Infrared Physics & Technology 150 (2025) 106056

Contents lists available at ScienceDirect

Infrared Physics and Technology

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

Research Paper

Deep spectral metric learning with Siamese network for hyperspectral target detection

Yulei Wang [a], Chao Deng [a], Hongzhou Wang [a], Enyu Zhao [a,*], Qiongqiong Lan [b,*]

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

ARTICLE INFO

Keywords:
Hyperspectral imagery
Target detection
Spectral metric learning
Siamese network
Generative adversarial network

ABSTRACT

Hyperspectral target detection (HTD) is a critical methodology characterized by its extensive applications, particularly noted for its efficacy in target identification based on limited prior spectral information, thereby reducing dependence on texture and geometric features. Although recent advancements in HTD have largely embraced deep learning-based approaches, they often hindered by the significant challenge of data scarcity, particularly in obtaining labeled datasets. To address this challenge, this paper proposes a novel approach employing spectral metric learning within a Siamese network framework, named as SN-HTD. Leveraging the metric learning capabilities of the Siamese network architecture, the proposed method strategically employs unlabeled samples to train a model with the ability of target and background spectral discrimination, aiming to minimize the distance between homogeneous features while maximizing the separation between heterogeneous features. Spectral data argumentation is firstly conducted by modulating the priori target spectra with Gaussian white noise to effectively address the issue of insufficient target samples. Then the training procedure of the proposed deep learning model is bifurcated into pre-training and spectral metric learning phases, aiming to optimize resource utilization and computational efficiency. The core of the proposed model is a spectral metric learning with Siamese network, constructed atop the discriminator of a pre-trained one-dimensional generative adversarial network (GAN), which is fed by positive and negative sample pairs derived from prior target spectrum against the augmented data and unlabeled background samples, respectively. Additionally, a guided image filter is incorporated for spatial information exploitation, thereby improving the detection performance of the method. Comparative experiments have been conducted on real hyperspectral images captured by different sensors in various scenes, demonstrating the superiority of the proposed SN-HTD method against the state-of-the-art methods, positioning it as a notable advancement in the field of HTD.

1. Introduction

Hyperspectral imaging, distinct from traditional single-band and multi-spectral imaging systems, captures the reflectance of objects across hundreds of narrow and contiguous bands, yielding to three-dimensional hyperspectral images (HSIs) that exhibit rich spectral and spatial information. The exceptional spectral band resolution allows each pixel in HSIs to be represented as a spectral curve, providing detailed insights into substance properties. This characteristic positions hyperspectral target detection (HTD) as an advanced remote sensing technology with extensively applications in both civil and military domains, including urban target detection, mineral surveying, medical diagnostics, environment detection, and military camouflage target identification [1–5].

HTD, as an approach proficient in identifying targets independent of texture and geometric features, relies on the discernment of spectral differences for precise recognition of ground objects. Many hyperspectral target detection methods have appeared in the past research literature. Among the traditional HTD methods, adaptive coherence estimation (ACE) [6,7] is a classical HTD method based on probability statistics by assuming that the background conforms to the multivariate Gaussian distribution. In signal filtering-based methods, constrained energy minimization (CEM) [8] is a notable approach, being instrumental in highlighting targets while suppressing background through a custom-designed finite pulse filter. The CEM detector has attracted increasing attention owing to its outstanding performance, aligning with

* Corresponding authors.
E-mail addresses: zhaoenyu@dlmu.edu.cn (E. Zhao), lanqiongqiong@126.com (Q. Lan).

https://doi.org/10.1016/j.infrared.2025.106056
Received 8 April 2025; Received in revised form 31 July 2025; Accepted 2 August 2025
Available online 5 August 2025

the burgeoning development of CEM-based methodologies, exemplified by notable advancements such as the hierarchical CEM (hCEM) [9], the ensemble-based CEM (E-CEM) [10], and so on. In light of the extensive integration of sparsity theory, the HTD has been advanced by proposing the sparsity-based target detector (STD) (Chen et al.) [11] and the combined sparse and cooperative representation detector (CSCR) (Li et al.) [12]. Within the STD framework, the detection process involves computing the reconstruction error of a vector, linearly represented by atomic vectors in a complete dictionary, with respect to the pixel to be detected, while CSCR achieves detection by representing the image element to be detected by targets library and backgrounds library.

The ascent of artificial intelligence technology in recent years has propelled deep learning to the forefront of image analysis in remote sensing [13–16]. In the context of HTD, perceived as an image binary classification problem, the potency of deep learning in generalization and high-level semantic feature extraction has ignited novel research avenues. Given the inherent limitations of prior information, where only the target spectrum of interest is known, transfer learning has been proved an effective strategy for deep learning-based HTD. Examples include convolutional neural network-based target detection (CNND) [17], and *meta*-learning with Siamese network-based target detection (MLSN) [18]. CNND augments training samples for HTD by pairing and labeling pixels from hyperspectral datasets with known labels in the source domain, and then trains a binary classification multi-layer CNN network for HTD. However, the challenge of domain mismatch caused by different sensor poses a notable hindrance. In response, MLSN incorporates *meta*-learning to enhance the adaptability of deep transfer learning models to HTD. Acknowledging the data-intensive nature of deep neural networks and the limited availability of training data for HTD, hyperspectral target detection based on deep network (HTD-Net) [19] adopts a U-net structure for generating potential target samples, and employs a linear prediction algorithm for identifying background samples significantly different from targets. Another approach is hyperspectral target detection method based on two-stream convolutional network proposed in [20], which employs the mixed sparse representation to acquire background pixels, and subsequently blends the prior target spectra and background pixels with selected typical background pixels. This process aims to generate a substantial volume of training data, pairing them with prior spectra to construct positive and negative training samples, which are then fed into the two-stream convolutional network to accomplish the detection task. In the realm of unsupervised learning, Xie et al. introduce the background learning based on a target suppression constraint (BLTSC) detector [21], employing a rough detection method to identify background samples, and then feeding them into an adversarial auto encoder (AAE) with target suppression constraints to reconstruct the pure background, thereby achieving target detection by comparing the reconstructed image with the image to be detected. To leverage spatial information for the improvement of the detection performance, a 3-D macro–micro-residual auto encoder is designed and used to extract macro- and micro-features, which are fused and sent to a hierarchical radial basis function (hRBF) detector for background suppression and target preservation [22]. From the perspective of self-supervised learning, a hyperspectral target detection method based on self-supervised spectral level contrast learning (SCLHTD) is proposed in [23,24]. SCLHTD uses the spectral data enhancement method based on odd and even band sampling to extract supervisory information from the image itself, followed with the learning of the spectral similarities and differences through spectral level contrast learning to achieve the target detection task.

This paper aims to address the performance bottleneck in hyperspectral target detection caused by the scarcity of labeled samples, while leveraging limited prior target information to achieve high-quality detection. To this end, inspired by the principles of metric learning [25], a deep spectral metric learning framework is proposed based on a Siamese network (SN-HTD). The training process is divided into two stages: pre-training and metric learning. In the pre-training stage, a one-dimensional generative adversarial network (1D-GAN) is employed to model the hyperspectral image under test, enabling the extraction of supervisory signals inherent to the image itself and training a discriminator with a certain level of spectral discrimination capability. In the metric learning stage, a Gaussian noise-based spectral data augmentation strategy is firstly employed to expand the limited target samples. This strategy preserves the consistency of spectral features while simulating variations under different environmental conditions, thereby enhancing generalization without introducing significant distortions. The pre-trained discriminator is then transferred to a Siamese network to extract spectral features. Metric learning is performed using positive and negative sample pairs to model subtle spectral differences between targets and background, thereby enhancing the model's sensitivity to fine-grained variations, improving generalization, and accelerating convergence. Finally, spatial information is integrated via a guided image filter to enhance spectral-spatial consistency and further improve detection performance. The main contributions of this paper are summarized as follows:

1) A two-stage training strategy is proposed, involving pretraining and spectral metric learning. Specifically, a 1D-GAN is employed in the pretraining phase to extract discriminative spectral features, which are subsequently transferred to a Siamese network for efficient metric learning.
2) To address the challenge of limited annotated samples, a spectral data augmentation based on Gaussian noise is proposed, enhancing sample diversity while preserving spectral characteristics.
3) Hyperspectral target detection is reformulated as a spectral metric learning problem. A Siamese-based spectral metric network is designed to learn discriminative spectral difference features from positive and negative pairs, enabling accurate similarity measurement between pixels under test and the prior target spectra.

The remainder of this paper is organized as follows: Section 2 provides a comprehensive description of the proposed SN-HTD method, and Section 3 presents experimental studies and analysis to validate the proposed method. Finally, conclusions are drawn in Section 4.

2. Proposed method

This section shows the details of the proposed deep spectral metric Siamese network for hyperspectral target detection (SN-HTD). The approach consists of three key stages: pre-training of the 1D-GAN, spectral metric learning with spectral metric Siamese network and spectral data augmentation, and spectral-spatial target detection, with the flowchart shown in Fig. 1.

2.1. Pre-training for 1D-Generative adversarial network

In the first stage of model training, the hyperspectral image under test is utilized to pretrain a 1D-GAN, as illustrated in the top of Fig. 1. The goal of this pre-training process is to enhance the spectral discrimination ability of the discriminator, which serves as a strong foundation for the subsequent spectral metric learning phase.

Generative Adversarial Networks (GANs), which combines generative and adversarial principles, have been widely applied in natural image processing [27,28]. A typical GAN consists of two neural networks: a generator G and a discriminator D. The generator attempts to produce synthetic ("fake") samples that resemble real data, while the discriminator strives to distinguish between real and generated samples. Through adversarial training, both networks iteratively improve, with the generator learning to produce more realistic outputs and the discriminator learning more refined decision boundaries. This adversarial process can be mathematically expressed as follow:

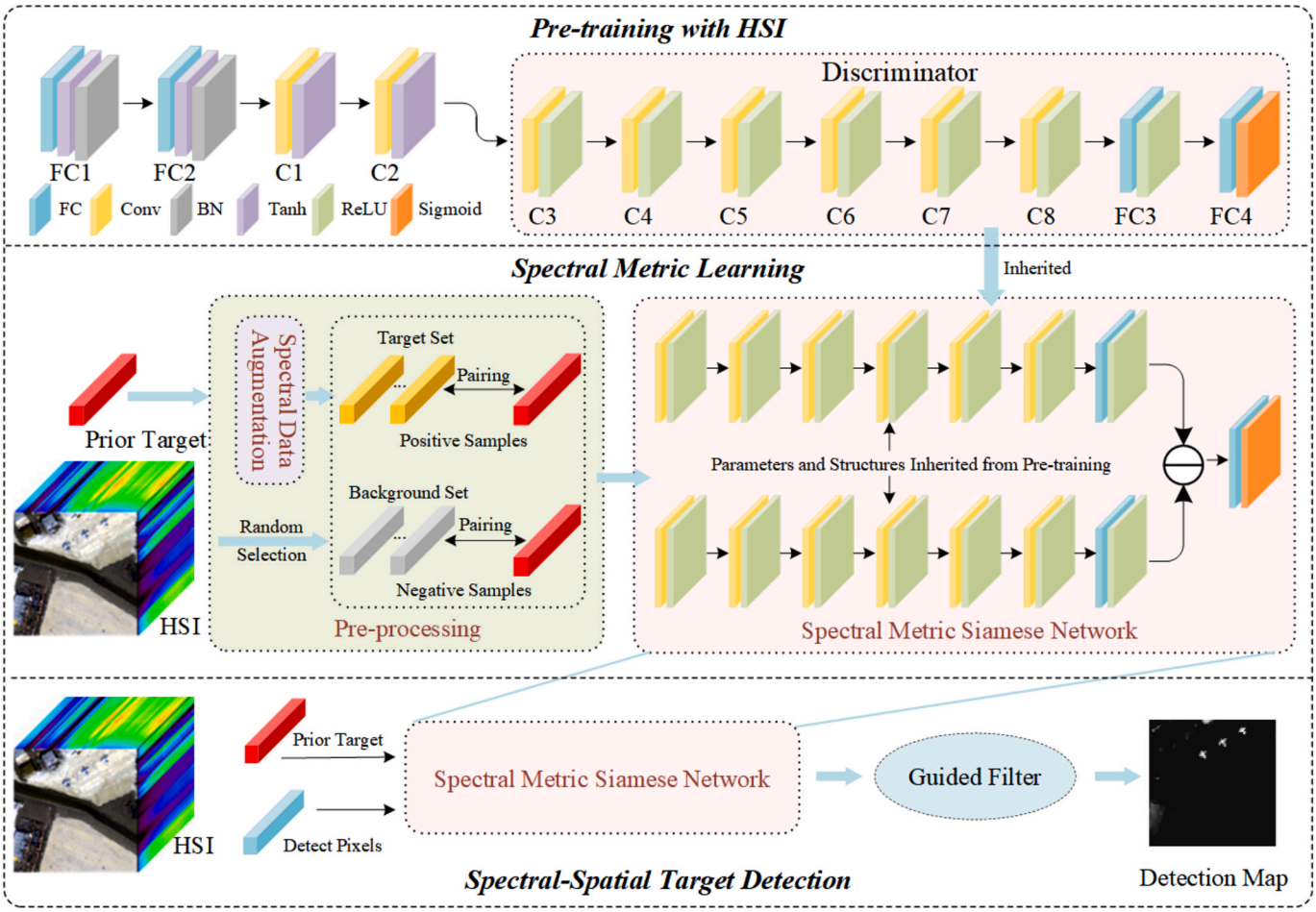

Fig. 1. Flowchart of the proposed SN-HTD method.

$$\min_{G}\max_{D} V(D,G) = \mathrm{E}_{x\sim p_{data}}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))] \tag{1}$$

where the generator $G(\cdot)$ builds a mapping $p_z : G(z) \rightarrow p_g$, p_g denotes the distribution of generated samples and p_z denotes the distribution of radom noise. As for the discriminator, let $D(x, \theta)$ denote the probability of x from the distribution p_{data} of real data rather than p_g. θ demotes whether the input of D is true or not. V (D, G) represents the process where the generator and the discriminator alternate and iteratively train during the training stage, which can be expressed as:

$$\min_{G} V(D,G) = \min_{G} \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))] \tag{2}$$

$$\max_{D} V(D,G) = \max_{D}\left(\mathrm{E}_{x\sim p_{data}}[\log D(x)] + \mathrm{E}_{z\sim p_z}[\log(1 - D(G(z)))]\right) \tag{3}$$

In natural image processing, GANs typically use two-dimensional convolutional layers to act on the spatial dimensions of images. However, hyperspectral target detection relies on the spectral dimension of hyperspectral images (HSIs). Therefore, two-dimensional convolutional layers are not directly applicable to the task of hyperspectral target detection. To address this, 1D-GAN is specifically designed for the spectral dimension, utilizing one-dimensional convolutional layers in place of two-dimensional convolutions. It is important to note that different criteria are followed in constructing both the structures of generator and discriminator. To further improve the stability of the generator in modeling the complex distributions of target and background spectra, the spectral normalization (BN) layer [29] is added after each linear fully connected layer of the generator. This helps prevent training degradation due to poor initialization. For the discriminator, in order to make it more focused on extracting spectral feature information, it is primarily constructed using one-dimensional convolutional layers. To prevent the loss of detailed spectral semantic features caused by pooling operation, a one-dimensional convolutional layer with a stride of 2 is used instead of the average pooling layer. This enables the discriminator to more effectively capture relevant spectral information and adapt quickly to the detection task. Additionally, to stabilize the training of the 1D-GAN, the generator employs a Tanh activation function.

Empirical observations show that the generator can effectively synthesize both target-like and background-like spectral samples, while the discriminator successfully learns to differentiate them. After convergence, the discriminator is fine-tuned using a small number of labeled samples, transforming it into a simple yet effective detector. This detector achieves performance comparable to, or even exceeding, that of traditional methods, demonstrating the strong spectral feature extraction capability of the pretrained discriminator. Building upon this capability, the pretrained discriminator serves as the backbone of a spectral metric Siamese network in the subsequent learning stage, further enhancing detection performance.

2.2. *Deep spectral metric learning with Siamese network*

To efficiently assess the similarity between spectral pixels and the prior target spectra for hyperspectral target detection, the conventional binary classification problem of target detection is reformulated into a similarity metric learning problem. Metric learning is a well-established machine learning technique that aims to learn a distance function capable of distinguishing between similar and dissimilar samples. Its core objective is to minimize the distance between homogeneous

(similar) features while maximizing the separation between heterogeneous (dissimilar) ones. In other words, metric learning seeks to bring similar samples closer together in feature space, while pushing dissimilar ones further apart. This principle enables models to generalize better, particularly when labeled data are scarce. Due to the success of Siamese network in various visual tasks [30–34], which a widely used method in metric learning, it has gained significant popularity. A Siamese network is composed of two or more identical subnetworks (with shared weights), which process input pairs to learn whether they are similar or not. This architecture has been widely applied in tasks such as detection, tracking, and face verification [35–37]. A key advantage of Siamese networks is their inductive bias toward invariance—meaning that two observations of the same class should yield the same output. This characteristic has contributed to its success in modeling complex transformations, similar to how convolution operations model translational invariance.

In this section, the spectral data augmentation and spectral metric learning with Siamese network is introduced into hyperspectral target detection, transforming the target detection task into a deep spectral metric learning problem.

2.2.1. *Spectral data augmentation*

In hyperspectral target detection, it is common to have only one prior target spectrum, with no additional labeled data. This scarcity is compounded by the difficulty and cost of acquiring labeled hyperspectral samples. To address this issue, this paper proposes a new data augmentation approach aimed at overcoming the scarcity of labeled training data for hyperspectral target detection. The proposed data augmentation not only enhances the models' generalization ability, but also acts as a regularization technique to avoid overfitting, thereby improving the quality of the learned representations.

Traditional data augmentation methods for RGB images typically include operations such as random flipping, random cropping, and random rotation, which enhance image symmetry, reduce positional dependence, and increase viewpoint diversity, respectively. However, in hyperspectral target detection, which relies primarily on the spectral information, applying these conventional image operations would significantly disrupt the spectral features. Currently, target sample augmentation methods mainly involve mixing the a priori target spectra with background spectra, which addresses issues related to subpixels in hyperspectral target detection.

To overcome the aforementioned challenges, this paper proposes a spectral data augmentation approach specifically designed for hyperspectral data. In this approach, a sufficient number of target samples are generated by modulating the known priori target spectrum with Gaussian white noise at varying signal-to-noise ratios (SNRs). This strategy simulates the aberrant target spectra that arise from different environmental conditions. As shown in Fig. 2, the augmented spectra maintain the overall shape and trend of the original spectral curve at the global level, ensuring semantic consistency. At the local level, the added noise introduces moderate perturbations that reflect practical spectral variability. This augmentation strategy offers two main advantages: On one hand, it preserves spectral consistency, ensuring that augmented samples remain representative of the true target class; On the other hand, it introduces controlled variability, enhancing the model's ability to generalize to real-world target conditions. It is evident that this strategy of spectral data augmentation has two advantages. This can, to some extent, alleviate the issue of spectral variability for the same object.

It should be noted that, the proposed model utilizes the unlabeled pixels, which are considered as background samples, along with the known prior target spectra to form negative sample pairs. While treating all unlabeled spectra as background may seem overly optimistic, particularly given the presence of target pixels within the image scence, the experimental results demonstrate that pre-training of generative adversarial network effectively mitigates this misallocation [26]. This

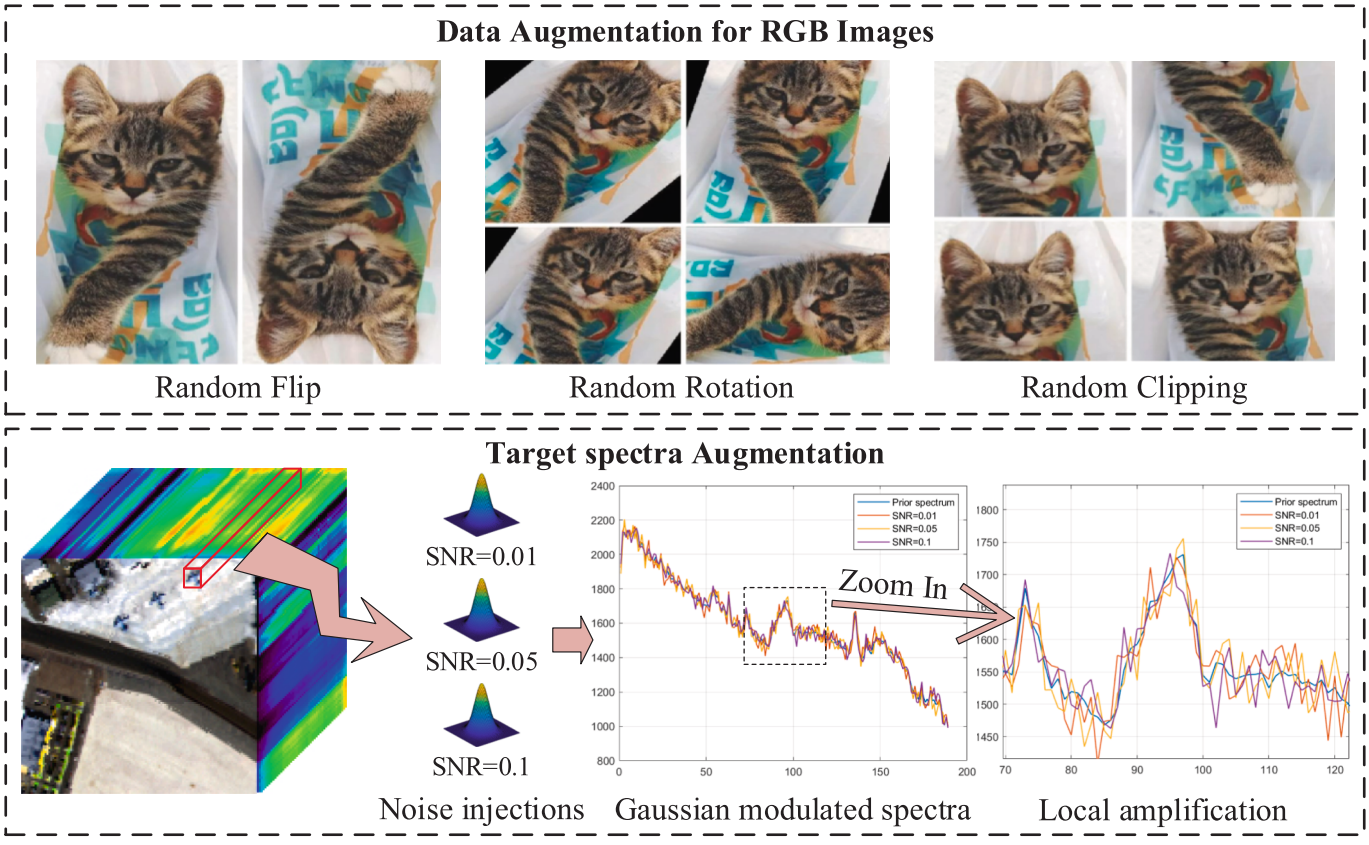

Fig. 2. Different approaches for data augmentation.

can be largely attributed to the fact that the target pixels in hyperspectral images typically account for less than 1 % of the total pixels, making such a treatment feasible.

2.2.2. Spectral metric learning with Siamese network

As shown in Fig. 3, a spectral metric Siamese network is designed for extracting spectral feature difference. This network is constructed based on the discriminator pretrained via the 1D-GAN introduced in Section 2.1. Both the upper and lower branches of the Siamese architecture are instantiated using the same structure as the discriminator, ensuring shared weights and identical feature extraction pathways. By transferring the pretrained parameters from the GAN discriminator, the Siamese network benefits from a strong initialization and can quickly adapt to the downstream detection task. At the end of the network, a spectral classification head—composed of a linear fully connected layer followed by a Sigmoid activation function—is appended to produce a similarity score.

Although the 1D-GAN discriminator is effective at extracting spectral features, it is not directly optimized for binary classification. Therefore, the Siamese network is fine-tuned using supervised sample pairs. In this process, positive sample pairs are formed by pairing target samples with the prior target spectra, while negative sample pairs are formed by pairing background samples with the prior target spectra. These pairs are used to train the network to distinguish between pixels that match the target signature and those that do not.

The output features from the twin branches are passed through the spectral classification head to produce a similarity score, which is interpreted as the probability that a given pixel belongs to either the "target" or the "background". The final output of the spectral metric Siamese network is presented as a score or label. To optimize the training process, the Binary Cross Entropy (BCE) function is used as the loss function, which is defined as follows:

$$Loss_{BCE} = -\frac{1}{B}\sum_{i=1}^{B}[y_i \cdot \log f_i + (1-y_i)\cdot \log(1-f_i)] \tag{4}$$

where f_i is the output of the Sigmoid layer of the spectral metric Siamese network and y_i denotes the label (1 or 0).

This loss function is well-suited to our binary target detection objective, enabling the Siamese network to learn a robust similarity metric that differentiates between target and background spectra based on pairwise spectral relationships.

2.3. Spectral-Spatial target detection

Following the spectral metric learning phase, each spectral pixel in the hyperspectral image (HSI) is paired with the prior target spectrum to form spectral pixel pairs. These pairs are then input into the trained spectral metric Siamese network to compute their similarity scores relative to the target, resulting in a spectral detection map.

Although the hyperspectral image is inherently a three-dimensional data cube containing both spatial and spectral information, the SN-HTD framework primarily focuses on spectral features during the detection phase. This may lead to suboptimal performance in complex scenes where spatial consistency is crucial. To address this limitation, a guide image filter [38] is incorporated to exploit the spatial information contained in the HSI. The guided image filter is an adaptive filtering technique that computes the filtering operation based on the content of a guiding image. Mathematically, the filter can be expressed as follows:

$$O_i = \sum_j W_{ij}(I)K_j \tag{5}$$

In essence, the guide image filter assumes a linear model between the image K to be detected and the output O of filter. The filter weight can be mathematically expressed as:

$$W_{ij}(I) = \frac{1}{|e|^2}\sum_{k:\,(i,\,j)\in e_k}\left(1+\frac{(I_i-\mu_k)(I_j-\mu_k)}{\sigma_k^2+\varepsilon}\right) \tag{6}$$

where e_k represents the window centered on the k^{th} pixel, with a window size of $(2r+1)\times(2r+2)$, and r represents the radius of the window. The mean and variance of the filter are represented by μ_k and σ_k^2, respectively. The penalty value is represented by ε, and $|e|$ represents the number of pixels in e_k. I_i and I_j represent two neighboring pixels in the bootstrap image.

The guide image filter is a smoothing operator that preserves edges and is more efficient than bilateral filters, particularly near image boundaries. To integrate spatial information into the detection process, the spectral detection result in fed into the guide image filter, using the first principal component of the HSI as the guide image, so as to obtain the final spectral-spatial detection result, where the first principal component is obtained through the principal components analysis (PCA) of HSI.

3. Experimental results and analysis

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

The experiments were performed on a system equipped with an Intel Core i5-8300H 8-core CPU and a NVIDIA GeForce RTX 1050ti GPU. The proposed SN-HTD, as well as the deep learning-based comparison methods, were implemented using Python 3.8.0 and PyTorch 1.12, with ROC analysis and result evaluation conducted in MATLAB R2022a, while other traditional comparison methods were implemented in MATLAB R2022a.

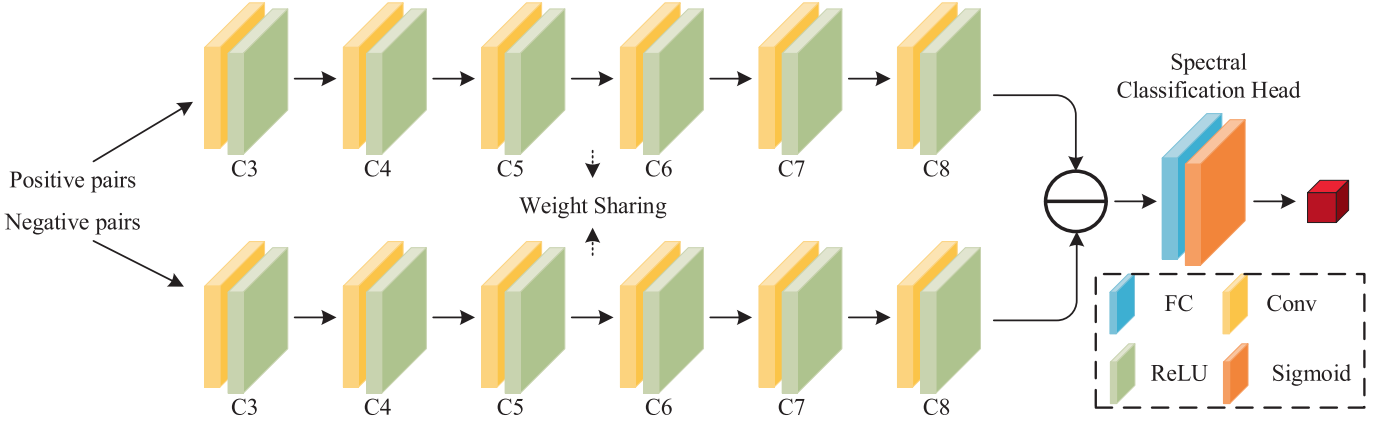

Fig. 3. The framework of spectral metric Siamese network.

3.1. Hyperspectral datasets

San Diego Dataset: The San Diego dataset was collected by AVIRIS over the San Diego Airport area, California, USA. It has 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. Two images from this dataset were used for experiments: San Diego1 (120 × 120) and San Diego2 (100 × 100), taken from the center and upper-left corner of the scene, respectively. The pseudo-color images and corresponding ground truth maps are shown in Fig. 4 (a)-(b) and Fig. 5 (a)-(b). The target, identified as aircraft in both images, contains 58 and 134 pixels, respectively.

Beach Dataset: The Beach dataset was captured by the AVIRIS sensor on Cat Island, with a spatial resolution of 17.2 m. The image used for the experiment has a size of 90 × 90 × 188 after removing the noise bands. The pseudo-color image and corresponding ground truth map, which includes 19 anomaly pixels, are shown in Fig. 6 (a) and (b).

Segundo Dataset: The Segundo dataset, also captured by the AVIRIS sensors in the EI Segundo region of California, USA, has a spatial resolution of 7.1 m and a wavelength range of 400–2500 nm. The whole image has 250 × 300 pixels, with a total of 224 bands. In the experiment, it is named the captured scene with the shape of 100 × 100 × 224 as Segundo. Its pseudo-color image and corresponding ground truth map are shown in Fig. 7 (a) and (b). There are 715 target pixels in the scene, including facilities such as oil storage tanks and towers.

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. 8 (a) and (b), including 21 target pixels of the types of roofs and cars.

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. 9 (a) and (b), including 39 target pixels.

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 under the curve (AUC) [39]. 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:

$$P_D(\tau) = \frac{n_{D,\tau}}{n_{D,\tau} + n_{FN,\tau}} \tag{7}$$

$$P_F(\tau) = \frac{n_{F,\tau}}{n_{F,\tau} + n_{TN,\tau}} \tag{8}$$

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 [39] 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 quanti-

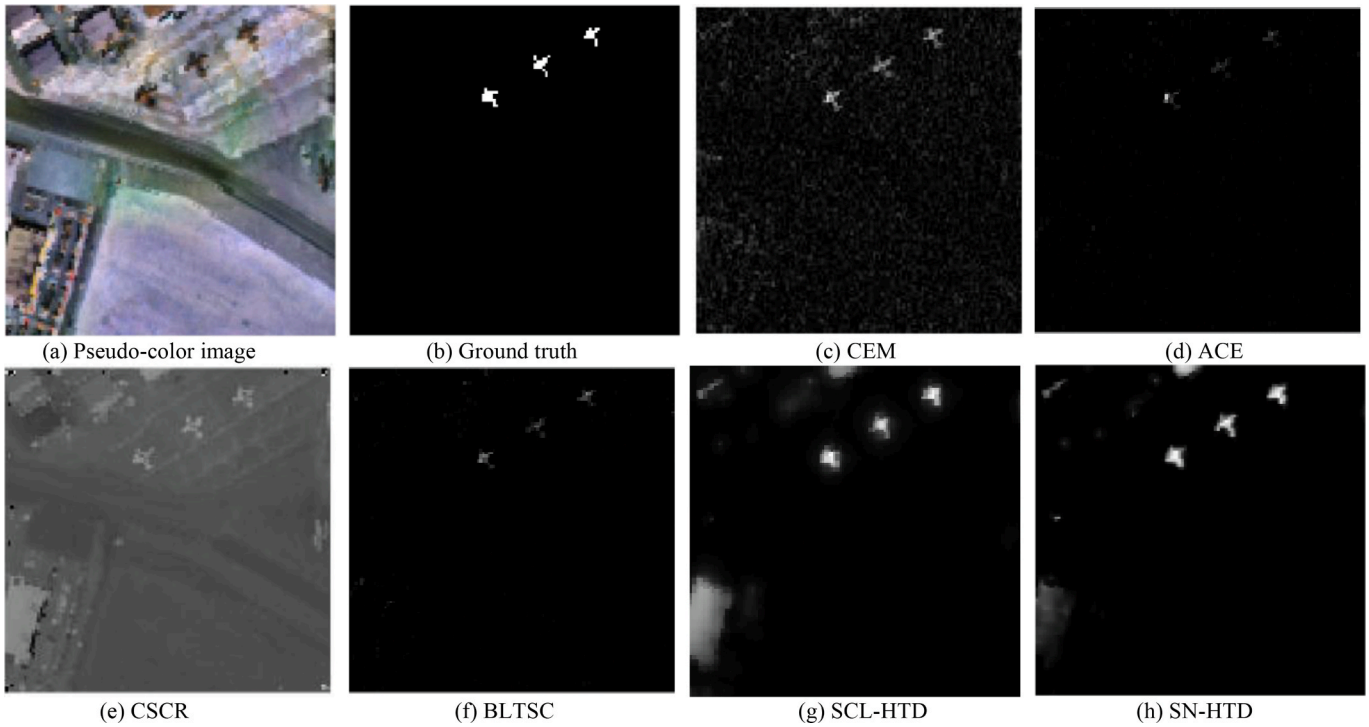

Fig. 4. Detection maps for San Diego1 dataset.

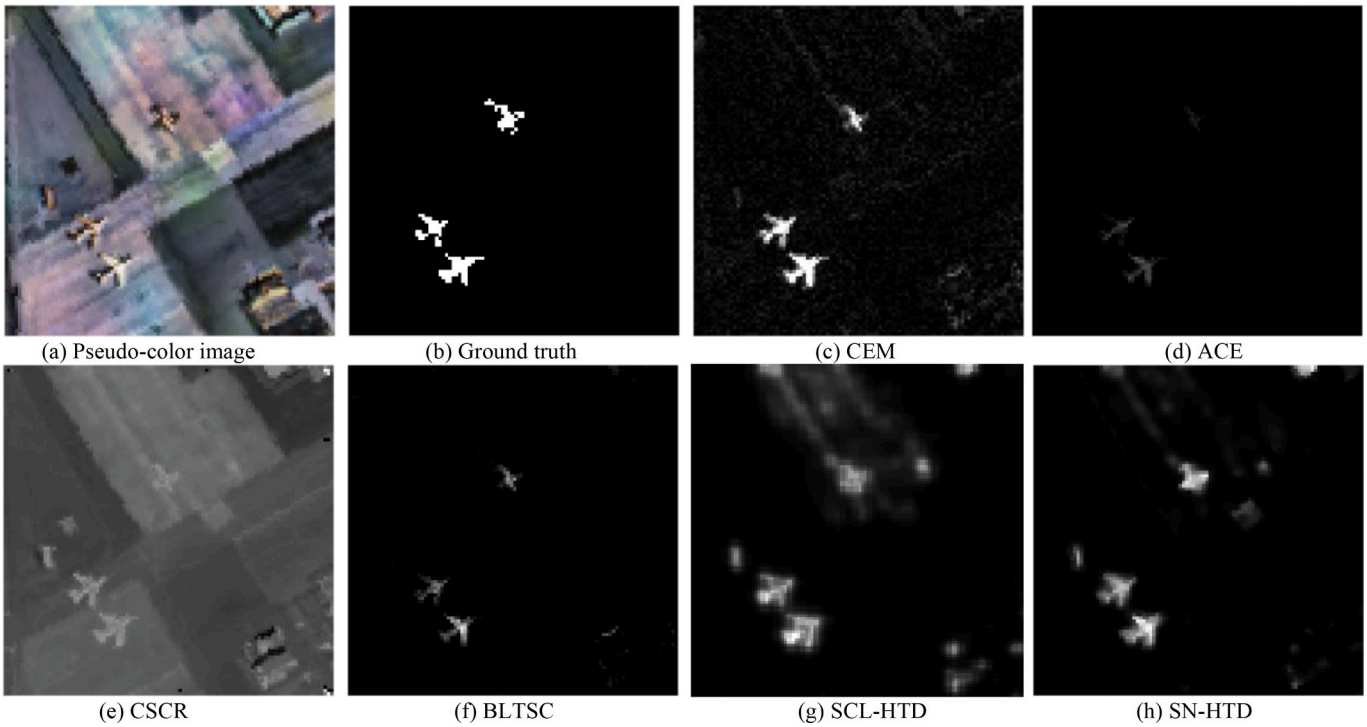

Fig. 5. Detection maps for San Diego2 dataset.

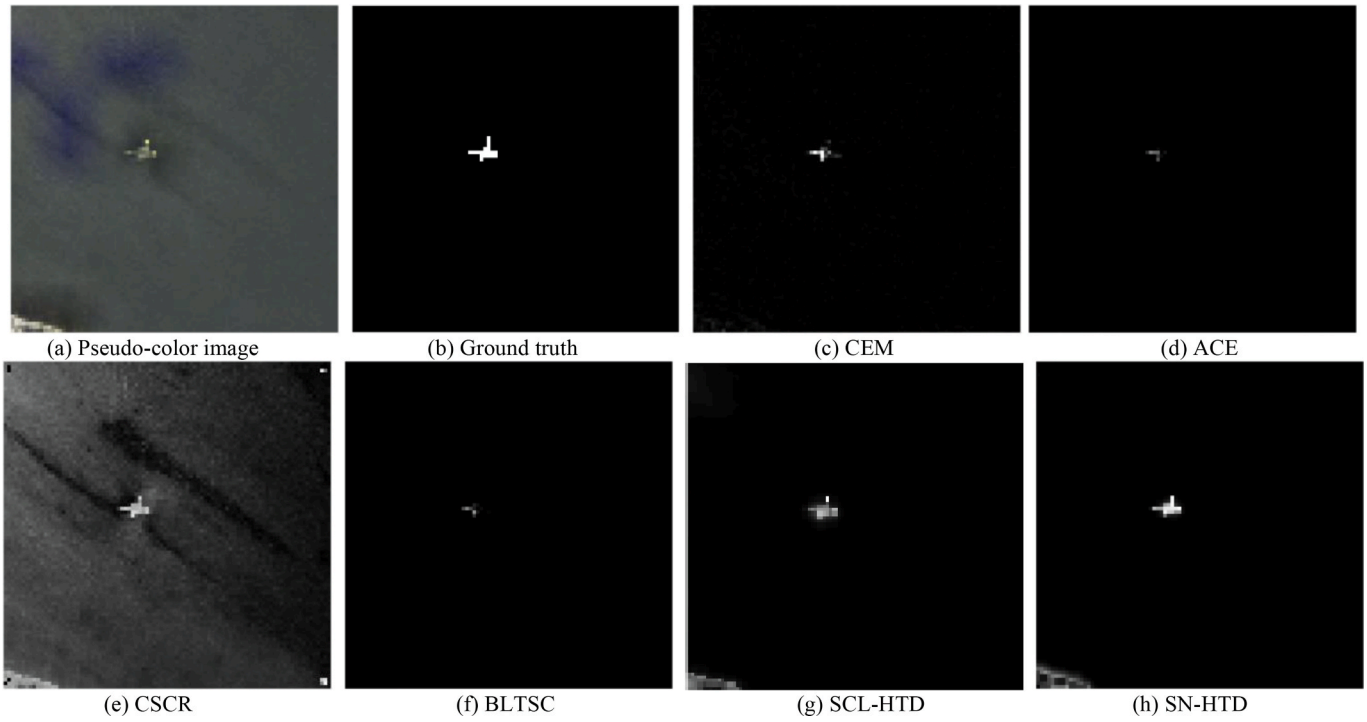

Fig. 6. Detection maps for Beach dataset.

tatively 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) 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. Besides, a new quantitative detection index designed in [39] 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:

$$AUC_{OD} = AUC(P_D, P_F) + AUC(P_D, \tau) - AUC(P_F, \tau) \quad (9)$$

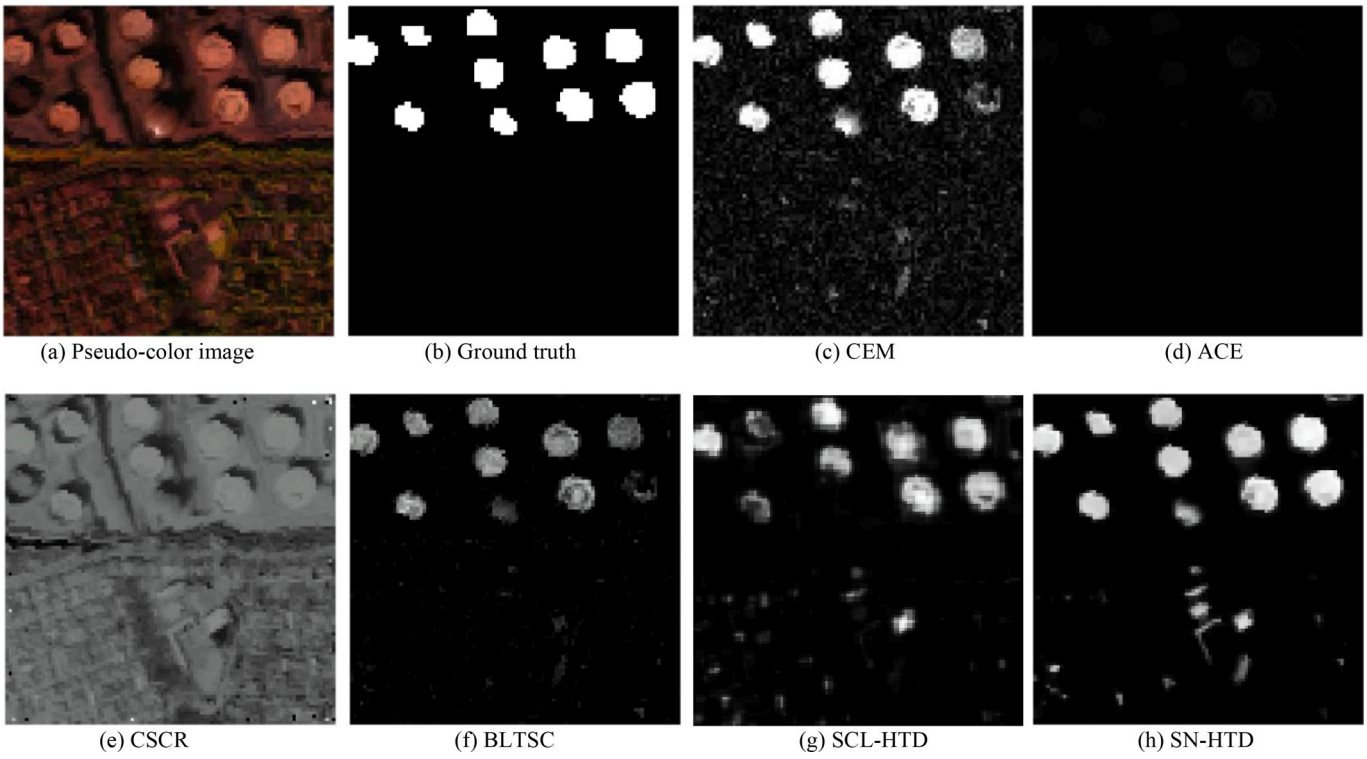

Fig. 7. Detection maps for Segundo dataset.

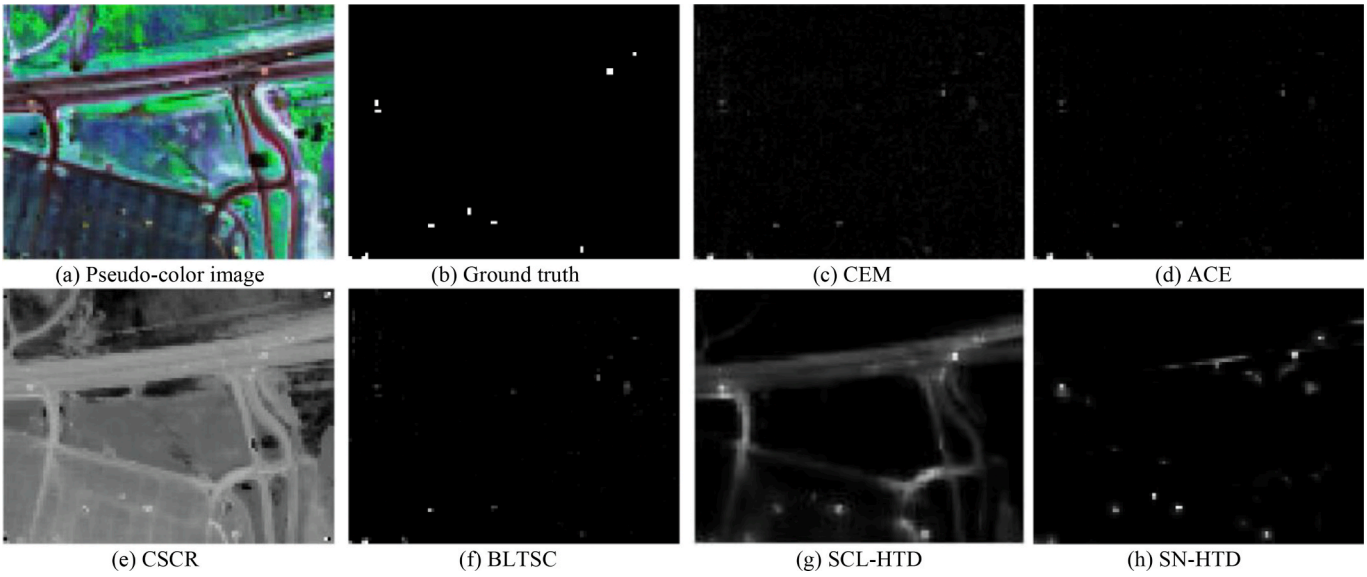

Fig. 8. Detection maps for HYDICE dataset.

3.3. Experimental Setup

This section mainly introduces the parameter setting used of the SN-HTD method, as well as the comparison methods.

3.3.1. Parameter settings of comparison methods

To evaluate the performance of the proposed SN-HTD method in the experiments, the following detection methods are compared with the proposed SN-HTD method: the classical detection method CEM [8] and ACE [6], the representation-based target detectors CSCR [12], and two deep learning-based methods the transfer learning-based BLTSC [21] and the SCLHTD [24] only using the background training samples. CEM and ACE do not have any parameters that need to be set artificially. For the CSCR detector, the outer and inner windows sizes are (11, 5) for SanDiego1 and Segundo datasets. For SanDiego2, Beach, HYDICE and Cuprite datasets, the outer and inner windows sizes are (11, 3). For the contrastive learning-based SCLHTD detection method, the training of the AAE is conducted in two stages. First, the encoder and decoder are optimized using the Adam optimizer with a learning rate of 1e-3. Subsequently, the generator and discriminator are trained separately: the generator is optimized using SGD with a learning rate of 1e-4, while the discriminator is trained with a learning rate of 1e-5. The AAE is trained for 20 epochs in total. The batch sizes for the San Diego 1, San Diego 2, Beach, Segundo, HYDICE, and Cuprite datasets are set to 240, 200, 180,

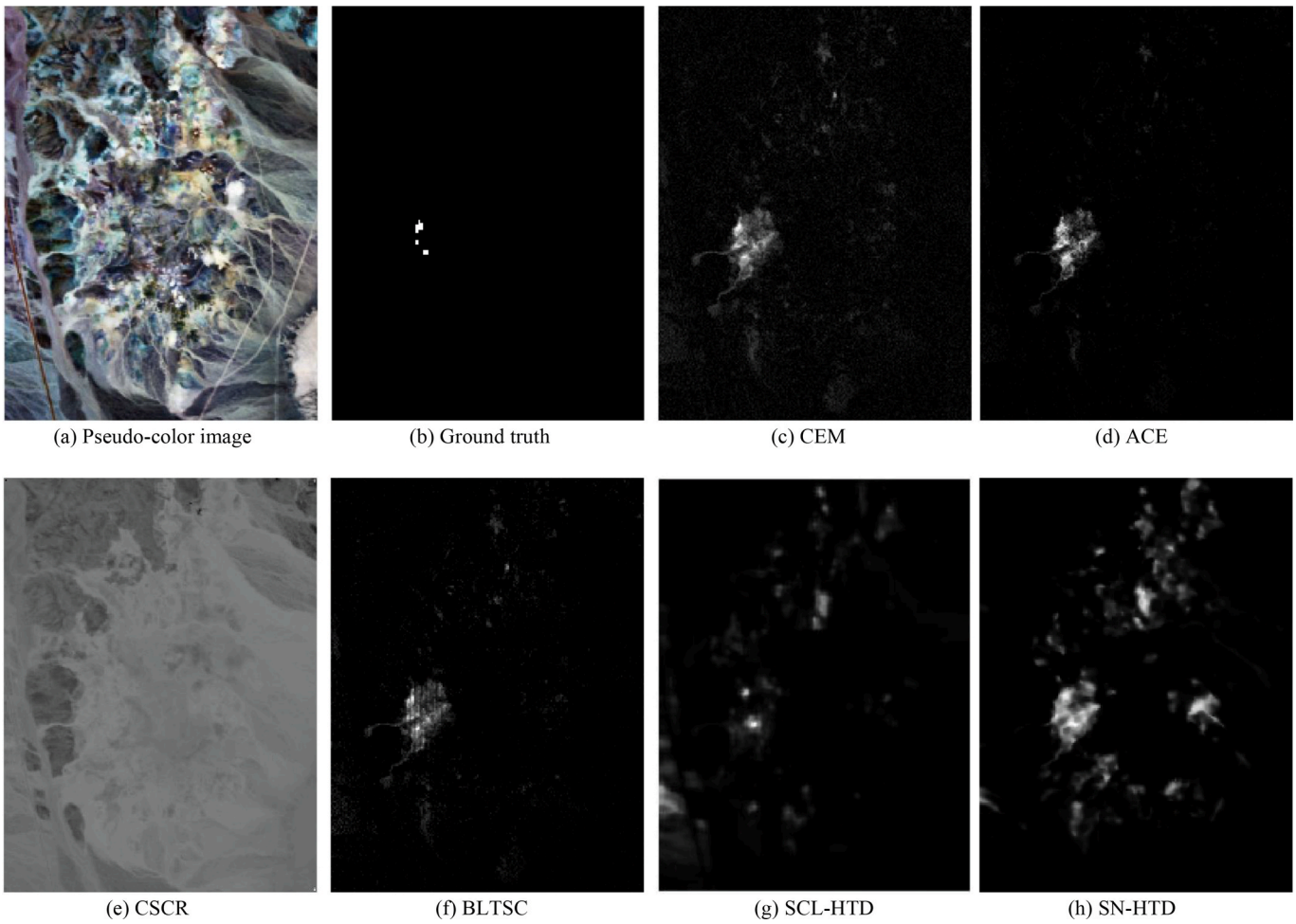

Fig. 9. Detection maps for Cuprite dataset.

200, 250, and 250, respectively. The output dimensionality of the latent code from the AAE encoder is set to 32, while the feature vector extracted by the encoder is fixed at a dimensionality of 64. For spectral-level contrastive learning, the number of training epochs, learning rate, and temperature coefficient are uniformly set to 100, 0.05, and 0.1, respectively, across all six datasets. The batch sizes for the San Diego A and San Diego B datasets are set to 240, 200, 180, 200, 250, and 250, respectively. For the final spectral–spatial joint target detection stage, the filter parameters are configured following the original settings described in the manuscript. For the BLTSC method only using the background training samples, the coarse detection is performed using the classical CEM method to gain sufficient background training data. It uses a learning rate and epoch set to 1e-4 and 500, during training for the five real datasets in this experiment, respectively.

3.3.2. Parameter settings of SN-HTD

The proposed SN-HTD method is implemented in four steps, including pre-training of 1D-GAN, spectral data augmentation, deep spectral metric learning and spectral–spatial target detection. The 1D-GAN is pre-trained using the training data. For the six real HSI datasets, when pre-training the 1D-GAN, the network is optimized by the Adam optimizer, and the learning rate is set to 1e-4. The batch sizes of SanDiego1, SanDiego2, Beach, Segundo, HYDICE and Cuprite datasets are set to 240, 200, 200, 240, 200 and 250, respectively. For spectral data augmentation, the prior target spectra are modulated using Gaussian white noise with random signal-to-noise ratios to gain a sufficiently large number of target samples. And the training data for target-background consists of these target samples and unlabeled pixels considered as background samples. When 1D-GAN pre-training converges, its discriminator is utilised to construct the spectral metric Siamese network. The priori target spectra are then paired with the target and background samples from the training data for target-background to obtain the positive and negative sample pairs, respectively. During spectral metric learning, the positive and negative sample pairs are fed into the spectral metric Siamese network to learn the more robust spectral difference characterization. The epoch, learning rate and batch size are all set to 100, 1e-4 and 256 for all HSI datasets. Finally, the spectral-spatial target detection is performed using the guide image filter with the penalty value of 0.04.

3.4. Results and analysis

For performance evaluation of the proposed SN-HTD method, five different state-of-the-art detection methods are used for comparison, which are the classical detection method CEM and ACE, the representation-based target detectors CSCR, and two deep learning-based methods including the transfer learning-based BLTSC and SCLHTD the only using the background training samples. Figs. 4–9 show the detection maps by the above six methods for the SanDiego1, SanDiego2, Beach, Segundo, HYDICE and Cuprite datasets.

It can be seen from the detection maps and ground truth maps that CEM, ACE and BLTSC miss many target pixels. However, hyperspectral data in real scenes exhibit usually show strong non-Gaussianity and nonlinearity, leading to a decrease in target detection accuracy of CEM and ACE. The CSCR can detect the most of targets, but there is poor background suppression and small separation between target and

background, resulting in the inability to visually identify targets, and the detection performance decreases when the background of the detection scene becomes complex. The SCLHTD method is inspired by self-supervised learning and aims to reduce the HTD model's dependence on high-quality prior information. It achieves spectral similarity and dissimilarity discrimination by constructing a spectral-level contrastive learning task and extracting features via a backbone network. Specifically, the original HSI is sampled into odd and even bands, each augmented using an adversarial convolutional autoencoder with spectral residual channel attention. Two augmented samples from the same pixel form positive pairs, while samples from other pixels serve as negative pairs for contrastive learning. To suppress background interference, an edge-preserving filter is applied. Although this approach enhances spectral discrimination, the sample pair generation based on band sampling and augmentation may cause some spectral detail loss, and the edge filtering has limited ability to model complex backgrounds, which may reduce detection performance on weak targets. The BLTSC performs a coarse detection of the HSI to be detected through CEM and finds reliable background samples for training AAE. After reconstructing the original HSI using the trained AAE, the background of the reconstructed HSI was reconstructed relatively accurately, and the target was reconstructed poorly. The difference between the reconstructed and original HSI was considered the target. The detection performance of BLTSC will be affected when CEM is not good enough to detect HSI. The proposed SN-HTD method shows excellent detection performance with high target detection accuracy, and visually obvious identification of the target in the detection maps obtained on real HSI datasets.

Subjective evaluation of the detection maps visually has limitations, and to quantitatively evaluate the performance of the SN-HTD method, 3D ROC curves and their corresponding the 2D ROC curves (P_D, P_F), (P_D, τ), and (P_F, τ) with the AUC of 2D ROC curves are used for quantitative evaluation. The 3D ROC curve is used to indicates the comprehensive detection capability of detectors, as shown in Figs. 10–15(a). The 2D ROC curve of (P_D, P_F) is used to demonstrate the effectiveness of detectors, as shown in Figs. 10–15(b). For the six real HSI datasets in the experiment, the ROC curve of the SN-HTD outperforms the curves of other detectors. The 2D ROC curve of (P_D, τ) is used to evaluate the preservation ability of the detector for the target, as shown in Figs. 10–15(c). The SN-HTD outperforms CEM, ACE and BLTSC, but CSCR and SCLHTD performs not weaker than SN-HTD on some of datasets. However, for the 2-D ROC curve of (P_F, τ), which evaluates the detector background suppression ability. The SN-HTD has relatively weak performance, but better than CEM, CSCR and SCLHTD.

The specific values of AUC (P_D, P_F), AUC (P_D, τ), AUC (P_F, τ), and AUC_{OD} for different detectors on the real datasets are given in Tables 1–6. The optimal results are shown in bold, and the suboptimal results are underlined. As can be seen from the tables, BLTSC performs the best in background suppression but being worse in target preservation. CSCR perform good in target preservation, but its background suppression ability is much weaker than SN-HTD. The AUC (P_D, P_F) and AUC_{OD} values of the proposed SN-HTD remain optimal on the HSI datasets in the experiment which exhibit better comprehensive detection ability. The AUC (P_D, τ) values remain suboptimal on the HYDICE datasets, but there are the suboptimal results. The AUC (P_F, τ) values remain inferior to ACE and BLTSC, but better than other methods.

To evaluate the effectiveness of SN-HTD in separating target from background, the target–background separability boxplot is used to show the separation degree of target and background. Fig. 16 shows the target–background separability boxplot for the six compared methods and the proposed SN-HTD method on the real HSI datasets. 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. Although the background suppression ability of SN-HTD is not the best among the comparison detection methods, it displays the excellent target–background separability, which indicates that the spectral metric learning enables the model to effectively learn the ability to discriminate spectral differences.

3.5. Ablation Study

3.5.1. Impact of pre-training for 1D-GAN

To assess the role of the pre-training 1D-GAN during the spectral metric learning, this subsection conducts a set of ablation experiments to demonstrate the effect of the pre-training 1D-GAN on target detection accuracy.

The first experiment uses a small amount of labeled data to fine-tune the discriminator of the training convergence 1D-GAN to obtain a target detector. The second experiment directly uses the main structure of the discriminator to construct a spectral metric Siamese network, which does not inherit the parameters of the discriminator obtained through the pre-training. The positive and negative samples from the training of target-background are then used for the spectral metric learning. And the third experiment is the proposed SN-HTD method. Table 7 illustrates the effect of the pre-training for 1D-GAN on the detection accuracy of HTD. The AUC (P_D, P_F) values in Table 7 are a direct measure of the similarity between the pixel spectrum to be detected and the prior target spectrum. It can be seen from Table VII that the performance of the first experiment is not weaker or even better than other ones on a few datasets, which shows that the discriminator of the training convergence 1D-GAN is able to extract useful information for detection and quickly adapt to the target detection task by fine-tuning. And the detection accuracy of the third experiment is higher than other ones on almost all datasets. It proves that without the pre-training or the spectral metric learning, the SN-HTD method cannot achieve the optimal detection performance.

3.5.2. Impact of proportion for target and background samples

To investigate the effect of proportion for target and background

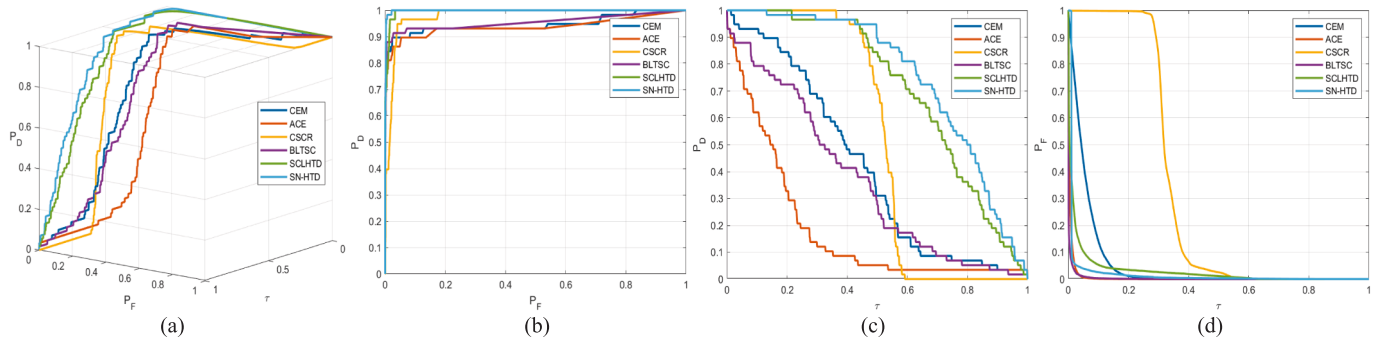

Fig. 10. ROC curves for San Diego1 dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

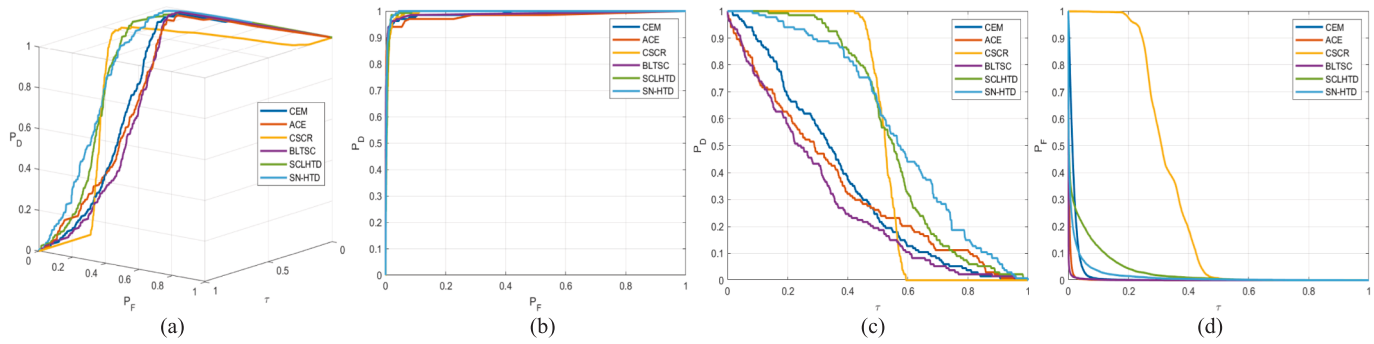

Fig. 11. ROC curves for San Diego2 dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

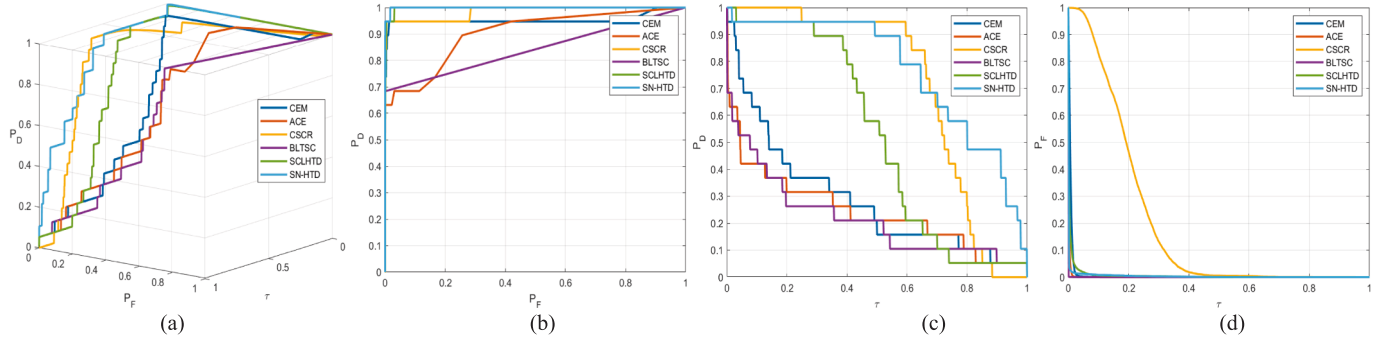

Fig. 12. ROC curves for Beach dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

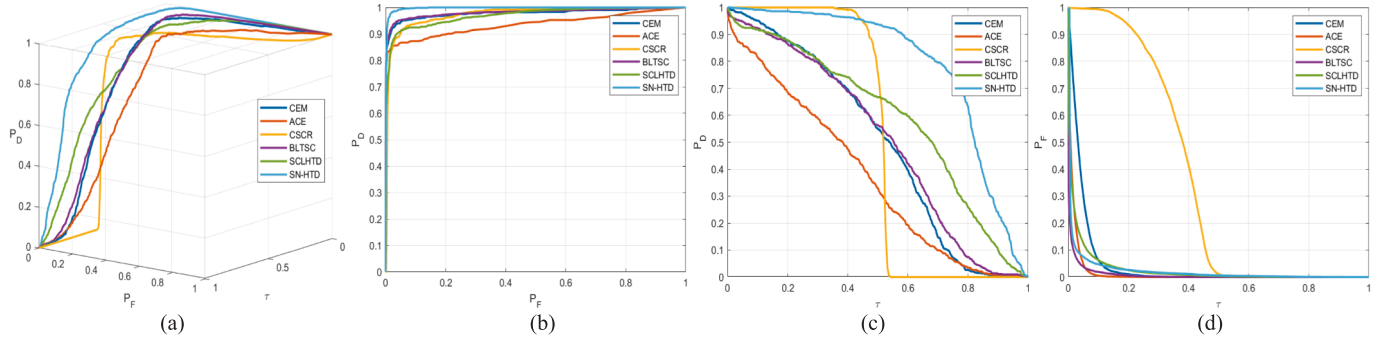

Fig. 13. ROC curves for Segundo dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

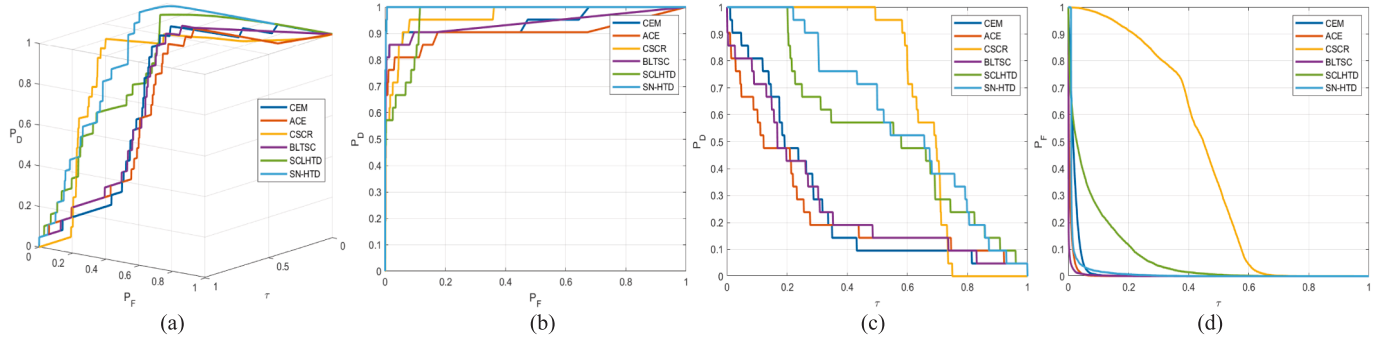

Fig. 14. ROC curves for HYDICE dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

samples on the detection accuracy of HTD, different proportions are used to conduct a series of repetitive experiments.

As illustrated in Fig. 17 (San Diego1), the blue, orange and yellow bars represent the detection performance of the first, second and third experiments, respectively. And the line with stars indicates the average detection result of three experiments. Only the proportions are different in three experiments, and the other conditions are the same. The experimental results unequivocally demonstrate that when the

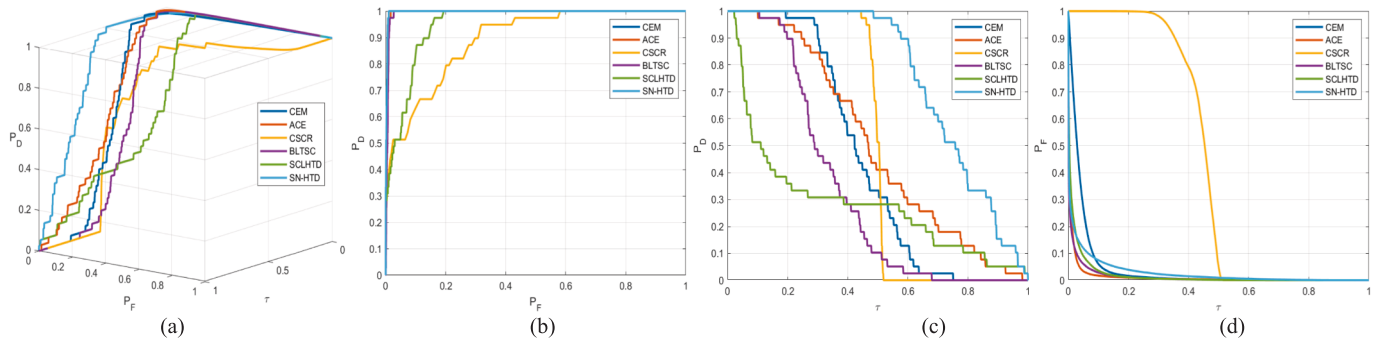

Fig. 15. ROC curves for Cuprite dataset. (a) 3D ROC curve. (b) 2D ROC of (P_D, P_F). (c) 2D ROC of (P_D, τ). (d) 2D ROC of (P_F, τ).

Table 1
Accuracy Comparison of Different Methods for San Diego1 Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9457	0.9376	0.9779	0.9551	0.9960	**0.9988**
AUC (P_D, τ)	0.4131	0.1817	0.5189	0.3602	0.7089	**0.7566**
AUC (P_F, τ)	0.0554	0.0068	0.3338	**0.0042**	0.0294	0.0193
AUC_{OD}	1.3034	1.1126	1.1630	1.3110	1.6755	**1.7360**

*The best results are in bold, while the second-best results are underlined.

Table 2
Accuracy Comparison of Different Methods for San Diego2 Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9909	0.9818	0.9923	0.9891	0.9945	**0.9963**
AUC (P_D, τ)	0.3568	0.3387	0.5240	0.2870	0.5585	**0.5824**
AUC (P_F, τ)	0.0186	0.0043	0.3252	**0.0026**	0.0367	0.0176
AUC_{OD}	1.3291	1.3161	1.1911	1.2735	1.5163	**1.5611**

*The best results are in bold, while the second-best results are underlined.

Table 3
Accuracy Comparison of Different Methods for Beach Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9534	0.9026	0.9832	0.8418	0.9978	**0.9991**
AUC (P_D, τ)	0.2875	0.2411	0.7139	0.2160	0.5193	**0.7694**
AUC (P_F, τ)	0.0080	0.0028	0.1989	**0.0015**	0.0072	0.0051
AUC_{OD}	1.2330	1.1409	1.4981	1.0563	1.5098	**1.7633**

*The best results are in bold, while the second-best results are underlined.

Table 4
Accuracy Comparison of Different Methods for Segundo Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9785	0.9359	0.9731	0.9813	0.9698	**0.9974**
AUC (P_D, τ)	0.4969	0.3692	0.5092	0.5067	0.5905	**0.7892**
AUC (P_F, τ)	0.0425	0.0153	0.3614	**0.0093**	0.0271	0.0234
AUC_{OD}	1.4330	1.2898	1.1209	1.4787	1.5333	**1.7633**

*The best results are in bold, while the second-best results are underlined.

Table 5
Accuracy Comparison of Different Methods for Hydice Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9425	0.9039	0.9661	0.9385	0.9637	**0.9991**
AUC (P_D, τ)	0.2675	0.2428	**0.6645**	0.2733	0.5399	0.6065
AUC (P_F, τ)	0.0210	0.0072	0.4261	**0.0035**	0.0727	0.0170
AUC_{OD}	1.1890	1.1395	1.2045	1.2083	1.4308	**1.5886**

*The best results are in bold, while the second-best results are underlined.

Table 6
Accuracy Comparison of Different Methods for Cuprite Dataset.

Method	CEM	ACE	CSCR	BLTSC	SCLHTD	Proposed
AUC (P_D, P_F)	0.9954	0.9961	0.8916	0.9937	0.9488	**0.9973**
AUC (P_D, τ)	0.4426	0.5051	0.4979	0.3322	0.2830	**0.7561**
AUC (P_F, τ)	0.0426	0.0138	0.4402	**0.0136**	0.0188	0.0289
AUC_{OD}	1.3954	1.4874	0.9493	1.3123	1.2129	**1.7245**

*The best results are in bold, while the second-best results are underlined.

proportion reaches 1:6, the result reaches the peak and then begins to gradually decline. This change proves that an excessive amount of training data for deep networks can lead to performance degradation due to the overfitting of the training. The three repeated experimental results for the same conditions are due to the unstable training. This is a common flaw of the HTD methods, which are through generating pixel labels.

3.6. Time cost

Table 8. lists the time consumption of the comparison methods and the proposed SN-HTD method. The time consumptions of the classical HTD method and the machine learning-based HTD method are much less than those of the deep learning-based HTD method. This is reasonable since the deep learning-based methods need to be trained to obtain the parameters of the networks. Among three deep learning-based methods, the training time for BLTSC includes the time to find reliable background samples using coarse detection and the time to train the AAE using the background samples. And the time for SN-HTD includes the time for spectral data augmentation, pre-training and spectral metric learning. In terms of training time of the deep learning-based HTD method, SN-HTD consumes less training time than BLTSC and SCLHTD. The training time for SN-HTD itself is approximately consistent across different datasets used. This is because the experiments use the same proportion of training data for target and background, with only the size of spectral dimension varying. Once the model has been trained well, the detective efficiency relies on the detection time. The detection time of the deep learning-based detection methods starts with loading the model and ends with the detection results. The detection time of the proposed SN-HTD is less than that of the other two deep learning-based methods (BLTSC and SCLHTD) using the same HSI datasets.

4. Conclusion

To address the problem of insufficient target samples in deep learning, a deep spectral metric Siamese network for hyperspectral target detection is proposed in this paper. For expanding the target samples, spectral data augmentation is proposed to mine the supervision information of HSIs to be detected, and spectral metric learning is then designed to make the model learn the difference between spectra. Specifically, the 1D-GAN is firstly pre-trained through the hyperspectral

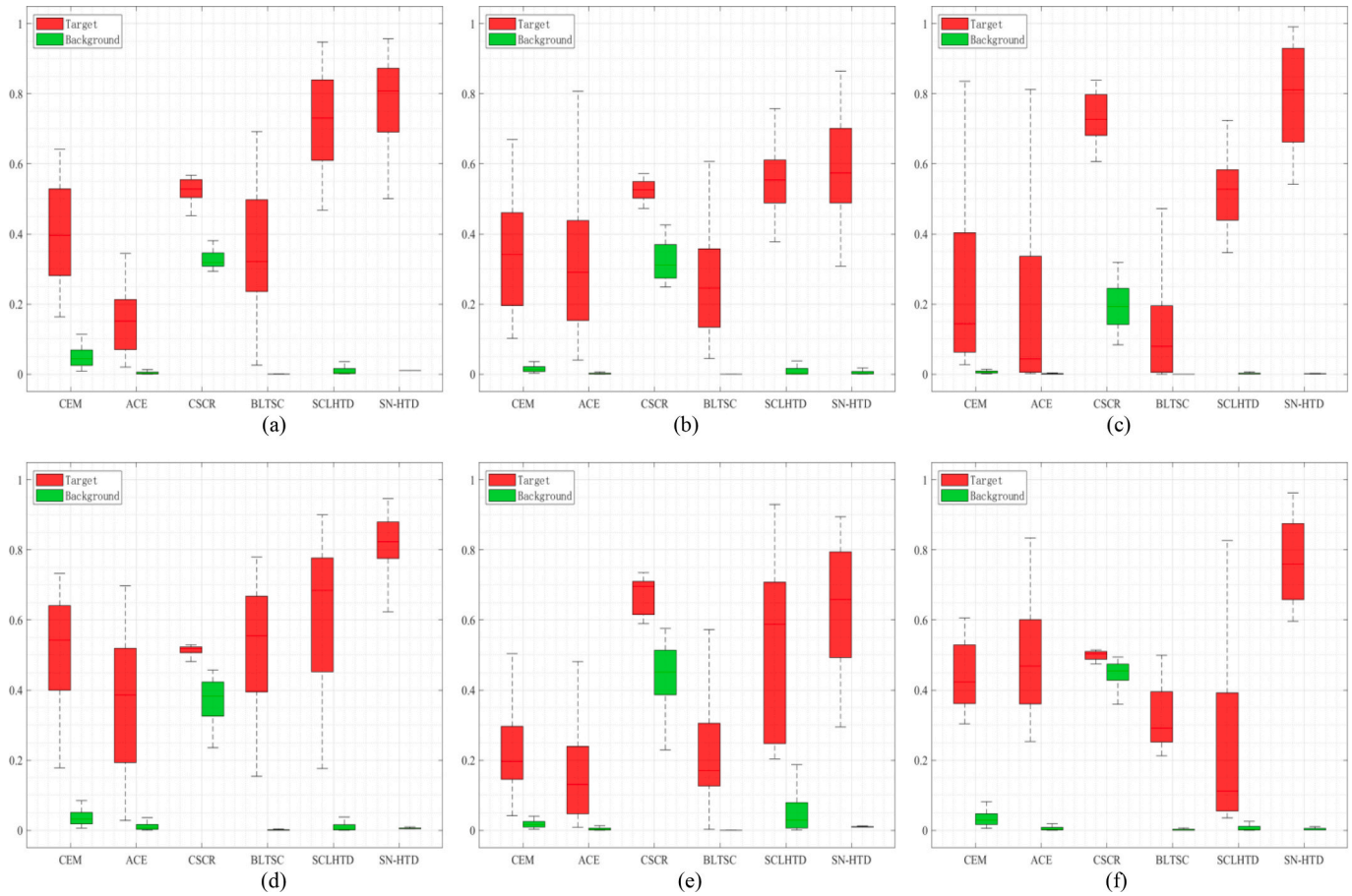

Fig. 16. Target-background separability boxplots for different datasets. (a) San Diego1, (b) San Diego2, (c) Beach, (d) Segundo, (e) HYDICE, (f) Cuprite.

Table 7
The Effect of Ablation Experiments.

Options	1st	2nd	3rd
Pre-training	√	×	√
Spectral metric learning	×	√	√
San Diego1	0.9824	0.9818	**0.9937**
San Diego2	0.9906	0.9733	**0.9932**
Beach	0.9768	0.9864	**0.9910**
Segundo	0.9976	**0.9982**	0.9980
HYDICE	0.9548	0.9845	**0.9923**
Cuprite	0.8194	0.9896	**0.9930**

*The best results are in bold, while the second-best results are underlined.

image. Then, a spectral data augmentation method for hyperspectral data is designed so as to simulate spectral aberrations due to different environmental factors. Through this data augmentation, there is a sufficiently large number of target samples. And these samples are combined with unlabeled pixels considered as background samples to form the training data of target-background. The binary classification problem for the HTD is then converted into the spectral similarity metric problem. With the primary structure of the discriminator of 1D-GAN obtained through pre-training, a spectral metric Siamese network is constructed, and inherits the parameters of the discriminator for adapting quickly to target detection. Next, the priori target spectra are paired with the samples from the training data of target-background to obtain positive and negative sample pairs, respectively. And these sample pairs are fed into the metric Siamese network for the spectral difference metric learning. The SN-HTD obtains detection result utilizing spectral information by measuring the difference between the spectra to be detected and the priori target spectra. Finally, combining

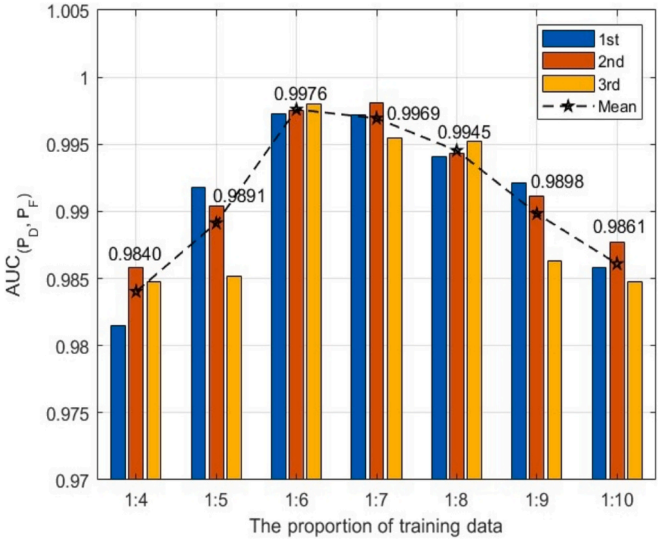

Fig. 17. Different proportion for target and background samples.

the spatial information, the spectral detection result is filtered by using the first principal component of the to-be-detected HSI to obtain the final target detection result. Comprehensive experiments show that the SN-HTD method is superior to other comparison detectors.

Although the proposed SN-HTD demonstrates promising performance in hyperspectral target detection, particularly in limited labeled samples, it still has certain limitations. While guided image filtering is

Table 8
Time Consumption of Different Methods.

Method		AVIRIS 1	AVIRIS 2	Beach	Segundo	HYDICE	Cuprite
CEM		0.1459	0.1112	0.0944	0.1184	0.9967	0.4859
ACE		0.3008	0.1861	0.1637	0.2116	0.1540	0.8824
CSCR		8.7746	4.9812	4.2521	5.4585	4.0291	26.5119
BLTSC	Train	2190.0375	1460.1186	1299.2160	1344.4637	1228.7092	6277.6399
	Detect	9.6767	7.6201	7.4708	9.5323	7.5766	43.9728
SCLHTD	Train	630.3452	579.4321	618.2576	674.3249	594.7715	2973.2121
	Detect	4.9342	4.4521	4.1832	5.7219	3.7213	30.2132
Proposed	Train	**125.2423**	**119.4939**	**122.8407**	**134.1689**	**119.8846**	**122.8450**
	Detect	**7.1359**	**3.6935**	**3.0481**	**4.5664**	**2.9522**	**27.0587**

*The best training and testing time using deep learning-based methods are in bold.

used to incorporate spatial information and enhance detection accuracy, the overall framework lacks an explicit background suppression mechanism. As a result, its ability to suppress interference from complex backgrounds remains limited. This limitation is particularly evident in the AUC (P_F, τ) metric, which reflects the effectiveness of background suppression. To address this in future work, integrating explicit background modeling or suppression modules into the current framework would be considered. For example, spatial weighting strategies based on attention mechanisms could be incorporated to improve robustness and generalization in complex scenes.

CRediT authorship contribution statement

Yulei Wang: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Conceptualization. **Chao Deng:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Hongzhou Wang:** Writing – original draft, Software, Methodology, Data curation. **Enyu Zhao:** Visualization, Validation, Project administration, Investigation, Formal analysis. **Qiongqiong Lan:** Visualization, Validation, Resources.

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 (61801075, 42271355), Natural Science Foundation of Liaoning Province (2022-MS-160), and the Fundamental Research Funds for the Central Universities (3132025251).

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. 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.
[3] 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.
[4] 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.
[5] Y. Wang, L. Wang, C. Yu, E. Zhao, M. Song, C.-H. Wen, C.-I. Chang, 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.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] R. Zhao, Z. Shi, Z. Zou, Z. Zhang, Ensemble-based cascaded constrained energy minimization for hyperspectral target detection, Remote Sens. 11 (11) (2019) 1310, https://doi.org/10.3390/rs 11111310.
[11] 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.
[12] 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.
[13] C. Yu, Y. Zhu, Y. Wang, E. Zhao, Q. Zhang, X. Lu, Concern with center-pixel labeling: center-specific perception transformer network for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens. 63 (2025) 5514614, https://doi.org/10.1109/TGRS. 2025. 3573233.
[14] 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.
[15] Y. Yang, Y. Wang, H. Wang, L. Zhang, E. Zhao, M. Song, C. Yun, 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.
[16] E. Zhao, N. Qu, Y. Wang, C. Gao, J. Zeng, TEBS: temperature–emissivity–driven band selection for thermal infrared hyperspectral image classification with structured state-space model and gated attention, Int. J. Appl. Earth Obs. Geoinf. 142 (2025) 104710, https://doi.org/10.1016/j.jag.2025.104710.
[17] 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.
[18] 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.
[19] 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.
[20] D. Zhu, B. Du, L. Zhang L, 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.
[21] 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.
[22] Y. Shi, J. Li, Y. Yin, B. Xi, Y. Li, Hyperspectral target detection with macro-micro feature extracted by 3-D residual autoencoder, IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 12 (12) (2019) 4907–4919, https://doi.org/10.1109/JSTARS.2019.2939833.
[23] X. Chen, Y. Wang, Z. Che, Contrastive learning for hyperspectral target detection, Proc. IEEE Int. Geosci. Remote Sens. Symp. (2022) 887–890, https://doi.org/10.1109/IGARSS46834.2022.9883439.
[24] 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.

[25] J. Bromley, I. Guyon, Y. LeCun, E. Sckinger, R. Shah, Signature verification using a "Siamese" time delay neural network, Int. J. Pattern Recognit. Artif Intell. 7 (4) (1993) 669–688, https://doi.org/10.1142/S0218001493000339.
[26] Y. Gao, Y. Feng, X. Yu, Hyperspectral target detection with an auxiliary generative adversarial network, Remote Sens. 13 (21) (2021) 4454, https://doi.org/10.3390/rs13214454.
[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, A. Courville, Y. Bengio, Generative adversarial networks, Adv. Neural Inf. Process. Syst. 27 (2014) 2672–2680, https://doi.org/10.1145/3422622.
[28] J. Ji, Y. Zhao, Y. Zhang, C. Wang, X. Ma, F. Huang, J. Yao, Infrared and visible image fusion of generative adversarial network based on multi-channel encoding and decoding, Infrared Phys Techn. 134 (2023) 104853, https://doi.org/10.1016/j.infrared.2023.104853.
[29] Z. Lin, V. Sekar, G. Fanti, Why spectral normalization stabilizes gans: analysis and improvements, Adv. Neural Inf. Process. Syst. 34 (2021) 9625–9638.
[30] C. Fang, H. Ma, J. Li, A finger vein authentication method based on the lightweight Siamese network with the self-attention mechanism, Infrared Phys Techn. 128 (2023) 104483, https://doi.org/10.1016/j.infrared.2022.104483.
[31] X. Yang, Y. Li, D. Li, S. Wang, Z. Yang, Siam-AUnet: an end-to-end infrared and visible image fusion network based on gray histogram, Infrared Phys. Techn. 141 (2024) 105488, https://doi.org/10.1016/j.infrared.2024.105488.
[32] P. Wang, H. Sun, X. Bai, S. Guo, D. Jin, Traffic thermal infrared texture generation based on siamese semantic CycleGAN, Infrared Phys. Techn. 116 (2021) 103748, https://doi.org/10.1016/j.infrared. 2021.103748.
[33] C. Yu, Y. Liu, S. Wu, Z. Hu, X. Xia, D. Lan, X. Liu, Infrared small target detection based on multiscale local contrast learning networks, Infrared Phys. Techn. 123 (2022) 104104, https://doi.org/10.1016/j.infrared.2022.104107.
[34] M. Caron, I. Misra, J. Mairal, P. Bojanowski, A. Joulin, Unsupervised learning of visual features by contrasting cluster assignments, Int. Conf. Neural Inf. Process. Syst. 33 (2020) 9912–9924.
[35] 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) 3387985, https://doi.org/10.1109/JSTARS.2024.
[36] R. Chen, S. Liu, Z. Miao, F. Li, GFSNet: Generalization-friendly siamese network for thermal infrared object tracking, Infrared Phys Techn. 123 (2022) 104190, https://doi.org/10.1016/j.infrared. 2022.104190.
[37] Y. Wang, S. Ma, X. Shen, A novel video face verification algorithm based on TPLBP and the 3D Siamese-CNN, Electronics 8 (12) (2019) 1544, https://doi.org/10.3390/electronics8121544.
[38] K. He, J. Sun, X. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 35 (6) (2013) 1397–1409, https://doi.org/10.1109/TPAMI.2012.21.
[39] C.-I. Chang, An effective evaluation tool for hyperspectral target tetection: 3D receiver operating characteristic curve analysis, IEEE Trans. Geosci. Remote Sens. 59 (6) (2021) 5131–5153, https://doi.org/10.1109/TGRS.2020.3021671.