# Hyperspectral Target Detection via Global Spatial–Spectral Attention Network and Background Suppression

Xiaoyi Wang, Liguo Wang, Qunming Wang, Anna Vizziello, *Senior Member, IEEE*, and Paolo Gamba, *Fellow, IEEE*

***Abstract*—The accuracy of hyperspectral target detection is often affected by the problems of spectral variation and complex background distribution. Inspired by the powerful representational ability of deep learning, we proposed a three-dimensional (3-D) convolution-based global spatial–spectral attention network (GS$^2$A-Net) to deal with spectral variation in hyperspectral images (HSIs). GS$^2$A-Net uses 3-D convolution kernels of different sizes to capture local spatial and spectral features to achieve multiscale information interaction. Different from the previous 2-D attention mechanisms, GS$^2$A-Net simultaneously considers the information in the spatial and spectral dimensions, and creates a weight map consistent with the size of the original HSI. Furthermore, we proposed a new background suppression strategy based on the spectral angle mapping to achieve more accurate target detection, which can preserve the targets as much as possible when suppressing the background. The method was validated through experiments on five real-world HSI datasets. Compared with several classical and deep-learning-based methods, the proposed method exhibits greater detection accuracy.**

***Index Terms*—Background suppression, global spatial–spectral attention network (GS$^2$A-Net), hyperspectral target detection (HTD), spectral variation.**

## I. Introduction

HYPERSPECTRAL sensors record hyperspectral images (HSIs) with hundreds of continuous and narrow bands, reaching a spectral resolution of around 10 nm. Due to its powerful representational ability, HSIs have been widely used in land cover classification [1], [2], target detection [3], [4], anomaly detection [5], [6], image unmixing [7], [8], and change detection [9]. Hyperspectral target detection (HTD) is one of the most challenging issues in these applications due to the limited amount of known information about target spectra and the background. Indeed, HTD can be regarded as a problem of weakly supervised binary classification.

Over the past few decades, HTD has received extensive attention. The simplest methods are distance-based ones, such as spectral angle mapping (SAM). Other classical HTD methods include statistic-based, subspace-based, and representation-based methods.

In statistic-based methods, the most typical methods are the spectral matched filter (SMF) [10], the adaptive coherence estimator (ACE) [11], and the constrained energy minimization (CEM) [12]. Both SMF and ACE first estimate the covariance matrix and mean value of HSI, and then use the generalized likelihood ratio test to achieve target detection. The core of the CEM method is to constrain the energy in the target direction by a designed linear filter while minimizing the energy in the other directions. To effectively distinguish between the background and targets, several enhanced versions of CEM were developed. For example, Zou and Shi [13] proposed a hierarchical CEM that uses a layer-by-layer filtering strategy to suppress background. Chen et al. [14] used an extended morphological attribute profile to initially separate the background and the targets, and proposed a diverse-direction CEM to further reduce the interference from the background.

With respect to subspace-based methods, Chang [15] proposed an orthogonal subspace projection (OSP) method. It assumes that the target subspace is orthogonal to the background subspace and maximizes the signal-to-noise ratio (SNR) of the target subspace to achieve target detection. Based on OSP, Capobianco et al. [16] proposed a semisupervised graph-based kernel OSP method, which uses the contextual selection of unlabeled samples to approximate the marginal distribution. Without inverting matrices, Song and Chang [17] used a recursive technique to perform OSP. To solve the interference problem of complex background, Chang and Chen [18] integrated data segmentation and low-rank and sparse matrix decomposition (LRSMD) to extend OSP for performance enhancement. Subsequently, Chen and Chang

Manuscript received 10 June 2023; revised 28 July 2023; accepted 19 August 2023. Date of publication 30 August 2023; date of current version 5 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 41971297 and Grant 62071084, and in part by the Fundamental Research Funds for the Central Universities under Grant 3072022GIP0801. *(Corresponding author: Qunming Wang.)*

Xiaoyi Wang is with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: wangxyi11@126.com).

Liguo Wang is with the College of Information and Communications Engineering, Dalian Minzu University, Dalian 116600, China, and also with the College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China (e-mail: wangliguo@hrbeu.edu.cn).

Qunming Wang is with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: wqm11111@126.com).

Anna Vizziello and Paolo Gamba are with the Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy (e-mail: anna.vizziello@unipv.it; paolo.gamba@unipv.it).

Digital Object Identifier 10.1109/JSTARS.2023.3310189

[19] proposed a background-annihilated target-constrained interference-minimized filter. Specifically, data sphering, LRSMD, and component decomposition analysis are first used to annihilate the background. Then, OSP is used to enhance target detectability, while CEM is exploited to suppress the background.

In recent years, the representation-based theory has shown reliable performances in HTD. Chen et al. [20] used the global target and local background dictionaries to sparsely represent each pixel, and the reconstruction residual was used to detect the targets. Zhang et al. [21] considered two situations when the target is absent or present and proposed a sparse representation-based binary hypothesis model. Li et al. [22] proposed a combined sparse and collaborative representation (CSCR) method for HTD, which considers that the targets have global sparsity and the background can be cooperatively represented by adjacent pixels. Bitar et al. [23] applied LRSMD theory for HTD and proposed a sparse and low-rank matrix decomposition (SLRMD) method. Cheng and Wang [24] used a locality-constrained linear coding method to create a compact background dictionary and proposed a union dictionary-based target detector. Zhao et al. [25] proposed a weighted Cauchy distance graph and local adaptive collaborative representation detection method, which fully considered the spatial information of HSI.

Due to the powerful feature extraction ability, deep-learning-based methods have received extensive attention. Based on autoencoder and SAM, Xie et al. [26] proposed a band selection method to remove the redundant information in HSI. Specifically, the authors matched the compressed latent feature with the original HSI band-by-band and selected the bands with smaller spectral angles to form the optimal subset. Zhang et al. [27] proposed an HTD network (HTD-Net). It uses an autoencoder to generate pseudotarget pixels and linear prediction to select background pixels. Then, the background–target and target–target pixel pairs are input into a similarity discrimination convolutional network, and the similarity score is used as the detection result. Zhu et al. [28] proposed a sparse representation-based strategy for background sample selection and linearly mixed the prior target and background samples to generate sufficient target samples. Furthermore, they proposed a two-stream convolution-based network (denoted as TSCNTD) to learn the differences between the background and targets. Similarly, Rao et al. [29] proposed a two-stream transformer-based network. In the sample construction method, the authors considered the subpixel targets and mixed background pixels. To reduce the negative impact of spatial and spectral redundant information, Shi et al. [30] introduced region-of-interest feature transformation and multiscale-spectral-attention module for HTD. Meanwhile, the shortcut and long-term connections are applied to improve the training ability of the network.

Due to mixed pixel effect, atmospheric attenuation, adjacent pixel effect, and other factors, the spectral variation is a common issue in HSI processing. Spectral variation may weaken the spectral information of ground objects, making it difficult to recognize them based on their spectral characteristics. Currently, only a few studies have focused on this issue. For instance, Ren et al. [31] used the orthogonal subspace unmixing method to address spectral variation. To solve the spectral registration problem between multisource datasets, Ye et al. [32] proposed a Bayesian-based super-resolution model. Li et al. [33] introduced the first-order neighborhood information into a graph convolutional network to alleviate the initial feature deviation caused by spectral variation. The above methods reduce the impact of spectral variation to a certain extent, but they fail to fully utilize the abundant spatial and spectral information of the original dataset. Additionally, although the distance-based HTD methods are simple to operate, they are not able to separate the background and target satisfactorily.

To solve the above problems, this article proposes a novel global spatial–spectral attention network ($GS^2A$-Net), coupled with a SAM-based background suppression strategy. Specifically, to fully utilize the correlation of the adjacent space and band in HSIs, the three-dimensional (3-D) convolution kernels with different sizes are used to capture multiscale local spatial–spectral information. The 3-D convolutional neural networks (CNNs) have been shown to be effective in the field of HSI processing. For example, Mei et al. [34] proposed a 3-D convolutional autoencoder that maximizes the extraction of spatial and spectral information from HSIs, eliminating the need for labeled training samples. Roy et al. [35] devised a hybrid spectral CNN by combining 2-D and 3-D CNNs. They considered that the 3-D CNN enables the joint representation of spatial and spectral information, while the 2-D CNN can enhance the representation of spatial characteristics. Based on 3-D CNN, Ahmad et al. [36] integrated transfer learning and active learning into a unified framework for HSI classification. Simultaneously, attention networks have gained substantial recognition and have been widely applied to remote sensing image processing. Hang et al. [37] employed spatial and spectral attention subnetworks to individually represent the spatial and spectral features of HSIs. Wu et al. [38] proposed a residual attention mechanism that integrated channel and feature map attentions for impedance inversion in seismic data. Guo et al. [39] sequentially injected spectral and spatial attention modules into a CNN to enhance the discrimination ability. Different from the previous attention mechanisms by 2-D convolution and pooling layers, the proposed $GS^2A$-Net consists of 3-D convolution layer, 3-D dropout layer, 3-D batch normalization (BN) layer, and the activation function. The 3-D convolution layer enables the interaction of local spatial–spectral information. The pooling layer is discarded to avoid losing useful features. The $GS^2A$-Net creates a 3-D attention map consistent with the size of the original HSI, which is generated by a *Sigmoid* function [40], and a skip-connection method is introduced to enhance the stability of the network. The proposed $GS^2A$-Net can be regarded as a feature extraction process from local to global. In the proposed SAM-based background suppression strategy, a simple SAM strategy is used to generate two versions of the initial detection maps. Subsequently, guided filtering is used to combine the advantages of different initial detection maps, which can suppress the background while preserving the targets. Finally, a nonlinear exponential function is used to further suppress the background.

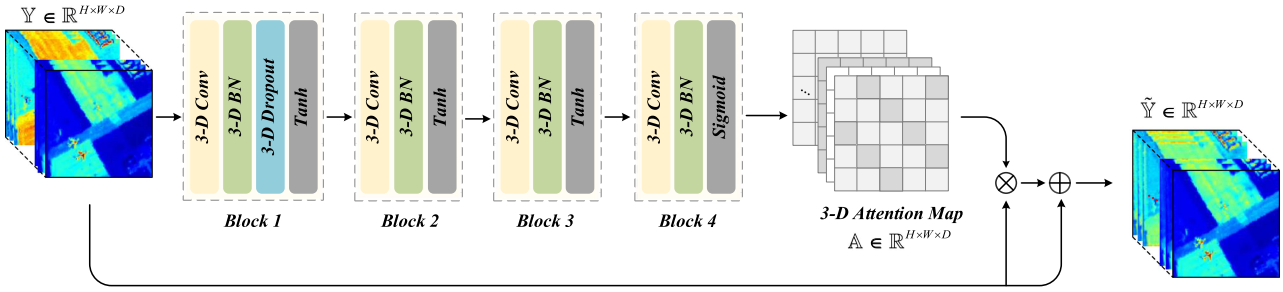



Fig. 1. Framework of the proposed GS$^2$A-Net.

In summary, this article aims at adding to the state-of-the-art of HTD. The two main contributions are as follows.

1) A GS$^2$A-Net: Using 3-D convolution layers to integrate spatial and spectral attention mechanisms into a unified network, the network extracts spatial–spectral information to a larger extent. Moreover, convolution kernels of different sizes are used to adapt to multiscale local spatial information. This network can effectively solve the spectral variation problem in HSIs.
2) A background suppression strategy based on SAM: This strategy inherits the brevity of SAM, but with the introduction of guided filtering and a nonlinear exponential function, it preserves the targets while jointly suppressing the background.

The rest of this article is organized as follows. Section II introduced the proposed GS$^2$A-Net and SAM-based background suppression strategy in detail. Section III provides the experimental results based on five real HSI datasets. In Section IV, the proposed method and results are further discussed. Finally, Section V concludes this article.

## II. Proposed Method

In this section, we provide a solution to deal with the spectral variation issue by GS$^2$A-Net, which is able to enhance the ability to identify targets. Furthermore, a SAM-based background suppression strategy is proposed to suppress the background while preserving the targets.

Let us define a 3-D HSI $\mathbb{Y} \in \mathbb{R}^{H\times W\times D}$, where $H$, $W$, and $D$ represent the height, width, and the number of bands, respectively. There are $L = H \times W$ pixels in the HSI, and the 2-D form of $\mathbb{Y}$ can be expressed as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_i, \ldots, \mathbf{y}_l] \in \mathbf{R}^{D\times L}$. The prior target is $\mathbf{t} \in \mathbf{R}^{D\times 1}$. In this article, all methods are preceded by the max–min normalization for datasets. The specific calculation method is given as follows:

$$\mathbf{Y} = \frac{\mathbf{Y} - \min(\mathbf{Y})}{\max(\mathbf{Y}) - \min(\mathbf{Y})} \tag{1}$$

where $\min(\mathbf{Y})$ and $\max(\mathbf{Y})$ are the minimum and maximum values of $\mathbf{Y}$, respectively.

### A. Global Spatial–Spectral Attention Network

The framework and architecture of the proposed GS$^2$A-Net are shown in Fig. 1 and Table I, respectively. The GS$^2$A-Net is an unsupervised end-to-end network, which includes four blocks. The 3-D convolution is used to extract local spatial–spectral information. As the name suggests, a 3-D convolution is a small cube that can slide in three directions (i.e., height, width, and band) on the 3-D HSI. There are several important parameters of the 3-D convolution that need to be set, including the length $h$, the width $w$, the depth $d$, the stride $s$, and the padding $p$ for different directions. Therefore, the output feature by 3-D convolution can be expressed as follows:

TABLE I
Architecture Features of the Proposed GS$^2$A-Net

| Pathway | Component |
|---|---|
| Block 1 | 3-D Conv<br>3-D BN<br>3-D Dropout<br>*Tanh* |
| Block 2 | 3-D Conv<br>3-D BN<br>*Tanh* |
| Block 3 | 3-D Conv<br>3-D BN<br>*Tanh* |
| Block 4 | 3-D Conv<br>3-D BN<br>*Sigmoid* |

$$\begin{cases} h_{\text{out}} = (h_{\text{in}} - h + 2p_h)/s_h + 1 \\ w_{\text{out}} = (w_{\text{in}} - w + 2p_w)/s_w + 1 \\ d_{\text{out}} = (d_{\text{in}} - d + 2p_d)/s_d + 1 \end{cases} \tag{2}$$

in which subscripts "in" and "out" indicate the input and output size of features in different directions. The parameters $p_h$ and $s_h$ are the padding and stride along the length direction, and the other two directions are similar. In this article, the 3-D convolution kernels of different sizes are used to capture local spatial–spectral features at different scales and achieve multiscale information interaction. Specifically, the kernel sizes of $7 \times 7$, $5 \times 5$, $3 \times 3$, and $1 \times 1$ are used to gradually extract local spatial features and adapt to spatial correlations at different scales. The corresponding kernel sizes in the spectral direction are 5, 5, 5, and 3 in the four 3-D convolutions, respectively (the parameters for 3-D convolutions in the four blocks are shown in Table II). To take full advantage of the abundant spatial and spectral information of HSI, we set the stride to 1 in all directions, adjusting the padding so that the output features of each layer are consistent with the size of the original HSI.

TABLE II
PARAMETER SETTINGS OF THE FOUR 3-D CONVOLUTIONS IN THE PROPOSED GS$^2$A-NET

| Pathway | Kernel size | Stride | Padding |
|---|---|---|---|
| Block 1 | [7, 7, 5] | [1, 1, 1] | [3, 3, 2] |
| Block 2 | [5, 5, 5] | [1, 1, 1] | [2, 2, 2] |
| Block 3 | [3, 3, 5] | [1, 1, 1] | [1, 1, 2] |
| Block 4 | [1, 1, 3] | [1, 1, 1] | [0, 0, 1] |

To avoid the problems of solution gradient explosion and disappearance caused by the internal covariance shift and accelerate the convergence speed of network, a 3-D BN is performed after each layer of 3-D convolution. The output of 3-D BN layer can be expressed as follows:

$$\tilde{\mathbf{z}} = \gamma(\mathbf{z}) + \beta \tag{3}$$

where $\mathbf{z}$ and $\tilde{\mathbf{z}}$ represent the input and output features of 3-D BN layer, and $\gamma$ and $\beta$ are the leachable parameters. The 3-D BN layer is also conducted in the three dimensions of the data. There are fewer parameters in the convolutional network; hence, the dropout layer is not as effective as that in a fully connected network. However, introducing the dropout layer in a lower layer of network may cause noisy input to subsequent networks while avoiding overfitting [7]. Hence, we only apply the 3-D dropout layer after the BN layer in the first block, with a dropout rate of 0.4.

As shown in Fig. 1, the Tanh function and the Sigmoid function are used as activation functions to increase the nonlinear expressive ability of the network

$$\text{Tanh}(\mathbf{z}) = \frac{e^{\mathbf{z}} - e^{-\mathbf{z}}}{e^{\mathbf{z}} + e^{-\mathbf{z}}} \tag{4}$$

$$\text{Sigmoid}(\mathbf{z}) = \frac{1}{1 + e^{-\mathbf{z}}}. \tag{5}$$

It is worth noting that the Tanh function is used in the hidden layers (i.e., blocks 1–3), while the Sigmoid function is used in the output layer (i.e., block 4). Using the Tanh function in the hidden layer helps to accelerate the convergence of the network and to partially avoid the gradient disappearance problem. The Sigmoid function ensures that the range of the generated 3-D attention map is within [0, 1]. After the four blocks, a 3-D attention map $\mathbb{A} \in \mathbb{R}^{W \times H \times D}$ is obtained. Subsequently, the attention map is multiplied by the original HSI pixel-by-pixel, and a skip connection is introduced to obtain the final output. The skip connection not only makes the network easier to train but also improves the network stability. The output of the GS$^2$A-Net can then be regarded as follows:

$$\tilde{\mathbb{Y}} = \mathbb{Y} \cdot \mathbb{A} + \mathbb{Y}\cdot \tag{6}$$

To train the network, one needs to minimize the reconstruction error. In this article, the classical mean square error is used to measure the loss between $\mathbb{Y}$ and $\tilde{\mathbb{Y}}$

$$\text{loss}_{\text{mse}} = \left\| \mathbb{Y} - \tilde{\mathbb{Y}} \right\|_{\mathbf{2}}^{\mathbf{2}}. \tag{7}$$

Taking dataset I (specific features are described in Section IV-A) as an example, Fig. 2 shows the normalized target spectral curves of the original and reconstructed HSIs. It can be obviously seen that the reconstructed target spectral curves are more compact compared with that of the original HSI.

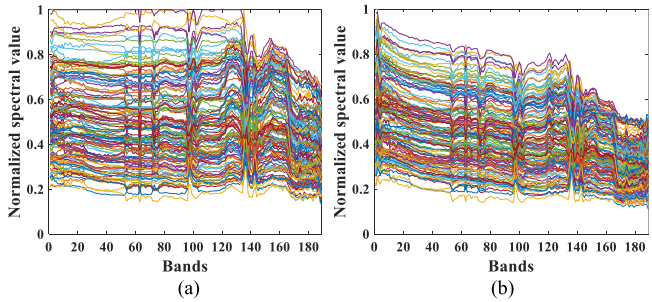



Fig. 2. (a) Original and (b) reconstructed normalized target spectral curves of an HSI.

### B. Background Suppression Strategy Based on SAM

Through the above GS$^2$A-Net, a new HSI $\tilde{\mathbb{Y}}$ is obtained, which has been corrected for the spectral variation problem, and the new target spectrum $\tilde{\mathbf{t}}$ is extracted from $\tilde{\mathbb{Y}}$. In this article, we use the simple and effective SAM procedure as the detection approach. However, SAM is not able to suppress the background well. To solve this problem, an improved strategy is considered, with the following three steps.

1) *Initial Detection Maps:* Recalling that the formula of SAM

$$\text{SAM}(\tilde{\mathbf{y}}_i) = \mathbf{cos}^{-\mathbf{1}} \frac{\tilde{\mathbf{y}}_i \tilde{\mathbf{t}}}{\sqrt{(\tilde{\mathbf{t}}^T \tilde{\mathbf{t}})(\tilde{\mathbf{y}}_i^T \tilde{\mathbf{y}}_i)}}. \tag{8}$$

Targets are expected to have smaller spectral angular distances from the prior target, while background pixels result in larger values. Therefore, we use the reciprocal method and the exponential method to obtain two initial detection maps

$$\mathbf{I}_1 = \frac{1}{\text{SAM}(\tilde{\mathbf{y}}_i)} \tag{9}$$

$$\mathbf{I}_2 = e^{(-\text{SAM}(\tilde{\mathbf{y}}_i))}. \tag{10}$$

Looking at Fig. 3, one may see that $\mathbf{I}_1$ has a lower false alarm rate, while the targets in $\mathbf{I}_2$ are brighter.

2) *Target Enhancement:* In this step, the guided filtering approach [41] is used to combine the advantages of these two initial detection maps. Specifically, to enhance the targets without increasing the false alarm rate as much as possible, we use $\mathbf{I}_1$ as the input map and $\mathbf{I}_2$ as the guided map. The two initial detection maps are first normalized. Moreover, two important paraments need to be preset: the local window radius (set to $r = \mathbf{1}$) and regularization parameter ($\varepsilon = \mathbf{4} \times \mathbf{10}^{-\mathbf{4}}$) (see Section IV-E). The output of the guided filtering procedure is

$$\mathbf{q}(i) = a_k \mathbf{I}_2(i) + b_k, i \in w_k \tag{11}$$

where $w_k$ is the number of pixels in the local window, (equal to $(2r+1)^2$). $a_k$ and $b_k$ are two linear coefficients, which are

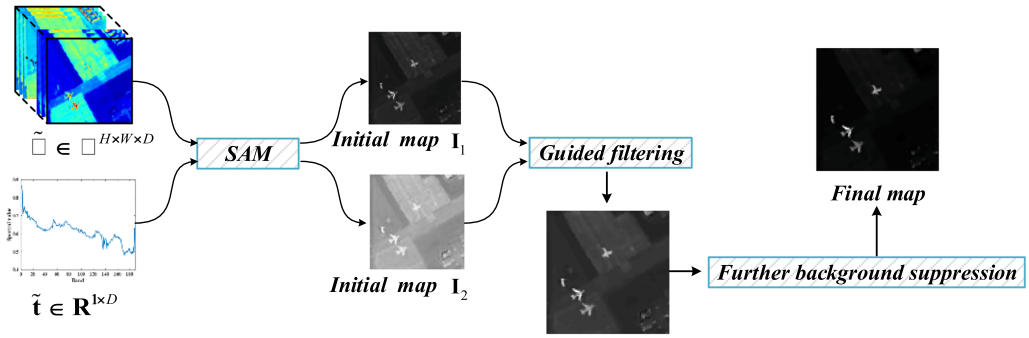



Fig. 3. Framework of the proposed SAM-based background suppression strategy.

computed as follows:

$$a_k = \frac{\mathrm{cov}_k(\mathbf{I_1}, \mathbf{I_2})}{\sigma^{\mathbf{2}} + \varepsilon} \tag{12}$$

$$b_k = \mu_{\mathbf{1}} - a_k \mu_{\mathbf{2}} \tag{13}$$

where $\mu_1$ and $\mu_2$ are the mean values of the local window in $\mathbf{I}_1$ and $\mathbf{I}_2$, respectively, while $\sigma$ is the variance of the local window in $\mathbf{I}_2$, and $\mathrm{cov}_k(\cdot)$ represents the covariance operation.

Because each pixel may be contained in multiple windows, the average is used to obtain the final output

$$\mathbf{q}(i) = \bar{a}_i \mathbf{I_2}(i) + \bar{b}_i \tag{14}$$

where $\bar{a}_i$ and $\bar{b}_i$ are the average coefficients of all the local windows, including the $i$th pixel.

3) *Background suppression:* Although the second step enhances the targets, it also introduces an undesired false alarm rate. Therefore, an exponential nonlinear function is used to further suppress the background. The final detection map is

$$\mathbf{D} = (\mathbf{1} - e^{-\mathbf{q}})\mathbf{q} \tag{15}$$

where $1 - e^{-\mathbf{q}}$ can be regarded as a weight map. The background pixels are assigned smaller weights and the targets are assigned larger ones.

## III. Experiments

### A. Data Description

Five real-world HSIs were used in the experiments to prove the performances of the proposed techniques. False-color images as well as the corresponding ground-reference maps for these five HSI datasets are shown in Fig. 4.

1) *Datasets I and II:* Datasets I and II were collected by airborne visible/infrared imaged spectrometer (AVIRIS) from San Diego Airport, California, USA. Both datasets I and II have a spatial size of 100 × 100 pixels, and the spatial resolution is 3.5 m. After discarding lower SNR and water absorption bands, the number of considered bands is 189. There are three aircraft as the targets, which cover 134 and 58 pixels in datasets I and II, respectively. In datasets I and II, the 11th and 21st target pixels are selected as the priori spectrums, respectively.

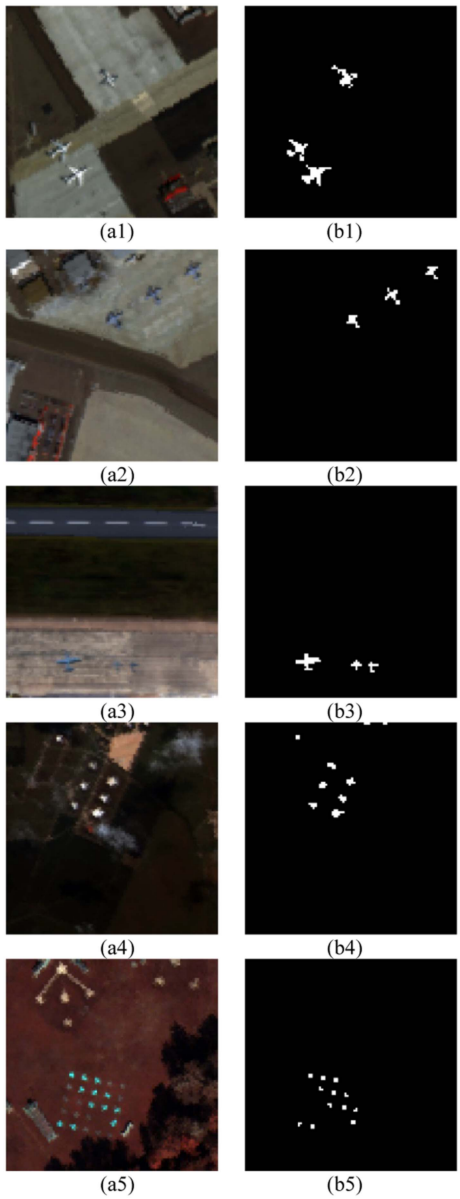



Fig. 4. Five real-world HSI datasets used in the experiments. (a1)–(a5) are false-color images (bands 37, 18, and 8 as RGB) for datasets I–V, respectively. (b1)–(b5) are the corresponding ground-reference maps of the five datasets.

2) *Dataset III:* Dataset III was collected by AVIRIS from Gulfport Airport, Mississippi, USA. It covers 100 × 100 pixels with a spatial resolution of 3.4 m. There are 191 high SNR bands. The targets are three aircraft containing 88 pixels. The 57th target pixel is selected as the priori spectrum.
3) *Dataset IV:* Dataset IV was collected by AVIRIS from Texas Coast, USA. It covers 100 × 100 pixels with a spatial resolution of 17.2 m and contains 204 high SNR bands. The targets are several buildings containing 67 pixels. The 25th target pixel is selected as the priori spectrum.
4) *Dataset V:* The last dataset was collected by ProSpecTIR-VS sensor from Avon, New York, USA. It covers 100 × 100 pixels with a spatial resolution of 1 m. There are 360 high SNR bands. The targets are several blue tarps containing 43 pixels. The 30th target pixel is selected as the priori spectrum.

### B. Experimental Setup

*1) Implementation Details:* Statistic-based methods ACE [11] and CEM [12], subspace-based method OSP [15], representation-based methods CSCR [22] and SLRMD [23], and deep-learning-based method HTD-Net [27] and TSCNTD [28] were taken as the benchmark methods. All these methods were implemented on an Intel Core (TM) i9-8950HK central processing unit with 32 GB RAM. The HTD-Net, TSCNTD, and the proposed method were implemented by Python 3.7.11 and Pytorch 1.10.0, while the other benchmark methods were implemented by MATLAB 2012. Both ACE and CEM are nonparametric methods. In the OSP, the density peak-based clustering method was applied. The cluster number was set to 8. By excluding the class closest to the prior target, we considered the remaining seven classes as the background. From each background class center, we selected the 20 nearest pixels to form the background endmembers. In the CSCR, the tradeoff parameters were set to $\lambda = 0.1$ and $\beta = 0.01$ for all datasets. The dual-window sizes of the CSCR were set to (5, 17), (9, 13), (3, 5), (13, 15), and (15, 19) for datasets I–V, respectively. The tradeoff parameters of SLRMD were set to $\lambda = 0.01$ and $\beta = 1$ for all datasets. There is a target generation network and a similarity discrimination convolutional network in HTD-Net. For all datasets, the number of pseudotarget samples was set to 100 with 5 epochs in the target generation network. The learning rare and iterations were set to 0.0001 and 200 in the similarity discrimination convolutional network. According to Zhu et al. [28], the TSCNTD used a batch size of 256 and a learning rate of 0.0001, and the number of the selected background samples was 1000. In the proposed GS$^2$A-Net, the learning rate and iterations were set to 0.0001 and 150 for all datasets.

*2) Evaluation Indices:* The first index is receiver operating characteristic (ROC) curve presenting the probability of false alarm $P_f$ and the probability of detection $P_d$ for a given threshold $\tau$. Accordingly, the 2-D ROC curves of ($P_f$, $P_d$) and ($\tau$, $P_f$) [42], and 3-D ROC curve of ($P_f$, $\tau$, $P_d$) [43] are used. The second index is the area under the ROC curve (AUC) value [44]. A well-performing detector should produce larger AUC ($P_f$, $P_d$) value and smaller AUC ($\tau$, $P_f$) value. In addition, to conduct a comprehensive analysis of the performances of different methods, we introduced $\mathrm{AUC_{BS}}$, $\mathrm{AUC_{TD}}$, and $\mathrm{AUC_{SNPR}}$ to measure the background suppressibility, target detectability, and signal-to-noise probability ratio, respectively [45].

Finally, the box-whisker plot [46] is used to visually show the degree of separation between the background and targets. The box-whisker plot is represented by the normalized detection statistical range of the background and targets. Ideally, there is a larger distance between the background and target boxes. Meanwhile, smaller background box can demonstrate that the background is well suppressed.

### C. Detection Performance of Different Methods

Fig. 5 shows the detection map of the different methods applied to the five selected datasets. For dataset I, OSP, CSCR, and SLRMD detect the three target aircraft, while they fail to adequately separate the background and targets. Although ACE can suppress the background well, the important target is also hidden. Due to heavy reliance on the reliability of constructed background–target samples, HTD-Net cannot preserve the complete shape of aircraft, and a large number of background pixels are incorrectly detected as targets. While TSCNTD shows a greater detection accuracy than HTD-Net, it still struggles to separate background from anomalies because of uncertainties in manually extracting background samples. Fig. 5(h1) illustrates that the topmost airplane is merged with the background and lacks clear visibility. The CEM method has weaker performance in terms of both target detection and background suppression.

For the remaining datasets, the representation-based methods, i.e., CSCR and SLRMD, cannot suppress the background. The main reason is that CSCR and SLRMD use local and global background dictionaries to represent the background, respectively. When the background dictionary is incomplete or mixed with target noise, the background cannot be well represented, and the false alarm rate becomes larger. ACE can barely separate the background and targets. The main reason for this is that ACE assumes that the background and targets have the same mean and covariance, and when there is spectral variation, the difference between the background and the prior target is small. For datasets II, IV, and V, OSP and the proposed method can detect most of the target pixels, while the proposed method can better suppress the background. Furthermore, some methods are sensitive to noise. For example, for dataset II, there is obvious horizontal noise in the detection maps by CEM, OSP, CSCR, HTD-Net, and TSCNTD. Overall, by visual inspection, it can be clearly seen the proposed method cannot only effectively extract the targets but also suppress the background.

Tables III and IV and Figs. 6 and 7 quantitatively show the detection performance of different methods. As shown in Fig. 6, the proposed method is represented by dark red line, which is closer to the upper left corner of the coordinate axis in the ROC ($P_f$, $P_d$) curve. Except for dataset I, the ROC ($\tau$, $P_f$) curve of the proposed method is next only to ACE (indicated in light blue) and closer to the lower left corner of the coordinate axis compared with the remaining methods. The 3-D ROC curves show similar results to the 2-D ROC curves. Table III presents the AUC values of the eight methods for the

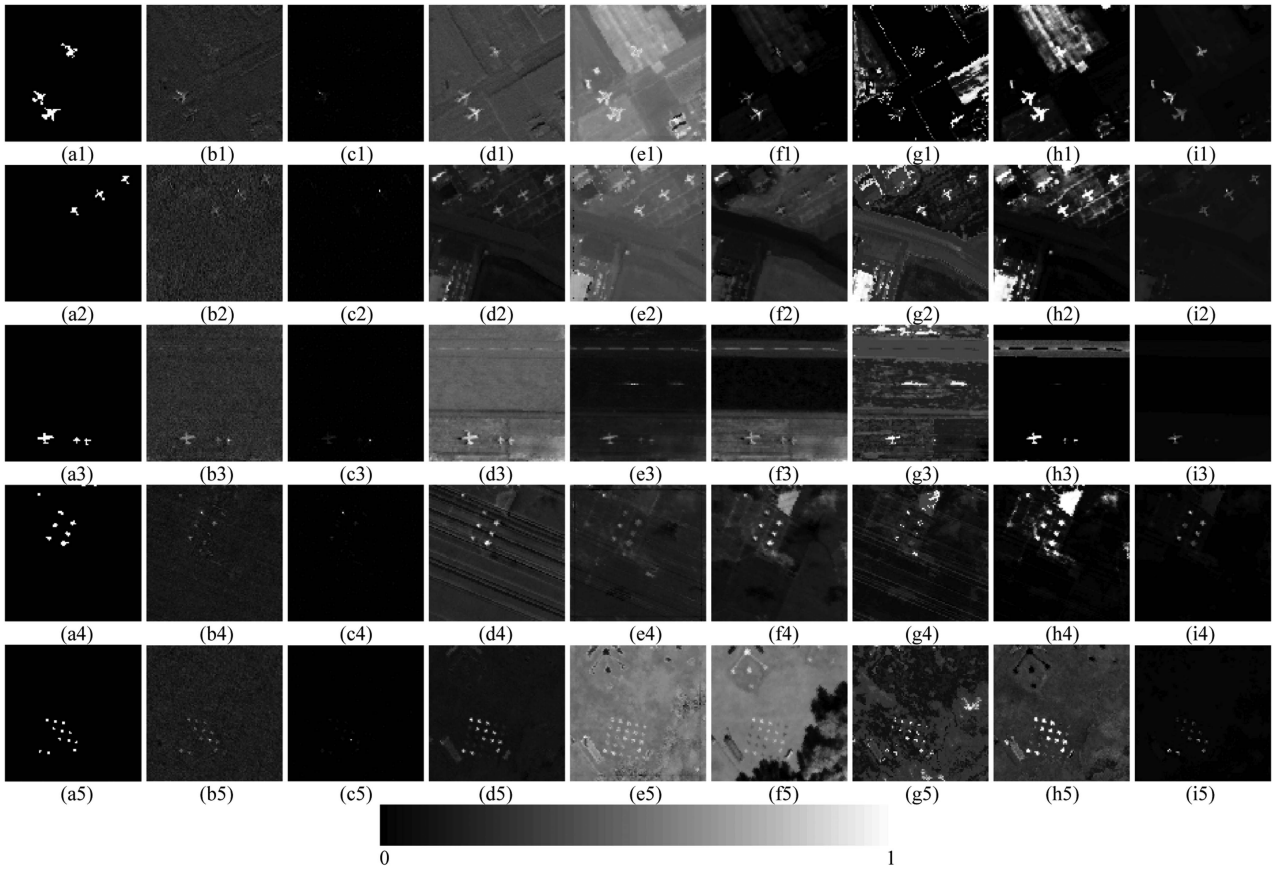



Fig. 5. Detection maps based on eight methods for the datasets I–V. (a) Corresponding ground-reference map. (b)–(i) Detection maps for CEM, ACE, OSP, CSCR, SLRMD, HTD-Net, TSCNTD, and proposed method, respectively. Lines 1–5 represent dataset I, dataset II, dataset III, dataset IV, and dataset V, respectively.

five datasets, where the first and second most accurate results in terms of the AUC values for each dataset are highlighted in boldface and underlined, respectively. In terms of the AUC $(P_f, P_d)$ value, the proposed method outperforms the benchmark methods. Taking the dataset I as an example, the AUC $(P_f, P_d)$ value of the proposed method is 0.9967, and compared with that of CEM, ACE, OSP, CSCR, SLRMD, HTD-Net, and TSCNTD, the accuracy gains are 0.2777, 0.2561, 0.0407, 0.0130, 0.1110, 0.1783, and 0.0115, respectively. In addition, the AUC $(\tau, P_f)$ value of the proposed method is relatively small. Taking datasets III and V as examples, the AUC $(\tau, P_f)$ values of the proposed method are 0.0133 and 0.0176, which are just 0.0067 and 0.0122 higher than that of ACE. Furthermore, the values of $\mathrm{AUC}_{\mathrm{BS}}$, $\mathrm{AUC}_{\mathrm{TD}}$, and $\mathrm{AUC}_{\mathrm{SNPR}}$ also demonstrate the effectiveness of the proposed method, especially in suppressing background and noise. Specifically, for all datasets, the $\mathrm{AUC}_{\mathrm{TD}}$ and $\mathrm{AUC}_{\mathrm{SNPR}}$ values of the proposed method are the highest.

Fig. 7 shows the ability of different methods to separate the background and targets, where the background and target boxes are represented by blue and orange, respectively. It can be clearly seen that the proposed method still exhibits good performance. Specifically, although for most datasets, the proposed method has a closer distance between the background box and the target box compared with those of OSP and TSCNTD; the proposed method has a larger background box compared with that of ACE. Table IV lists the running time of eight methods for the five datasets in s, where the first and second shortest running times are highlighted in boldface and underlined, respectively. It is not difficult to find that the classical methods generally have the advantage of low time-consuming, while deep-learning-based methods (i.e., HTD-Net, TSCNTD, and the proposed method) take longer time to train network. Especially for HTD-Net, constructing pseudotarget samples is time-consuming. In this article, we constructed 100 pseudotarget samples for each dataset, and the training time exceeded 1000 s.

### D. Performance Analysis of the Proposed GS$^2$A-Net

In this section, to validate the effectiveness of the proposed GS$^2$A-Net for spectral correction, we used the same detection method (i.e., SAM) with (w) and without (w/o) GS$^2$A-Net as preprocessing. The comparative AUC $(P_f, P_d)$ values for the five datasets are listed in Table V. It can be clearly seen that for all datasets, there are significant improvements in the detection accuracy after spectral correction. Specifically, compared with the case without GS$^2$A-Net, the accuracy gains with GS$^2$A-Net are 0.0284, 0.0378, 0.0295, 0.0195, and 0.0231 for datasets I–V, respectively. Moreover, taking dataset I as an example, Fig. 8 shows the AUC $(P_f, P_d)$ values using different target pixels as the prior information. There are 134 target pixels in dataset I. After spectral correction, the detection accuracy of different prior targets has significantly improved. Especially, when using

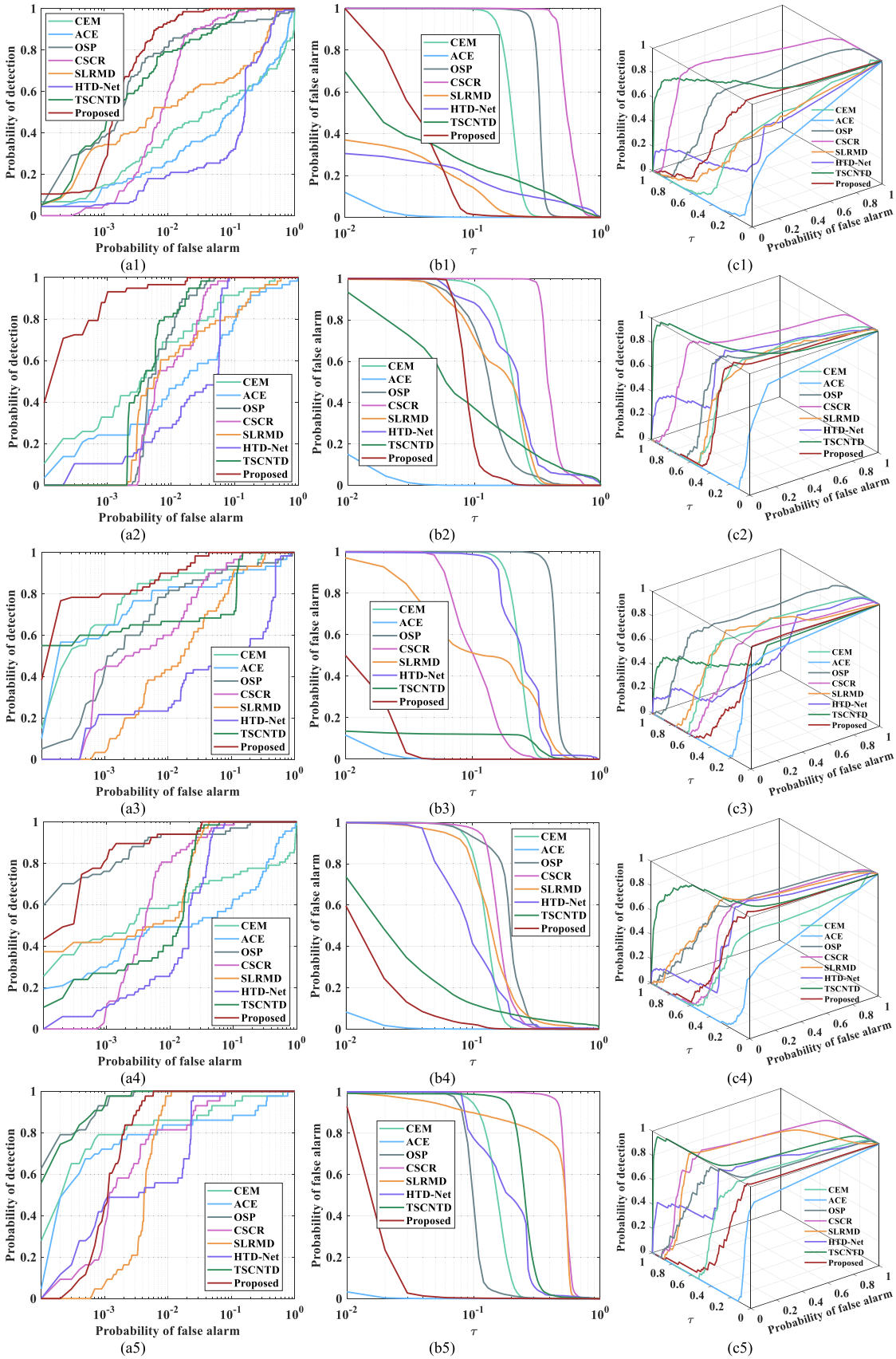



Fig. 6. ROC curves of the eight methods for the five real datasets. From left to right are (a) two-dimensional ROC curves of ($P_f$, $P_d$), (b) two-dimensional ROC curves of ($\tau$, $P_f$), and (c) three-dimensional ROC curves. Lines 1–5 are the results for datasets I–V, respectively.

TABLE III
MULTIVERSION AUC VALUES OF THE EIGHT METHODS FOR THE FIVE DATASETS (LINES 1–5 ARE AUC $(P_f, P_d)$, AUC $(\tau, P_f)$, $\text{AUC}_{\text{TD}}$, AND $\text{AUC}_{\text{SNPR}}$ VALUES, RESPECTIVELY)

| | CEM | ACE | OSP | CSCR | SLRMD | HTD-Net | TSCNTD | Proposed |
|---|---|---|---|---|---|---|---|---|
| Dataset I | 0.7190 | 0.7406 | 0.9560 | 0.9837 | 0.8857 | 0.8184 | 0.9852 | **0.9967** |
| | 0.2024 | **0.0067** | 0.3324 | 0.5607 | 0.0357 | 0.0918 | 0.1169 | 0.0416 |
| | 0.5166 | 0.7339 | 0.6236 | 0.4230 | 0.8500 | 0.7266 | 0.8683 | **0.9551** |
| | 0.9902 | 0.7688 | 1.5182 | 1.8023 | 1.1580 | 1.1414 | **1.8877** | 1.4088 |
| | 1.3399 | 4.2090 | 1.6913 | 1.4600 | 7.6275 | 3.5185 | 7.7203 | **9.9063** |
| Dataset II | 0.9671 | 0.9175 | 0.9914 | 0.9852 | 0.9478 | 0.9874 | **0.9934** | 0.9991 |
| | 0.2027 | **0.0072** | 0.1419 | 0.4055 | 0.1808 | 0.2386 | 0.1554 | 0.0906 |
| | 0.7644 | 0.9103 | 0.8495 | 0.5797 | 0.7670 | 0.7488 | 0.8380 | **0.9085** |
| | 1.3348 | 0.9529 | 1.4730 | 1.7369 | 1.3378 | 1.6564 | **1.9827** | 1.3436 |
| | 1.8140 | 4.9167 | 3.3939 | 1.8538 | 2.1571 | 2.8039 | **6.3662** | 3.8024 |
| Dataset III | 0.9771 | 0.9386 | 0.9533 | 0.9815 | 0.9495 | 0.7820 | 0.9605 | **0.9968** |
| | 0.2354 | **0.0066** | 0.4531 | 0.1129 | 0.1849 | 0.2653 | 0.0447 | 0.0133 |
| | 0.7417 | 0.9320 | 0.5002 | 0.8686 | 0.7646 | 0.5167 | 0.9158 | **0.9835** |
| | 1.4485 | 1.0221 | 1.1795 | 1.3413 | 1.4737 | 1.2753 | **1.5874** | 1.1882 |
| | 2.0025 | 12.6515 | 0.5622 | 3.1869 | 2.8350 | 1.8594 | 14.0246 | **14.3910** |
| Dataset IV | 0.8015 | 0.8311 | 0.9918 | 0.9912 | 0.9891 | 0.9782 | 0.9874 | **0.9978** |
| | 0.1292 | **0.0061** | 0.1968 | 0.1659 | 0.1556 | 0.1108 | 0.0685 | 0.0184 |
| | 0.6723 | 0.8250 | 0.7950 | 0.8253 | 0.8335 | 0.8674 | 0.9189 | **0.9794** |
| | 1.0701 | 0.8764 | 1.5938 | 1.3577 | 1.6182 | 1.3902 | **1.9021** | 1.3305 |
| | 2.0789 | 7.4262 | 3.0589 | 2.2092 | 4.0431 | 3.7184 | 13.3533 | **18.0815** |
| Dataset V | 0.9755 | 0.9499 | **0.9998** | 0.9917 | 0.9955 | 0.9887 | **0.9998** | 0.9982 |
| | 0.1541 | **0.0054** | 0.1020 | 0.5334 | 0.4498 | 0.1987 | 0.2582 | 0.0176 |
| | 0.8214 | 0.9445 | 0.8978 | 0.4583 | 0.5457 | 0.7900 | 0.7416 | **0.9806** |
| | 1.3453 | 1.0044 | 1.6582 | 1.7441 | 1.7226 | 1.6606 | **1.9875** | 1.2739 |
| | 2.3997 | 10.0926 | 6.4549 | 1.4106 | 1.6165 | 3.3815 | 3.8253 | **15.6648** |

The first and second most accurate results in terms of the AUC values for each dataset are highlighted in boldface and underlined, respectively.

TABLE IV
RUNNING TIME (IN S) OF THE EIGHT METHODS FOR THE FIVE DATASETS

| | CEM | ACE | OSP | CSCR | SLRMD | HTD-Net | TSCNTD | Proposed |
|---|---|---|---|---|---|---|---|---|
| Dataset I | **0.5268** | 1.4658 | 4.6568 | 17.3535 | 7.2582 | 1725.9486 | 374.9164 | 371.3756 |
| Dataset II | **0.5227** | 0.6260 | 1.3673 | 5.3697 | 8.4387 | 2138.9275 | 328.9865 | 342.9862 |
| Dataset III | **0.5772** | 0.8623 | 1.4234 | 1.8643 | 8.7933 | 1875.2947 | 417.0943 | 396.5578 |
| Dataset IV | **0.5599** | 0.8551 | 1.4207 | 4.9481 | 9.2190 | 1978.2948 | 273.4085 | 362.9795 |
| Dataset V | **1.4417** | 1.6626 | 1.5181 | 13.0766 | 9.2190 | 2648.1305 | 298.9851 | 641.5469 |

The first and second shortest running times are highlighted in boldface and underlined, respectively.

the 35th–45th and 90th–100th target pixels as the prior target, the increase in detection accuracy range of 0.0036–0.4946 occurs.

### *E. Performance and Parameter Analysis of the Proposed Background Suppression Strategy*

To analyze the effectiveness of the proposed SAM-based background suppression strategy, we compared the detection performance of the initial and final maps. As illustrated in Fig. 9, for most datasets, $\mathbf{I}_1$ has a lower false alarm rate and $\mathbf{I}_2$ has bright targets. However, as shown in Fig. 9(a3)–(e3), the results of the final map preserve the shape of targets as much as possible while suppressing the background. Table VI lists the AUC $(P_f, P_d)/(\tau, P_f)$ values of different detection results, where the first and second most accurate results are highlighted in boldface and underlined, respectively. There are three points

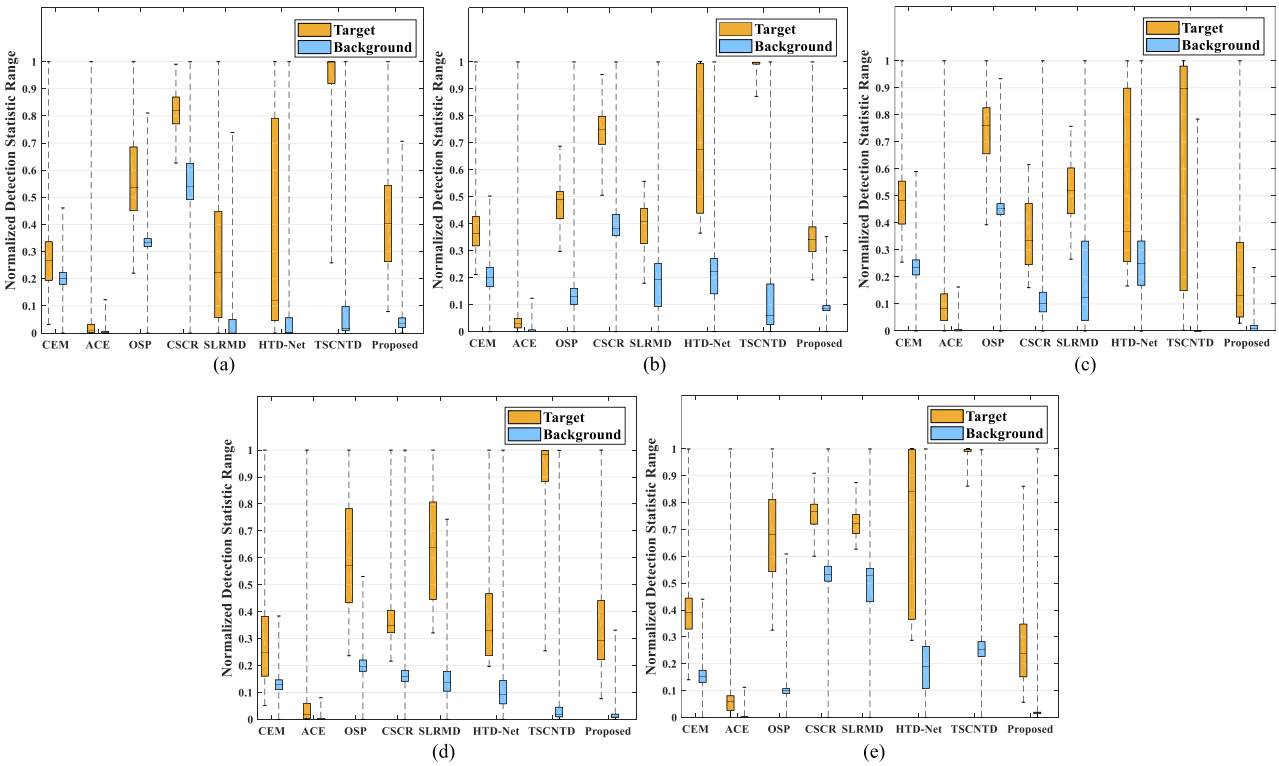



Fig. 7. Box-whisker plots for the eight target detection methods for the five datasets. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V.

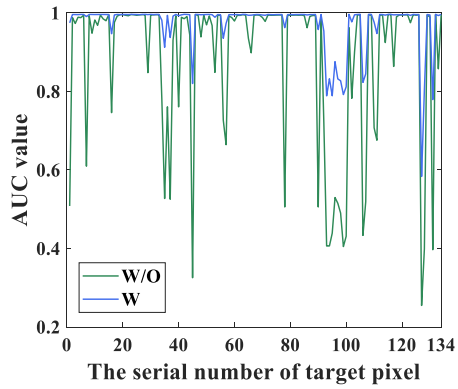



Fig. 8. AUC values generated by different prior targets for dataset I (the green and blue lines represent with (w) and without (w/o) spectral correction, respectively.

that need to be noted. First, the AUC ($P_f$, $P_d$) values of the two initial detection results obtained by (19) are consistent. Second, the detection accuracy is improved after guided filtering, and the accuracy gains for datasets I–V are 0.0007, 0.0027, 0.0024, 0.0027, and 0.0019, respectively. Third, after the nonlinear exponential function, the detection accuracy remains unchanged but the false alarm rate decreases. Taking dataset I as an example, the false alarm rates of the two initial detection maps and the final detection map are 0.0878, 0.6464, and 0.0416, respectively. This indicates that the proposed strategy is effective in suppressing background.

TABLE V
AUC ($P_f$, $P_d$) VALUES WITH (W) AND WITHOUT (W/O) SPECTRAL CORRECTION FOR THE FIVE DATASETS

| | Dataset I | Dataset II | Dataset III | Dataset IV | Dataset V |
|---|---|---|---|---|---|
| w/o | 0.9683 | 0.9613 | 0.9673 | 0.9783 | 0.9751 |
| w | **0.9967** | **0.9991** | **0.9968** | **0.9978** | **0.9982** |

The more accurate results for each dataset are highlighted in boldface.

There are two important parameters that need to be tuned in guided filtering: local window radius $r$ and regularization parameter $\varepsilon$. Fig. 10 shows the AUC ($P_f$, $P_d$) values under different values of $r$ and $\varepsilon$ for the five datasets. It can be clearly seen that when $\varepsilon$ is small, the AUC value decreases significantly when the window size increases. With $\varepsilon$ increases, the change of AUC ($P_f$, $P_d$) value is less affected by the window size. However, a larger $\varepsilon$ value will cause the image to become smooth, especially at the edge of the targets. To avoid excessive smoothing to the edge of targets, we set a smaller $\varepsilon$ value. When $\varepsilon$ is small, a smaller window is better, and thus, the window radius was set to 1.

### F. Influence of Convolution Kernel Size on Detection Accuracy

In this section, to validate the effectiveness of the hyperparameter setting in GS$^2$A-Net, we tested the detection results using a unified kernel size (i.e., $3 \times 3 \times 3$, $5 \times 5 \times 5$, or $7 \times 7 \times 7$) in the four modules. The corresponding AUC ($P_f$,

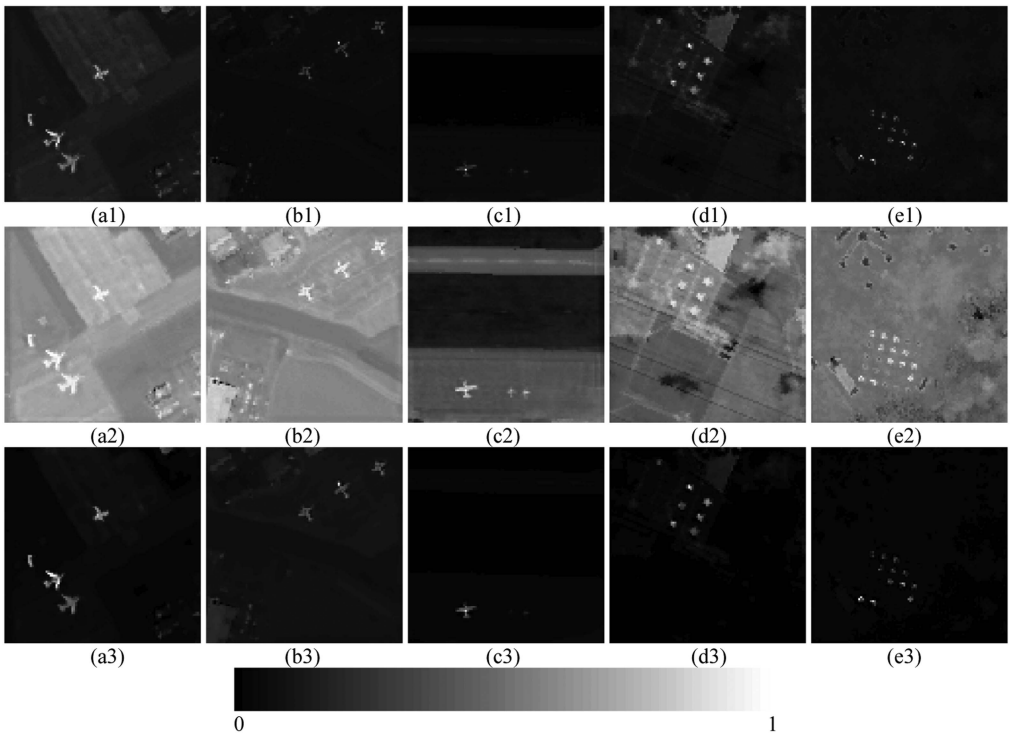



Fig. 9. Initial and final detection maps of the five datasets. Lines 1 and 2 express the initial detection maps $I_1$ and $I_2$, respectively, and line 3 is the final map. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V.

TABLE VI
AUC $(P_f,\ P_d)/(\tau,\ P_f)$ VALUES OF THE DIFFERENT DETECTION MAPS FOR THE FIVE DATASETS

| | Dataset I | Dataset II | Dataset III | Dataset IV | Dataset V |
|---|---|---|---|---|---|
| $\mathbf{I}_1$ | 0.9960/0.0878 | 0.9964/**0.0502** | 0.9944/0.0324 | 0.9951/0.0872 | 0.9963/0.0633 |
| $\mathbf{I}_2$ | 0.9960/0.6464 | 0.9964/0.6011 | 0.9944/0.2509 | 0.9951/0.4724 | 0.9963/0.4891 |
| Final map | **0.9967/0.0416** | **0.9991**/0.0906 | **0.9968/0.0133** | **0.9978/0.0184** | **0.9982/0.0176** |

The first and second most accurate results in terms of the AUC values for each dataset are highlighted in boldface and underlined, respectively.

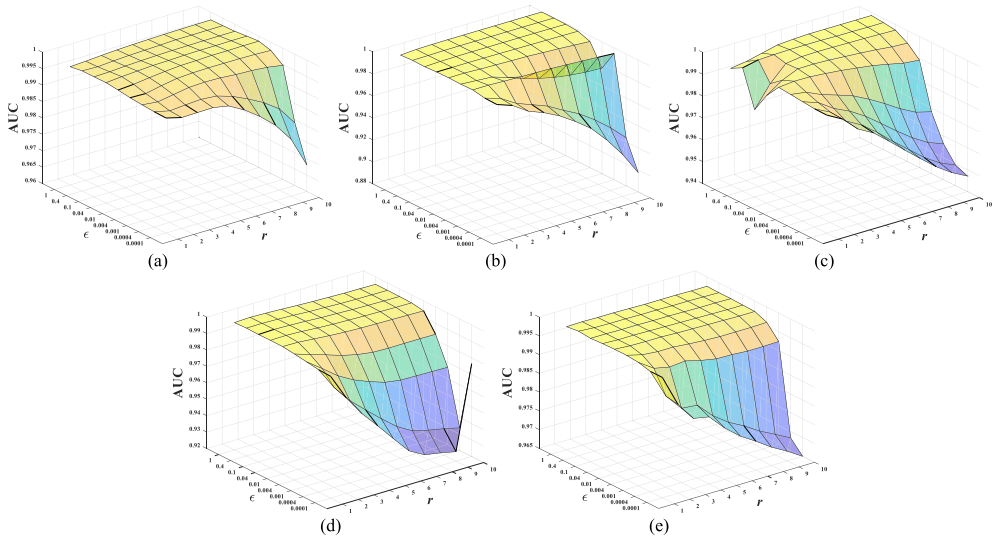



Fig. 10. AUC values of the background suppression strategy with different values of $r$ and $\varepsilon$. (a) Dataset I. (b) Dataset II. (c) Dataset III. (d) Dataset IV. (e) Dataset V.

TABLE VII
AUC ($P_f$, $P_d$) VALUES FOR DIFFERENT SIZES OF THE CONVOLUTION KERNEL FOR THE FIVE DATASETS

| | 3×3×3 | 5×5×5 | 7×7×7 | Proposed |
|---|---|---|---|---|
| Dataset I | 0.9824 | 0.9902 | 0.9896 | **0.9967** |
| Dataset II | 0.9921 | 0.9785 | 0.9933 | **0.9991** |
| Dataset III | 0.9654 | 0.9867 | 0.9769 | **0.9968** |
| Dataset IV | 0.9952 | 0.9905 | 0.9720 | **0.9978** |
| Dataset V | 0.9927 | 0.9837 | 0.9859 | **0.9982** |

The more accurate results for each dataset are highlighted in boldface.

$P_d$) values are shown in Table VII. It is clear that for all datasets, the AUC ($P_f$, $P_d$) values of unified kernel sizes are smaller than the proposed parameter setting. Taking dataset I as an example, compared with the case with unified kernel sizes, the AUC ($P_f$, $P_d$) value of the proposed parameter setting is 0.0143, 0.0065, and 0.0071 larger than for $3 \times 3 \times 3$, $5 \times 5 \times 5$, and $7 \times 7 \times 7$, respectively.

## IV. DISCUSSION

In this article, we use reconstructed HSI and prior target to achieve target detection. Nevertheless, in practice, prior targets usually originate from the spectral library rather than being directly extracted from HSIs. When facing such a situation, we can first use distance-based methods, such as SAM and ED to find the closest pixel in the HSI to the prior target, and use it to replace the prior spectrum. However, when there is a significant difference between the prior spectrum and the tested targets, utilizing the spatial and spectral information of HSI may not be sufficient to solve the problem of complex spectral variation. Hence, this requires more research, which may help further improve the recognition ability of ground objects under complex conditions.

The method proposed in this article is a two-step method. Although this strategy can achieve better performance for the tested datasets, it might lead to local optimal solution in theory. For convenience, in future research, we can attempt to design an end-to-end network to achieve joint optimization of feature extraction and target detection. Specifically, we can utilize advanced deep-learning networks, such as transformer and contrastive learning networks, to extract features at different levels. Subsequently, the classic methods could be used as detection networks. How to integrate classic methods into deep-learning networks will be an interesting question.

We use the same hyperparameters of GS$^2$A-Net for different datasets, which indicates that the GS$^2$A-Net has strong generalization ability. However, it still took a long time during the debugging process. Meanwhile, compared with classical methods, deep-learning-based methods are always more time-consuming, which is not conducive to real-time target detection. We will continue to explore the features of HSIs to aid parameter setting to reduce the complexity of parameter tuning. Furthermore, the network parameters (such as the weight matrix and bias of the convolution kernel) vary for different datasets. Thus, the network should be retrained when a new dataset is tested. Accordingly, it is worthwhile to design a unified network for the same target or sensor. In addition, more attention is still needed to study the spatial and spectral correlation in HSIs, as motivated by the encouraging performance in various tasks, such as super-resolution mapping [47], super-resolution [48], and image classification [49].

Last but not least, in this article, we proposed a background suppression strategy in which the core is to combine the advantages of the two versions (i.e., the reciprocal method and the exponential method) of detection results by SAM. This method may equally be applied to other detection methods, such as CSCR that has great detection accuracy but high false alarm rate. Specifically, we can use the detection map of CSCR as the guide map, thereby improving the visual recognition ability of the background and targets.

## V. CONCLUSION

To overcome the effect of spectral variation and complex background distribution in HTD, in this article, we proposed a two-step HTD method. Specifically, a 3-D convolution-based GS$^2$A-Net was first used to fuse multiscale local spatial and spectral features, thereby emphasizing the target features and facilitating further increase of the detection accuracy. Additionally, we designed a background suppression strategy based on SAM by introducing guided filtering to fully utilize the advantages of different versions of the detection maps in which the nonlinear exponential function is used to further suppress the background. Through visual and quantitative analysis of the experimental results on five HSI datasets, the proposed method outperforms six benchmark methods (including some popular and deep-learning-based methods).

## REFERENCES

[1] M. Seydgar, S. Rahnamayan, P. Ghamisi, and A. A. Bidgoli, "Semisupervised hyperspectral image classification using a probabilistic pseudo-label generation framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, Aug. 2022.
[2] M. Baisantry, A. K. Sao, and D. P. Shukla, "Discriminative spectral–spatial feature extraction-based band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Nov. 2022.
[3] J. Lei, S. Xu, W. Xie, J. Zhang, Y. Li, and Q. Du, "A semantic transferred priori for hyperspectral target detection with spatial–spectral association," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, Mar. 2023.
[4] H. Gao, Y. Zhang, Z. Chen, S. Xu, D. Hong, and B. Zhang, "A multidepth and multibranch network for hyperspectral target detection based on band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, Mar. 2023.
[5] L. Gao, X. Sun, X. Sun, L. Zhuang, Q. Du, and B. Zhang, "Hyperspectral anomaly detection based on chessboard topology," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, Feb. 2023.
[6] L. Gao, D. Wang, L. Zhuang, X. Sun, M. Huang, and A. Plaza, "BS$^3$LNet: A new blind-spot self-supervised learning network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–18, Feb. 2023.
[7] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022.
[8] Z. Han, D. Hong, L. Gao, B. Zhang, M. Huang, and J. Chanussot, "Auto NAS: Automatic neural architecture search for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Jun. 2022.
[9] F. Luo, T. Zhou, J. Liu, T. Guo, X. Gong, and J. Ren, "Multiscale diff-changed feature fusion network for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, Jan. 2023.

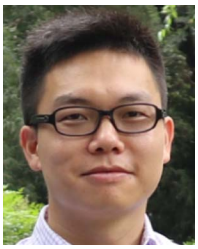
[10] N. M. Nasrabadi, "Regularized spectral matched filter for target recognition in hyperspectral imagery," *IEEE Signal Process. Lett.*, vol. 15, pp. 317–320, Mar. 2008.
[11] S. Kraut, L. L. Scharf, and R. W. Butler, "The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 427–438, Feb. 2005.
[12] W. H. Farrand and J. C. Harsanyi, "Mapping the distribution of mine tailings in the Coeur d'Alene river valley, Idaho, through the use of a constrained energy minimization technique," *Remote Sens. Environ.*, vol. 59, no. 1, pp. 64–76, Jan. 1997.
[13] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.
[14] Z. Chen et al., "Global to local: A hierarchical detection algorithm for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5544915.
[15] C.-I. Chang, "Orthogonal subspace projection (OSP) revisited: A comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, Mar. 2005.
[16] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3822–3833, Nov. 2009.
[17] M. Song and C.-I. Chang, "A theory of recursive orthogonal subspace projection for hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3055–3072, Jun. 2015.
[18] C.-I. Chang and J. Chen, "Orthogonal subspace projection using data sphering and low-rank and sparse matrix decomposition for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8704–8722, Oct. 2021.
[19] J. Chen and C.-I. Chang, "Background-annihilated target-constrained interference-minimized filter (TCIMF) for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–24, Sep. 2022.
[20] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.
[21] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1346–1354, Mar. 2015.
[22] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, 2015.
[23] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.
[24] T. Cheng and B. Wang, "Decomposition model with background dictionary learning for hyperspectral target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1872–1884, Jan. 2021.
[25] X. Zhao, W. Li, C. Zhao, and R. Tao, "Hyperspectral target detection based on weighted cauchy distance graph and local adaptive collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Apr. 2022.
[26] W. Xie, J. Lei, J. Yang, Y. Li, Q. Du, and Z. Li, "Deep latent spectral representation learning-based hyperspectral band selection for target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2015–2026, Mar. 2020.
[27] G. Zhang, S. Zhao, W. Li, Q. Du, Q. Ran, and R. Tao, "HTD-Net: A deep convolutional neural network for target detection in hyperspectral imagery," *Remote Sens.*, vol. 12, no. 9, 2020, Art. no. 1489.
[28] D. Zhu, B. Du, and L. Zhang, "Two-stream convolutional networks for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6907–6921, Aug. 2021.
[29] W. Rao, L. Gao, Y. Qu, X. Sun, B. Zhang, and J. Chanussot, "Siamese transformer network for hyperspectral image target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–19, Mar. 2022.
[30] Y. Shi, J. Li, Y. Zheng, B. Xi, and Y. Li, "Hyperspectral target detection with RoI feature transformation and multiscale spectral attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5071–5084, Jun. 2021.
[31] L. Ren, D. Hong, L. Gao, X. Sun, M. Huang, and J. Chanussot, "Orthogonal subspace unmixing to address spectral variability for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–13, Jan. 2023.
[32] F. Ye, Z. Wu, Y. Xu, H. Liu, and Z. Wei, "Bayesian hyperspectral image super-resolution in the presence of spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Dec. 2022.
[33] Y. Li, Y. Chong, S. Pan, and Y. Ding, "First-order smoothing-based deep graph network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–16, Mar. 2023.
[34] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Sep. 2019.
[35] K. Roy, G. Krishna, R. Dubey, and B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.
[36] M. Ahmad et al., "A disjoint samples-based 3D-CNN with active transfer learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Sep. 2022.
[37] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.
[38] B. Wu, Q. Xie, and B. Wu, "Seismic impedance inversion based on residual attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Jul. 2022.
[39] W. Guo, H. Ye, and F. Cao, "Feature-grouped network with spectral–spatial connected attention for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, Jan. 2022.
[40] S. M. Siniscalchi and V. M. Salerno, "Adaptation to new microphones using artificial neural networks with trainable activation functions," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1959–1965, Aug. 2017.
[41] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.
[42] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Feb. 2008.
[43] M. Song, X. Shang, and C.-I. Chang, "3-D receiver operating characteristic analysis for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8093–8115, Nov. 2020.
[44] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2021.
[45] C. Ferri, J. Hernández-Orallo, and P. Flach, "A coherent interpretation of AUC as a measure of aggregated classification performance," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 657–664.
[46] Z. Wu et al., "Hyperspectral anomaly detection with relaxed collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Jul. 2022.
[47] P. Wang, L. Wang, H. Leung, and G. Zhang, "Super-resolution mapping based on spatial–spectral correlation for spectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2256–2268, Mar. 2021.
[48] C. Yi, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral image super-resolution based on spatial and spectral correlation fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4165–4177, Jul. 2018.
[49] C. Shi, H. Wu, and L. Wang, "A positive feedback spatial-spectral correlation network based on spectral slice for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–17, Feb. 2023.

**Xiaoyi Wang** received the M.A. degree in signal and information processing from Heilongjiang University, Harbin, China, in 2018. She is currently working toward the Ph.D. degree information and communication engineering with the College of Information and Communication Engineering, Harbin Engineering University, Harbin, China.

Since 2022, she has been a visiting Ph.D. student with Telecommunications and Remote Sensing Laboratory, University of Pavia, Pavia, Italy, which is supported by China Scholarship Council. Her research interests include remote sensing image fusion and hyperspectral anomaly detection.

**Liguo Wang** received the M.S. degree in natural science and the Ph.D. degree in engineering from the Harbin Institute of Technology, Harbin, China, in 2002 and 2006, respectively.

He held a postdoctoral research position from 2006 to 2008 with the College of Information and Communications Engineering, Harbin Engineering University. He is currently a Professor with the College of Information and Communication Engineering, Dalian Minzu University, Dalian, China. His research interests include remote sensing image processing and machine learning. He has authored or coauthored four books, 40 patents, and more than 260 papers in journals and conference proceedings.



**Qunming Wang** received the Ph.D. degree in photogrammetry and remote sensing from Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. From 2017 to 2018, he was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., where he is currently a Visiting Professor. His three-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored more than 70 peer-reviewed articles in international journals, such as *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Dr. Wang is an Editorial Board member of Remote Sensing of Environment and serves as an Associate Editor for the *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering and Remote Sensing*. He was also an Associate Editor for *Computers and Geosciences* from 2017 to 2020.

**Anna Vizziello** (Senior Member, IEEE) received the Laurea degree in electronic engineering and the Ph.D. degree in electronics and computer science from the University of Pavia, Pavia, Italy, in 2007 and 2011, respectively.

She is an Assistant Professor with Telecommunications and Remote Sensing Laboratory, University of Pavia, Italy. From 2007 to 2009, she also collaborated with the European Centre for Training and Research in Earthquake Engineering working with Telecommunications and Remote Sensing Group. From 2009 to 2010, she was a Visiting Researcher with Broadband Wireless Networking Lab, Georgia Institute of Technology, Atlanta, GA, USA; in summer 2009 and 2010, with Universitat Politècnica de Catalunya, Barcelona, Spain; and in winter 2011 and summer 2016, with Northeastern University, Boston, MA, USA. She has been included in the 2018 list of "N2Women: Rising Stars in Computer Networking and Communications" for outstanding and impactful contributions in the area of networking/communications, supported by the IEEE Communication Society. Her research interests include signal processing, wireless communication, and sensor systems.

**Paolo Gamba** (Fellow, IEEE) received the Laurea degree in electronic engineering "*cum laude*" and the Ph.D. degree in electronic engineering from the University of Pavia, Pavia, Italy, in 1989 and 1993, respectively.

He is a Professor with Telecommunications and Remote Sensing Laboratory, University of Pavia, Italy. He served as an Editor-in-Chief for the *IEEE Geoscience and Remote Sensing Letters* from 2009 to 2013 and the Chair of the Data Fusion Committee of the *IEEE Geoscience and Remote Sensing Society* (GRSS) from October 2005 to May 2009. He was elected to the GRSS AdCom in 2014, served as GRSS President from 2019 to 2020, and is currently the GRSS Past President. He was the organizer and Technical Chair of the biennial GRSS/ISPRS Joint Workshops on "Remote Sensing and Data Fusion Over Urban Areas" from 2001 to 2015. He also served as the Technical Co-Chair of the 2010, 2015, and 2020 IGARSS conferences, Honolulu (Hawaii), Milan (Italy), and online, respectively. He has been invited to give keynote lectures and tutorials on several occasions about urban remote sensing, data fusion, and EO data for physical exposure and risk management. He authored or coauthored more than 180 papers in international peer-reviewed journals and presented 320 research works in workshops and conferences.