Infrared Physics & Technology 143 (2024) 105623

Contents lists available at ScienceDirect

Infrared Physics and Technology

journal homepage: www.elsevier.com/locate/infrared

Breaking dimensional barriers in hyperspectral target detection: Atrous convolution with Gramian Angular field representations

Hongzhou Wang [a], Yulei Wang [a,*], Yuchao Yang [a], Enyu Zhao [a], Jian Zeng [b,*]

[a] Information Science and Technology College, Dalian Maritime University, Dalian 116026, China
[b] China Center for Resources Satellite Data and Application, Beijing 100094, China

ARTICLE INFO

Keywords:
Hyperspectral imagery
Target detection
Deep learning
Gramian angular field
Atrous convolution

ABSTRACT

Hyperspectral images contain extensive spectral bands with rich spectral information that reflects object properties. Leveraging state-of-the-art deep learning techniques has proven to be effective in hyperspectral target detection. However, compared to two-dimensional matrix data, the one-dimensional nature of spectral sequence limits the information that can be extracted, posing a challenge for deep learning-based hyperspectral target detection methodologies. To address this issue, a novel hyperspectral target detection method employing atrous convolution with gramian angular field representations is proposed in this paper. This approach breaks the barrier between one-dimensional vector and two-dimensional matrix by gramian angular field, transforming the spectral sequences from one-dimensional vectors into two-dimensional matrices, enabling the exploration of multidimensional relationships within spectral band relations through an atrous convolution-based spectral feature extraction network. The proposed model transcends the traditional one-dimensional spectral target detection limitations, offering a new perspective for spectral-based hyperspectral target detection. Experimental results on four real-world hyperspectral datasets demonstrate that the proposed method significantly outperforms existing state-of-the-art methods in detection performance, showcasing its potential for advancing hyperspectral target detection.

1. Introduction

Hyperspectral images (HSIs) capture the spectral information of objects across different wavelengths, providing more detailed and richer spectral features, enabling more accurate description of the spectral characteristics and material composition of object [1,2]. Unlike the RGB images, which are limited to three spectral bands, HSIs contain hundreds of bands, making them applicable in various applications, such as classification [3], target detection [4,5], and band selection [6,7], et al. Among them, hyperspectral target detection focuses on extracting and identifying specific targets within an HSI by exploring the unique spectral features of different targets, which allow for effective distinguish and detect targets through spectral analysis. In terms of spectral feature extraction, hyperspectral target detection utilizes the variations in reflectance of objects across different spectral bands. By analyzing the spectral responses of targets with different spectral bands, the relevant feature information related to targets can be extracted. Therefore, hyperspectral target detection has wide-ranging applications in many fields [8–11]. Unlike RGB images where targets are often tagged using bounding boxes, HTD focuses on the targets of interest in the scene to be detected [12,13]. It considers all other scene components to be detected except the object of interest as background. However, for HTD, the prior knowledge generally possessed is only the spectral features of the target of interest, and there is no information about the class labels in the scene to be detected. Moreover, phenomena such as the complexity of the background and the inherent variation of the spectra of the same substance pose difficulties for HTD The main challenge of HTD is to accurately identify and localize targets from complex backgrounds based on a priori target spectrum and to be able to suppress the background effectively.

To utilize the spectral characteristics of materials to detect targets of interest in HSIs, many HTD methods have been developed in the past decades. The Spectral Angle Mapper (SAM) [14] calculates the spectral angle between a prior target spectrum and each pixel within the HSI. The Spectral Information Divergence (SID) [15] detects targets by calculating the probabilistic difference between two spectral pixel vectors. These are the simple and straightforward target detectors. The method based on Constraint Energy Minimization (CEM) [16]

* Corresponding authors.
E-mail addresses: wangyulei@dlmu.edu.cn (Y. Wang), zengjian@cresda.com (J. Zeng).

https://doi.org/10.1016/j.infrared.2024.105623
Received 25 July 2024; Received in revised form 27 October 2024; Accepted 4 November 2024
Available online 13 November 2024

minimizes the impact of background by constructing the finite impulse response (FIR) filters and constraining the features of target to a specific gain, showing excellent performance in hyperspectral target detection. Subsequently, there have been some variant methods of CEM. The hierarchical CEM (hCEM) [17] method adopts a hierarchical structure of different layers of CEM detectors, maintaining target spectra and suppressing background spectra through the process of layer-by-layer filtering, where gradually improving the detection performance. Adaptive coherent\cosine estimator (ACE) [18,19] is a probabilistic statistics-based HTD method that assumes that the background conforms to a multivariate Gaussian distribution and detects the target adaptively from the background based on the covariance matrix of the HSI. And a method similar to ACE is Adaptive Matched Filter (AMF) detector [20], which also is a probabilistic statistics-based HTD method. Subspace-based hyperspectral target detection algorithms have also been proposed, such as the Orthogonal Subspace Projection (OSP) detector [21] proposed by Chein-I Chang. OSP is an HTD method based on subspace modeling, which detects targets by projecting the pixel spectra into orthogonal subspaces of individual background end-elements to suppress the interference of the background. However, the above methods ideally define the spectral features of target of interest with a single target spectrum or the target subspace, which may not always align with the complexity of real-world objects and result in the poor detection performance. The above HTD methods are extended to corresponding kernel-based nonlinear versions to utilize the nonlinear relationship between spectral bands such as kernel-CEM (KCEM) [22], kernel adaptive coherence estimation (KACE) [23], and kernel OSP (KOSP) [24]. Kernel-based HTD methods implicitly map the data into a high-dimensional kernel feature space by using the corresponding kernel function to make the data well separated. However, kernel-based HTD methods rely on the assumption that the transformation into the kernel feature space becomes linearly separable. Representation-based HTD methods have emerged to avoid making any explicit assumptions about the statistical distribution [25], such as the sparse representation-based target detector (STD) [26], combined sparse and co-sparse representation-based HTD method (CSCR) [27], and decomposition model for HTD based on background dictionary learning (DM-BDL) [28]. Although the representation-based methods can achieve good detection results, obtaining a pure background dictionary is difficult due to factors such as noise, and the optimal number of dictionary atoms may differ for different data [29], requiring human experience to be set, significantly limiting its ability to adapt to different scenarios.

The concept of deep learning stems from the exploration of artificial neural networks, initially proposed by Hinton *et al.* in 2006. This methodology integrates low-level features to construct more abstract, high-level features, thereby uncovering distributed representations of data features. The significant advancements achieved by deep learning-based network models grounded in the realm of RGB image processing have introduced the novel research avenues for hyperspectral image processing, including classification [30], band selection [31], images fusion [32], and target detection [33,34]. However, challenges in labeling hyperspectral data, scarcity of labeled data, and the imbalance of positive and negative samples, have hindered the progress in deep learning-based hyperspectral target detection. Several deep learning-based hyperspectral target detection methods have been proposed, demonstrating the potential of this technique. Some researchers have approached the problem from the perspective of transfer learning in the hope of overcoming the inability to train detectors in a supervised learning manner due to the scarcity of prior information, such as the convolutional neural network-based HTD (CNND) method [35], the *meta*-learning and Siamese network-based HTD (MLSN) method [36], and the transfer-learning-based HSI spectral–spatial joint target detection method [37]. However, using the idea of transfer learning to transfer the model knowledge learned on a dataset with known label information in the source domain to the target domain to be detected scene will be limited by the adaptability of the model knowledge to the detection scene, resulting in a poor target detection performance. In addition to employing transfer learning, some researchers have helped train models by synthetically expanding target and background samples. In this regard, the HTD-Net method in [38] adopts a U-net structure to design a modified auto-encoder to generate target signatures, and then find background samples based on linear prediction. Finally, the known target pixels are paired with both target and background pixels to augment the training samples. The two-stream CNN-based detector [39] finds enough background pixels by hybrid sparse representation and classification-based pixel selection, and then blends a prior target spectrum with some typical background pixels to generate sufficient target samples. Then, the generated target and background samples are, respectively, constructed with the prior target spectrum into positive and negative training samples to be expended and sent to the two-stream CNN to learn the spectral difference discrimination ability. In contrast to utilizing spectral differences from pixel pairs as training samples, this method preserves discriminative spectral features to a greater extent. While these methods achieve better detection performance, they rely heavily on supervised learning and data augmentation to address the challenges posed by insufficient labeled data and sample imbalance. Unsupervised learning methods, such as the generative adversarial network-based approach proposed by Xie et al [40], map the original hyperspectral data into a deep spectral feature space, ultimately performing target detection on this feature space. However, assuming normal data distribution limits their effectiveness in real-world scenarios. To fully exploit the global spectral features, a Siamese transformer target detector (STTD) [41] uses the transformer encoder to extract spectral features, and then the paired features were subtracted and fed into MLP for the final similarity scores. Additionally, the band selection-based [42] and the robust signature-based [43] hyperspectral target detection methods have also shown competitive performance. While these above-mentioned deep learning-based hyperspectral target detection methods exploit spectral sequence data's potential, achieving better detection performance and robustness than traditional methods, they still face several challenges:

1. Most deep learning models for hyperspectral target detection are adapted from the state-of-the-art natural image processing models. The distinctiveness of the spectral sequence data for HSI often leads to substituting traditional two-dimensional convolution for spatial information extraction with one-dimensional convolution for spectral information extraction. This adaptation results in a one-dimensional network model optimized for the spectral feature extraction. While this substitution via one-dimensional convolution has yielded promising detection performance for HTD, exploring the effectiveness of two-dimensional convolution for spectral feature extraction remains an open question.
2. Current methods primarily mine the spectral sequence data of one-dimensional vectors to acquire the higher-order features corresponding to spectral pixels for target detection. However, spectral sequence data, despite being continuous spectral curves, essentially constitutes a one-dimensional vector type of data, with a limited amount of information to be extracted compared with two-dimensional matrix-type data.
3. As is widely known, hyperspectral images contain a wealth of information in the spectral dimension with numerous highly correlated spectral bands. This, in turn, results in the presence of information redundancy within bands. Given the inability to fully address the redundancy, whether it is feasible to harness this redundancy to enhance the detection performance of hyperspectral target detection.

To address these challenges, this paper proposes an atrous convolution-based hyperspectral target detection method using Gramian Angular Field (GAF). Leveraging GAF, this method explores the multidimensional relationship among spectral bands, transforming spectral sequence data of one-dimensional vectors into band relation

map of two-dimensional matrices for target detection. Specifically, target spectral data is firstly obtained through modulation with a prior target spectrum and Gaussian white noise. Subsequently, GAF is employed to generate band relation maps of target samples, where the band relation maps of pure background samples are obtained by GAF and SAM detector to complete the preparation of training data. An atrous convolution-based spectral feature extraction network is designed to extract the spectral band features from the band relation maps, enabling a more accurate target detection with more feature information. Finally, the proposed method incorporates exponential function and normalization operation to effectively suppress background interference. Comparative experiments with state-of-the-art detection methods on four real-world hyperspectral datasets validate the feasibility and effectiveness of the proposed method. The main contributions of this paper are summarized as follows:

- The proposed approach addresses the limitations of one-dimensional vectors of spectral sequence data by mining multidimensional relationships between spectral bands through using GAF, transforming one-dimensional vector data of spectral sequence data into two-dimensional matrix data of band relation map. This approach enables hyperspectral target detection without being limited to one-dimensional spectral dimensions, and converts the redundancy of information between spectral bands into a natural advantage for improving detection performance.
- Given that the data for HTD is transformed from one-dimensional vector data to two-dimensional matrix data, it is imperative that the corresponding feature extraction network should also be transitioned from one-dimensional to two-dimensional network models. Consequently, a spectral feature extraction network based on atrous convolution is proposed to extract features of spectral bands relation maps for target detection.

2. Proposed method

The proposed model, as illustrated in Fig. 1, is divided into three primary steps: training samples preparation, atrous convolution neural network training, and target detection with background suppression. The training samples preparation is a crucial step and is mainly achieved by GAF and Gaussian white noise modulation. Once the training samples are prepared, the next step is to train the atrous convolution neural network, where the obtained target spectral augmentation samples and pure background spectral samples are fed into the atrous convolution neural network to extract the spectral features to achieve target detection. The final step involves detecting targets using the trained network and suppressing the background to improve detection accuracy, where the exponential function and normalization operation are used to achieve background suppression and obtain the final detection results. The following sub-sections describes the key techniques of the proposed approach.

2.1. Gramian Angular field

The Gramian Angular Field (GAF) [44] is a method to represent the one-dimensional sequence data as polar coordinates, which are then used to generate the two-dimensional matrix data. For spectral pixel vector, GAF transforms the one-dimensional spectral sequence into two-dimensional band relation maps, thus to maximally describe the multi-dimensional relationship between different spectral bands.

As shown in Fig. 2, the fundamental principle of GAF is to transform one-dimensional sequence data from the Cartesian coordinate system to the polar coordinate system. This is achieved through the unique definition of the inner product, which allows the information characterizing the correlation between the spectral bands to be obtained through the trigonometric sum/difference. The information is then tiled into the matrix from the upper left to the lower right in order, generating the GASF and GADF two types of matrix data. And this paper mainly uses the GADF matrix data.

2.1.1. Transformation process

Assume the one-dimensional spectral sequence data $\mathbf{X} = [x_1, x_2, \cdots, x_N]$ contains N spectral sequence vectors. Before mapping to the Cartesian coordinate system, it is necessary to scale the original data x_i to the range of [-1, 1] using Eq. (1):

$$\tilde{x}_i = \frac{x_i - \max(\mathbf{X}) + x_i - \min(\mathbf{X})}{\max(\mathbf{X}) - \min(\mathbf{X})} \tag{1}$$

The GAF matrix is defined by Eq. (2). The inner product between different vectors can be used to quantify the degree of correlation between them, while the angle between vectors indicates the strength of correlation between different vectors.

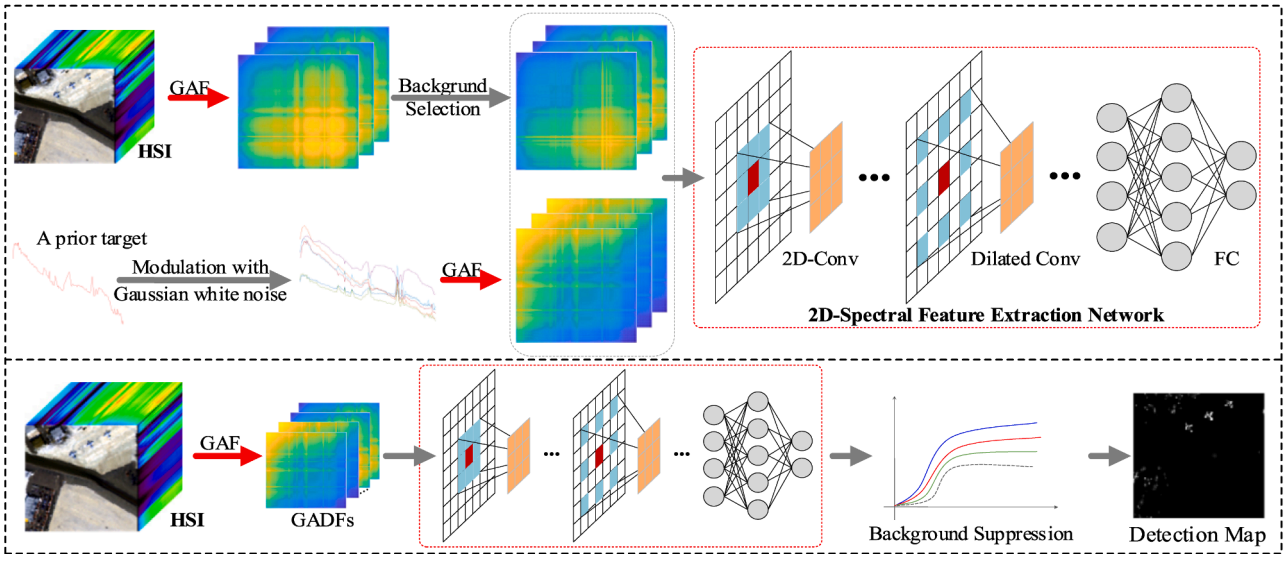

Fig. 1. Flowchart of the proposed method.

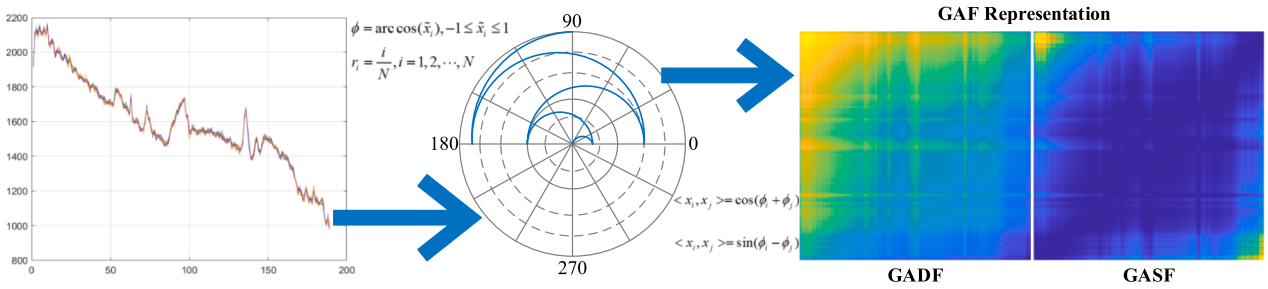

Fig. 2. Flowchart of the Gramian Angular Field representation.

$$G = \mathbf{X}^T\mathbf{X} = \begin{bmatrix} < x_1, x_1 > \cdots < x_1, x_n > \\ < x_2, x_1 > \cdots < x_2, x_n > \\ \vdots \\ < x_n, x_1 > \cdots < x_n, x_n > \end{bmatrix} \tag{2}$$

where G is the GAF matrix and $<\cdot>$ is the inner product operation.

Since the one-dimensional spectral sequence data are not inherently vectors, it is necessary to use the polar coordinate transformation to transform the spectral sequence data into vectors. The transformation formula is shown in Eq. (3).

$$\begin{aligned} &\phi = \arccos(\tilde{x}_i), -1 \le \tilde{x}_i \le 1 \\ &r_i = \frac{i}{N}, i = 1, 2, \cdots, N \end{aligned} \tag{3}$$

This mapping equation has two crucial properties. Firstly, it enables two-way mapping in accordance with the trigonometric function. This is demonstrated by the fact that $\cos(\phi)$ is monotonically decreasing when $\phi \in [0, \pi]$, thereby ensuring the uniqueness of its mapping to polar co-ordinates. Furthermore, the inverse mapping has uniqueness as well. Secondly, in polar coordinates, the absolute relationship between different bands is maintained. Consequently, GAF has defined two distinct forms of inner products with penalty terms designed to mitigate the impact of Gaussian noise. These are defined in Equations (4)-(5):

$$< x_i, x_j >= \cos(\phi_i + \phi_j) \tag{4}$$

$$< x_i, x_j >= \sin(\phi_i - \phi_j) \tag{5}$$

For the two different forms of definition of the inner product, by obtaining the sum or difference between each vector and presenting its correlation using a matrix, it is known as the GASF and GADF matrix data. The mathematical expressions for generating the GASF and GADF matrices based on the estimated values of the polar co-ordinates of each input are shown in Eqs. (6)–(7):

$$G_{\text{GASF}} = \begin{bmatrix} \cos(\phi_1 + \phi_1) \cdots \cos(\phi_1 + \phi_n) \\ \cos(\phi_2 + \phi_1) \cdots \cos(\phi_2 + \phi_n) \\ \vdots \\ \cos(\phi_n + \phi_1) \cdots \cos(\phi_n + \phi_n) \end{bmatrix} \tag{6}$$

$$G_{\text{GADF}} = \begin{bmatrix} \sin(\phi_1 - \phi_1) \cdots \sin(\phi_1 - \phi_n) \\ \sin(\phi_2 - \phi_1) \cdots \sin(\phi_2 - \phi_n) \\ \vdots \\ \sin(\phi_n - \phi_1) \cdots \sin(\phi_n - \phi_n) \end{bmatrix} \tag{7}$$

2.1.2. Application in hyperspectral target detection

In order to enhance the performance of HTD, it is crucial to utilize the relational features between different bands of pixel spectra. GAF achieves this by converting pixel spectra into spectral band relationship maps. These maps allow for a more comprehensive analysis of the spectral data, capturing the intricate relationships between spectral bands and improving the accuracy of target detection.

By leveraging GAF, the proposed method effectively transforms one-dimensional spectral sequence data into two-dimensional matrices, facilitating the extraction of valuable spectral features for hyperspectral target detection. This transformation is a key component of the atrous convolution-based network, enabling more accurate and robust target detection.

2.2. Training data preparation

Deep learning-based hyperspectral target detection often faces the challenge of insufficient labeled trainig data, apart from the priori target spectrum. To solve this ubiquitous problem, this section presents a two-step strategy for the training data preparation: target sample augmentation and GAF-based pure background samples acquisition.

2.2.1. Target sample augmentation

As shown in Fig. 3, to obtain a sufficient number of target samples, the known priori target spectra are modulated by Gaussian white noise with varying signal-to-noise ratios (SNRs).

€This approach simulates aberrant target spectra caused by different environmental factors. From the perspective of global information, it can be observed that the target samples obtained by modulation strictly obey the distribution of the target spectra. And the local positions of target samples exhibit varying degrees of variation. It is evident that this strategy of spectral data augmentation has two advantages. On the one hand, it avoids the situation in which there is a large spectrum shape difference between the image pixels used to expand the target samples and the prior target spectrum. On the other hand, it effectively simulates the differences between the target spectra while preserving the prior target spectral information.

2.2.2. GAF-based pure background samples acquisition

As shown in Fig. 4, To address the problem of impure background samples in spectral feature extraction networks, we extract pure background samples using a priori target spectrum, SAM detector and GAF.

Specifically, it is assumed that the hyperspectral image $\boldsymbol{X} \in^{W \times H \times l}$ contains l bands, and the corresponding spectral sequence data is $\widehat{\mathbf{X}} \in^{WH \times l}$. The priori target spectrum is denoted as $\mathbf{d} \in \mathbb{R}^{1 \times l}$. With GAF, the priori target spectrum $\mathbf{d}$ and the spectral sequence data $\widehat{\mathbf{X}}$ are transformed into the corresponding two-dimensional matrix data $\mathbf{d}_{\text{GADF}} \in^{1 \times l \times l}$ and $\mathbf{X}_{\text{GADF}} \in^{WH \times l \times l}$, respectively. Then, each feature map of $\mathbf{X}_{\text{GADF}}$ and $\mathbf{d}_{\text{GADF}}$ are fed into SAM detector to get the correlation coefficient of the priori target spectrum and each spectral pixel of the hyperspectral image. And these correlation coefficients form a correlation map w of the spectral pixels of the hyperspectral image with the priori target spectrum. Finally, it is necessary to scale w to $\tilde{w}$ with the range of [-1, 1] by Eq. (1). The values of $\tilde{w}$ are sorted from smallest to largest, and the spectral pixels corresponding to the first 20–30 % of the sorted values are taken as the pure background samples.

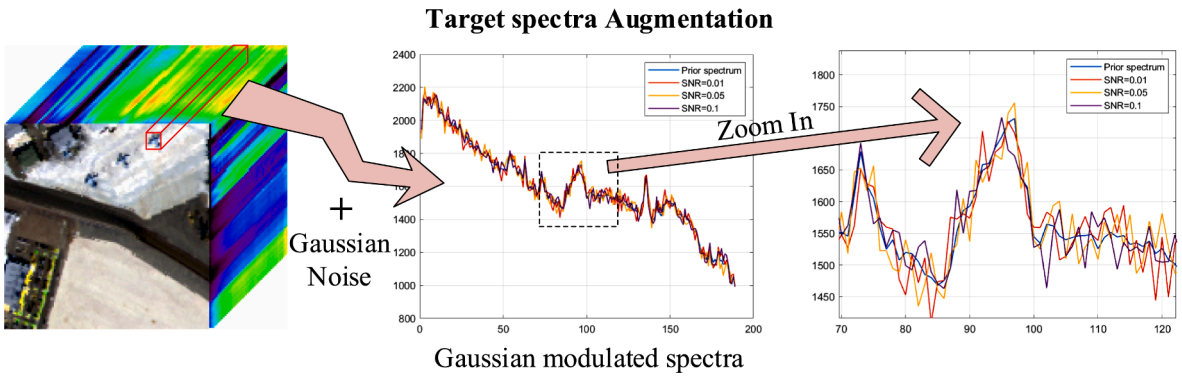

Fig. 3. Flowchart of the target sample augmentation.

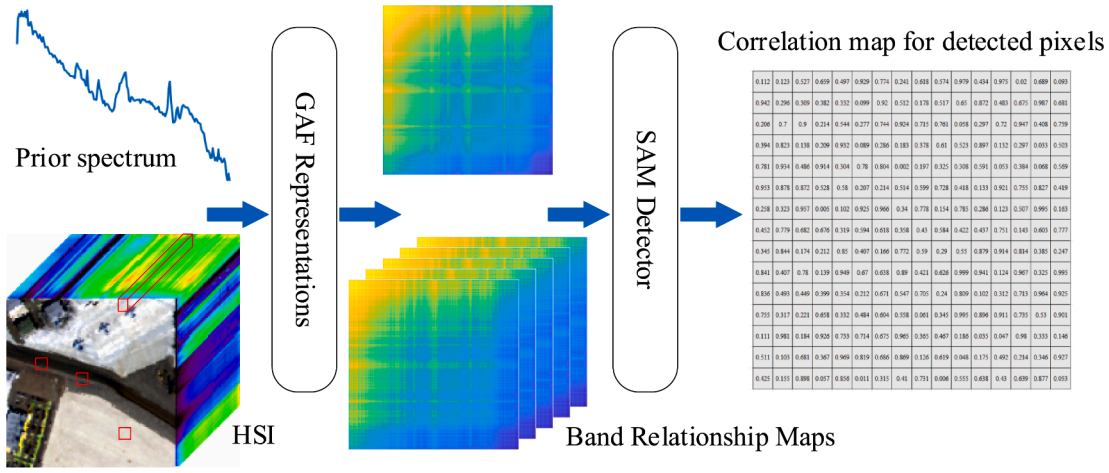

Fig. 4. Flowchart of the GAF-based pure background samples Acquisition.

2.3. Atrous convolution neural network training

Once the one-dimensional spectral sequence has been converted into two-dimensional band relation maps, it is necessary to transit the spectral feature extraction network from a one-dimensional convolution-based model to a two-dimensional convolution-based model. Consequently, a large number of high-performance feature extraction networks in the domain of natural image processing can be readily transferred to deep learning-based hyperspectral target detection methods. However, models that concentrate on the extraction of spatial information about target contours are also not particularly suitable for HTD. For instance, the VGG16 network results in a significant reduction for feature map size due to the utilization of numerous MaxPooling layers for down sampling the feature maps. And it results in the loss of some detailed information from the feature maps, which is irreducible to the lost information and targets. This can ultimately lead to unsatisfactory performance for HTD, which is essentially a binary classification problem. At the same time, the removal of the MaxPooling layer in a straightforward and unsophisticated way will result in the feature maps corresponding to the sensory fields of the original maps becoming much smaller, which will consequently impede the convolution from acquiring the deeper information. Therefore, considering that the two-dimensional image converted from the one-dimensional spectral sequence mainly describes the relationship between spectral bands rather than the spatial information of target contours, atrous convolution [45] is introduced to solve the above problem.

2.3.1. Atrous convolution

The atrous convolution (dilated convolution) was originally developed in the algorithm of wavelet decomposition [46]. The main idea of atrous convolution is to insert "holes (zeros)" between pixels in convolutional kernels to increase the resolution, thus enabling dense feature extraction in deep CNNs. Specifically, the atrous convolution is a special two-dimensional convolution that obtains the relatively large receptive field without reducing the image size, and its main advantage is that it allows flexible resizing of the receptive field to capture the multi-scale information and improve the performance of target classification and semantic segmentation. The two-dimensional atrous convolution operator can be defined in Eq. (8):

$$g_{i,j}(x_{\mathrm{GADF}}) = \sum \theta_{k,r}^{i,j} * x_{\mathrm{GADF}} \tag{8}$$

where $g_{i,j}$ is the convolution operation of the feature map, $*$ denotes the convolution operator, x_{GADF} is the feature map and $\theta_{k,r}^{i,j}$ is the atrous convolution kernel of size k and dilation rate $r \in Z^+$. For the atrous convolution, the convolution kernel size k is increased to $k+(k\text{-}1)\cdot(r\text{-}1)$, and when $r = 1$, it is equivalent to the two-dimensional standard convolution. The feeler field of a standard convolution is related to the convolution kernel size and step size of all convolutional layers preceding that layer in the network. Its feeler field grows linearly. In contrast, the feeler field of the atrous convolution is $(2^{r+1}\text{-}1)\times(2^{r+1}\text{-}1)$, a cascade of the atrous convolutions results in exponential growth of the feeler field, such that each convolutional output contains more information.

However, according to the description in [47], a fundamental issue arises in the dilated convolution, which is identified as "gridding," As shown in Fig. 5. As zeros are padded between two pixels in the convolutional kernel, the receptive field of this kernel only covers an area

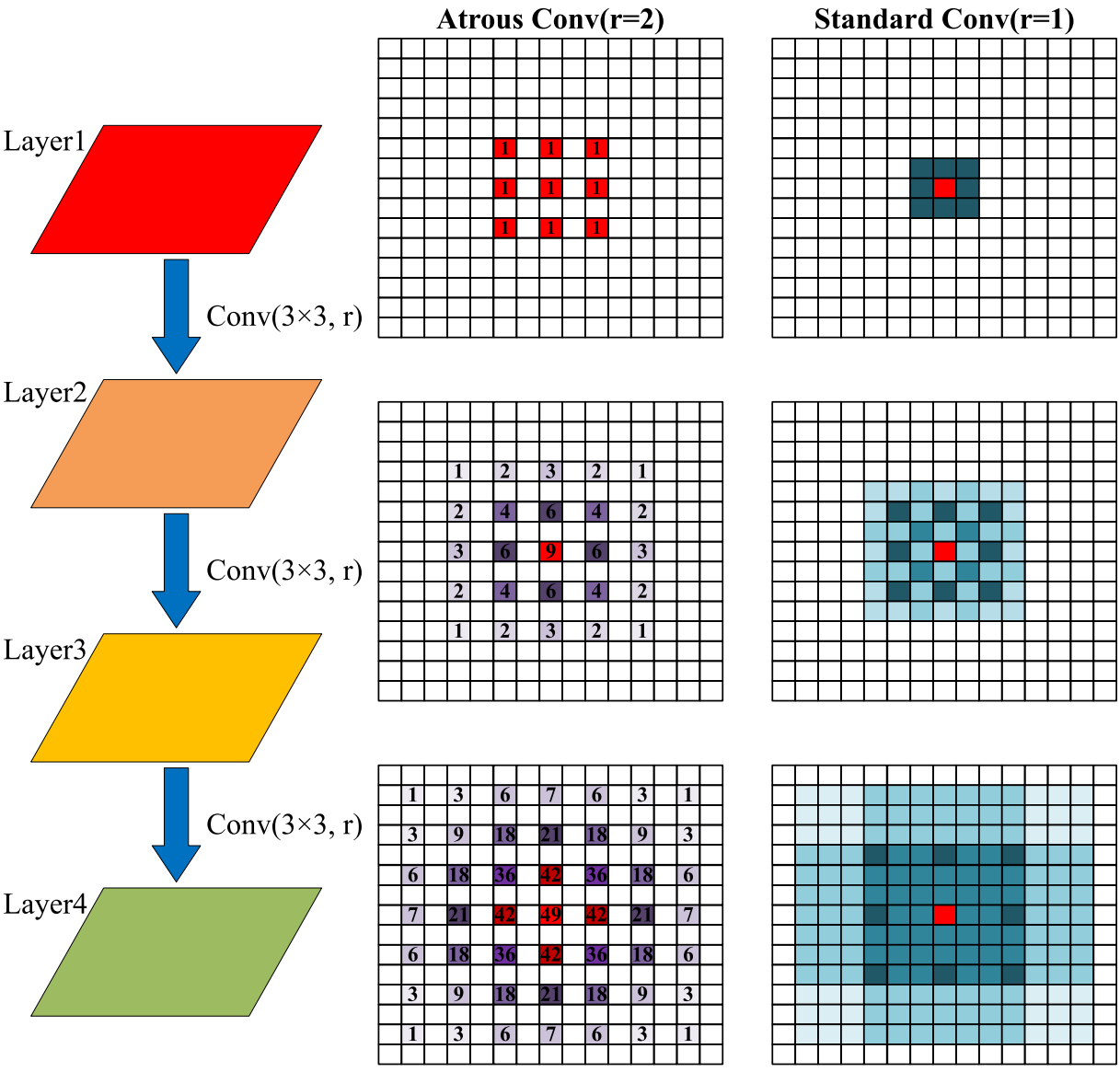

Fig. 5. The schematic diagram of "gridding".

with checkerboard patterns (only locations with non-zero values are sampled), resulting in the loss of some neighboring information. The issue becomes more pronounced as the rate of dilation increases, particularly in higher layers where the receptive field is extensive. The convolutional kernel becomes insufficiently dense to encompass local information, as the non-zero values are too widely dispersed. Information that contributes to a fixed pixel is always derived from the predefined gridding pattern, which results in the loss of a significant portion of information. Therefore, Wang *et al.* proposed a criterion called HDC to solve the problem of "gridding." Specifically, the goal of HDC is to ensure that the final size of the RF of a series of convolutional operations fully encompasses a square region without any holes or missing edges. The "maximum distance between two nonzero values" is defined in Eq. (9):

$$M_i = \max[M_{i+1} - 2r_i, M_{i+1} - 2(M_{i+1} - r_i), r_i] \tag{9}$$

where $M_n = r_n$. The goal is to let $M_2 \leq K$. For example, for kernel size $K = 3$, and $r = [1,2,5]$ pattern works as $M_2 = 2$; however, $r = [1,2,9]$ pattern does not work as $M_2 = 2$. Thus, for the HDC criterion, the assignment of dilation rate follows a sawtooth wave-like heuristic. This involves the grouping of several layers together to form the "rising edge" of the wave, which has an increasing dilation rate. The next group then repeats the same pattern. To illustrate, for all layers with a dilation rate of $r = 2$, three subsequent layers are grouped together and their dilation rates are altered to 1, 2, and 3, respectively. This approach enables the top layer to access information from a more extensive range of pixels within the same region as the original configuration.

Another benefit of HDC is that it can use arbitrary dilation rates through the process, thus naturally enlarging the receptive fields of the network without adding extra modules, which is important for recognizing objects that are relatively big. One important thing to note, however, is that the dilation rate within a group should not have a common factor relationship (like 2,4,8, etc.), otherwise the gridding problem will still hold for the top layer.

2.3.2. Atrous convolution neural network

Following the acquisition of two-dimensional band relation maps of target spectral augmentation samples and pure background samples, spectral feature extraction is conducted utilizing an atrous convolutional neural network.

The network is a feed-forward network, which encompasses two-dimensional standard convolutions, atrous convolutions, fully con-

nected layers, and activation function. It can extract high-level features between spectral bands through connections between neighboring layer neurons. As shown in Fig. 6, the mainly network includes 11 learnable two-dimensional convolutional layers. In the designed framework, the atrous convolution layers, namely C1, C2, C3, C4, C5, C6, C7, C8 and C9, are configured with a stride of 1, while the two-dimensional standard convolution layers P1, P2 and P3 replace the MaxPooling layer to better maintain the information between spectral bands and prevent the loss of important information. The convolution kernel size for the aforementioned 10 convolutional layers is set to 3 × 3, and the final learnable two-dimensional convolutional layer is a 1 × 1 two-dimensional standard convolutional layer. As shown in Fig. 6, it is worth noting that in order to avoid the problem of "gridding", we combine C1, C2, C3, C4, C5, C6, and C7, C8, C9 into 3 atrous convolution block with different dilation rates, respectively. For each atrous convolution block, the dilation rates of the first, middle and last layers are 1, 2 and 3 respectively, thus better following the HDC. The sigmoid activation function is served as the final layer of network to produce an output representation in terms of scores (or labels), deriving the probability that a given pixel belongs to target. Consequently, the optimization function employed for target detection is the Binary Cross Entropy (BCE), as expressed in Eq. (10):

$$Loss_{BCE} = -\frac{1}{B}\sum_{i=1}^{B}[y_i \cdot \log f_i + (1-y_i)\cdot \log(1-f_i)] \tag{10}$$

2.4. Target detection & background suppression

After training the atrous convolution neural network, each spectral pixel of the hyperspectral image is converted into a two-dimensional relationship map by GAF, which is then fed into the atrous convolution neural network to obtain the spectral detection result. This process utilizes the spectral information for target detection. This is followed by a further movement of the value of the background pixels away from the value of the target pixels by an exponential function and normalization operation, with the purpose of background suppression. The background suppression is achieved through the application of exponential function and normalization operation. The process can be formalized in Eq. (11):

$$S = \alpha^{B} \tag{11}$$

where B is the spectral detection result, S is the final detection result, and α is positive parameter for adjusting the background suppression performance, respectively.

3. Results and discussion

In this section, a comprehensive set of experiments is conducted on four real hyperspectral datasets to validate the effectiveness of the proposed method in terms of detection performance.

3.1. Hyperspectral datasets

SanDiego Dataset: The SanDiego dataset, collected by AVIRIS at the SanDiego Airport area, CA, USA, exhibits a spatial resolution of 3.5 m and a spectral resolution of 10 nm. It has 400 × 400 pixels with 224 bands 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. In the experiment, a image of size 120 × 120 was captured from the center of the SanDiego dataset. They are named SanDiego1 and SanDiego2, respectively. The pseudo-color image and corresponding ground truth map are shown in Fig. 7(a) and (b). The aircraft in the image scene is treated as target for detection and contained 58 target pixels.

Urban Dataset: The Urban dataset was captured by AVIRIS sensors off the coast of TX, USA, with a spatial resolution of 17.2 m. It has 100 × 100 pixels, and after removing the low signal-to-noise band the remaining 204 bands. The pseudo-color image and corresponding ground truth map are shown in Fig. 8(a) and (b). A total of 67 pixels are considered as targets for detection.

Beach Dataset: The Beach dataset is captured by the AVIRIS sensor on Cat Island with a spatial resolution of 17.2 m. In the experiment, the image of 90 × 90 × 188 size is obtained after removing the noise band. The pseudo-color image and corresponding ground truth map are shown in Fig. 9(a) and (b), including 19 anomaly points.

HYDICE Dataset: The HYDICE dataset is collected by HYDICE sensors at the urban area in California, USA, with the spectral resolution is 10 m. The whole image has a total of 307 × 307 pixels with a total of 210 bands, and the wavelength is from 400 nm to 2500 nm. In the experiment, we remove the band affected by dense water vapor and atmosphere, and intercept the scene with size of 80 × 100 × 175 for detection. Its pseudo-color image and corresponding ground truth map are shown in Fig. 10(a) and (b), including 21 target pixels of the types of roofs and cars.

3.2. 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

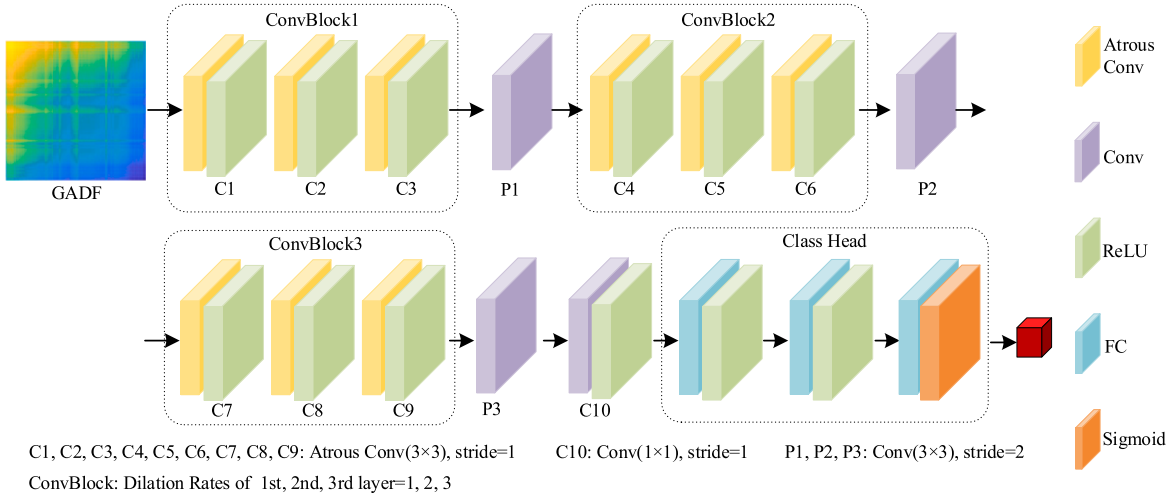

Fig. 6. The framework of atrous convolution neural network.

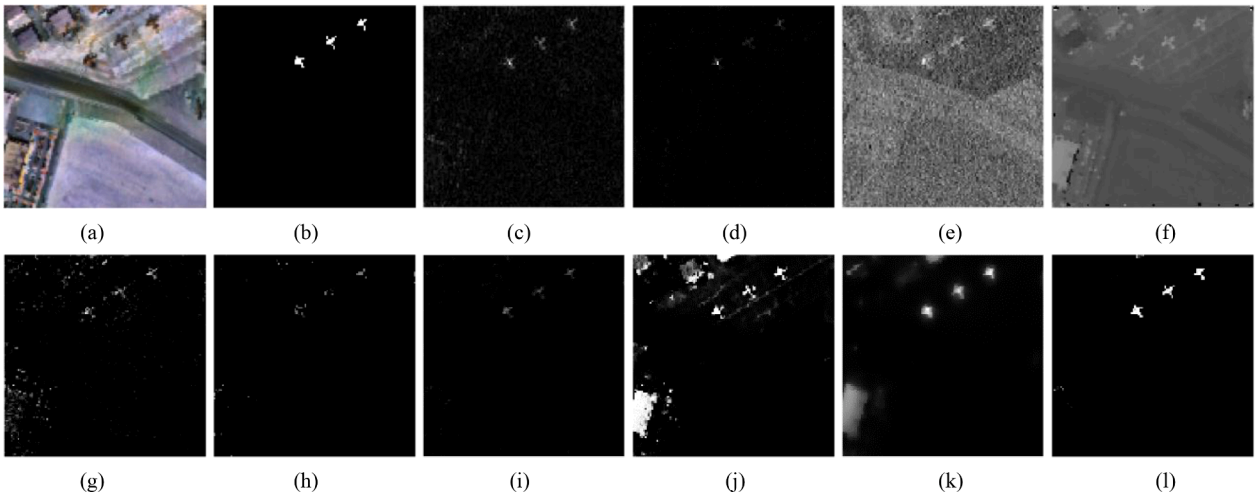

Fig. 7. Detection maps for the SanDiego dataset. (a) Pseudo-color image (b) Ground truth (c) CEM (d) ACE (e) OSP (f) CSCR (g) DM-BDL (h) CNND (i) BLTSC (j) STTD (k) SCLHTD (l) Proposed.

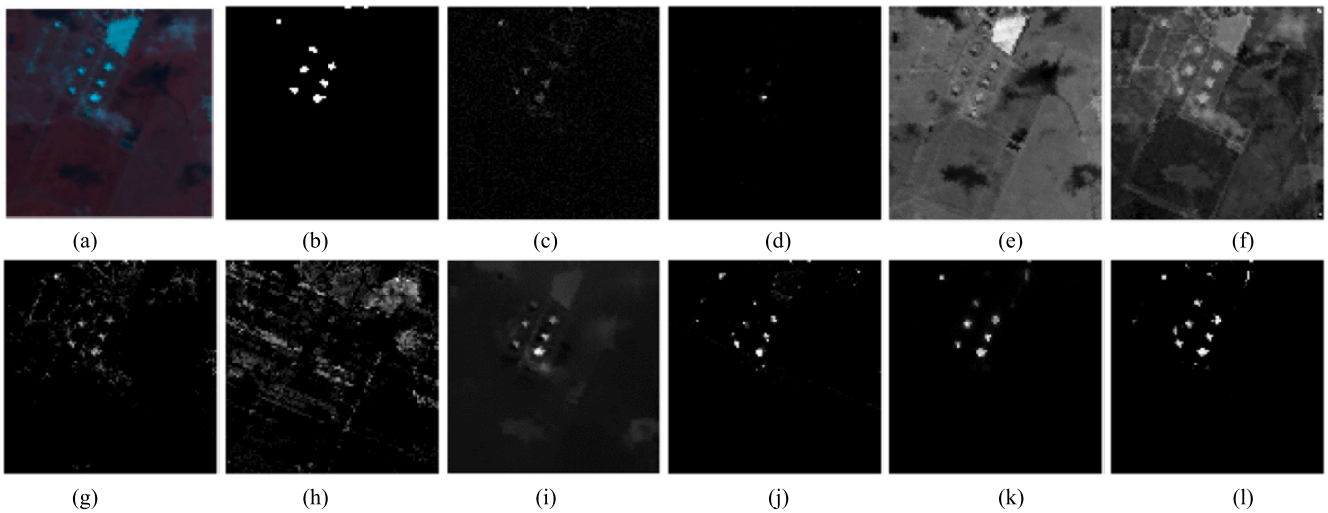

Fig. 8. Detection maps for the Urban dataset. (a) Pseudo-color image (b) Ground truth (c) CEM (d) ACE (e) OSP (f) CSCR (g) DM-BDL (h) CNND (i) BLTSC (j) STTD (k) SCLHTD (l) Proposed.

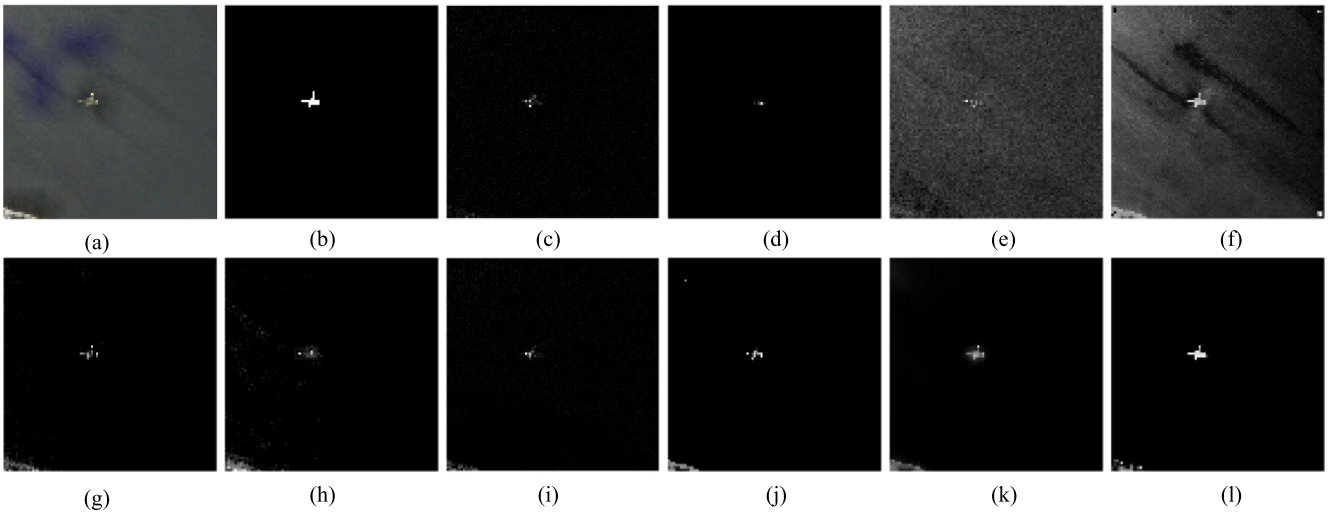

Fig. 9. Detection maps for the Beach dataset. (a) Pseudo-color image (b) Ground truth (c) CEM (d) ACE (e) OSP (f) CSCR (g) DM-BDL (h) CNND (i) BLTSC (j) STTD (k) SCLHTD (l) Proposed.

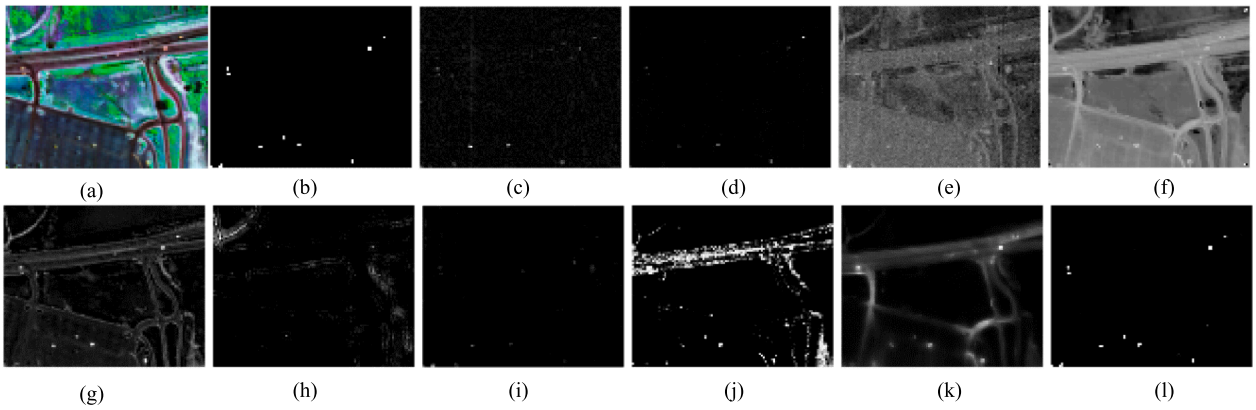

Fig. 10. Detection maps for the HYDICE dataset. (a) Pseudo-color image (b) Ground truth (c) CEM (d) ACE (e) OSP (f) CSCR (g) DM-BDL (h) CNND (i) BLTSC (j) STTD (k) SCLHTD (l) Proposed.

under the curve (AUC) [48,49]. 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 Eqs. (12)–(13), respectively:

$$P_D(\tau) = \frac{n_{D,\tau}}{n_{D,\tau} + n_{FN,\tau}} \tag{12}$$

$$P_F(\tau) = \frac{n_{F,\tau}}{n_{F,\tau} + n_{TN,\tau}} \tag{13}$$

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 paper uses 3D ROC curve [48] as the evaluation standard, and three 2D 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 2D 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 area under the curve of the 2D 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 area under the curve of the 2D ROC curve ($P_{F,}$ τ), measuring the ability of the background suppression, with smaller values indicating better suppression of the background. And the background suppression capability quantitative index AUC_{BS}, with a range of [-1,1], which combines the probability of detection and the probability of false alarm to thoroughly measure the background suppression capability of the detector, is defined as:

$$AUC_{BS} = AUC(P_D, P_F) - AUC(P_F, \tau) \tag{14}$$

The larger value of AUC_{BS} indicate the better background suppression for detectors. Besides, a new quantitative detection index designed in [48] takes the three AUC values as a whole to measure the total performance, named as AUC_{OD}, with a range of [-1,2], which is defined as:

$$AUC_{OD} = AUC(P_D, P_F) + AUC(P_D, \tau) - AUC(P_F, \tau) \tag{15}$$

3.3. Experimental Setup

To evaluate the performance of the proposed method, several existing state-of-the-art detection methods are compared, including five tradition methods and four deep learning-based methods. The traditional methods include the classical detection method CEM, the statistical distribution model-based detection methods ACE, the subspace model-based detection method OSP, the representation-based detection methods CSCR and decomposition model with background dictionary learning (DM-BDL). The deep learning-based methods include the transfer learning-based detection method (CNND), the background learning-based method (BLTSC), the siamese transformer network-based detection method (STTD) and the self-supervised spectral-level contrastive learning-based method (SCLHTD).

CEM and ACE do not have any parameters that need to be set artificially. For CSCR, the outer and inner windows sizes are (7, 3), (11, 3), (11, 3), and (3, 11) for the Sandiego, Urban, Beach, and HYDICE datasets, respectively. The regularization parameters λ_1 and λ_2 are set to 10^{-1} and 10^{-2} for all datasets in the experiment. The decay parameter in the DM-BDL detector was set to 0.982 for all datasets in the experiment, and the other parameters followed the settings in the original paper. For the transfer learning-based CNND detector, the training set is constructed by subtracting the spectra of similar pixels and subtracting the spectra of different classes of pixels using the dataset with known labels captured by the corresponding sensor when training the deep CNN. For all datasets in the experiment, the learning rate, batch size, and epoch of the CNND method during training are set to 10^{-3}, 256, and 50, respectively. For the background learning-based method BLTSC method, it only uses the background training samples, the coarse detection is performed using the classical CEM method to grain the sufficient background training data. It uses the learning rate and epoch set to 1e-4 and 500, during training for four real datasets in this experiment, respectively. For the Siamese transformer network-based detection method STTD, the threshold of extracting background pixels on abundance maps is set to 0.75. The threshold of the pre-detection filter and the spectral-angle-based filter are set to 0.975 and 0.375, respectively. For all datasets in the experiment, the learning rate and batch size during training are set to 10^{-4} and 100, respectively. For the SCLHTD method, the total parameters follow the settings in the original paper for all four real datasets in this experiment.

The proposed method is broadly divided into three steps: training data preparation, atrous convolution neural network training, and target detection with background suppression. The acquisition of target samples for the training data is mainly realized by GAF and Gaussian white noise modulation, which results in the band relation maps corresponding to target spectral augmentation samples. For the acquisition of pure background spectra, all the spectral pixels of the hyperspectral image

are converted into the corresponding band relation maps by the GAF, and then the pure background samples are obtained by using the SAM method and the characteristics of the data distribution with a small threshold. During the training of the atrous convolution neural network, the learning rate of all four HSI datasets are set to 10^{-4}. The epoch and the batch size are set to (100, 256), (100, 256), (150, 256) and (150, 256) for the Sandiego, Urban, Beach and HYDICE datasets, respectively. Regarding the background suppression process for all datasets in the experiment, the exponential function set α to 2e3, 4e3, 26e1, and 3e1 for Sandiego, Urban, Beach, and HYDICE datasets, respectively.

The experimental hardware environment includes an AMD Ryzen Threadripper 3990X 64-core CPU and Quadro RTX 8000 48 GB GPU. The implementation of both the proposed method and the deep learning-based method is carried out using Python 3.8.0, PyTorch 1.12, and MATLAB R2022a. And the other traditional comparison methods are carried out using MATLAB R2022a.

3.4. Results and analysis comparison by different methods

3.4.1. Subjective Assessment of detection results

As described in section 3.3.2, nine state-of-the-art detection methods are used for comparison in the Experiments to verify the effectiveness of the proposed method. Figs. 7-10 show the detection maps by the above methods and the proposed method for the SanDiego, Urban, Beach, and HYDICE datasets.

It can be seen from the detection maps that CEM, ACE, CNND and BLTSC miss many target pixels, and they have very low tolerance for target spectral variations. This is due to the strong non-Gaussianity and nonlinearity of the real scenes for HSIs, where leading to a decrease in target detection accuracy of CEM and ACE. OSP and CSCR can detect the most of targets, but there is poor background suppression and small separation between target and background, where resulting in the inability to visually identify targets, and the detection performance decreases when the background of the scene for detecting becomes more complex. Compared with CSCR and OSP, the target for the detection map of DM-BDL is detectable and has relatively good background suppression, but it requires more prior target spectra to improve the detection accuracy. The CNND method expands the training samples for training by pairing pixels of the similarity class and pairing pixels of dissimilarity classes based on known labelled classified datasets of the corresponding sensor for each dataset, which enables the deep CNN to learn spectral difference for target detection. Since the spectral pairing is performed by the difference of the pixels, it leads to the loss of detailed spectral information of the original image, and the transfer knowledge is not so well adapted to the detection task in the target domain, which makes many target pixels are not detected. The STTD and SCLHTD method detect most of targets, but also have higher false detection rates and poor background suppression for datasets with more complex background. And the performance of SCLHTD is limited by the quality of the prior target spectra used for detecting. The proposed method shows excellent detection performance with both high target detection accuracy, good background suppression, and visually obvious identification of target in the detection maps obtained on four real datasets, showing the best performance among all comparable methods.

3.4.2. Quantitative Assessment of detection results

Subjective evaluation of the detection maps visually has limitations, and to quantitatively evaluate the performance of the proposed method, 3-D ROC curves and their corresponding 2-D ROC curves ($P_{D,}$ P_F), ($P_{D,}$ τ), and ($P_{F,}$ τ) with the AUCs of ($P_{D,}$ P_F), ($P_{D,}$ τ), and ($P_{F,}$ τ) are used for quantitative evaluation. The 2-D ROC curve of ($P_{D,}$ P_F) is used to demonstrate the effectiveness of different methods, as shown in Figs. 11–14(b). For the four real HSI datasets in the experiment, the red curve is the ROC curve of the proposed method, which outperforms the curves of other comparison methods. The 2-D ROC curve of ($P_{D,}$ τ) is used to evaluate the preservation ability of the method for the target. As shown in Figs. 11–14 (c), the proposed method outperforms other methods for all datasets. For the 2-D ROC curve of ($P_{F,}$ τ), which evaluates the background suppression ability, as shown in Figs. 11–14 (d), the proposed method has a significantly better background suppression ability than other methods, and it shows the strong background suppression ability for all datasets.

The specific values of AUC ($P_{D,}$ P_F), AUC ($P_{D,}$ τ), AUC ($P_{F,}$ τ), AUC_{BS} and AUC_{OD} for different methods on the SanDiego, Urban, Beach, and HYDICE datasets are given in Tables 1-4. The optimal results are shown in bold, and the suboptimal results are underlined. For the SanDiego dataset, AUC ($P_{D,}$ P_F), AUC ($P_{D,}$ τ), AUC ($P_{F,}$ τ), and AUC_{BS} are the optimal results. Therefore, the proposed method has the optimal integrated detection ability on the SanDiego dataset. AUCBS is a more reasonable indicator of the background suppression capacity for the method. For the Urban dataset, AUC ($P_{D,}$ P_F), AUC ($P_{D,}$ τ), AUC ($P_{F,}$ τ), and AUC_{BS} are also the optimal results. The proposed method has the better integrated detection ability on the Urban dataset. For Beach dataset, the proposed method also performs optimally on all evaluation indicators. For HYDICE dataset, AUC ($P_{D,}$ P_F), AUC ($P_{D,}$ τ), AUC ($P_{F,}$ τ), and AUC_{BS} are also the optimal results. And AUC ($P_{F,}$ τ) is far superior to other methods. Therefore, the proposed method has the optimal integrated detection ability on the HYDICE dataset.

3.4.3. Target-background separability comparison

To evaluate the effectiveness of the proposed method in separating target from the background, target-background separability boxplot [50] is used to show the separation degree of target and background by different methods.

Fig. 15 shows the target-background separability boxplot for nine comparison methods and the proposed method on the four real datasets (SanDiego, Urban, Beach, and HYDICE). The boxes in the target-background separability boxplot represent pixels with statistically distributed values, removing the highest and lowest 10 % of data in the target and background. The red box and green box represent the target and background, respectively. The horizontal line in the middle of each box indicates the median value, and the upper and lower horizontal lines indicate the maximum and minimum values. The proposed method displays good background suppression performance for all datasets in the experiment and can better separate the target from the background. The excellent target-background separability indicates that the

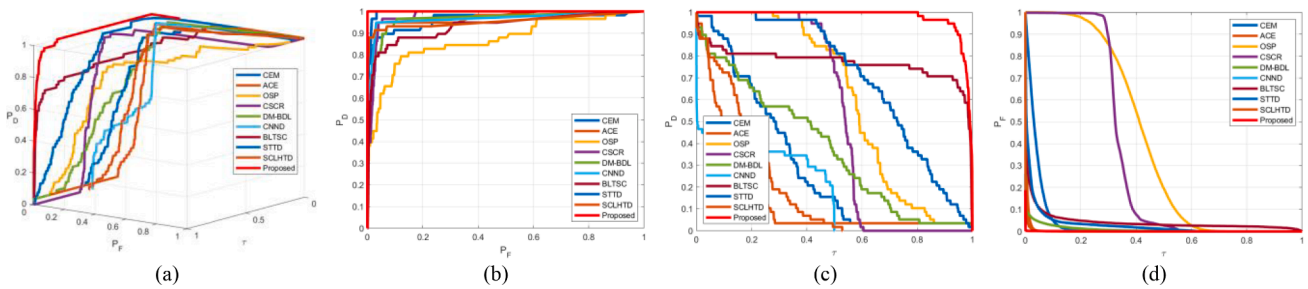

Fig. 11. ROC curves for SanDiego dataset. (a) 3D ROC curve. (b) 2D ROC curve of (P_D, P_F). (c) 2D ROC curve of (P_D, τ). (d) 2D ROC curve of (P_F, τ).

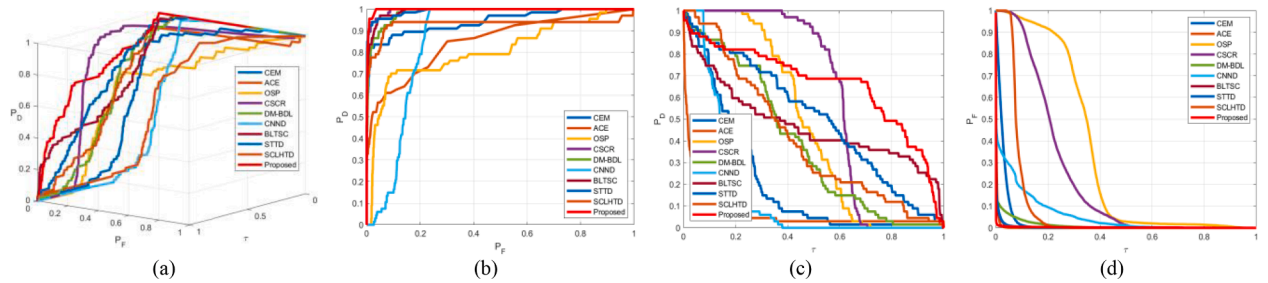

Fig. 12. ROC curves for Urban dataset. (a) 3D ROC curve. (b) 2D ROC curve of (P_D, P_F). (c) 2D ROC curve of (P_D, τ). (d) 2D ROC curve of (P_F, τ).

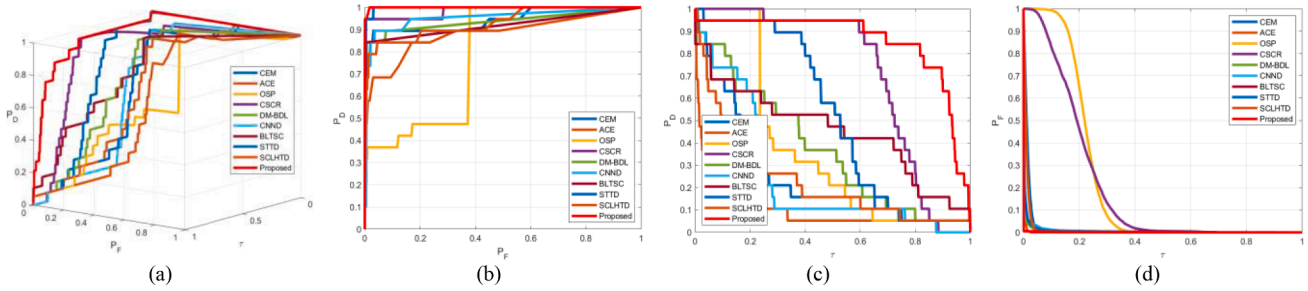

Fig. 13. ROC curves for Beach dataset. (a) 3D ROC curve. (b) 2D ROC curve of (P_D, P_F). (c) 2D ROC curve of (P_D, τ). (d) 2D ROC curve of (P_F, τ).

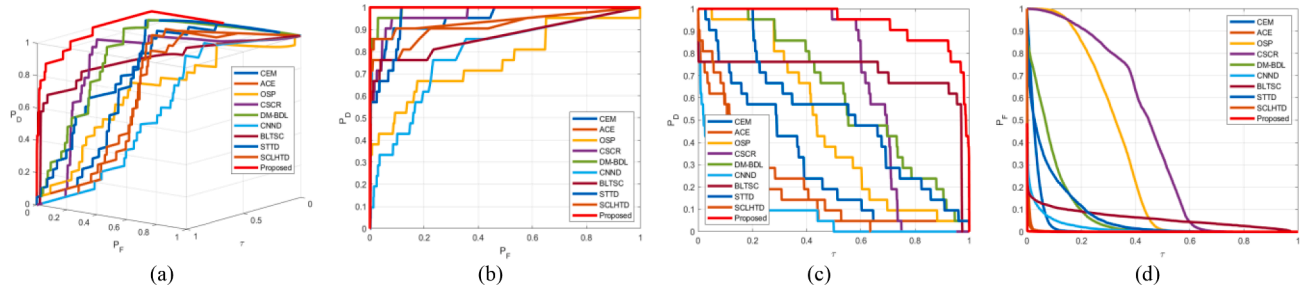

Fig. 14. ROC curves for HYDICE dataset. (a) 3D ROC curve. (b) 2D ROC curve of (P_D, P_F). (c) 2D ROC curve of (P_D, τ). (d) 2D ROC curve of (P_F, τ).

enhanced spectral information present in two-dimensional data enables the model to a more accurately distinction between target and background spectra. By utilizing the target-background separability boxplot, we can visually and statistically assess the performance of different methods, providing clear evidence of the superior capability of the proposed method in accurately separating targets from the background.

3.5. Ablation study of the proposed model

3.5.1. Ablation experiments of GASF and GADF

With GAF, one-dimensional spectral sequence data can be transformed into two distinct two-dimensional data: GASF and GADF. And this paper utilizes GADF as the two-dimensional data input to the spectral feature extraction network. To investigate the effect of GASF and GADF on the detection accuracy of HTD, GASF and GADF are used as the inputs for training and detection of the spectral feature extraction network, respectively. Table 5 illustrates the effect of GASF and GADF on the detection accuracy of HTD, and the AUC (P_D, P_F) values in Table 5 demonstrate the effectiveness of methods.

It can be seen from Table 5 that the detection accuracy of the method using GADF is almost identical for the method using GASF on all four real datasets. It demonstrates that the detection accuracy of the method is independent of the form in which the two-dimensional data is acquired, and is related to the data type of the one-dimensional and two-dimensional data. This is because the one-dimensional spectral sequence data contains limited spectral feature information.

3.5.2. Ablation experiments of atrous convolution

For the proposed method, an ablation study of the traditional applied two-dimensional standard and atrous convolution is conducted to verify whether the type of two-dimensional convolution used by the spectral feature extraction network has any effect on the detection performance. Considering that the main body of the spectral feature extraction network is the cascade structure of 3 convolutional blocks, we set up two sets of experiments, including the network with convolutional block using standard convolution and the network with convolutional block using atrous convolution, respectively. For both experiments, all other settings are kept the same. Table 6 illustrates the effect of standard and atrous convolution on the detection performance of HTD. The AUC (P_D, P_F) values in Table 6 demonstrate the effectiveness of methods, and the times in Table 6 indicate the time for the spectral feature extraction network training. The parameters in Table 6 demonstrate the parameter quantity of two networks, and the FLOPs indicate the floating-point operations per second of two networks.

As can be seen from Table 6, the AUC (P_D, P_F) values for the same dataset are equivalently close, which demonstrates that the detection accuracy of the method is independent of the type of two-dimensional convolution. However, it can be seen from Table 6 that the times of network training are markedly disparate. The time of convolutional block using atrous convolution is less than the time of convolutional

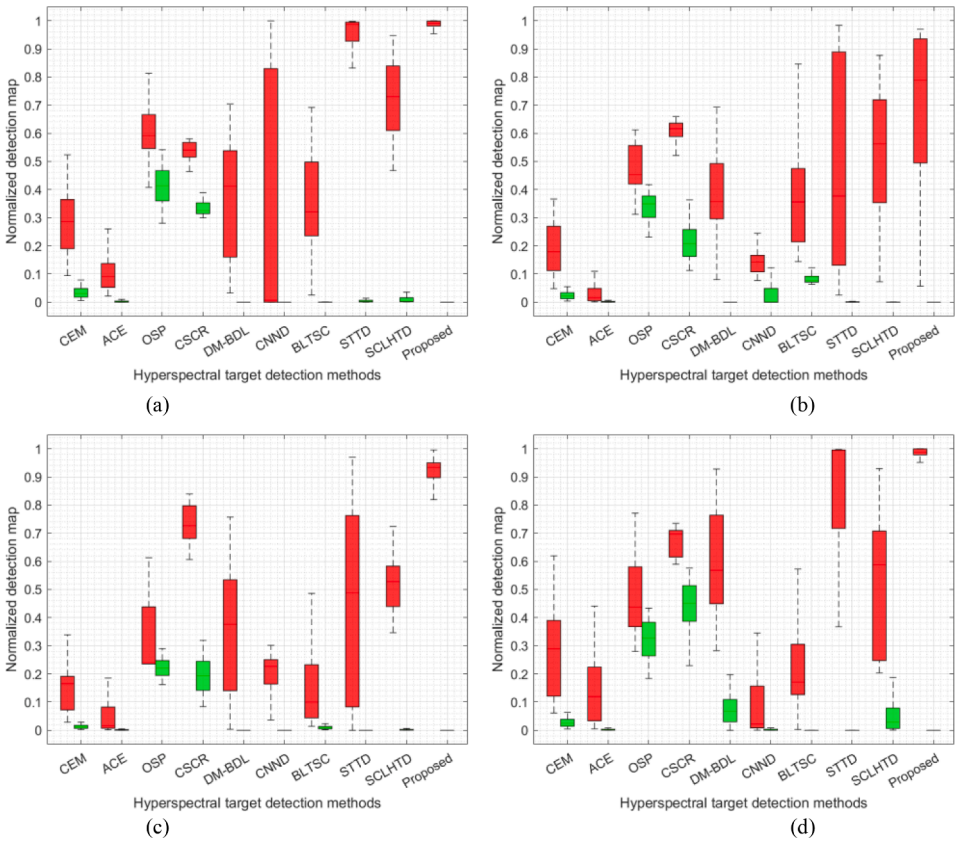

Fig. 15. Target-background separability boxplots for different datasets. (a) San Diego. (b) Urban. (c) Beach. (d)HYDICE. (The red boxes represent target and the green boxes represent background.). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Accuracy comparison of different methods for SanDiego dataset.

Method	AUC (P_D, P_F)	AUC (P_D, τ)	AUC (P_F, τ)	AUC_{BS}	AUC_{OD}
CEM	0.9616	0.2968	0.0388	0.9228	1.2196
ACE	0.9746	0.1336	0.0045	0.9701	1.1038
OSP	0.8682	0.6052	0.4102	0.4580	1.0632
CSCR	0.9774	0.5298	0.3394	0.6380	1.1678
DM-BDL	0.9653	0.3727	0.0095	0.9558	1.3285
CNND	0.9657	0.1832	0.0007	0.9650	1.1482
BLTSC	0.9551	0.1900	0.0017	0.9534	1.1434
STTD	0.9308	0.7694	0.0435	0.8873	1.6568
SCLHTD	0.9960	0.7079	0.0284	0.9676	1.6755
Proposed	**0.9998**	**0.9787**	**0.0006**	**0.9992**	**1.9779**

*Boldface highlights the best results while underline shows the second-best.

Table 2
Accuracy comparison of different methods for Urban dataset.

Method	AUC (P_D, P_F)	AUC (P_D, τ)	AUC (P_F, τ)	AUC_{BS}	AUC_{OD}
CEM	0.9430	0.2082	0.0270	0.9160	1.1242
ACE	0.8505	0.0678	0.0030	0.8475	0.9152
OSP	0.8066	0.4680	0.3368	0.4698	0.9378
CSCR	0.9935	0.5996	0.2224	0.7711	1.3707
DM-BDL	0.9856	0.3846	0.0111	0.9745	1.3592
CNND	0.8556	0.1506	0.0557	0.7999	0.9505
BLTSC	0.9310	0.4066	0.0881	0.8428	1.2494
STTD	0.9907	0.4700	0.0042	0.9865	1.4566
SCLHTD	0.9917	0.5058	0.0039	0.9878	1.4937
Proposed	**0.9973**	**0.6585**	**0.0029**	**0.9944**	**1.6529**

*Boldface highlights the best results while underline shows the second-best.

Table 3
Accuracy comparison of different methods for Beach dataset.

Method	AUC (P_D, P_F)	AUC (P_D, τ)	AUC (P_F, τ)	AUC_{BS}	AUC_{OD}
CEM	0.9433	0.2391	0.0141	0.9292	1.1683
ACE	0.8821	0.1040	0.0023	0.8798	0.9838
OSP	0.7864	0.3663	0.2223	0.5641	0.9304
CSCR	0.9832	0.7129	0.1979	0.7852	1.4981
DM-BDL	0.9363	0.3645	0.0039	0.9324	1.2969
CNND	0.9537	0.2419	0.0065	0.9472	1.1891
BLTSC	0.9318	0.2273	0.0116	0.9202	1.1476
STTD	0.9191	0.4537	0.0025	0.9166	1.3703
SCLHTD	0.9978	0.5183	0.0062	0.9915	1.5098
Proposed	**0.9992**	**0.8560**	**0.0020**	**0.9972**	**1.8532**

*Boldface highlights the best results while underline shows the second-best.

Table 4
Accuracy comparison of different methods for HYDICE dataset.

Method	AUC (P_D, P_F)	AUC (P_D, τ)	AUC (P_F, τ)	AUC_{BS}	AUC_{OD}
CEM	0.9549	0.3077	0.0306	0.9246	1.2323
ACE	0.9145	0.1945	0.0035	0.9110	1.1055
OSP	0.7617	0.4758	0.3166	0.4451	0.9209
CSCR	0.9661	0.6635	0.4251	0.5410	1.2045
DM-BDL	0.9908	0.6174	0.0848	0.9061	1.5234
CNND	0.7868	0.1023	0.0187	0.7681	0.8704
BLTSC	0.9385	0.1729	0.0016	0.9369	1.1098
STTD	0.8658	0.7277	0.0618	0.8040	1.5318
SCLHTD	0.9637	0.5389	0.0717	0.8920	1.4308
Proposed	**0.9999**	**0.9385**	**0.0002**	**0.9997**	**1.9382**

*Boldface highlights the best results while underline shows the second-best.

Table 5
Effect of GASF&GADF for detection accuracy on four datasets.

Methods Datasets	SanDiego	Urban	Beach	HYDICE
Network with GASF	0.9987	0.9972	0.9993	0.9997
Network with GADF	0.9998	0.9973	0.9992	0.9999

Table 6
Ablation of atrous convolution for detection accuracy.

Datasets Methods		Standard Convolution	Atrous Convolution
SanDiego	AUC (PD, PF)	0.9973	0.9998
	Parameters(M)	1.0851	0.4843
	FLOPs(G)	4.4832	3.6341
	Time(S)	1332.7782	1183.5091
Urban	AUC (PD, PF)	0.9965	0.9973
	Parameters(M)	1.3357	0.5088
	FLOPs(G)	5.1941	4.2653
	Time(S)	1278.9656	1032.2841
Beach	AUC (PD, PF)	0.9990	0.9992
	Parameters(M)	1.0851	0.4645
	FLOPs(G)	4.4112	3.5718
	Time(S)	1633.1364	1282.7389
HYDICE	AUC (PD, PF)	0.9998	0.9999
	Parameters(M)	0.8897	0.4488
	FLOPs(G)	3.8382	3.0681
	Time(S)	1589.5463	1220.0082

block using standard convolution, which is due to the ability of atrous convolution to flexibly adjust the size of the receptive field to capture multi-scale information. The shorter time is required for the spectral feature extraction network with atrous convolution training to convergence. And for the same dataset, the number of network parameters and FLOPs of standard convolution are higher than those of atrous convolution. Comparing to the standard convolution-based network, this also proves that atrous convolution-based network requires less training time.

3.6. Training time consumption

As can be seen from Table 7, for the STTD method, in the self-attention mechanism of Transformer, when the attention weight of each position is calculated for the spectral input sequence length n, the similarity of all positions needs to be calculated in pairs. Then calculate the attention score, which is $O(n^2)$ in computational complexity. This calculation process requires the dot product of the vectors of each pair of positions, and the amount of computation increases sharply as the length of the sequence n increases. At the same time, the multi-head attention mechanism is one of the core components of Transformer. It consists of self-attention calculations for multiple heads (such as h heads). For each head, the computational process of the self-attention mechanism described above needs to be carried out. Although the computation of each head can be performed in parallel, the total computation amount is still h times that of a single self-attention mechanism. Therefore, more parameters mean that more computation is required to update these parameters during training and inference, both forward and back-propagation, thus increasing the time cost. In addition, the STTD method uses the Siamese network structure, and the computational complexity of the model will be doubled. Therefore, the training time of the proposed method is lower than that of the STTD method.

4. Conclusion

Considering that hyperspectral images have more spectral bands with high correlation, the one-dimensional vector of spectral sequence data contains limited information that can be extracted compared to the two-dimensional matrix data, posing challenges to spectral-based deep learning model for hyperspectral target detection. To address these challenges, this paper proposes a novel hyperspectral target detection method that leverages atrous convolutional neural network and GAF. By transforming the one-dimensional vector of spectral sequence data into two-dimensional matrix data, the proposed method overcome the limitations of one-dimensional vector of spectral features extraction and improve the performance of target detection. Specifically, a two-step strategy is firstly employed to solve the common issue of lacking labelled samples for deep learning-based model, followed by the GAF. A priori target spectrum is modulated using Gaussian white noise with different SNRs, and the two-dimensional band relation maps of sufficient target samples are obtained by GAF. The two-dimensional band relation maps of the pure background samples are obtained by GAF and SAM detector to complete the preparation of training data. Subsequently, according to the characteristics of the two-dimensional spectral band relation maps, a spectral feature extraction network based on atrous convolution is proposed, where the training data are fed into the network to distinguish target and background spectra. Finally, background suppression is achieved by the exponential function, the power function, and the normalization operation. Comprehensive experiments conducted on four real-world hyperspectral datasets demonstrate that the proposed method significantly outperforms 9 existing state-of-the-art detection methods.

It should be noted that, since the proposed method is achieved on two-dimensional matrix data for feature extraction and network training, it inherently requires more training time compared to methods implemented on one-dimensional data. Despite this increase in training time, the use of atrous convolution helped mitigate the issue by expanding the receptive field and accelerating convergence. However, the training time of network remains slightly longer than that of one-dimensional networks. The proposed method leverages the redundancy inherent in the strong spectral band correlations of hyperspectral data, effectively trading efficiency for improved accuracy in hyperspectral target detection.

Table 7
Training time consumption of different methods on four datasets.

Method	SanDiego	Urban	Beach	HYDICE
CNND	348.2861	342.4716	350.6986.	329.4984
BLTSC	965.6516	722.1543	735.1299	703.6953
STTD	1522.1083	1647.8313	1766.6989	1495.4649
SCLHTD	301.4326	232.4835	256.9689	216.3185
Proposed	1183.5091	1032.2841	1282.7389	1220.0082

CRediT authorship contribution statement

Hongzhou Wang: Writing – original draft, Software, Methodology, Data curation, Conceptualization. **Yulei Wang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Yuchao Yang:** Visualization, Validation, Software, Investigation. **Enyu Zhao:** Validation, Project administration, Investigation. **Jian Zeng:** Investigation, Validation, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported in part by National Nature Science Foundation of China (42271355, 61801075), Natural Science Foundation of

Liaoning Province (2022-MS-160), China Postdoctoral Science Foundation (2020M670723), and the Fundamental Research Funds for the Central Universities (3132024234).

Data availability

Data will be made available on request.

References

[1] T. Chen, C. Leng, Z. Pei, J. Peng, A. Basu, Multimanifold Bistructured Low Rank Representation of hyperspectral images, Infrared Phys. Techn. 136 (2024) 105039, https://doi.org/10.1016/j.infrared.2023.105039.

[2] Y. Li, Q. Xu, Z. Kong, W. Li, MULS-Net: a multilevel supervised network for ship tracking from low-resolution remote-sensing image sequences, IEEE Trans. Geosci. Remote Sens. 61 (2023) 5624214, https://doi.org/10.1109/TGRS.2023.3326613.

[3] S. Feng, H. Zhang, B. Xi, C. Zhao, Y. Li, J. Chanussot, Cross-domain few-shot learning based on decoupled knowledge distillation for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 62 (2024) 5534414, https://doi.org/10.1109/TGRS.2024.3476116.

[4] Y. Wang, L. Wang, C. Yu, E. Zhao, M. Song, C.-H. Wen, et al., Constrained-target band selection for multiple-target detection, IEEE Trans. Geosci. Remote Sens. 57 (2019) 6079–6103, https://doi.org/10.1109/TGRS.2019.2904264.

[5] Y. Li, Q. Xu, Z. He, W. Li, Progressive task-based universal network for raw infrared remote sensing imagery ship detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 5610013, https://doi.org/10.1109/TGRS. 2023.3275619.

[6] A.A. Hameed, A. Jamil, A. Seyyedabbasi, An optimized feature selection approach using sand Cat Swarm optimization for hyperspectral image classification, Infrared Phys. Techn. 141 (2024) 105449, https://doi.org/10.1016/j.infrared.2024.105449.

[7] Y. Wang, Q. Zhu, H. Ma, H. Yu, A hybrid gray wolf optimizer for hyperspectral image band selection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13, https://doi.org/10.1109/TGRS.2022.3167888.

[8] Y. Wang, H. Wang, E. Zhao, M. Song, C. Zhao, Tucker decomposition-based network compression for anomaly detection with large-scale hyperspectral images, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 17 (2024) 10674–10689, https://doi.org/10.1109/JSTARS.2024.3404607.

[9] S. Yang, Z. Song, H. Yuan, Z. Zou, Z. Shi, Fast high-order matched filter for hyperspectral image target detection, Infrared Phys. Techn. 94 (2018) 151–155, https://doi.org/10.1016/j.infrared.2018.09.018.

[10] Y. Yang, Y. Wang, H. Wang, L. Zhang, E. Zhao, M. Song, et al., Spectral-enhanced sparse transformer network for hyperspectral super-resolution reconstruction, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 17 (2024) 17278–17291, https://doi.org/10.1109/JSTARS.2024.3457814.

[11] H. Sun, Z. Luo, D. Ren, B. Du, L. Chang, J. Wan, Unsupervised multi-branch with high-frequency enhancement network for image dehazing, Pattern Recognit. 56 (2024) 110763, https://doi.org/10.1016/j.patcog.2024.110763.

[12] Y. Wang, X. Chen, E. Zhao, M. Song, Self-supervised spectral-level contrastive learning for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 61 (2023) 1–16, https://doi.org/10.1109/TGRS.2023.3270324.

[13] Y. Li, Y. Shi, K. Wang, B. Xi, J. Li, P. Gamba, Target detection with unconstrained linear mixture model and hierarchical denoising autoencoder in hyperspectral imagery, IEEE Trans. Image Process. 31 (2022) 1418–1432, https://doi.org/10.1109/TIP.2022.3141843.

[14] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, et al., The spectral image processing system (SIPS)interactive visualization and analysis of imaging spectrometer data, Remote Sens. Environ. 44 (1993) 145–163, https://doi.org/10.1016/0034-4257(93)90013-N.

[15] C.-I. Chang, An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis, IEEE Trans. Inf. Theory 46 (5) (2000) 1927–1932, https://doi.org/10.1109/18.857802.

[16] C.-I. Chang, D. Heinz, Constrained subpixel target detection for remotely sensed imagery, IEEE Trans. Geosci. Remote Sens. 38 (3) (2000) 1144–1159, https://doi.org/10.1109/36.843007.

[17] Z. Zou, Z. Shi, Hierarchical suppression method for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 54 (1) (2016) 330–342, https://doi.org/10.1109/TGRS.2015.2456957.

[18] X. Jin, S. Paswaters, H. Cline, A comparative study of target detection algorithms for hyperspectral imagery, Proc. SPIE 7334 (2009) 682–693, https://doi.org/10.1117/12.818790.

[19] S. Kraut, L. Scharf, The CFAR adaptive subspace detector is a scale-invariant GLRT, IEEE Trans. Signal Process 47 (9) (1999) 2538–2541, https://doi.org/10.1109/78.782198.

[20] F. Robey, D. Fuhrmann, E. Kelly, R. Nitzberg, A CFAR adaptive matched filter detector, IEEE Trans. Aerosp. Electron. Syst. 28 (1) (1992) 208–216, https://doi.org/10.1109/7.135446.

[21] Q. Du, H. Ren, C.-I. Chang, A comparative study for orthogonal subspace projection and constrained energy minimization, IEEE Trans. Geosci. Remote Sens. 41 (6) (2003) 1525–1529, https://doi.org/10.1109/TGRS.2003.813704.

[22] X. Jiao, C.-I. Chang, Kernel-based constrained energy minimization (K-CEM), Proc. SPIE 6966 (2008) 523–533.

[23] H. Kwon, N.M. Nasrabadi, Kernel adaptive subspace detector for hyperspectral imagery, IEEE Geosci. Remote Sens. Lett. 3 (2) (2006) 271–275.

[24] H. Kwon, N.M. Nasrabadi, Kernel orthogonal subspace projection for hyperspectral signal classification, IEEE Trans. Geosci. Remote Sens. 43 (12) (2005) 2952–2962.

[25] Y. Zhang, B. Du, L. Zhang, A sparse representation-based binary hypothesis model for target detection in hyperspectral images, IEEE Trans. Geosci. Remote Sens. 53 (3) (2015) 1346–1354.

[26] Y. Chen, N. Nasrabadi, T. Tran, Sparse representation for target detection in hyperspectral imagery, IEEE J. Sel. Topics Signal Process. 5 (3) (2011) 629–640, https://doi.org/10.1109/JSTSP.2011.2113170.

[27] W. Li, Q. Du, B. Zhang, Combined sparse and collaborative representation for hyperspectral target detection, Pattern Recognit. 48 (12) (2015) 3904–3916, https://doi.org/10.1016/j.patcog.2015.05.024.

[28] T. Cheng, B. Wang, Decomposition model with background dictionary learning for hyperspectral target detection, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 14 (2021) 1872–1884.

[29] P. Xiang, J. Song, H. Qin, W. Tan, H. Li, H. Zhou, Visual attention and background subtraction with adaptive weight for hyperspectral anomaly detection, IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens. 14 (2021) 2270–2283.

[30] C. Yu, X. Zhao, B. Gong, Y. Hu, M. Song, H. Yu, C.-I. Chang, Distillation-constrained prototype representation network for hyperspectral image incremental classification, IEEE Trans. Geosci. Remote Sens. 62 (2024) 5507414, https://doi.org/10.1109/TGRS.2024.3359629.

[31] Y. Wang, H. Ma, Y. Yang, E. Zhao, M. Song, C. Yu, Self-supervised deep multi-level representation learning fusion-based maximum entropy subspace clustering for hyperspectral band selection, Remote Sens. 16 (2) (2024) 224, https://doi.org/10.3390/rs16020224.

[32] C. Liu, J. Feng, Amsin,, An adaptive multi-scale input network for hyperspectral image fusion, Infrared Phys. Techn. 140 (2024) 105347, https://doi.org/10.1016/j.infrared.2024.105347.

[33] Y. Wang, X. Chen, E. Zhao, M. Song, C. Yu, An unsupervised momentum contrastive learning based transformer network for hyperspectral target detection, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. (2024), https://doi.org/10.1109/JSTARS.2024.3387985.

[34] L. Zhang, B. Cheng, A stacked autoencoders-based adaptive subspace model for hyperspectral anomaly detection, Infrared Phys Techn. 96 (2019) 52–60, https://doi.org/10.1016/j.infrared.2018.11.015.

[35] W. Li, G. Wu, Q. Du, Transferred deep learning for hyperspectral target detection, Proc. IEEE Int. Geosci. Remote Sens. Symp. (2017), https://doi.org/10.1109/IGARSS.2017.8128168.

[36] Y. Wang, X. Chen, F. Wang, M. Song, C. Yu, Meta-learning based hyperspectral target detection using siamese network, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–13, https://doi.org/10.1109/TGRS.2022.3169970.

[37] Z. Feng, J. Zhang, J. Feng, Spectral–spatial joint target detection of hyperspectral image based on transfer learning, Proc. IEEE Int. Geosci. Remote Sens. Symp. (2020) 1770–1773.

[38] G. Zhang, S. Zhao, W. Li, Q. Du, Q. Ran, R. Tao, HTD-Net: A deep convolutional neural network for target detection in hyperspectral imagery, Remote Sens. 12 (9) (2020), https://doi.org/10.3390/rs12091489.

[39] D. Zhu, B. Du, L. Zhang, Two-stream convolutional networks for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 59 (8) (2021) 6907–6921, https://doi.org/10.1109/TGRS.2020.3031902.

[40] W. Xie, X. Zhang, Y. Li, K. Wang, Q. Du, Background learning based on target suppression constraint for hyperspectral target detection, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 13 (2020) 5887–5897, https://doi.org/10.1109/JSTARS. 2020.3024903.

[41] W. Rao, L. Gao, Y. Qu, X. Sun, B. Zhang, J. Chanussoot, Siamese transformer network for hyperspectral image target detection, IEEE Trans. Geosci. Remote Sens. 60 (5526419) (2022), https://doi.org/10.1109/TGRS.2022.3163173.

[42] M. Song, S. Liu, D. Xu, H. Yu, Multiobjective optimization-based hyperspectral band selection for target detection, IEEE Trans. Geosci. Remote Sens. 60 (2022) 5529022, https://doi.org/10.1109/TGRS.2022.3176856.

[43] Y. Gao, Y. Feng, X. Yu, S. Mei, Robust signature-based hyperspectral target detection using dual networks, IEEE Geosci. Remote Sens. Lett. 20 (2023) 5500605, https://doi.org/10.1109/LGRS.2023.3237746.

[44] Z. Wang, T. Oates, Imaging Time-Series to Improve Classification and Imputation, arXiv preprint (2015), Doi: 10.48550/ arXiv.1506. 00327.

[45] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv preprint (2015), https://arxiv.org/abs/1511.07122.

[46] M. Holschneider, R. Kronland-Martinet, J. Morlet, P. Tchamitchian, A real-time algorithm for signal analysis with the help of the wavelet transform, In Wavelets (1990).

[47] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, G. Cottrell, Understanding Convolution for Semantic Segmentation, arXiv preprint (2018), https://arxiv.org/abs/1702.08502.

[48] C.-I. Chang, An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis, IEEE Trans. Geosci. Remote Sens. 59 (6) (2021) 5131–5153, https://doi.org/10.1109/TGRS.2020.3021671.

[49] C.-I. Chang, J. Chen, Orthogonal subspace projection using data sphering and low-rank and sparse matrix decomposition for hyperspectral target detection, IEEE Trans. Geosci. Remote Sens. 59 (10) (2021) 8704–8722, https://doi.org/10.1109/TGRS.2021.3053201.

[50] L. Zhang, B. Cheng, Fractional fourier transform and transferred CNN based on tensor for hyperspectral anomaly detection, IEEE Geosci. Re-Mote Sens. Lett. 19 (2022), https://doi.org/10.1109/LGRS.2021.3072249.