND method expands training samples by subtracting spectral pixel pairs, leading to the drawback of losing intricate spectral details inherent in the original HSIs. In addition, the composition of spectral pixel pairs, involving the central pixel and its surrounding counterparts, poses challenges for the effective

TABLE V
VALUES OF DIFFERENT DUAL WINDOWS (W_{in} , W_{out}) FOR AUC (PD, PF)

(W_{in}, W_{out})	SanDiego	Beach	Urban	Cuprite
(3, 5)	0.9660	0.9964	0.9966	0.9976
(3, 7)	0.9793	0.9960	0.9964	0.9983
(3, 9)	0.9851	0.9962	0.9961	0.9987
(3, 11)	0.9880	0.9965	0.9951	0.9988
(3, 15)	0.9933	0.9959	0.9913	0.9988
(5, 7)	0.9794	0.9962	0.9943	0.9982
(5, 9)	0.9850	0.9963	0.9944	0.9987
(5, 11)	0.9878	0.9966	0.9934	0.9988
(5, 15)	0.9934	0.9959	0.9894	0.9988

adaptation of transfer learning knowledge to the target domain. The experimental analysis reveals that, though anomalies are detected using CNND, there is a notable deficiency in the separation between anomalies and backgrounds, leading to poor background suppression with complex background distribution in the detection maps. In contrast, the proposed TD-CNND, employing Tucker decomposition into the framework, exhibits more prominent anomalies compared to CNND. The introduction of Tucker decomposition into the network presents two advantages, eliminating redundancy in network parameters and deepening the network, enabling it to focus on more concealed details. Thus, the experimental results of TD-CNND outperform those of CNND in terms of detection performance. However, it is crucial to acknowledge that TD-CNND inherits certain design limitations from the original CNND, such as poor background suppression.

In summary, both visual analysis and quantitative evaluation consistently indicate the superior comprehensive detection ability of TD-CNND across various datasets with diverse scenarios. While there is a decline in performance with increasing background complexity, the TD-CNND exhibits robustness and efficiency, demonstrating its effectiveness in practical scenarios.

E. Parameters Analysis

In this section, we focus on the effect of dual windows (W_{in} , W_{out}) on detection performance AUC (PD, PF). Table V shows that AUC (PD, PF) varies for different datasets when the dual windows (W_{in} , W_{out}) take different values. For the SanDiego dataset, the detection performance of it is strongly influenced by its high background complexity. For the remaining three datasets, the impact on detection performance AUC (PD, PF) is less significant when the dual window (W_{in} , W_{out}) is different. When the disparity between the values of W_{in} and W_{out} is significant, the dual window gets more pairs of pixels. For the Cuprite dataset, when (W_{in} , W_{out}) are (3, 7) and (5, 7), the former has 753 448 more pixel pairs than the latter. Differing from both the AE and GAN-based methods indirectly detecting anomalies by reconstructing background, TD-CNND is a pixel-level detection method that directly detects pixels belonging to an anomaly or background. Therefore, the selection of the dual window (W_{in} , W_{out}) needs to consider the performance AUC (PD, PF) and time cost of detection.

TABLE VI
MODEL COMPLEXITY FOR DNN-BASED METHODS

Cuprite	Parameters/M	FLOPS/G	Running Time/S	AUC (PD, PF)
Auto-AD	3.2262	28.1454	45.3600	0.8638
CNND	3.5430	89.0302	54.7286	0.9945
TD-CNND	0.5803	15.2692	28.4618	0.9983

The optimal performance are marked with bold.

F. Model Complexity

In this section, we mainly evaluate TD-CNND in terms of efficiency. Since DNN-based and non-DNN methods are implemented on different hardware platforms and software, we mainly compare TD-CNND with CNND and Auto-AD for the sake of fairness. Table VI gives the model complexity and AUC (P_D , P_F) of the involved DNN-based methods, and the experimental data utilizes the Cuprite dataset. Compared with auto-AD and CNND, TD-CNND has the fewest parameters and the better performance. Due to the network compression scheme, TD-CNND has a huge advantage in terms of spatial memory. The running time of the unsupervised auto-AD method comprises the whole process. However, for the transfer learning-based methods CNND and TD-CNND, the models that contain transfer knowledge from source domains can be pretrained well in advance. Consequently, for the datasets captured by the same sensor, the running time can just comprise the detection process. And auto-AD, in which the background is reconstructed by the network while the anomalies appear as reconstruction errors, is an indirect detection method, whereas CNND and TD-CNND are direct detection methods. For CNND and TD-CNND, the probability of each pixel in the image belonging to the anomaly or background is calculated individually. Therefore, this section will only discuss the running time and FLOPS of CNND and TD-CNND. TD-CNND has fewer floating point operations (FLOPs) than CNND, which benefits from the help of Tucker decomposition-based network compression. Ignoring transfer learning, TD-CNND has higher detection accuracy than auto-AD, despite its longer running time. However, the advantages of transfer learning cannot be overlooked for practical purposes. Model training is delegated to other devices, where it can be used multiple times for various scenarios, leaving only the specific detection process for the detection device to complete. In summary, the TD-CNND exhibits superior detection performance and demonstrates lower model complexity.

G. Ablation Study

To validate the effectiveness of the proposed TD-CNND method, this section undertakes two ablation studies. The first aims to reduce computational costs by employing Tucker decomposition, while the second focuses on enhancing performance through SFRA.

1) *Impact of Network Compression by Tucker Decomposition:* To access the network compression by the proposed TD-CNND method, especially for large-scale HSI datasets anomaly detection, this section conducts two interrelated sets of ablation experiments.

TABLE VII
TIME COST OF PASSING THROUGH LAYERS WITH TUCKER DECOMPOSITION AND ORIGINAL LAYERS

Computational costs (<i>second</i>)	Conv2	Conv4	Conv6	Conv8	Conv10	Conv12	Conv14
Original Layer without Tucker Decomposition	0.1007	0.0689	0.0981	0.0606	0.0642	0.0472	0.0977
Layer with Tucker decomposition	First Layer	0.0483	0.0251	0.0311	0.0308	0.0286	0.0446
	Core Layer	0.0297	0.0009	0.0011	0.0025	0.0015	0.0009
	Last Layer	0.0022	0.0025	0.0011	0.0033	0.0011	0.0005
Total time cost with Tucker decomposition	0.0802	0.0285	0.0333	0.0366	0.0312	0.0462	0.0606

The optimal performance are marked with bold.

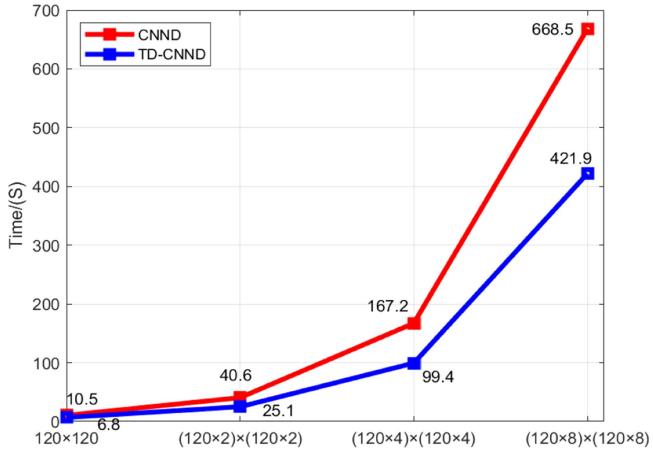


Fig. 13. Time cost for different sizes of datasets, where blue lines and red lines represent TD-CNND and CNND, respectively.

The first experiment compares the time required for the same size of data to pass through the convolutional layers with/without tucker decomposition, as detailed in Table VII. Notably, in Tucker decomposition, the first layer of the revised structure consumes a relatively significant amount of time compared to the core layer and last layer. This is attributed to the first layer's role in compressing the data to derive feature vectors for an easier processing. The results in Table VI indicate that the time required for the decomposed structures is shorter than that for the undecomposed convolutional layers. The computational cost reduction is expected to be more obvious with the increase in image size. Consequently, the network compression scheme proposed by the TD-CNND method holds significant advantages for large-scale HSIs.

To validate the hypothesis that the TD-CNND method is particularly advantageous for large-scale HSI datasets, another experiment is conducted on expanded HSI datasets from the San Diego dataset. The original San Diego dataset undergoes an enlargement operation with different factors, resulting in artificially generated datasets expanded by factors of 4 (obtained size of 240×240), 16 (obtained size of 480×480), and 64 (obtained size of 960×960), respectively. Fig. 13 depicts the computational costs associated with the original CNND and the proposed TD-CNND methods across datasets of different sizes. The red lines represent the time cost of CNND, while the blue lines represent the time cost of TD-CNND. The

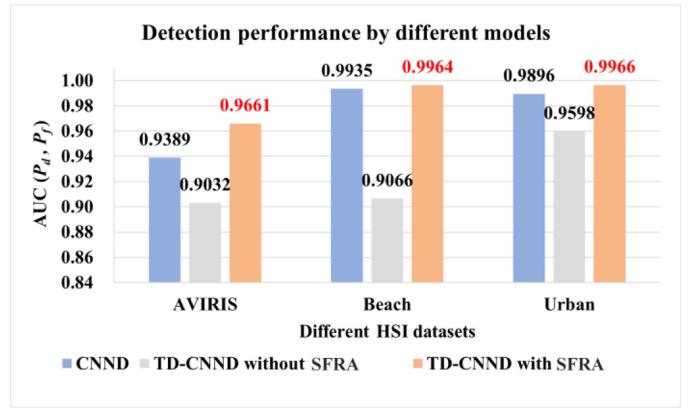


Fig. 14. Detection accuracy under different condition.

experimental results unequivocally demonstrate the superior computational efficiency of the proposed TD-CNND compared with the original CNND. Specifically, Fig. 13 illustrates that the computational time of TD-CNND consistently outperforms that of CNND for datasets of the same sizes, moreover, as the dataset size increases, the time reduction becomes more pronounced. In summary, the results lead to the conclusion that the proposed TD-CNND method is more efficient in reducing computational cost, especially for HSIs with larger spatial dimensions.

2) *Impact of Performance Improvement by SFRA*: To substantiate the efficacy of SFRA, comparative experiments are conducted on three real hyperspectral datasets, where the original CNND, the TD-CNND without SFRA, and the TD-CNND with SFRA are conducted for HAD. The area under curves (AUCs) serve as indicators of the detection performance for quantitative comparisons.

As illustrated in Fig. 14, the blue bar represents the detection performance of the traditional CNND model, the gray bar represents the detection performance of TD-CNND without SFRA, while the orange bar represents the detection performance of TD-CNND with SFRA as described in this section. The experimental results unequivocally demonstrate that the decomposition process for network compression leads to a degradation in detection performance, while the SFRA emerges as the pivotal role in recovering the detection accuracy of the decomposed network. In addition, it is evident that the performance of the TD-CNND method is comparable to or even surpasses that of

the original CNND method across different real hyperspectral datasets.

IV. CONCLUSION

To address the time cost challenges associated with deep learning-based methods for HAD with large-scale HSIs, this article presents a HAD method with Tucker decomposition. The proposed method optimizes a CNN-based framework through Tucker tensor decomposition for the purpose of network compression. Spectral pixel pairs are generated using labeled HSI from a source domain, and transfer learning is employed to train the CNN to discern feature differences among different spectra. Subsequently, Tucker decomposition is performed on selected convolution layers of the trained network, decomposing each 1-D convolution and replacing it with a structure of three-layer 1-D convolution. This process results in a new CNN structure devoid of network parameter redundancy. Finally, by employing sliding dual-windows, spatial information is combined and the under-test image of the target domain is sent into the new CNN with SFRA to accomplish anomaly detection. The comprehensive experimental results demonstrate the comparative advantage of the proposed TD-CNND in terms of quantity of parameters and detection accuracy, particularly in the time efficacy of HSIs with large spatial domains.

And something important writing at last, it is crucial to demonstrate why 1-D networks are not initially compressed using the Tucker decomposition and then trained with all the training samples on the source domain. The primary reason for applying Tucker decomposition after the initial training phase, rather than before, relates to the initial learning capabilities and integrity of the convolutional layers. In CNNs, especially when dealing with hyperspectral data, the convolutional layers are tightly connected to each other before decomposition, allowing the network to capture intricate spectral difference features more effectively, which is crucial for accurate anomaly detection.

REFERENCES

- [1] 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.
- [2] C. Zhou, Z. He, A. Lou, and A. Plaza, "RGB-to-HSV: A frequency-spectrum unfolding network for spectral super-resolution of RGB videos," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5609318.
- [3] B. Lu, P. Dao, and J. Liu, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, Aug. 2020, Art. no. 2659.
- [4] Y. Wang et al., "Constrained-target band selection for multiple-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6079–6103, Aug. 2019.
- [5] Y. Wang, X. Chen, F. Wang, M. Song, and C. Yu, "Meta-learning based hyperspectral target detection using siamese network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527913.
- [6] B. Tu, Z. Wang, H. Ouyang, X. Yang, J. Li, and A. Plaza, "Hyperspectral anomaly detection using the spectral-spatial graph," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5542814.
- [7] X. Yang, B. Tu, Q. Li, J. Li, and A. Plaza, "Graph evolution-based vertex extraction for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2023.3303273](https://doi.org/10.1109/TNNLS.2023.3303273).
- [8] B. Tu, X. Yang, W. He, J. Li, and A. Plaza, "Hyperspectral anomaly detection using reconstruction fusion of quaternion frequency domain analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: [10.1109/TNNLS.2022.3227167](https://doi.org/10.1109/TNNLS.2022.3227167).
- [9] C. Yu, B. Xue, M. Song, Y. Wang, S. Li, and C.-I. Chang, "Iterative target-constrained interference-minimized classifier for hyperspectral classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1095–1117, Apr. 2018.
- [10] Y. Wang, Q. Zhu, H. Ma, and H. Yu, "A hybrid gray wolf optimizer for hyperspectral image band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5527713.
- [11] Y. Wang, Q. Zhu, and M. Song, "A spatial-enhanced LSE-SFIM algorithm for hyperspectral and multispectral images fusion," *Remote Sens.*, vol. 13, no. 24, 2021, Art. no. 4967.
- [12] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [13] 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.
- [14] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [15] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.
- [16] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, pp. 1–12, Apr. 2008.
- [17] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [18] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [19] 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.
- [20] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [21] Y. Chen, W. Lai, W. He, X.-L. Zhao, and J. Zeng, "Hyperspectral compressive snapshot reconstruction via coupled low-rank subspace representation and self-supervised deep network," *IEEE Trans. Image Process.*, vol. 33, pp. 926–941, 2024.
- [22] 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.
- [23] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.
- [24] T. Guo, L. He, F. Luo, X. Gong, Y. Li, and L. Zhang, "Anomaly detection of hyperspectral image with hierarchical antinoise mutual-incoherence-induced low-rank representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510213.
- [25] Y. Wang, X. Chen, E. Zhao, C. Zhao, M. Song, and C. Yu, "An unsupervised momentum contrastive learning based transformer network for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 9053–9068, 2024.
- [26] X. Wang, L. Wang, Q. Wang, A. Vizziello, and P. Gamba, "Hyperspectral target detection via global spatial-spectral attention network and background suppression," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1283–1292, 2023.
- [27] L. Sun, Z. Ma, and Y. Zhang, "ABLAL: Adaptive background latent space adversarial learning algorithm for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 411–427, 2024.
- [28] Y. Wang, X. Chen, E. Zhao, and M. Song, "Self-supervised spectral-level contrastive learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510515.
- [29] C. Yu et al., "Hyperspectral image classification method based on CNN architecture embedding with hashing semantic feature," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1866–1881, Jun. 2019.
- [30] W. Li, G. Wu, and Q. Du, "Transferred deep learning for hyperspectral target detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5177–5180.

- [31] D. Wang, L. Zhuang, L. Gao, X. Sun, M. Huang, and A. J. Plaza, "PDB-SNet: Pixel-shuffle down-sampling blind-spot reconstruction network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5511914.
- [32] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511314.
- [33] S. Wang, X. Wang, L. Zhang, and Y. Zhong, "Auto-AD: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503314.
- [34] W. Xie, S. Fan, J. Qu, X. Wu, Y. Lu, and Q. Du, "Spectral distribution aware estimation network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512312.
- [35] L. Wang, X. Wang, A. Vizziello, and P. Gamba, "RSAAE: Residual self-attention-based autoencoder for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5510614.
- [36] P. Xiang, S. Ali, S. K. Jung, and H. Zhou, "Hyperspectral anomaly detection with guided autoencoder," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538818.
- [37] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, 1966.
- [38] L. De Lathauwer, B. De Moor, and V. Joos, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
- [39] Y.-D. Kim and S. Choi, "Nonnegative Tucker decomposition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
- [40] J. Ye, "Generalized low rank approximations of matrices," *Mach. Learn.*, vol. 61, no. 1–3, pp. 167–191, 2005.
- [41] S. Nakajima, M. Sugiyama, S. D. Babacan, and R. Tomioka, "Global analytic solution of fully-observed variational Bayesian matrix factorization," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 1–37, 2013.
- [42] S. Nakajima, R. Tomioka, M. Sugiyama, and S. D. Babacan, "Perfect dimensionality recovery by variational Bayesian PCA," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 971–979.
- [43] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
- [44] C.-I. Chang and J. Chen, "Orthogonal subspace projection using data spherling and low-rank and sparse matrix decomposition for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8704–8722, Oct. 2021.
- [45] H. Qin, Q. Shen, H. Zeng, Y. Chen, and G. Lu, "Generalized nonconvex low-rank tensor representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5526612.



Yulei Wang (Member, IEEE) was born in Yantai, Shandong Province, China, in 1986. She received the B.S. and Ph.D. degrees in signal and information processing from Harbin Engineering University, Harbin, China, in 2009 and 2015, respectively.

She was a joint Ph.D. student in Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County, Baltimore, MD, USA, in 2011–2013. From 2011 to 2013, she was a Research Assistant with the Shock, Trauma and Anesthesiology Research Organized Research Center (STAR-ORC), University of Maryland, School of Medicine. She is currently an Associate Professor and Doctoral Supervisor in Hyperspectral Imaging in Remote Sensing (CHIRS), Information Science and Technology College, Dalian Maritime University, Dalian, China. Her current research interests include hyperspectral image processing, multisource remote sensing fusion, and vital signs signal processing.



Hongzhou Wang was born in Shenyang, Liaoning, China, in 2000. He received the B.S. degree in communication engineering from the Information Science and Technology College, Dalian Maritime University, Dalian, China, in 2022, where he is currently working the M.S. degree in information and communication engineering.

His research interests include hyperspectral target detection and deep learning.



Enyu Zhao was born in Dalian, Liaoning Province, China, in 1987. He received the Ph.D. degree in cartography and geographic information system from the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing, China, in 2017.

He was a joint Ph.D. Student with Engineering Science, Computer Science and Imaging Laboratory, University of Strasbourg, Strasbourg, France, from 2014 to 2016. He is currently an Associate Professor with the College of Information Science and Technology, Dalian Maritime University, Dalian, China. His research interests include quantitative remote sensing and hyperspectral image processing.



Meiping Song received the Ph.D. degree from the College of Computer Science and Technology, Harbin Engineering University, Harbin, China, in 2006.

From 2013–2014, she was a visiting Associate Research Scholar with Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore County, Baltimore, MD, USA. She is currently an Associate Professor with the College of Information Science and Technology, Dalian Maritime University. Her research interests include remote sensing and hyperspectral image processing.



Chunhui Zhao received the B.S. and M.S. degrees from Harbin Engineering University, Harbin, China, in 1986 and 1989, respectively, and the Ph.D. degree from the Department of Automatic Measure and Control, Harbin Institute of Technology, Harbin, China, in 1998.

He was a Postdoctoral Research Fellow with the College of Underwater Acoustical Engineering, Harbin Engineering University. He is currently a Professor and a Doctoral Supervisor with the College of Information and Communication Engineering, Harbin Engineering University. His research interests include digital signal and image processing, mathematical morphology, and hyperspectral remote sensing image processing.

Dr. Zhao is a Senior Member of Chinese Electronics Academy.