# Target Detection With Unconstrained Linear Mixture Model and Hierarchical Denoising Autoencoder in Hyperspectral Imagery

Yunsong Li, *Member, IEEE*, Yanzi Shi, *Graduate Student Member, IEEE,* Keyan Wang, Bobo Xi, *Graduate Student Member, IEEE,* Jiaojiao Li, *Member, IEEE,* and Paolo Gamba, *Fellow, IEEE*

***Abstract*—Hyperspectral imagery with very high spectral resolution provides a new insight for subtle nuances identification of similar substances. However, hyperspectral target detection faces significant challenges of intraclass dissimilarity and interclass similarity due to the unavoidable interference caused by atmosphere, illumination, and sensor noise. In order to effectively alleviate these spectral inconsistencies, this paper proposes a novel target detection method without strict assumptions on data distribution based on an unconstrained linear mixture model and deep learning. Our proposed detector firstly reduces interference via a specifically designed deep-learning-based hierarchical denoising autoencoder, and then carries out accurate detection with a two-step subspace projection, aiming at background suppression and target enhancement. Additionally, to generate representative background and reliable target samples required in the detection procedure, an efficient spatial-spectral unified endmember extraction method has been developed. Performance comparison with several state-of-the-art detection methods and further analysis on four real-world hyperspectral images demonstrate the effectiveness and efficiency of our proposed target detector.**

***Index Terms*—Unconstrained linear mixture model, hierarchical denoising autoencoder, spatial-spectral unification, hyperspectral target detection.**

## I. Introduction

With the development of spectroscopic imaging technology, optical remote sensing, mainly including grayscale panchromatic (PAN), multispectral, superspectral and hyperspectral images (HSIs), could cover an increasingly broader electromagnetic spectrum. In comparison with the former three types of images, HSIs provide much more abundant spectral information for better material components characterization, usually ranging from visible to infrared response spectrum. This characteristic quickly attracted extensive attention in the research domains of classification, detection, and image quality enhancement [1], and aroused interests in various practical applications, such as camouflage identification [2], man-made object detection [3], mineral exploration [4], land cover mapping [5], etc.

Hyperspectral target detection (HTD) aims at detecting interested objects in HSIs from a joint spatial-spectral perspective. However, due to the limitations of current spectroscopic techniques, fine spatial and spectral resolutions cannot be simultaneously obtained, usually leading to a low target occupation in the hyperspectral scene. Therefore, HTD produces pixel-level or subpixel-level soft-decision maps, i.e., it assigns larger (smaller) response values to target (background) pixels, a result which is significantly different from box-level detection and pixel-level segmentation in PAN/RGB images. It should be noted that the prior knowledge required for HTD is quite limited, and usually one target spectrum is adequate.

Exploiting HSIs to perform target detection can be dated back to the 1980s. So far, the existing target detection methods can be roughly divided into five families as follows.

(1) Structured background model based detector (SBMbD): these algorithms physically characterize HSIs with a linear mixture model (LMM), which considers that each pixel can be approximately represented as a linear combination of several endmembers with different fractional abundances. Typical structured approaches include adaptive matched subspace detector (AMSD) [6], constrained energy minimization (CEM) [7], orthogonal subspace projection (OSP) [8], fully/non-negativity/sum-to-one constrained least squares (FCLS/NCLS/SCLS) [9], etc.

(2) Unstructured background model based detector (UBMbD): different from SBMbDs, UMBbDs statistically model background with a generic distribution (e.g., multivariate normal distribution), and then construct a target detector via binary hypothesis testing. Generalized likelihood ratio test (GLRT) [10], adaptive matched filter (AMF) [11], and adaptive cosine/coherent estimator (ACE) [12] are kinds of unstructured detectors obtained under different hypothesis models.

(3) Hybrid background model based detector (HBMbD): the HBMbD methods integrate unstructured and structured

This work was supported in part by the National Key Research and Development Program of China under Grant 2018AAA0102702(no. 2018AAA0102702), the National Nature Science Foundation of China (no. 61901343), the China Postdoctoral Science Foundation (no. 2017M623124), the China Postdoctoral Science Special Foundation (no.2018T111019) and the China Scholarship Council Program under Grant 202006960028. The project was also partially supported by the Open Research Fund of CAS Key Laboratory of Spectral Imaging Technology (no.LSIT201924W) and the Fundamental Research Funds for the Central Universities JB190107. It was also partially supported by the National Nature Science Foundation of China (no. 61571345, 61671383, 91538101, 61501346 and 61502367), Yangtse Rive Scholar Bonus Schemes (No. CJT160102), Ten Thousand Talent Program, and the 111 project (B08038). *(Corresponding authors: Yanzi Shi; Keyan Wang.)*

Y. Li, Y. Shi, K. Wang, B. Xi, and J. Li are with the State Key Laboratory of Integrated Service Network, Xidian University, Xi'an 710071, China (e-mail: ysli@mail.xidian.edu.cn; yzshi_xidian@163.com; kywang@mail.xidian.edu.cn; xibobo1301@foxmail.com; jjli@xidian.edu.cn).

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

background model for target detection with both physical and statistical significance. In [6], Broadwater *et al.* developed two hybrid detectors, called hybrid unstructured detector (HUD) and hybrid structured detector (HSD), to provide a non-negative and sum-to-one abundance estimation using the MSD and ACE approaches, respectively. Considering that not all endmembers are contained in each pixel, Zhang *et al.* further improved HUD and HSD via adaptively selecting endmembers for different pixels [13]. The above statistical detection methods have truly formed the cornerstone for the development of HTD, and could achieve well performance under specific assumptions. However, there is often a mismatch between actual data distribution and theoretical model description, leading to unavoidable performance degradation. Moreover, spatial structures of the cubic HSIs are totally ignored in such methods, which offer opportunities to build a more discriminative detector.

(4) Machine learning based detector (MLbD): machine learning has been successfully applied in HTD, especially kernel-based machine learning, and sparse and collaborative representation. Several kernel-based target detectors have been proposed by extending classical statistical methods, including kernel orthogonal subspace projection (KOSP) [14], kernel adaptive subspace detector (KASD) [15], and kernel target-constrained interference-minimized filter (KTCIMF) [16]. Unfortunately, these methods also rely heavily on certain assumptions. As for sparse and collaborative representations, since Chen *et al.* firstly presented a sparsity-based target detector (STD) [17], there have been several other effective works, such as sparse representation-based binary hypothesis-based detector (SRBBHD) [18], hybrid sparsity and statistics detector (HSSD) [19], and combined sparse and collaborative representation (CSCR) [20]. However, in order to meet the requirements of dictionary construction, the representation-based methods need more information about target signatures. If the prior target information is insufficient, the performance of such detectors would be seriously limited.

(5) Deep learning based detector (DLbD): pixel/subpixel-level DLbD in HSI is highly different from box-level object detection in RGB/PAN images due to the insufficient labeled training sample and target-background class imbalance. The application of deep learning to perform HTD is relatively new, and only a few methods have emerged. Supervised detectors mainly enlarge target samples via synthesis and then construct an end-to-end detector by training on large amounts of pixel-pairs. Popular algorithms include convolutional neural network target detector (CNNTD) [21], deep network-based HTD (denoted as HTD-Net) [22], and two-stream convolutional network-based target detector (TSCNTD) [23]. Moreover, a domain adaptive learning model has been proposed to transfer knowledge learned from a large-sample source domain to the small-sample target domain [24]. Unsupervised methods usually enhance the discrimination capacity with unsupervised networks under certain constraints, and then perform target detection with simple matching strategy [25]–[27]. These advanced DLbDs really bring better detection performance and stronger robustness compared with the classical ones.

Despite significant progress in target detection techniques for HSIs, most of these methods are sensitive to target prior due to spectral variations. As an example, in the case of CEM, the performance would be easily deteriorated as the target prior used for detection could be highly different from that in the actual image caused by adjacency effects and illumination variations. This issue should be addressed to achieve target detection with higher accuracy and stronger robustness. Although some methods, like OSP, could minimize the interference, the assumption of Gaussian noise/background is usually violated in reality. Therefore, a more robust detector independent of specific distributions would be of great interest.

In this paper, a novel hyperspectral target detector integrating classical unconstrained LMM (ULMM) and state-of-the-art deep learning technique is proposed to address the issues of spectral variation and inaccurate data distribution assumption. The proposed detector (denoted as ULMMDL) regards each pixel as a linear combination of representative endmembers. Both non-negative and sum-to-one constraints are ignored for a more concise solution of the target detection task, following the work in technical literature [28]. Importantly, ULMMDL is preceded by a hierarchical denoising autoencoder (HDAE) network to reduce interference. Moreover, an automatic weighting optimization strategy is exploited to tackle target-background imbalance issue, thereby making the HDAE processing step more effective.

The main contribution of our paper is therefore the design of a novel HDAE network, which consists of several layers of denosing autoencoders (DAEs) and that can gradually reduce spectral variation layer by layer. Secondly, a simple but distribution-free two-step subspace projection (TSP) detector is developed on the premise of ULMM, where no constraints on fractional abundance are made. Lastly, as the necessity before TSP is to obtain representative background and highly reliable target endmembers, we propose a spatial-spectral unified endmember extraction ($S^2UE^2$) module. Compared with endmember extraction only using spectral information, $S^2UE^2$ has much less computational cost and a better performance.

The remainder of this paper is organized as follows. The ULMM and DAE network are introduced in Section II. Section III provides a description of ULMMDL in detail. In Section IV, experiments and further discussion on four real-world HSIs are given. Finally, Section V reports the conclusions of the research work presented in this paper.

## II. Background

### A. Linear Mixture Model

The linear mixture model (LMM) is as a mature model, and has been widely applied in hyperspectral unmixing and detection due to its simplicity and competitive performance.

LMM represents each spectral vector as a linear mixture of several endmembers with different fractional abundances, as

$$\mathbf{x}_k = \underbrace{\mathbf{U}\boldsymbol{\alpha}_k + \mathbf{I}_k}_{\text{unmixing}} = \underbrace{\mathbf{B}\boldsymbol{\alpha}_{b_k} + \mathbf{D}\boldsymbol{\alpha}_{d_k} + \mathbf{I}_k}_{\text{detection}}$$
$$\mathbf{U} = [\mathbf{B} \quad \mathbf{D}], \boldsymbol{\alpha}_k = [\boldsymbol{\alpha}_{b_k}; \boldsymbol{\alpha}_{d_k}] \tag{1}$$

where

| | |
|---|---|
| $\mathbf{x}_k \in \mathbb{R}^{L\times 1}$ | spectrum of the $k$th pixel. |
| $\mathbf{B} \in \mathbb{R}^{L\times n_b}$ | spectra of background endmembers. |
| $\mathbf{D} \in \mathbb{R}^{L\times n_d}$ | spectra of target endmembers. |
| $\boldsymbol{\alpha}_{b_k} \in \mathbb{R}^{n_b\times 1}$ | fractional abundance of background. |
| $\boldsymbol{\alpha}_{d_k} \in \mathbb{R}^{n_d\times 1}$ | fractional abundance of target. |
| $\mathbf{I}_k \in \mathbb{R}^{L\times 1}$ | inevitable interference. |
| $L$ | the number of bands. |
| $n_b$ | background endmember number. |
| $n_d$ | target endmember number. |

In unmixing, there is no need distincting target and background. However, as for HTD, endmembers are grouped into target and background for better modeling. LMM in the remainder of this paper refers to that for detection.

For physical modeling, popular non-negative and sum-to-one constraints are applied for various HSI processing techniques, which can be denoted as follows

$$\begin{aligned} &\text{non-negativity: } \boldsymbol{\alpha}_{b_k} \geq 0, \boldsymbol{\alpha}_{d_k} \geq 0 \\ &\text{sum-to-one: } \sum_{i=1}^{n_b} \alpha_{b_k}^i + \sum_{j=1}^{n_d} \alpha_{d_k}^j = 1 \end{aligned} \tag{2}$$

Although the physical significance of LMM is guaranteed under equation (2), it would impose high complexity for accurate abundance estimation. Unlike precise quantification of a specific signature in hyperspectral unmixing, HTD pays more attention to its existence, which can be measured by the relative value of abundance, i.e., the larger abundance indicates a target sample and the smaller abundance indicates a background sample. Therefore, we ignore both constraints and build our detector according to the unconstrained LMM (ULMM).

Suppose the target and background endmembers are given, the objective of ULMM-based HTD is to perform unconstrained abundance estimation of target endmembers for each tested pixel, i.e., calculating $\alpha_d^k$. Thus, we firstly annihilate unnecessary $\mathbf{B}$ before carrying out detection of $\mathbf{D}$ with an orthogonal complement space as in OSP

$$\mathbf{S}_{\mathbf{B}}^{\perp} = \mathbf{E}_{L\times L} - \mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T \tag{3}$$

where $\mathbf{E}_{L\times L}$ is an $L$-order identity matrix, and $\mathbf{S}_{\mathbf{B}}^{\perp}$ denotes the desired orthogonal complement space, which is orthogonal to the background subspace. Projecting each tested pixel $\mathbf{x}_k$ into space $\mathbf{S}_{\mathbf{B}}^{\perp}$ will lead to a background-suppressed HSI as follows

$$\mathbf{x}_k^{bs} = \mathbf{S}_{\mathbf{B}}^{\perp}\mathbf{x}_k = \mathbf{S}_{\mathbf{B}}^{\perp}\mathbf{D}\boldsymbol{\alpha}_{d_k} + \mathbf{S}_{\mathbf{B}}^{\perp}\mathbf{I}_k \tag{4}$$

It is obvious that the interference term would cause serious side effects on the detection performance. In order to obtain an effective detector insensitive to interference, OSP designed a linear filter to maximize SNR. However, this approach only works when the interference conforms to Gaussian distribution, which may be hardly satisfied in real scenes.

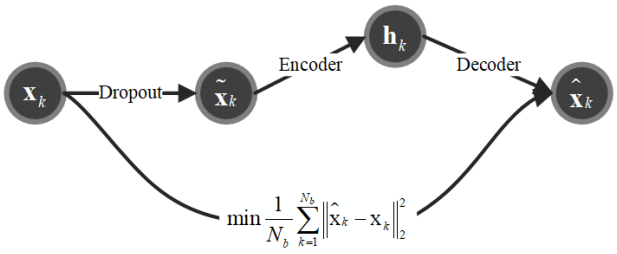



Fig. 1. The typical structure of DAE model.

### B. Denoising Autoencoder

The autoencoder (AE) neural network is designed to reconstruct input through a symmetric structure (encoder-decoder), i.e., minimizing the error between input and output. Differently, denoising autoencoders (DAEs) aim at obtaining clean output from noisy input, which would decrease the sensitivity to interferential or largely variated data, e.g., spectral variations in HSIs. Given the original data $\mathbf{x}_k \in \mathbb{R}^{L\times 1}$, DAE first generates the noisy vector $\widetilde{\mathbf{x}}_k$ as training data by randomly deactivating several input neurons. In this paper, dropout trick [29] is employed to corrupt the input data. The typical structure of DAE model is shown in Fig. 1. The encoding process of DAE can be represented as follows:

$$\mathbf{h}_k = f\left(\boldsymbol{\Omega}_e\widetilde{\mathbf{x}}_k + \boldsymbol{\beta}_e\right) \tag{5}$$

where $\mathbf{h}_k \in \mathbb{R}^{N_h\times 1}$ denotes the hidden high-level representation of $\widetilde{\mathbf{x}}_k$, $N_h$ is the number of hidden neurons, $\boldsymbol{\Omega}_e$ and $\boldsymbol{\beta}_e$ are learnable parameters of encoder, and $f(\cdot)$ is a nonlinear activation function, which can be set as sigmoid, tanh, ReLU, etc. Note that the sigmoid function is used in our work, i.e., $f(a) = \frac{1}{1+e^{-a}}$.

Symmetrically, the decoder part of DAE can be formulated as

$$\hat{\mathbf{x}}_k = f\left(\boldsymbol{\Omega}_d\mathbf{h}_k + \boldsymbol{\beta}_d\right) \tag{6}$$

where $\hat{\mathbf{x}}_k$ is the approximate reconstruction of clean $\mathbf{x}_k$, $\boldsymbol{\Omega}_d$ and $\boldsymbol{\beta}_d$ are learnable parameters of the decoding part of DAE.

Based on the encoding and decoding processes, the loss function of DAE can be calculated as

$$\min_{\boldsymbol{\Omega}_\mathbf{e},\boldsymbol{\Omega}_\mathbf{d},\boldsymbol{\beta}_e,\boldsymbol{\beta}_d} \frac{1}{N_b}\sum_{k=1}^{N_b}\|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_2^2 \tag{7}$$

where $N_b$ is the minibatch size.

## III. Proposed Method

Fig. 2 shows the overall flowchart of our proposed ULMMDL method. ULMMDL takes a three-dimensional HSI and a prior-known target spectral vector as inputs, and is then performed by the following three submodules.

1) *HDAE*. The HDAE network is specifically designed to enhance spectral consistency via layer-by-layer iterations of DAE, thereby facilitating subsequent target detection.

2) *$S^2UE^2$*. Considering that the background endmembers, supposed to be available in ULMM (see equation (3)), must be extracted from the HSI itself, $S^2UE^2$ associating spatial

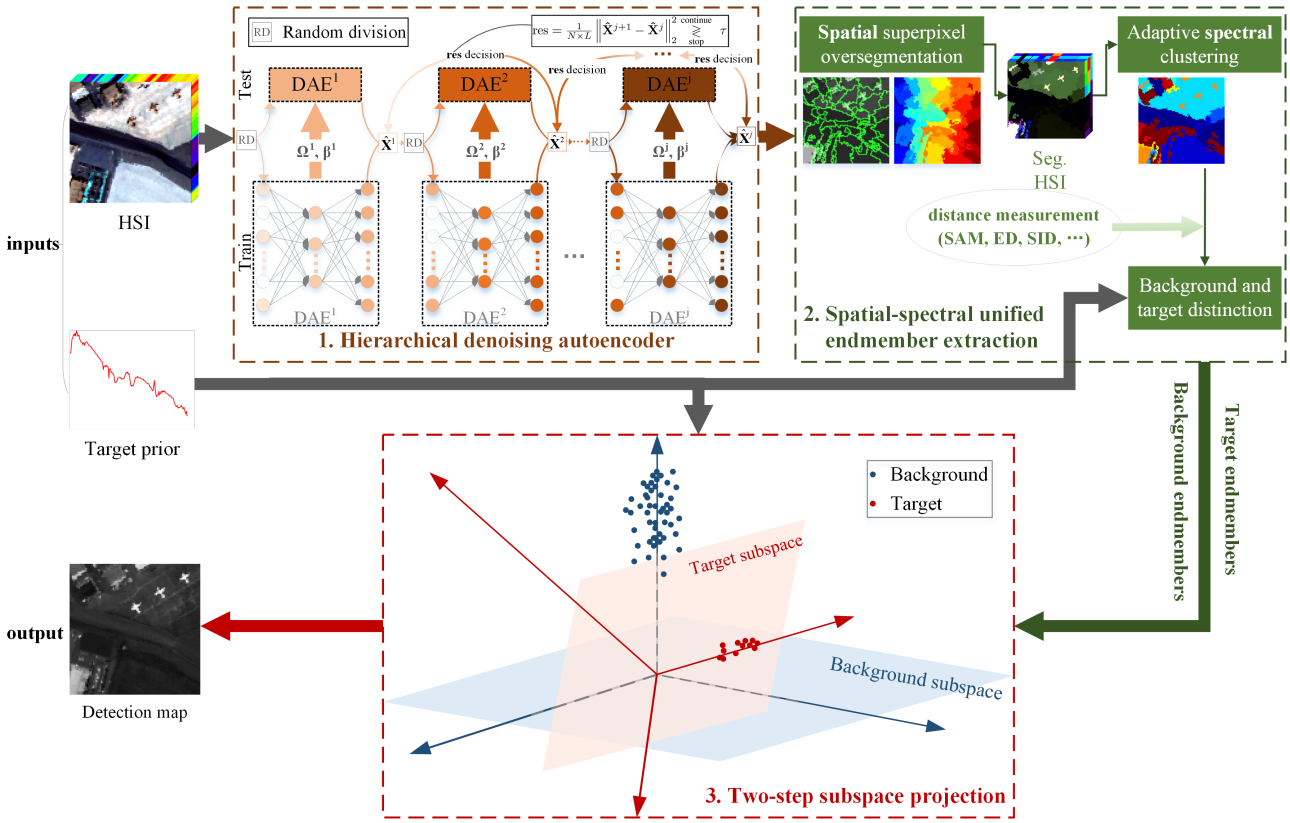



Fig. 2. The overall flowchart of the proposed ULMMDL, where TSP based on ULMM, as in equation (1) is the main branch to perform HTD, HDAE and S$^2$UE$^2$ aim to reduce $\mathbf{I}$ and obtain $\mathbf{B}$ and $\mathbf{D}$ for follow-up detection, respectively.

segmentation and spectral clustering is presented to obtain not only representative background endmembers, but also reliable target endmembers as a complement to the target prior, which can further improve robustness of our detection method.

3) *TSP*. With the interference-reduced HSI and necessary endmembers, a simple TSP detector is developed to generate a soft-decision detection output. Moreover, there is no need for TSP making implicit assumptions on data distribution, so the performance degrades less when a mismatch between actual data distribution and theoretical model description occurs.

### A. Hierarchical denoising autoencoder

The use of AEs for hyperspectral data denoising and analysis (e.g., dimension reduction and pansharpening) has been introduced in [30]–[32]. Different from the widely used stacked DAE (SDAE) focusing more on extracting useful features by applying a series of DAEs on the hidden layer, we construct a novel HDAE neural network to alleviate spectral interference, i.e., $\mathbf{I}_k$ in equation (1), by applying several DAEs on the output layer, as shown in Fig. 2. HDAE consists of a series of simple DAEs, where the output of the former DAE is sent to the input layer of the next DAE until the DAE's reconstruction remains roughly unchanged.

Consider an HSI with $N$ pixels and $L$ bands denoted as $\{\mathbf{x}_n \in \mathbb{R}^{L\times 1}, n = 1, 2, \cdots, N\}$, and a target prior denoted as $\mathbf{d} \in \mathbb{R}^{L\times 1}$. We randomly divide the HSI preprocessed by band normalization into two non-overlapped sets for training and testing. As introduced in Section II-B, the reconstructed output of DAE for the $k$th training sample $\mathbf{x}_k$ can be represented as

$$\begin{aligned} \widetilde{\mathbf{x}}_k &= \text{dropout}\left(\mathbf{x}_k, p\right) \\ \mathbf{h}_k &= f\left(\boldsymbol{\Omega}_e \widetilde{\mathbf{x}}_k + \boldsymbol{\beta}_e\right) \\ \hat{\mathbf{x}}_k &= f\left(\boldsymbol{\Omega}_d \mathbf{h}_k + \boldsymbol{\beta}_d\right) \end{aligned} \tag{8}$$

where $p$ is a probability parameter of dropout.

However, considering that solely error measurement between spectral curves with Euclidean distance (ED) in equation (7) is insufficient, we further introduce spectral angle mapper (SAM) to supplement the loss function, so that HDAE could better preserve spectral consistency from the viewpoint of amplitude and slope

$$\begin{aligned} \min_{\boldsymbol{\Omega}_\mathbf{e}, \boldsymbol{\Omega}_\mathbf{d}, \boldsymbol{\beta}_e, \boldsymbol{\beta}_d} & \frac{1}{N_b}\sum_{k=1}^{N_b} \left\|\hat{\mathbf{x}}_k - \mathbf{x}_k\right\|_2^2 \\ & + \lambda \frac{1}{N_b}\sum_{k=1}^{N_b} \arccos \frac{\hat{\mathbf{x}}_k^T \mathbf{x}_k}{\left\|\hat{\mathbf{x}}_k\right\|_2 \left\|\mathbf{x}_k\right\|_2} \end{aligned} \tag{9}$$

Additionally, the phenomenon of high imbalance between target and background signatures is quite usual in HTD, and may cause error reconstruction of small-sample target class, i.e., falsely reconstructing target as background. To deal with this issue, we apply an automatic weighting strategy to

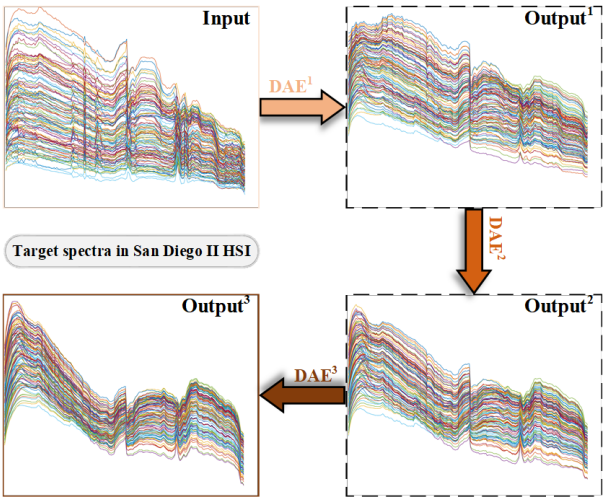



Fig. 3. All output target spectra of HDAE with different layers on the San Diego II hyperspectral data set.

establish an improved loss, formulated as follows:

$$\min_{\mathbf{\Omega_e},\mathbf{\Omega_d},\beta_e,\beta_d} \frac{1}{N_b}\sum_{k=1}^{N_b}\sigma_k^1\left\|\hat{\mathbf{x}}_k-\mathbf{x}_k\right\|_2^2 + \lambda\frac{1}{N_b}\sum_{k=1}^{N_b}\sigma_k^2\arccos\frac{\hat{\mathbf{x}}_k^T\mathbf{x}_k}{\left\|\hat{\mathbf{x}}_k\right\|_2\left\|\mathbf{x}_k\right\|_2} \quad (10)$$

in which $\sigma_k^1=e^{-\left\|\mathbf{x}_k-\mathbf{d}\right\|_2^2}$ and $\sigma_k^2=\frac{\mathbf{x}_k^T\mathbf{d}}{\left\|\mathbf{x}_k\right\|_2\left\|\mathbf{d}\right\|_2}$ assign larger weight to target samples, so that it can avoid small-sample target to be overly reconstructed as large-sample background, and $\lambda$ is a pre-defined trade-off coefficient to control the importance of SAM.

Now, for the $j$th DAE, its output of each pixel in the HSI, represented as $\hat{\mathbf{X}}^j=\left[\hat{\mathbf{x}}_1^j \quad \hat{\mathbf{x}}_2^j \quad \cdots \quad \hat{\mathbf{x}}_N^j\right]$ would be sent forward to the $(j+1)$th DAE for deeper inference alleviation. The aforementioned processes would be terminated until the average residual between $\hat{\mathbf{X}}^j$ and $\hat{\mathbf{X}}^{j+1}$ is less than a specific number, which can be calculated as follows

$$\mathrm{res}=\frac{1}{N\times L}\left\|\hat{\mathbf{X}}^{j+1}-\hat{\mathbf{X}}^j\right\|_2^2 \quad (11)$$

If res is less than a very small number (set as $0.001$ in this paper), the training of HDAE will stop, and the $j$th DAE's output $\hat{\mathbf{X}}^j$ would be regarded as the final clean HSI and used as a basis for follow-up processing.

For the example of San Diego II, Fig. 3 shows the target spectral curves of the reconstructed HSI with different layers to verify the effectiveness of the HDAE approach. It can be visually observed that the deviation among target spectra gradually decreases as the layer goes deeper. Further quantitative analysis is provided in the Experimental Results section.

### *B. Spatial-spectral unified endmember extraction*

The typical spectral endmember extraction methods are usually restricted by either expensive computational load (e.g., pixel purity index [33]) or high sensitivity to noise/outliers (e.g., automatic target generation process [34]). To address this issue, applying superpixel segmentation before endmember extraction has been proposed [35]–[38]. In this paper, we also reduces the number of endmember candidates by superpixel oversegmentation in spatial domain, and then apply an adaptive spectral clustering (ASC) method to obtain the final endmembers. Different from the existing superpixel-based endmember extraction methods, the number of endmembers is automatically determined according to the substance diversity in the scene, rather than a predefined fixed value. This spatial-spectral unified strategy will significantly decrease time consumption, and exploiting spatial information could avoid treating noisy samples as endmembers to a certain extent. Besides, considering that the endmembers extracted by $S^2UE^2$ might include both background and target spectra, we need to separate them to facilitate the subsequent TSP detection. Extraction of representative background and reliable target endmembers consists of three steps as follows.

**Step 1** Spatial superpixel oversegmentation.
*Input*: clean HSI $\hat{\mathbf{X}}$.

(1) Adopt principle component analysis (PCA) to obtain the first principle component of HSI $\hat{\mathbf{X}}_{\mathrm{pc}_1}$;
(2) Oversegment $\hat{\mathbf{X}}_{\mathrm{pc}_1}$ into $N_s$ continuous areas, i.e., superpixel, with entropy rate superpixel (ERS) method [39], where $N_s$ is a large number to avoid omission of rarely occurred endmembers;
(3) Calculate the average spectra of each superpixel as endmember candidates, denoted as $\hat{\mathbf{X}}^{\mathrm{ec}}=\left[\hat{\mathbf{x}}_1^{\mathrm{ec}} \quad \hat{\mathbf{x}}_2^{\mathrm{ec}} \quad \cdots \quad \hat{\mathbf{x}}_{N_s}^{\mathrm{ec}}\right]$.

*Output*: endmember candidates $\hat{\mathbf{X}}^{\mathrm{ec}}$.
**Step 2** ASC.
*Input*: endmember candidates $\hat{\mathbf{X}}^{\mathrm{ec}}$.

(1) Set stop criterion for ASC;
(2) Initialize cluster centers;
(3) Classify each endmember candidates to the closest cluster under a specific distance measurement, e.g., SAM, ED, and spectral information divergence (SID) (SAM applied in this paper), and discard clusters with relatively small samples;
(4) Combine two clusters with small interclass distance/divide one cluster with large intraclass deviation into two, and update the cluster centers;
(5) Determine whether the stop criterion is met. If yes, output the cluster centers as refined endmembers; else, go back to (3).

*Output*: refined endmembers $\hat{\mathbf{X}}^{\mathrm{re}}$.
**Step 3** Background and target endmembers distinction.
*Input*: target prior $\mathbf{d}$, refined endmembers $\hat{\mathbf{X}}^{\mathrm{re}}$.

(1) Calculate SAM (use the same distance measurement with Step 2) between $\mathbf{d}$ and each endmember $\hat{\mathbf{X}}^{\mathrm{re}}$;
(2) The endmembers with SAM smaller than $\tau_1$ are selected as background, while those with SAM larger than $\tau_2$ are regarded as target. In order to guarantee that no target (background) is mistaken as background (target), we set $\tau_1<\tau_2$. In this paper, they are set as $0.98$ and $0.99$, respectively.

*Output*: representative background endmembers $\hat{\mathbf{B}}$ and reliable

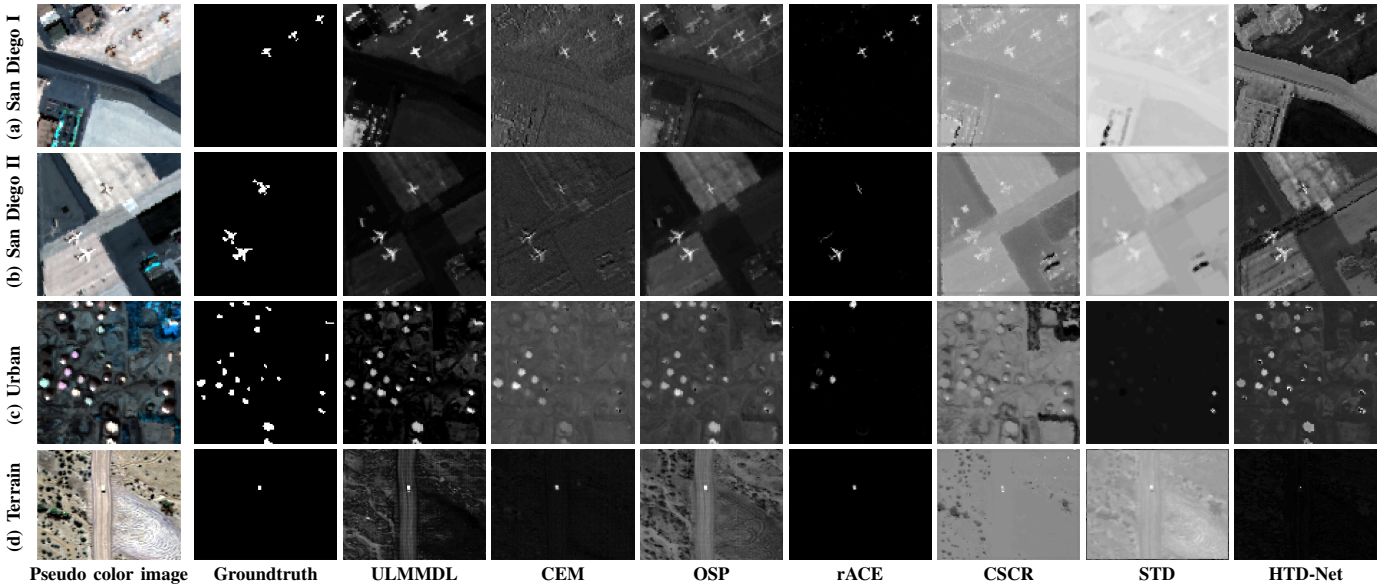



Fig. 4. Experimental HSIs and detection maps of different competitors. From left to right: the pseudo color image, groundtruth, detection maps of ULMMDL, CEM, OSP, rACE, CSCR, STD, and HTD-Net.

target endmembers $\hat{\mathbf{D}}$.

### C. Two-step subspace projection

Once the clean HSI $\hat{\mathbf{X}}$, representative background $\hat{\mathbf{B}}$ and target signature, including supplementary $\hat{\mathbf{D}}$ and prior $\mathbf{d}$, are completely prepared, ULMM can be formulated as

$$\hat{\mathbf{x}}_k = \hat{\mathbf{B}}\boldsymbol{\alpha}_{b_k} + \hat{\mathbf{d}}_j\alpha_{d_k} + \hat{\mathbf{I}}_k, k = 1, 2, \cdots, N \tag{12}$$

where only one target spectrum, denoted as $\hat{\mathbf{d}}_j$, is selected in ULMM for model stability, and $\hat{\mathbf{I}}_k \ll \mathbf{I}_k$, which will lead to much less negative impacts on detection performance.

As illustrated in Section II-A, our TSP detector first constructs an orthogonal complement space $\mathbf{S}^{\perp}_{\hat{\mathbf{B}}}$, and then maps the interference-reduced HSI $\hat{\mathbf{X}}$ into $\mathbf{S}^{\perp}_{\hat{\mathbf{B}}}$, which can effectively improve background suppressibility. The background-suppressed data $\hat{\mathbf{X}}^{bs}$ can be calculated as follows

$$\hat{\mathbf{X}}^{bs} = \mathbf{S}^{\perp}_{\hat{\mathbf{B}}}\hat{\mathbf{X}} = \mathbf{S}^{\perp}_{\hat{\mathbf{B}}}\hat{\mathbf{d}}_j\alpha_d = \hat{\mathbf{d}}'_j\alpha_d \tag{13}$$

in which the interference term is ignored here, and $\hat{\mathbf{d}}'_j$ represents the target signatures corresponding to the background-suppressed HSI.

For target enhancement, we design multiple target subspace detectors using each signatures $\hat{\mathbf{d}}'_j$ in $\left[\mathbf{S}^{\perp}_{\hat{\mathbf{B}}}\hat{\mathbf{D}} \quad \mathbf{S}^{\perp}_{\hat{\mathbf{B}}}\mathbf{d}\right]$, and then apply a simple winner-take-all operation to combine all these detectors. It can be defined as

$$\mathbf{Y} = \max\left\{\left(\hat{\mathbf{d}}'^T_j\hat{\mathbf{d}}'_j\right)^{-1}\hat{\mathbf{d}}'^T_j\hat{\mathbf{X}}_{bs}\right\}_{j=1}^{N_d} \tag{14}$$

in which $\mathbf{Y}$ is the final detection map, $N_d$ is the number of target spectral signatures. The multiple targets strategy could further deal with spectral variation, and provide ULMMDL with much stronger robustness.

TABLE I
DETAILED INFORMATION OF FOUR REAL-WORLD HSIS. "-" DENOTES THE CORRESPONDING TERM IS UNKNOWN.

| HSIs | Sensor | SpeRes[3] | SpaRes[3] | Size (H, W, L)[1] | Target |
|---|---|---|---|---|---|
| **San Diego I** | AVIRIS[2] | 10nm | 3.5m | 100, 100, 189 | airplane |
| **San Diego II** | AVIRIS[2] | 10nm | 3.5m | 100, 100, 189 | airplane |
| **Urban** | AVIRIS[2] | 10nm | 7.1m | 100, 100, 205 | storage tank |
| **Terrain** | - | - | - | 103, 134, 167 | car |

[1] H, W, and L represent the height, width, and the number of bands in the corresponding hyperspectral scene, respectively.
[2] AVIRIS is abbreviated from Airborne Visible/ InfraRed Imaging Spectrometer, and it covers wavelengths ranging from 371nm to 2510nm.
[3] SpeRes and SpaRes are the abbreviations for spectral and spatial resolution, respectively.

## IV. EXPERIMENTAL RESULTS

### A. Experimental setup

*1) Datasets description:* Four real-world HSIs, named San Diego I, San Diego II, Urban, and Terrain, are applied for performance evaluation of our proposed ULMMDL detector, and the detailed information of these datasets are provided in Table I. Except for the first two frequently used datasets, the Urban HSI can be downloaded on this website[1] [40], and the last data is provided by Dr. Ahmad W. Bitar [41]. It should also be noted that all these HSIs have been converted from radiance to reflectance by atmospheric correction and the low-quality bands are removed. Fig. 4 gives the pseudo color image and corresponding groundtruth map where the location of target pixels is marked.

*2) Evaluation criterion:* In order to measure detection performance in hyperspectral imagery, besides solely visualization, receiver operating characteristic (ROC) analysis is widely applied, including a 3D ROC curve specified by threshold $\tau$, detection probability $\mathrm{P_d}$ and false alarm rate $\mathrm{P_f}$, and three 2D

[1] http://xudongkang.weebly.com/

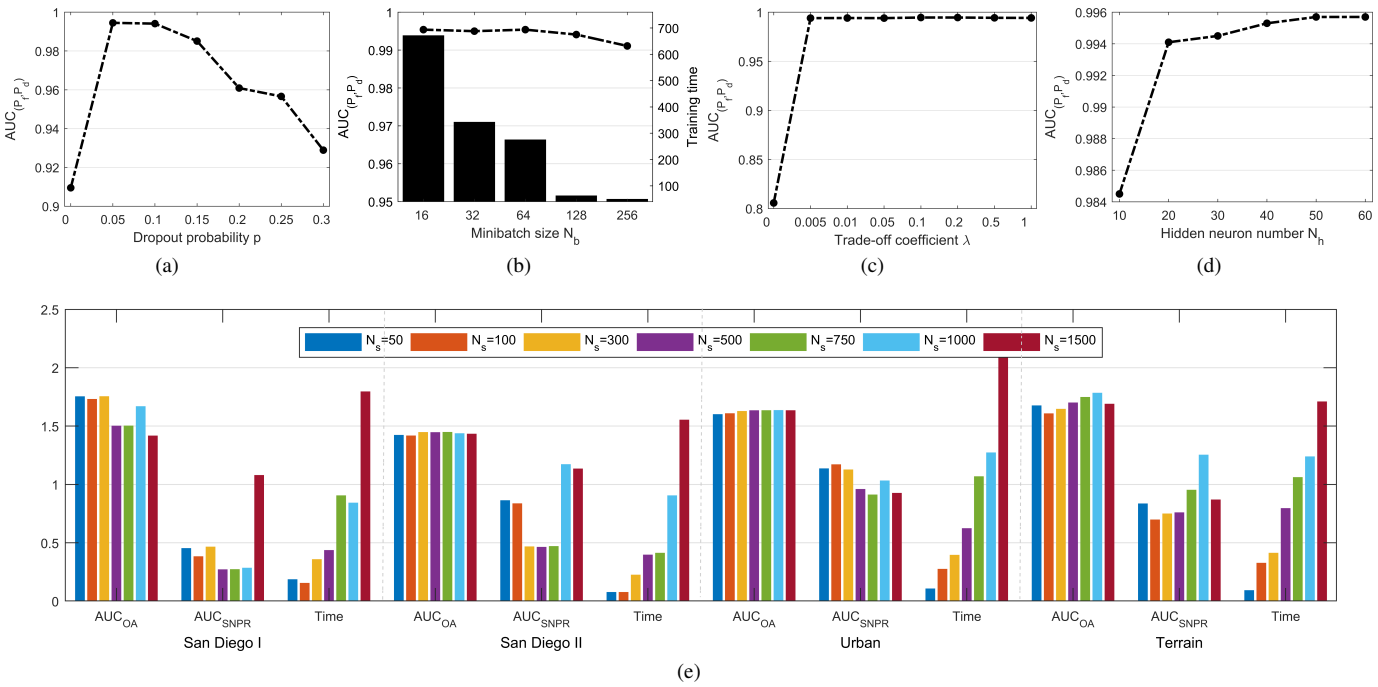



Fig. 5. Impacts of parameter selection on detection performance. (a) Dropout probability $p$, (b) minibatch size $N_b$, (c) trade-off coefficient $\lambda$, (d) the number of hidden neurons $N_h$, and (e) the number of oversegmented superpixels $N_s$ (Noted that the accuracy and time are scaled by different factors for better illustration).

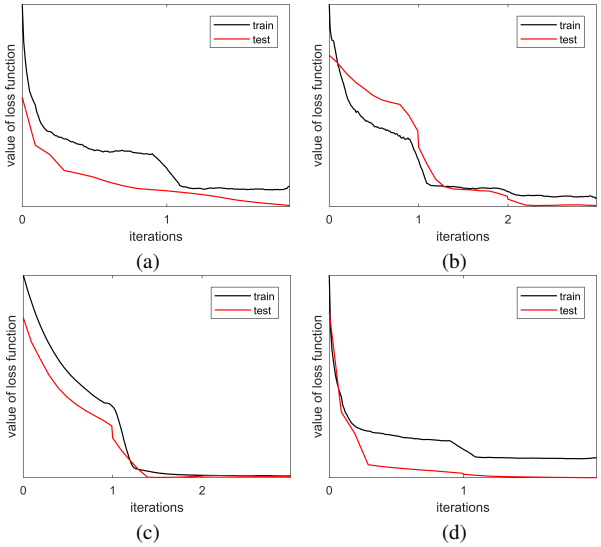



Fig. 6. Loss curves of HDAE on four HSIs over different iterations. (a) San Diego I, (b) San Diego II, (c) Urban, and (d) Terrain.

ROC curves of $(\mathrm{P_f}, \mathrm{P_d})$, $(\tau, \mathrm{P_d})$, and $(\tau, \mathrm{P_f})$, where both $\mathrm{P_f}$ and $\mathrm{P_d}$ are functions of $\tau$. They can be obtained by

$$\mathrm{P_d^{(\tau)}} = \frac{\mathrm{TP^{(\tau)}}}{\mathrm{N_{tgt}}}, \mathrm{P_f^{(\tau)}} = \frac{\mathrm{FP^{(\tau)}}}{\mathrm{N_{bkg}}} \tag{15}$$

where $\mathrm{TP}^{(\tau)}$ and $\mathrm{FP}^{(\tau)}$ represent the number of target and background pixels whose detection value is larger than the given threshold $\tau$. $N_{tgt}$ and $N_{bkg}$ are the total number of target and background pixels, respectively. The 2D ROC curves of $(\mathrm{P_f}, \mathrm{P_d})$, $(\tau, \mathrm{P_d})$, and $(\tau, \mathrm{P_f})$ could indicate overall detection effectiveness, target detectability, and background suppressibility of different HTD methods, respectively. In practice, a detector whose 2D ROC curves of $(\mathrm{P_f}, \mathrm{P_d})$, $(\tau, \mathrm{P_d})$, and $(\tau, \mathrm{P_f})$ are close to the upper left, upper right, and lower left corner of the coordinate axis is actually desired. However, there generally exists intersection between two ROC curves, so that it is difficult to tell which one has better performance. To provide a quantitative analysis of ULMMDL, the area under curve (AUC) of three 2D ROCs, denoted as $\mathrm{AUC}_{(\mathrm{P_f},\mathrm{P_d})}$, $\mathrm{AUC}_{(\tau,\mathrm{P_d})}$, and $\mathrm{AUC}_{(\tau,\mathrm{P_f})}$ are particularly used. A well performed detector should have larger $\mathrm{AUC}_{(\mathrm{P_f},\mathrm{P_d})}$ $(\to 1)$ and $\mathrm{AUC}_{(\tau,\mathrm{P_d})}$ $(\to 1)$, but smaller $\mathrm{AUC}_{(\tau,\mathrm{P_f})}$ $(\to 0)$. Moreover, Chang developed two more indicators based on the aforementioned three AUCs for all-around evaluation [42], calculated as follows

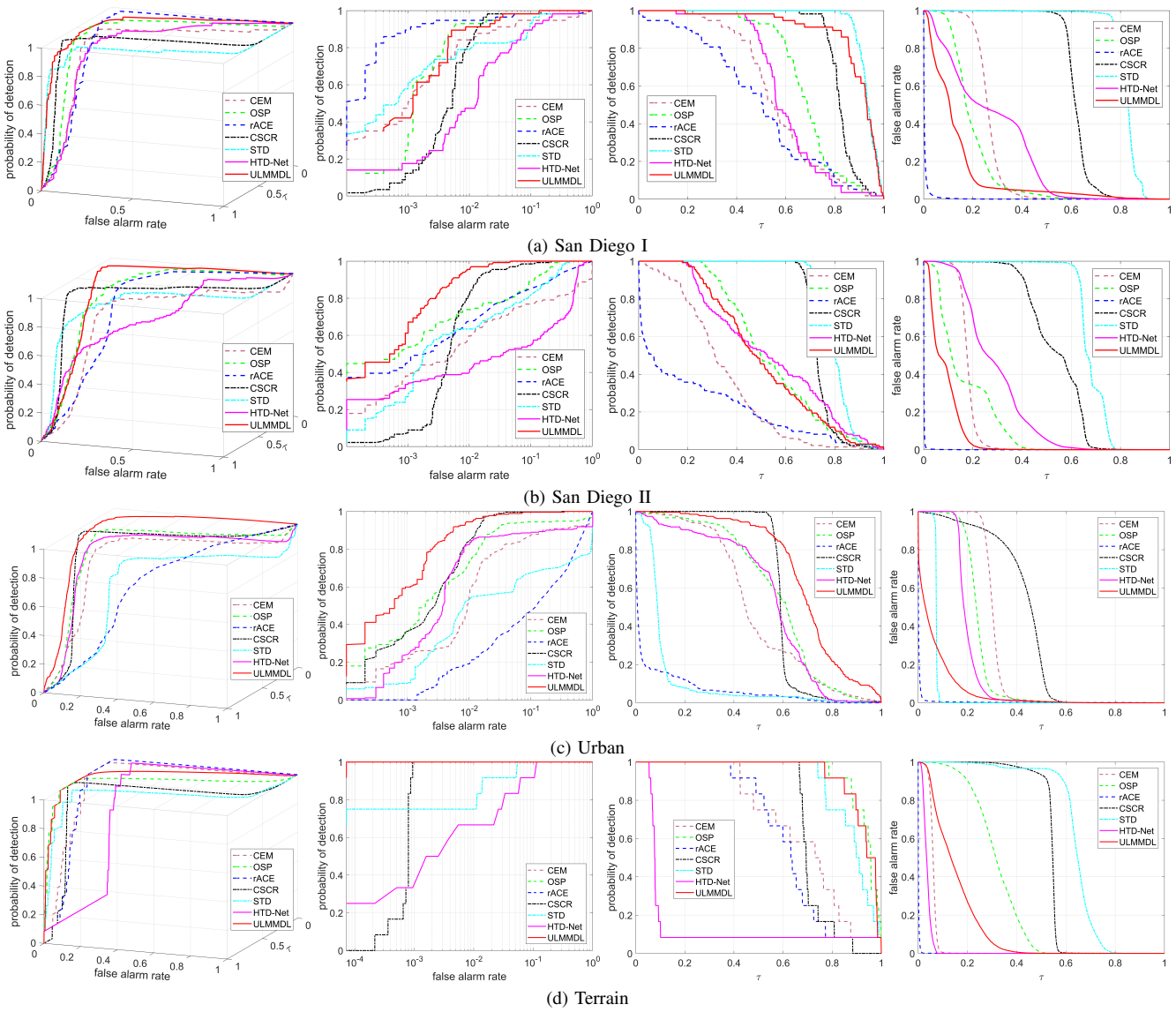
$$\mathrm{AUC_{OA}} = \mathrm{AUC}_{(\mathrm{P_f},\mathrm{P_d})} + \mathrm{AUC}_{(\tau,\mathrm{P_d})} - \mathrm{AUC}_{(\tau,\mathrm{P_f})} \tag{16}$$

$$\mathrm{AUC_{SNPR}} = \frac{\mathrm{AUC}_{(\tau,\mathrm{P_d})}}{\mathrm{AUC}_{(\tau,\mathrm{P_f})}} \tag{17}$$

in which the larger $\mathrm{AUC_{OA}}$ $(\to 2)$ and $\mathrm{AUC_{SNPR}}$ $(\to \infty)$, the better the detector.

In order to obtain a global insight into the performance of different detection approaches, Friedman test [43] and Nemenyi post hoc test [44] are applied to aggregate the results obtained over all test HSIs, performed as follows.

Firstly, assign a rank to each method on each HSI according to the evaluation accuracy. Let $r_i^j$ denote the rank of the $i$th $(i = 1, 2, \cdots, M)$ method on the $j$th $(j = 1, 2, \cdots, N)$ HSI, where $M$ and $N$ are the number of comparing methods and HSIs, respectively. In this way, an average rank for each method can be obtained by $r_i = \sum_j r_i^j / N$.




(a) San Diego I

(b) San Diego II

(c) Urban

(d) Terrain

Fig. 7. 3D-ROC and three corresponding 2D-ROC curves of different comparing methods on four HSIs. (Left to right) 3D-ROC curve, 2D-ROC curve of $(\mathrm{P_f}, \mathrm{P_d})$, 2D-ROC curve of $(\tau, \mathrm{P_d})$, and 2D-ROC curve of $(\tau, \mathrm{P_f})$.

Secondly, carry out Friedman test to determine if these competitors have the same performance as follows:

$$\begin{aligned} \tau_{F^2} &= \frac{(N-1)\,\tau_{\chi^2}}{N(M-1)-\tau_{\chi^2}} \\ \tau_{\chi^2} &= \frac{12N}{M(M+1)}\left(\sum_{i=1}^{M} r_i^2 - \frac{M(M+1)^2}{4}\right) \end{aligned} \tag{18}$$

in which $\tau_{F^2}$ conforms to a $F$-distribution with $M-1$ and $(M-1)(N-1)$ degrees of freedom. If the $p$-value, obtained by subtracting the cumulative density function of F-distribution at $\tau_{F^2}$ from 1, is less than a significance threshold $\alpha$, the null hypothesis that there is no difference between the comparing methods would be rejected, and the Nemenyi test is then performed to evaluate the significance of performance difference between various detectors.

In Nemenyi test, if the average rank of two methods differs by more than a critical difference (CD) value, it is considered that these two methods are significantly different. The CD value can be calculated using the following equation:

$$CD = q_{\alpha,\infty,M}\sqrt{\frac{M(M+1)}{12N}} \tag{19}$$

where $q_{\alpha,\infty,M}$ denotes the Studentized range distribution.

*3) Comparison methods:* In this paper, there are totally four kinds of target detection methods applied for performance comparison, including two SBMbDs: CEM and OSP, one UBMbD called reweighted ACE (rACE) [45], two MLbD methods: STD[2] and CSCR[3], and one DLbD method: HTD-Net[3]. The code of our ULMMDL method has been made available on the Github website[4]. The target prior applied in CEM, OSP, and rACE is the same as that in ULMMDL. For

[2]https://github.com/ShoupingShan/Hyperspectral-image-target-detection-based-on-sparse-representation

[3]http://fdss.bit.edu.cn/yjdw/js/153191.htm

[4]https://github.com/syz88/HDAE

TABLE II
DETECTION ACCURACY AND COMPUTATIONAL COSTS (IN SECONDS) COMPARISON OF DIFFERENT COMPARING APPROACHES ON **SAN DIEGO I** HSI. BOLDFACE HIGHLIGHTS THE BEST RESULTS WHILE UNDERLINE THE SECOND.

| Method | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ | Time |
|---|---|---|---|---|---|---|
| **CEM** | 0.9631 | 0.5725 | 0.2636 | 1.2720 | 2.1719 | **0.0097** |
| **OSP** | **0.9944** | 0.6860 | 0.1967 | 1.4837 | 3.4875 | 1.7712 |
| **rACE** | 0.9844 | 0.4970 | **0.0062** | 1.4752 | **80.1613** | 1.0788 |
| **CSCR** | 0.9918 | 0.8270 | 0.6233 | 1.1955 | 1.3268 | 8.2777 |
| **STD** | 0.9735 | **0.9353** | 0.8254 | 1.0834 | 1.1331 | 209.1822 |
| **HTD-Net** | 0.9613 | 0.6031 | 0.2652 | 1.2992 | 2.2741 | 17.7267 |
| **ULMMDL** | 0.9941 | 0.8919 | 0.1309 | **1.7551** | 6.8136 | 0.1937 |

TABLE III
DETECTION ACCURACY AND COMPUTATIONAL COSTS (IN SECONDS) COMPARISON OF DIFFERENT COMPARING APPROACHES ON **SAN DIEGO II** HSI. BOLDFACE HIGHLIGHTS THE BEST RESULTS WHILE UNDERLINE THE SECOND.

| Method | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ | Time |
|---|---|---|---|---|---|---|
| **CEM** | 0.8368 | 0.3415 | 0.1800 | 0.9983 | 1.8972 | **0.0152** |
| **OSP** | 0.9709 | 0.5417 | 0.1646 | 1.3480 | 3.2910 | 1.4486 |
| **rACE** | 0.9335 | 0.2121 | **0.0025** | 1.1431 | **84.8400** | 1.8146 |
| **CSCR** | 0.9906 | 0.7364 | 0.5373 | 1.1897 | 1.3706 | 9.5041 |
| **STD** | 0.9492 | **0.8033** | 0.6804 | 1.0721 | 1.1806 | 208.2840 |
| **HTD-Net** | 0.8057 | 0.5337 | 0.2877 | 1.0517 | 1.8551 | 17.7632 |
| **ULMMDL** | **0.9978** | 0.5107 | 0.0844 | **1.4241** | 6.0509 | 0.1652 |

TABLE IV
DETECTION ACCURACY AND COMPUTATIONAL COSTS (IN SECONDS) COMPARISON OF DIFFERENT COMPARING APPROACHES ON **URBAN** HSI. BOLDFACE HIGHLIGHTS THE BEST RESULTS WHILE UNDERLINE THE SECOND.

| Method | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ | Time |
|---|---|---|---|---|---|---|
| **CEM** | 0.9032 | 0.4961 | 0.3056 | 1.0937 | 1.6234 | **0.0193** |
| **OSP** | 0.9474 | 0.5769 | 0.2418 | 1.2825 | 2.3859 | 2.6289 |
| **rACE** | 0.7386 | 0.0669 | **0.0064** | 0.7991 | 10.4531 | 0.9648 |
| **CSCR** | 0.9941 | 0.5991 | 0.4253 | 1.1679 | 1.4087 | 5.6194 |
| **STD** | 0.7306 | 0.1192 | 0.0751 | 0.7747 | 1.5872 | 16.8533 |
| **HTD-Net** | 0.9043 | 0.5423 | 0.2012 | 1.2454 | 2.6953 | 18.0360 |
| **ULMMDL** | **0.9961** | **0.7027** | 0.0618 | **1.6370** | **11.3706** | 3.5208 |

STD, CSCR, and HTD-Net, multiple target priors are selected as recommended in the literature.

*4) Implementation details and equipment:* The implementation of ULMMDL can be divided into three parts as illustrated in Section III. As for HDAE training, we follow a $50\%/50\%$ split criterion to randomly divide each HSI into training and testing sets, in which the target prior is added to the training set. Note that the training-test dataset splits are repeatedly performed for each DAE. The training set randomly corrupted by dropout would be reconstructed as clean samples after layer-by-layer optimization. As for testing, each pixel is directly sent forward to the trained model without dropout. There are several main parameters required to be set in advance. Fig. 5 shows the impacts of various parameter selection on detection performance. Take San Diego I as an example, for dropout probability $p$, it can be observed that HDAE performs better than regular HAE ($p = 0$) for a wide range of corruption level. For minibatch size $N_b$, the advantage of large $N_b$ in accuracy tends to decrease due to the sharp minima, but the time consumption also significantly decreases. For trade-off coefficient $\lambda$, there is no obvious performance difference between non-zero $\lambda$, indicating the importance of spectral shape loss. As the number of hidden neurons $N_h$ increases, the detection accuracy increases first and then stabilizes. The parameter selection for other datasets is done by using this coarse grid search method. In this paper, taking both accuracy and efficiency into account, we set $N_b$, learning rate, and $N_h$ as 128, 0.001, and 20 for all HSIs, $\lambda$ as 0.01 for San Diego I, San Diego II, and Urban, and 0.1 for Terrain. The number of oversegmented superpixels in S$^2$UE$^2$ $N_s$ is determined according to the size of the detected object. Generally, we assign large $N_s$ for small objects to make sure that each superpixel solely includes target or background. Specifically, for each dataset, $N_s$ is chosen from the set $\{50, 100, 300, 500, 750, 1000, 1500\}$. Fig. 5(e) shows the impacts of $N_s$ on the detection performance and computational cost of ULMMDL for four datasets. It can be observed that ULMMDL could perform effectively and efficiently with small $N_s$ for San Diego I and San Diego II with large size targets. Moreover, we found that when $N_s = 1500$, $\mathrm{AUC_{SNPR}}$ on San Diego I and San Diego II are much larger in comparison with $N_s = 50$. The possible reason is that large $N_s$ provides help to extract very small size background, and make suppression of this kind of background possible, i.e., smaller $\mathrm{AUC}_{(\tau,\mathrm{P_f})}$ and larger $\mathrm{AUC_{SNPR}}$. As for the datasets Urban and Terrain with smaller target size, ULMMDL achieves better detection accuracy with large $N_s$. Additionally, the computational cost increases dramatically as $N_s$ increases. Therefore, as a trade off between effectiveness and efficiency, we set $N_s$ as $50, 50, 1000, 1000$ for San Diego I, San Diego II, Urban and Terrain, respectively. Lastly, no parameters in TSP module are required to pre-define.

All experiments in this paper are carried out on a computer equipped with an Intel Xeon E5-1620 v4 processor and an NVIDIA GeForce GTX 1080Ti GPU, and the requisite software environments include MatlabR 2017a and Python 3.6 with Tensorflow library. In order to demonstrate the convergence of HDAE, Fig. 6 displays the loss curves of the proposed HDAE network during training and testing phases on four datasets, where the training loss is recorded every epoch and the testing loss is recorded every 10 epochs. It can be observed that during each iteration, the training loss drops drastically in the first few epochs, and then tends to a continuous small decrease due to the learning rate decay. Moreover, the training and testing loss curves remain close, indicating that there is no overfitting problem in HDAE.

### *B. Comparison with other methods*

Fig. 4 visualizes the detection maps of ULMMDL and other comparing algorithms. It is obvious that MLbD and DLbD methods perform much worse than others in background suppression, especially CSCR and STD. Furthermore,

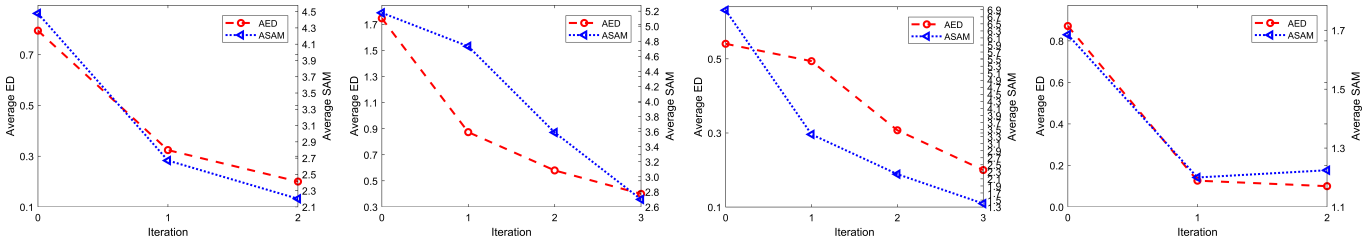



Fig. 8. Average ED and SAM distances between all target spectra and its mean spectrum as the layer of HDAE increases on four HSIs. (a) San Diego I HSI, (b) San Diego II HSI, (c) Urban HSI, (d) Terrain HSI.

TABLE V
DETECTION ACCURACY AND COMPUTATIONAL COSTS (IN SECONDS) COMPARISON OF DIFFERENT COMPARING APPROACHES ON **TERRAIN** HSI. BOLDFACE HIGHLIGHTS THE BEST RESULTS WHILE UNDERLINE THE SECOND.

| Method | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ | Time |
|---|---|---|---|---|---|---|
| **CEM** | **1.0000** | 0.6913 | 0.0614 | 1.6299 | 11.2590 | **0.0220** |
| **OSP** | **1.0000** | **0.9401** | 0.3172 | 1.6229 | 2.9637 | 1.4491 |
| **rACE** | **1.0000** | 0.6373 | **0.0030** | 1.6343 | **212.4333** | 0.2666 |
| **CSCR** | 0.9993 | 0.7179 | 0.5390 | 1.1782 | 1.3319 | 8.0420 |
| **STD** | 0.9935 | 0.8897 | 0.6436 | 1.2396 | 1.3824 | 436.6608 |
| **HTD-Net** | 0.9805 | 0.1539 | 0.0367 | 1.0977 | 4.1935 | 21.8910 |
| **ULMMDL** | **1.0000** | 0.9342 | 0.1489 | **1.7853** | 6.2740 | 2.4864 |

TABLE VI
AVERAGE RANK AND $p$-VALUE COMPARISON EVALUATED OVER THE FOUR HSIS USING FRIEDMAN TEST AND NEMENYI POST HOC TEST WITH $\alpha = 0.05$. THE SUPERSCRIPT NUMBERS INDICATE THE RANKING OF EACH METHOD AFTER NEMENYI TEST.

| Method | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ |
|---|---|---|---|---|---|
| **CEM** | $4.875^{2}$ | $5.500^{1,2}$ | $4.250^{2}$ | $5.000^{2}$ | $4.000^{1,2}$ |
| **OSP** | $2.375^{1,2}$ | $2.750^{1}$ | $4.000^{1,2}$ | $2.500^{1,2}$ | $3.750^{1,2}$ |
| **rACE** | $4.375^{1,2}$ | $6.750^{2}$ | $1.000^{1}$ | $3.750^{1,2}$ | $1.250^{1}$ |
| **CSCR** | $3.000^{1,2}$ | $2.750^{1}$ | $6.250^{2}$ | $4.750^{2}$ | $6.500^{2}$ |
| **STD** | $5.500^{2}$ | $2.750^{1}$ | $6.000^{2}$ | $6.000^{2}$ | $6.500^{2}$ |
| **HTD-Net** | $6.250^{2}$ | $5.000^{1,2}$ | $4.000^{1,2}$ | $5.000^{2}$ | $4.000^{1,2}$ |
| **ULMMDL** | $1.625^{1}$ | $2.500^{1}$ | $2.500^{1,2}$ | $1.000^{1}$ | $2.000^{1}$ |
| $p$-**value** | 3.55e-03 | 2.93e-03 | 2.58e-04 | 2.65e-03 | 7.85e-07 |

there exists extensive target pixels omitted in STD on Urban HSI. Although CEM and rACE could obtain really good performance in background suppression, they cannot detect all targets in every scene, i.e., inferior target detectability. As for ULMM-based subspace projection ULMMDL and OSP, we can observe that the distinction between target and background of our proposed ULMMDL method is much larger than that of OSP, especially on hyperspectral dataset San Diego II. In comparison with other methods, ULMMDL could output detection results much more similar to groundtruth.

Fig. 7 shows four kinds of ROC curves corresponding to the detection maps. It can be observed that 2D ROC curve of $(P_f, P_d)$ of ULMMDL is much closer to the upper left corner than that of other methods on San Diego II, Ubran, and Terrain, but performs slightly worse, i.e., lower detection probability than rACE and OSP when $P_f$ is less than $2 \times 10^{-2}$ on San Diego I, which could also be reflected in Fig. 4. However, ULMMDL obtains much more convincing target detection results than rACE and better background suppression performance than OSP on all testing HSIs, and also yields competitive performance compared to others. Lastly, there is no doubt that rACE provides the best results in background suppression, but it omits too many interested target pixels as shown in Fig. 4, while the proposed ULMMDL method shows a better capability to suppress background than most competitors under high detection reliance.

Obviously, ROC curves of different methods intersect with each other in Fig. 7. To deal with this issue, Tables II-V provide five AUC indicators of all comparing methods on four hyperspectral datasets for quantitative evaluation. ULMMDL always obtains the highest or the second $AUC_{(P_f,P_d)}$ on four HSIs, verifying its overall effectiveness on target detection. As for $AUC_{(\tau,P_d)}$ and $AUC_{(\tau,P_f)}$, ULMMDL still offers good competitiveness. Since it is hard to tell which method is better via any of $AUC_{(P_f,P_d)}$, $AUC_{(\tau,P_d)}$, and $AUC_{(\tau,P_f)}$, $AUC_{OA}$ and $AUC_{SNPR}$ are applied for an all-around evaluation. With respect to $AUC_{OA}$, it can be seen that ULMMDL performs superior to other comparing approaches, and brings an improvement of $27.14\%$, $7.61\%$, $35.45\%$, and $15.10\%$ on San Diego I, San Diego II, Urban, and Terrain, respectively. For $AUC_{SNPR}$, rACE outperforms all the other methods at the expense of certain loss of target preservation except on Urban HSI, which is consistent with the visualization maps and 2D ROC curves of $(\tau, P_f)$. However, the low detection probability of rACE can hardly meet the practical demands. Among the remaining competitors, our proposed ULMMDL detector could achieve higher accuracy in most cases. Moreover, on the basis of Friedman and Nemenyi test, Table VI provides a global insight into performance evaluation by aggregating the detection results over all datasets. It is clear that the null hypothesis under each evaluation index is rejected with $\alpha = 0.05$, indicating that all comparing methods are different. The rankings after the post hoc test are given by the superscripts. It can be observed that ULMMDL is definitely the first for $AUC_{(P_f,P_d)}$, $AUC_{(\tau,P_d)}$, $AUC_{OA}$ and $AUC_{SNPR}$. For $AUC_{(\tau,P_f)}$, rACE ranked first, followed by the proposed ULMMDL. But the extreme background suppression capability of rACE leads to serious target misses, while ULMMDL could achieve outstanding target identification with satisfying background suppression. The total experimental results and

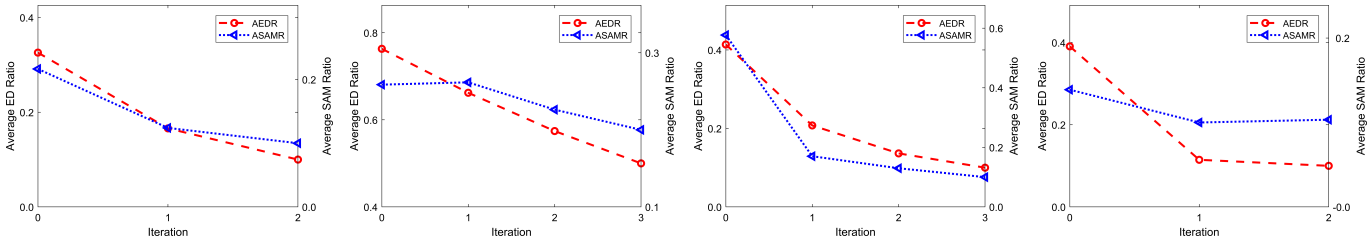



Fig. 9. Ratio of intraclass distance and interclass distance measured by AED and ASAM as the layer of HDAE increases on four HSIs. (a) San Diego I HSI, (b) San Diego II HSI, (c) Urban HSI, (d) Terrain HSI.

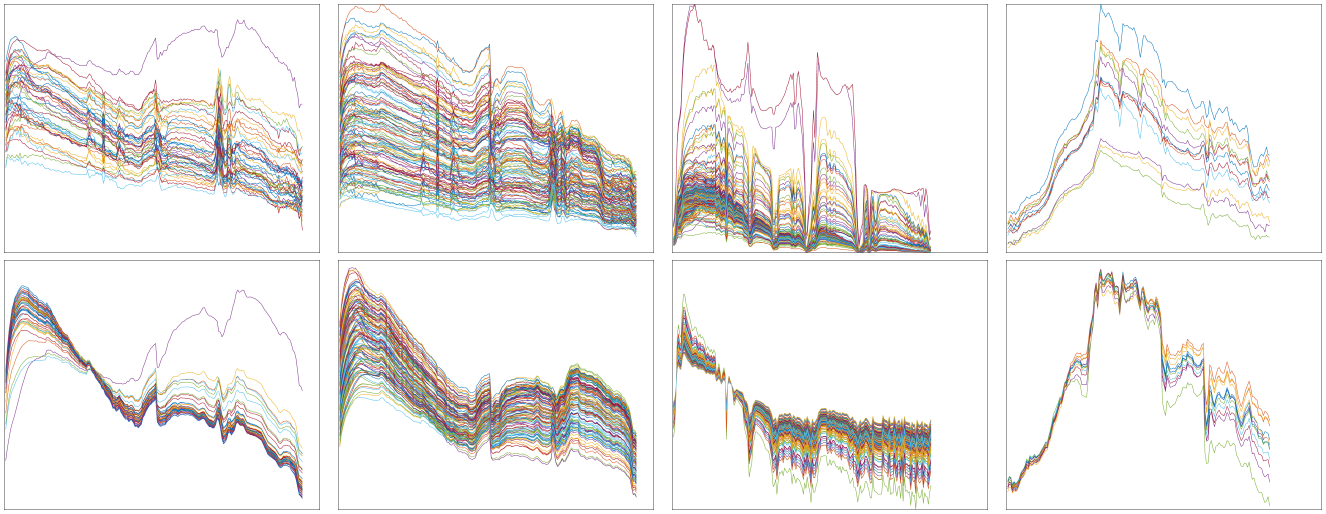



Fig. 10. Visualization of target spectra curves before (first row) and after (second row) training HDAE on four HSIs. (a) San Diego I HSI, (b) San Diego II HSI, (c) Urban HSI, (d) Terrain HSI.

TABLE VII
TRAINING TIME COMPARISON (IN SECONDS) OF DEEP-LEARNING BASED HTD-NET AND ULMMDL SPENT ON FOUR HSIS.

| Method | San Diego I | San Diego II | Urban | Terrain |
|---|---|---|---|---|
| **ULMMDL** | 62 | 103 | 136 | 113 |
| **HTD-Net** | 228 | 248 | 100 | 97 |
| | Moreover, it takes about **300** seconds for HTD-Net to construct target-target and target-background pixel-pairs. | | | |

aforementioned analysis verify the effectiveness of ULMMDL.

Furthermore, the computation cost of ULMMDL and of the compared approaches on four datasets are also provided in Tables II-V, where only testing time is contained for deep-learning-based HTD-Net and ULMMDL. Compared with SBMbD and UBMbD methods, MLbD and DLbD consume more time due to the large amount of parameters and the complex optimization. However, these algorithms could strongly improve detection robustness due to their independence to specific distributions. For HTD-Net and ULMMDL, Table VII additionally tabulates time required for neural network training. It should be noted that the training time of ULMMDL refers to time spent on all DAEs, while that of HTD-Net does not include time spent on complicated data preprocessing, such as target and background sample generation, and pixel-pairs construction. Therefore, ULMMDL is more efficient than HTD-Net in terms of both training and testing. Additionally, there is no need to retrain HDAE for the same hyperspectral scene. In this case, the total time spent on HDAE test, $S^2UE^2$, and TSP is dominated by $S^2UE^2$, also indicating the necessity to reduce the computational cost of $S^2UE^2$.

### C. Discussion

In this section, we further analyze two main components of our proposed ULMMDL method, i.e., HDAE and $S^2UE^2$ to help researchers better understand their effects on target detection.

*1) Performance analysis of HDAE:* As Section II-A illustrated, HDAE could provide stronger interference alleviation than single DAE. In order to verify this, Fig. 8 describes the average ED (AED) and average SAM (ASAM) distances between all target spectra and its mean spectrum, i.e., intraclass distance, as the layer of HDAE goes deeper. It can be observed that both AED and ASAM decrease when the iteration increases, which indicates the effectiveness of HDAE, mainly owing to the joint application of ED and SAM in loss function. Furthermore, we calculate AED and ASAM between all background spectra and the average target

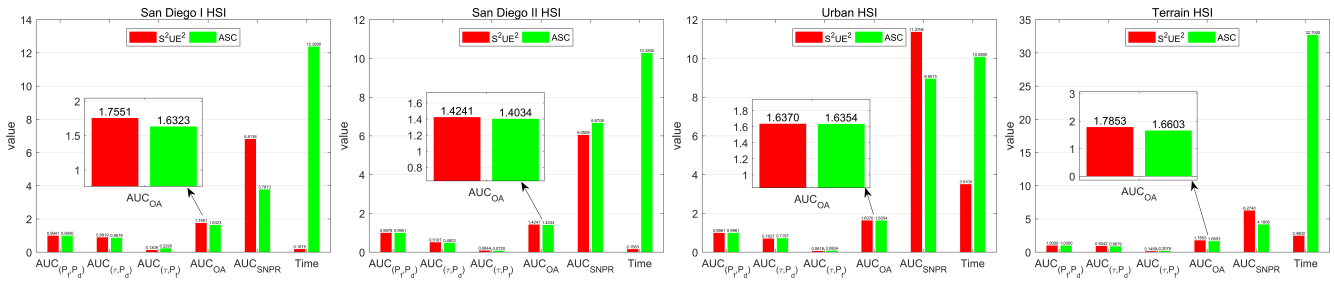



Fig. 11. Detection accuracy and computational cost comparison under different endmember extraction strategies on four HSIs. Note time consumption on training HDAE is not included. Best zoomed-in view.

TABLE VIII
EFFECTIVENESS ANALYSIS OF HDAE ON FOUR HSIS. BOLDFACE HIGHLIGHTS THE BEST RESULTS.

| Dataset | HDAE? | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ |
|---|---|---|---|---|---|---|
| San Diego I | ✓ | **0.9941** | **0.8919** | **0.1309** | **1.7551** | **6.8136** |
| | × | 0.9714 | 0.6074 | 0.2640 | 1.3148 | 2.3008 |
| San Diego II | ✓ | **0.9978** | **0.5107** | **0.0844** | **1.4241** | **6.0509** |
| | × | 0.9164 | 0.3178 | 0.0911 | 1.1431 | 3.4885 |
| Urban | ✓ | **0.9961** | **0.7027** | 0.0618 | **1.6370** | **11.3706** |
| | × | 0.9940 | 0.1669 | **0.0406** | 1.1203 | 4.1108 |
| Terrain | ✓ | **1.0000** | **0.9342** | 0.1489 | **1.7853** | 6.2740 |
| | × | **1.0000** | 0.6689 | **0.0820** | 1.5869 | **8.1573** |

spectrum as interclass distance, the ratio of intraclass distance and interclass distance measured by AED and ASAM under different iterations is given in Fig. 9. As expected, both ratios decreases with the iteration increases, indicating that HDAE could enhance intraclass similarity without deteriorating interclass dissimilarity. Fig. 10 shows the input and output target spectral curves of HDAE network. The visual inspection of target spectra further validates HDAE's positive impacts on spectral consistency enhancement. For San Diego I HSI, there is obviously one curve very different from others. By locating its position, we found that this pixel is exactly on the edge and severely mixed with the neighboring background pixels. Therefore, it is quite difficult to make this pixel similar to other target spectra due to its low fractional abundance of target endmember.

Moreover, we compare the detection results of ULMMDL (with HDAE) and ULMM (without HDAE), where ULMM directly sends the initial severely interferential HSI to $S^2UE^2$ module as input. Table VIII tabulates five kinds of AUC indicators on four HSIs. From the perspectives of target detectability, ULMMDL yields a better performance than ULMM on all datasets. For background suppressibility, it is slightly inferior to ULMM on data Urban and Terrain. This is mainly due to HDAE improperly reconstructs the hardest background samples (a very small proportion in general) as target to alleviate spectral variation. As a whole, ULMMDL could achieve higher overall accuracy in most cases. The experimental results demonstrate that spectral variation reduction through HDAE plays a critical role in HTD.

*2) Performance analysis of noise type:* In addition to dropout used in our method for data corruption, we further tested Gaussian noise with various standard deviations. Fig. 12 shows the detection accuracy obtained with different noise levels. Similar to experimental results obtained by dropout in Fig. 5(a), ULMMDL with low-level Gaussian noise also reconstructs a new HSI more appropriate for target detection than that with zero noise or high-level noise. However, there is a significant accuracy drop for ULMMDL with Gaussian noise when the standard deviation is larger than 0.08. In order to understand the reason behind this phenomenon, Fig. 13 displays target spectra corrupted by dropout and Gaussian noise for comparison. Obviously, Gaussian noise corrupts every band, causing serious deterioration of spectral correlation. On the other hand, dropout only corrupts a subset of input spectra, and HDAE could restore the corrupted parts with spectral correlation information. Therefore, for the task of denoising, we suggest to choose dropout for data corruption.

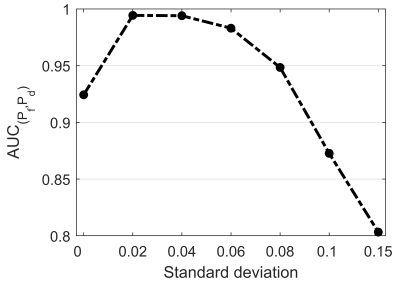



Fig. 12. Detection accuracy obtained on San Diego I by ULMMDL using Gaussian noise for data corruption in HDAE.

*3) Performance analysis of $S^2UE^2$:* In order to validate the effectiveness and efficiency of $S^2UE^2$, we replace $S^2UE^2$ with ASC for endmember extraction. Fig. 11 gives the detection accuracy and computational cost comparison on four HSIs. Note that the runtime does not include that on HDAE. It can be easily concluded that $S^2UE^2$ consumes much less time, but obtains better detection accuracy. This comparison verifies that endmember candidates selection via spatial superpixel oversegmentation could significantly reduce computational burden for subsequent ASC, and spatial-spectral information integration also provides reliable endmembers for effective detection.

*4) Performance analysis of TSP:* As introduced in Section II-A, our proposed TSP detector is developed without consideration of the non-negative and sum-to-one constraints for simplicity. To verify the effectiveness and efficiency of

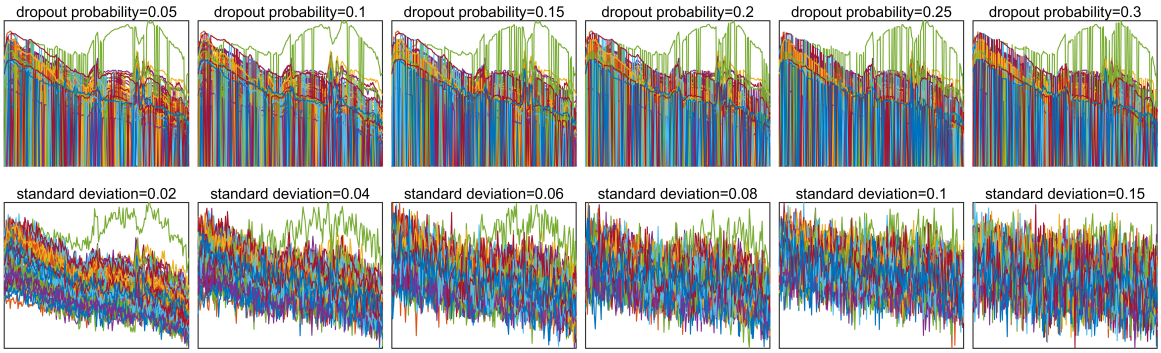



Fig. 13. Visualization of target spectra corrupted by dropout and Gaussian noise on San Diego I.

TABLE IX
ACCURACY AND TIME CONSUMPTION OF TSP AND MFCLS ON FOUR HSIS. NOTE ONLY DETECTION TIME IS INCLUDED.

| Dataset | eq(2) | $AUC_{(P_f,P_d)}$ | $AUC_{(\tau,P_d)}$ | $AUC_{(\tau,P_f)}$ | $AUC_{OA}$ | $AUC_{SNPR}$ | Time |
|---|---|---|---|---|---|---|---|
| San Diego I | × | 0.9941 | 0.8919 | 0.1309 | 1.7551 | 6.8136 | 0.0781 |
| | ✓ | 0.9775 | 0.9831 | 0.0803 | 1.8803 | 12.2428 | 9.6250 |
| San Diego II | × | 0.9978 | 0.5107 | 0.0844 | 1.4241 | 6.0509 | 0.0625 |
| | ✓ | 0.9879 | 0.6265 | 0.2649 | 1.3495 | 2.3650 | 8.2656 |
| Urban | × | 0.9961 | 0.7027 | 0.0618 | 1.6370 | 11.3706 | 0.0469 |
| | ✓ | 0.9838 | 0.9966 | 0.3404 | 1.6400 | 2.9277 | 43.0469 |
| Terrain | × | 1.0000 | 0.9342 | 0.1489 | 1.7853 | 6.2740 | 0.04125 |
| | ✓ | 0.9991 | 0.9347 | 0.0782 | 1.8556 | 11.9527 | 46.6250 |

ULMM-based TSP, we further construct a multiple target detector based on constrained LMM and combine all detection maps by a winner-take-all operation the same as that in TSP. According to equation (14), the output of each TSP detector is the unconstrained target abundance contained in the test pixel. In order to obtain target abundance satisfying constraints in equation (2), background abundance estimation are also required, and the problem can be described as follows:

$$\begin{aligned} &\min \left(\hat{\mathbf{B}}\boldsymbol{\alpha}_{b_k} + \hat{\mathbf{d}}_j\alpha_{d_k} - \hat{\mathbf{x}}_k\right)^T \left(\hat{\mathbf{B}}\boldsymbol{\alpha}_{b_k} + \hat{\mathbf{d}}_j\alpha_{d_k} - \hat{\mathbf{x}}_k\right) \\ &s.t. \quad \boldsymbol{\alpha}_{b_k} \geq 0, \alpha_{d_k} \geq 0, \sum_{i=1}^{n_3} \alpha_{b_k}^i + \alpha_{d_k} = 1 \end{aligned} \tag{20}$$

which is known as FCLS. Therefore, the constrained multiple target detector, denoted as MFCLS, is represented as

$$y_k = \max\left\{\alpha_{d_k}^j\right\}_{j=1}^{N_d} \tag{21}$$

where $y_k$ is the target abundance in the $k$th pixel.

Table IX tabulates the accuracy and time consumption of TSP and MFCLS on four HSIs. It can be observed that the constrained MFCLS costs much more time than TSP due to its iterative optimization process. Furthermore, MFCLS did not bring significant improvements to target detection. Therefore, it is reasonable to use the less stringent ULMM to deal with detection problems as an excellent trade off between accuracy and complexity.

## V. CONCLUSION

In this paper, we propose an effective target detection method based on unconstrained linear mixture model (ULMM) and hierarchical denoising autoencoder (HDAE) in hyperspectral imagery, which applies deep learning framework to solve existed problems in classical statistical analysis. To alleviate unavoidable interference contained in HSIs, LMM-based detectors generally depends on specific data distributions, which might restrict detection performance in real applications. This paper proposes an HDAE network to reduce interference in advance so that there is no need making assumptions on data distribution in the subsequent two-step subspace projection (TSP) detection. Moreover, a spatial-spectral unified endmember extraction ($S^2UE^2$) strategy is developed to effectively and efficiently obtain reliable endmembers necessary for TSP, which could simultaneously preserve desired target and suppress undesired background. Experimental results and detailed analysis on four real-world HSIs demonstrate the effectiveness of our proposed hyperspectral target detector in comparison with other state-of-the-art methods.

## REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, June 2013.

[2] Y. Chen, X. Chen, J. Zhou, Y. Ji, and W. Shen, "Camouflage target detection via hyperspectral imaging plus information divergence measurement," in *International Conference on Optoelectronics and Microelectronics Technology and Application*, vol. 10244, International Society for Optics and Photonics. SPIE, 2017, pp. 88 – 92. [Online]. Available: https://doi.org/10.1117/12.2257938

[3] D. Manolakis, E. Truslow, M. Pieper, T. Cooley, and M. Brueggeman, "Detection algorithms in hyperspectral imaging systems: An overview of practical algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 24–33, Jan 2014.

[4] J. Theiler, A. Ziemann, S. Matteoli, and M. Diani, "Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 8–30, June 2019.

[5] S. Chakraborty, J. Phukan, M. Roy, and B. B. Chaudhuri, "Handling the class imbalance in land-cover classification using bagging-based semisupervised neural approach," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 9, pp. 1493–1497, Sep. 2020.

[6] J. Broadwater and R. Chellappa, "Hybrid detectors for subpixel targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1891–1903, Nov 2007.

[7] C.-I. Chang, *Hyperspectral data processing: algorithm design and analysis*. John Wiley & Sons, Ltd, 2013, ch. 2, pp. 33–62. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118269787.ch2

[8] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 779–785, July 1994.

[9] C. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer US, 2013.

[10] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 1–16, Jan 2001.

[11] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A cfar adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 1, pp. 208–216, Jan 1992.

[12] S. Kraut and L. L. Scharf, "The cfar adaptive subspace detector is a scale-invariant glrt," *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.

[13] L. Zhang, B. Du, and Y. Zhong, "Hybrid detectors based on selective endmembers," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 6, pp. 2633–2646, June 2010.

[14] H. Kwon and N. M. Nasrabadi, "Kernel orthogonal subspace projection for hyperspectral signal classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 12, pp. 2952–2962, Dec 2005.

[15] H. Kwon and N. M. Nasrabadi, "Kernel adaptive subspace detector for hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 2, pp. 271–275, April 2006.

[16] T. Wang, B. Du, and L. Zhang, "A kernel-based target-constrained interference-minimized filter for hyperspectral sub-pixel target detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 2, pp. 626–637, April 2013.

[17] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Sparse representation for target detection in hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 3, pp. 629–640, June 2011.

[18] Y. Zhang, B. Du, and L. Zhang, "A sparse representation-based binary hypothesis model for target detection in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1346–1354, March 2015.

[19] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics-based detector for hyperspectral images," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5345–5357, Nov 2016.

[20] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognition*, vol. 48, no. 12, pp. 3904–3916, 2015.

[21] J. Du and Z. Li, "A hyperspectral target detection framework with subtraction pixel pair features," *IEEE Access*, vol. 6, pp. 45 562–45 577, 2018.

[22] 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 Sensing*, vol. 12, no. 9, 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/9/1489

[23] D. Zhu, B. Du, and L. Zhang, "Two-stream convolutional networks for hyperspectral target detection," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[24] Y. Shi, J. Li, Y. Li, and Q. Du, "Sensor-independent hyperspectral target detection with semisupervised domain adaptive few-shot learning," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2020.

[25] Y. Shi, J. Lei, Y. Yin, K. Cao, Y. Li, and C. Chang, "Discriminative feature learning with distance constrained stacked sparse autoencoder for hyperspectral target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 9, pp. 1462–1466, Sep. 2019.

[26] Y. Shi, J. Li, Y. Yin, B. Xi, and Y. Li, "Hyperspectral target detection with macro-micro feature extracted by 3-d residual autoencoder," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 4907–4919, Dec 2019.

[27] Y. Shi, J. Li, Y. Zheng, B. Xi, and Y. Li, "Hyperspectral target detection with roi feature transformation and multiscale spectral attention," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–1, 2020.

[28] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392–1409, July 2001.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, p. 1929–1958, Jan. 2014.

[30] G. A. Licciardi and J. Chanussot, "Nonlinear pca for visible and thermal hyperspectral images quality enhancement," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 6, pp. 1228–1231, 2015.

[31] G. Licciardi and J. Chanussot, "Spectral transformation based on nonlinear principal component analysis for dimensionality reduction of hyperspectral images," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 375–390, 2018.

[32] G. Licciardi, G. Vivone, M. Mura, R. Restaino, and J. Chanussot, "Multi-resolution analysis techniques and nonlinear pca for hybrid pansharpening applications," *Multidimensional Systems and Signal Processing*, vol. 27, no. 4, pp. 807–830, Oct 2016. [Online]. Available: https://doi.org/10.1007/s11045-015-0359-y

[33] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of aviris data," in *Summaries of the Fifth Annual JPL Ai_ne Earth Science Workshop ry 23-26, 1995*, 1995, p. 23.

[34] H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1232–1249, 2003.

[35] D. R. Thompson, L. Mandrake, M. S. Gilmore, and R. Castaño, "Superpixel endmember detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4023–4033, 2010.

[36] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial preprocessing for hyperspectral unmixing," *Remote Sensing of Environment*, vol. 204, pp. 333–346, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425717304832

[37] X. Shen, W. Bao, K. Qu, and H. Liang, "Superpixel-guided preprocessing algorithm for accelerating hyperspectral endmember extraction based on spatial–spectral analysis," *Journal of Applied Remote Sensing*, vol. 15, no. 2, pp. 1 – 17, 2021. [Online]. Available: https://doi.org/10.1117/1.JRS.15.026514

[38] M. Q. Alkhatib and M. Velez-Reyes, "Improved spatial-spectral superpixel hyperspectral unmixing," *Remote Sensing*, vol. 11, no. 20, 2019. [Online]. Available: https://www.mdpi.com/2072-4292/11/20/2374

[39] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *CVPR 2011*, June 2011, pp. 2097–2104.

[40] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5600–5611, Oct 2017.

[41] A. W. Bitar, L. Cheong, and J. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5239–5251, Aug 2019.

[42] C. I. Chang, "An effective evaluation tool for hyperspectral target detection: 3d receiver operating characteristic curve analysis," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–23, 2020.

[43] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.

[44] P. B. Nemenyi, *Distribution-free multiple comparisons.* Princeton University, 1963.

[45] T. Wang, B. Du, and L. Zhang, "An automatic robust iteratively reweighted unstructured detector for hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 6, pp. 2367–2382, June 2014.


**Yunsong Li** received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, China, in 1999 and 2002, respectively.

He joined the school of telecommunications Engineering, Xidian University in 1999 where he is currently a Professor. Prof. Li is the director of the image coding and processing center at the State Key Laboratory of Integrated Service Networks. His research interests focus on image and video processing, hyperspectral image processing and high-performance computing.

**Yanzi Shi** received the B.S. degree from the School of Telecommunications Engineering, Xidian University, China, in 2017. She is currently working toward the Ph.D. degree in the State Key Laboratory of Integrated Service Networks of Xidian University, and a visiting Ph.D. student with the Telecommunications & Remote Sensing Laboratory, University of Pavia, Italy from 2021 to 2022, which is supported by the China Scholarship Council.

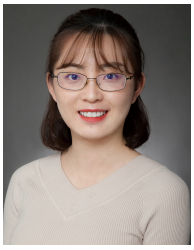
Her research interests include hyperspectral target detection, feature extraction, deep learning, and urban remote sensing.

**Keyan Wang** received the M.S. and Ph.D. degree in information and telecommunication engineering from Xidian University, Xi'an, China, in 2005 and 2008, respectively. She worked as a visiting scholar at the Department of Electrical and Computer Engineering of McMaster University, Hamilton, Canada, from September 2014 to September 2015. Currently, she is an Associate Professor at the school of Telecommunications Engineering, Xidian University, Xi'an, China.

Her current research interests focus on image/video compression and processing, computer vision, and deep learning.

**Bobo Xi** received the B.S. degree from the School of Telecommunications Engineering, Xidian University, China, in 2017. He is currently pursuing the Ph.D. degree in the State Key Laboratory of Integrated Service Networks of Xidian University.

His research interests include hyperspectral image process, machine learning and deep learning.



**Jiaojiao Li** received the B.E. degree in computer science and technology, M. S. degree in software engineering and Ph.D. degree in communication and information systems from Xidian University in 2009, 2012 and 2016, respectively. She was an exchange Ph.D. Student of Mississippi State University.

She is currently a Postdoctoral Researcher and a lecturer of Xidian University. Her research interests include hyperspectral remote sensing image analysis and processing, pattern recognition and data compression.

**Paolo Gamba** 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 in 1993, respectively.

He is a Professor with The University of Pavia, where he leads the Telecommunications and Remote Sensing Laboratory. He has been invited to give keynote lectures and tutorials in several occasions about urban remote sensing, data fusion, EO data for physical exposure and risk management. He has authored or coauthored more than 140 articles in international peer- review journals and presented nearly 300 research works in workshops and conferences.

Dr. Gamba served as an Editor-in-Chief for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS from 2009 to 2013, and as the Chair of the Data Fusion Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) from October 2005 to May 2009. He has been elected in the GRSS AdCom, since 2014, and he is the GRSS President. He has been the Organizer and the 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 a Technical Co-Chair for the 2010 and 2015 IGARSS Conferences, in Honolulu (Hawaii), and Milan (Italy), respectively.