UnDAT: Double-Aware Transformer for Hyperspectral Unmixing

Yuexin Duan, Xia Xu, Tao Li, Bin Pan, *Member, IEEE*, and Zhenwei Shi, *Senior Member, IEEE*

***Abstract*— Deep-learning-based methods have attracted increasing attention on hyperspectral unmixing, where the transformer models have shown promising performance. However, recently proposed deep-learning-based hyperspectral unmixing methods usually tend to directly apply visual models, while ignoring the characteristics of hyperspectral imagery. In this article, we propose a novel double-aware transformer for hyperspectral Unmixing (UnDAT), which aims at simultaneously exploiting the region homogeneity and spectral correlation of hyperspectral imagery. One of the major assumptions of UnDAT is that hyperspectral remote-sensing images involve many homogeneous regions. Pixels inside a homogeneous region usually present similar spectral features, and the edge pixels are just the reverse. Another observation is that the pixel spectra are continuous and correlated. Based on the above assumption and observation, we construct the UnDAT by developing two modules: Score-based homogeneous-aware (SHA) module and the spectral group-aware (SGA) module. In the SHA module, a feature map rearrangement (FMR) approach is proposed to split the shallow feature maps from a linear encoder into an ordered homogeneous map (HomoMap) and an edge map and develop a homogenous region-aware strategy for deep feature representation. In the SGA module, the dependency among neighboring bands is described by dividing the hyperspectral image into multiple spectral groups and calculating the spectral similarity among bands within each group. Experiments on both real and synthetic datasets indicate the effectiveness of our model. We will publish the code of our approach if the article has the honor to be accepted.**

***Index Terms*— Deep learning, homogeneous, hyperspectral unmixing, transformer network.**

I. Introduction

A HYPERSPECTRAL image can be regarded as a 3-D cube, which is constructed by a stack of images obtained by imaging a given scene in respective bands [1]. Owing to the extensive spectral information, hyperspectral images, offering a huge potential to distinguish materials on the ground, are frequently used in target detection and tracking, image recognition, remote sensing, medical imaging, and other domains.

However, due to the low spatial resolution of the sensor, multiple pixels in a hyperspectral image contain several materials, which leads to inaccurate recognition of ground objects, bringing big challenges for high-level tasks. Many hyperspectral unmixing methods have been proposed to address the mixed pixels problem by determining the spectrally pure components (*endmembers*) and estimating their corresponding percentage (*abundances*) simultaneously [2], [3], [4], [5], [6], [7].

The linear mixing model (LMM) is widely used in unmixing methods due to its simplicity in comparison to more complex non-LMM [8], [9]. This model assumes that the observed image is a linear combination of end members and their corresponding abundances. In recent years, there have been a lot of methods devoted to hyperspectral unmixing under the assumption of LMM, which can be categorized into geometry-based, statistical-based, and sparse regression-based [10]. Geometrical-based methods are based on the assumption that the observed spectra vector lies within a simplex, where the vertices correspond to the endmembers. These methods use the geometry of the simplex to identify the constituent endmembers and their corresponding abundances [11], [12], [13], [14]. Statistical-based methods for hyperspectral unmixing typically rely on Bayesian inference to estimate the posterior distribution over endmember abundances given the observed spectral data. These methods often incorporate prior knowledge to improve the accuracy and robustness of the unmixing results [15], [16], [17]. Sparse regression methods select endmember candidates from a spectral library to model each pixel in the hyperspectral scene. These methods try to find the optimal endmembers to minimize the difference between the observed and modeled spectra [18], [19], [20], [21], [22], [23].

Deep-learning-based approaches for hyperspectral unmixing have undergone extensive research in recent years [24], [25], [26], [27], [28], [29], which are broadly categorized into two types: deep convolutional neural network (CNN) frameworks and deep autoencoder (AE) frameworks [30]. CNN-based methods involve stacking layers of linear convolution filters and nonlinear logistic functions [31], [32], [33]. Tao et al.

Manuscript received 23 May 2023; revised 6 August 2023; accepted 14 August 2023. Date of current version 11 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2022ZD0160401 and Grant 2022YFA1003803; in part by the National Natural Science Foundation of China under Grant 62001251, Grant 62125102, Grant 62001252, and Grant 62272248; and in part by the Beijing-Tianjin-Hebei Basic Research Cooperation Project under Grant F2021203109. *(Yuexin Duan and Xia Xu are co-first authors.) (Corresponding author: Tao Li.)*

Yuexin Duan, Xia Xu, and Tao Li are with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: duanyuexin@mail.nankai.edu.cn; xuxia@nankai.edu.cn; litao@nankai.edu.cn).

Bin Pan is with the School of Statistics and Data Science, KLMDASR, LEBPS, and LPMC, Nankai University, Tianjin 300071, China (e-mail: panbin@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/TGRS.2023.3310155

proposed a method that combines 3-D and 2-D convolutions to effectively extract features by leveraging CNNs' feature extraction ability [32]. Gao et al. [33] suggested an efficient method that enhances unmixing performance by learning two cascaded AEs in a cycle-consistent manner. AE-based unmixing methods use fully connected and activation layers, with the decoder representing endmembers and the hidden layer outputs providing the corresponding abundances [34], [35], [36], [37], [38], [39]. Jin et al. [40], a two-stream model is proposed, where the first stream maps endmember bundles to corresponding abundances and the second stream employs an untied-weighted AE to reduce reconstruction errors. The encoder output of paper [41] is analyzed by two modules to explore local homogeneity and global self-similarity in hyperspectral imagery. In [42], the first two stages are utilized to increase the receptive field for learning multiscale information, while the final stage is employed to preserve local details. Hua et al. [43] proposed an adaptive regularization term to smooth the center pixel with its surrounding pixels, exploiting the spatial correlation of HSIs in AE architecture.

Recently, motivated by the impressive ability of transformers [44] to capture long-range dependencies in images, Gh et al. presents an architecture that combinates an AE and a transformer to unmix the HSIs [45], which has shown promising performance. However, the method may overlook the band properties and treat all patches equally, neglecting spatial correlations within homogeneous regions, which fails to fully utilize the spectral and spatial characteristics of hyperspectral imagery.

To overcome the problem of underutilizing spectral similarity and avoiding the equal treatment of all patches, we propose a novel double-aware transformer for hyperspectral unmixing (UnDAT), which aims at simultaneously exploiting the region homogeneity and spectral correlation of hyperspectral imagery. The UnDAT is constructed by developing two modules: the score-based homogeneous-aware (SHA) module and the spectral group aware (SGA) module.

To exploit the spatial correlation among homogeneous regions, UnDAT incorporates an SHA module to split shallow feature maps from a linear encoder into an ordered homogeneous map (HomoMap) and an edge map, and a homogenous region-aware strategy is further developed to improve deep feature representation.

To comprehensively exploit the rich spectral information of hyperspectral imagery, the SGA module describes the dependency among adjacent bands by dividing the hyperspectral imagery into multiple spectral groups and calculating the spectral similarity among bands within each group.

We list our contributions as follows.

1) We propose a hyperspectral unmixing model based on the transformer, which aims at fully exploiting the rich spectral information and strong spatial correlation of HSIs.
2) In the proposed UnDAT, we design an SHA module to effectively leverage the strong correlation within homogeneous regions and enhance the accuracy of pixel classification at the edges of different materials.
3) We also design an SGA module to reduce the data-processing burden of downstream tasks and leverage the similarity among neighboring bands.

The article is structured as follows: Section II explains the proposed methodology, including the SHA module and the SGA module. Section III presents the experimental analysis on three actual and one simulated dataset, and Section IV summarizes the conclusions drawn from the study.

II. Methodology

This section presents a detailed exposition of the methodology employed in our research. First, we introduce the commonly used LMM. Subsequently, we introduce a novel framework for hyperspectral unmixing named UnDAT. We then provide a detailed exposition of the SHA module. Finally, we offer a comprehensive explanation of the SGA module.

A. Linear Unmixing Model

The LMM has been widely used in the past few decades to address the hyperspectral unmixing problem. This model assumes that each incident light packet interacts with only one material and neglects any interactions between materials. Based on this linear assumption, the unmixing model can be represented as follows:

$$\mathbf{Y} = MZ + E \tag{1}$$

where $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_L] \in \mathbb{R}^{L\times N}$ represent the original hyperspectral image with L bands and N pixels. The endmember matrix $M \in \mathbb{R}^{L\times P}$ contains P endmembers, and $Z \in \mathbb{R}^{P\times N}$ denotes the corresponding abundance map. Additive noise is represented by $E \in \mathbb{R}^{L\times N}$. To ensure physically meaningful abundance values, two conditions must be satisfied

$$\begin{aligned} \boldsymbol{Z} &\geq \mathbf{0} \\ \mathbf{1}_P^T \boldsymbol{Z} &= \mathbf{1}_N^T. \end{aligned} \tag{2}$$

B. Overall Architecture

To exploit the physical properties of hyperspectral images in both the spatial and spectral domains, namely the spectral similarity of adjacent spectra and the spatial correlation between homogeneous regions, a novel framework UnDAT is presented in Fig. 1. The UnDAT exploits the SGA module to divide the hyperspectral image into multiple spectral groups and calculate the spectral similarity among bands within each group. The multiple spectral groups are then aggregated into a feature map with the same scale as the input image. This feature map is passed through a multilayer linear encoder to extract the shallow features of the hyperspectral image, which are further analyzed by the SHA module to extract deep feature maps. The shallow and deep feature maps are combined using a residual connection and passed through a linear decoder to reconstruct the hyperspectral image. Our method achieves good unmixing results with a simple loss function, without requiring a well-designed regularization term. To handle the spectrum variability of reflection scaling, we select spectral

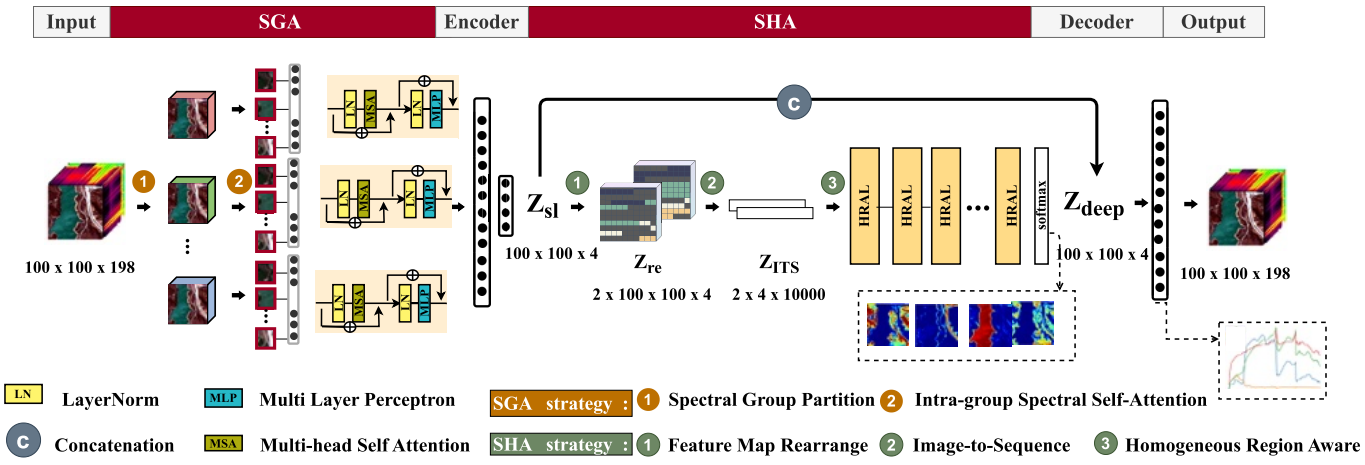


Fig. 1. Overall architecture of UnDAT. Note for ease of understanding, we have annotated the dimensions of each feature map in this figure and set the patch size to 1×1 within the ITS strategy.

angle distance (SAD) as the reconstruction loss function, defined as

$$L_{\text{SAD}} = \frac{1}{n}\sum_{i=1}^{n} \arccos\left(\frac{\hat{y}_i^T y_i}{\|\hat{y}_i\|_2 \|y_i\|_2}\right) \tag{3}$$

where n is the number of pixels in the hyperspectral image, and $\hat{y}_i$ and y_i are the reconstructed and original spectra of pixel i, respectively.

C. Score-Based Homogeneous-Aware

The unmixing method based on vision transformers [45] considers all pixels equally and calculates the correlations between each patch and all other patches in the image. However, hyperspectral images exhibit both homogeneous and heterogeneous regions, whereas the pixels in heterogeneous regions are typically uncorrelated. Consequently, calculating pixel correlations within these regions becomes redundant. To emphasize the correlations among homogeneous pixels in hyperspectral images, we introduce the SHA module. The SHA module incorporates three strategies: the feature map rearrangement (FMR) strategy, the image-to-sequence (ITS) strategy, and the homogeneous region-aware (HRA) strategy.

1) FMR Strategy: The FMR strategy shown in Fig. 2 adopts two approaches to exploit the spatial correlation among homogeneous regions. First, a superpixel segmentation algorithm (the Simple Linear Iterative Clustering algorithm is selected here) is utilized to divide the hyperspectral image into P superpixel blocks, where P represents the estimated number of endmembers obtained through the Hysime algorithm. Second, a scoring-based method is employed on the output of the shallow feature extractor $Z_{\text{sl}} \in \mathbb{R}^{1\times H\times W\times P}$ to rearrange the pixels in each superpixel block based on their Euclidean distances to the centroid. The Euclidean distance between a pixel p_i and its superpixel block centroid c is calculated, represented as (4), and pixels with shorter distances to the centroid are deemed more likely to belong to the same category

$$d(p_i, c) = \sqrt{(p_i - c)^2}. \tag{4}$$

Then the FMR strategy generates the (HomoMap for convenience) and the Edge Map from the sorted feature map by masking off the first and last α of the pixels in each sorted superpixel block and selecting only $(1-2\alpha)$ of the pixels to form the HomoMap, which can be represented as

$$\text{HomoMap}(p_i) = \begin{cases} p_i, & \text{if } i \in \left(\alpha N_{S_i}, (1-\alpha)N_{S_i}\right) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\text{HomoMap}(p_i)$ represents the pixel values in the HomoMap for the pixel p_i in the original hyperspectral image.

Concerning α selection, too low a value, such as 1%, engenders a HomoMap with 98% of superpixel block pixels, inevitably including edge pixels with potential negligible correlation to the homogeneous region. Conversely, a high α, like 40%, indicates merely 20% homogeneous pixels per block, possibly curtailing core pixel representation, thereby risking the accuracy of generated maps.

It is essential to note that the choice of α is not absolute and should be adapted according to specific application scenarios and datasets. Optimal α values may vary depending on the task at hand and the characteristics of the image data. In this work, the selection of 20% is an empirically justified and average choice for used datasets.

The edge map is created by masking the middle $(1-2\alpha)$ of pixels in each pixel and selecting only the first and last α of pixels. Both the HomoMap and Edge Map are feature maps with the same dimensions as the original image. These operations can be represented as

$$\text{EdgeMap}(p_i) = \begin{cases} p_i, & \text{if } i \in \left[0, \alpha N_{S_i}\right] \text{ or } \left[(1-\alpha)N_{S_i}, N_{S_i}\right] \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where $\text{EdgeMap}(p_i)$ denotes the pixel values in the edge map. And the whole procedure of FMR can be represented as follows:

$$Z_{\text{re}} = \text{FMR}(Z_{\text{sl}}) \tag{7}$$

where $Z_{\text{re}} \in \mathbb{R}^{2\times H\times W\times P}$ denotes the sorted feature maps, containing a HomoMap and an EdgeMap.

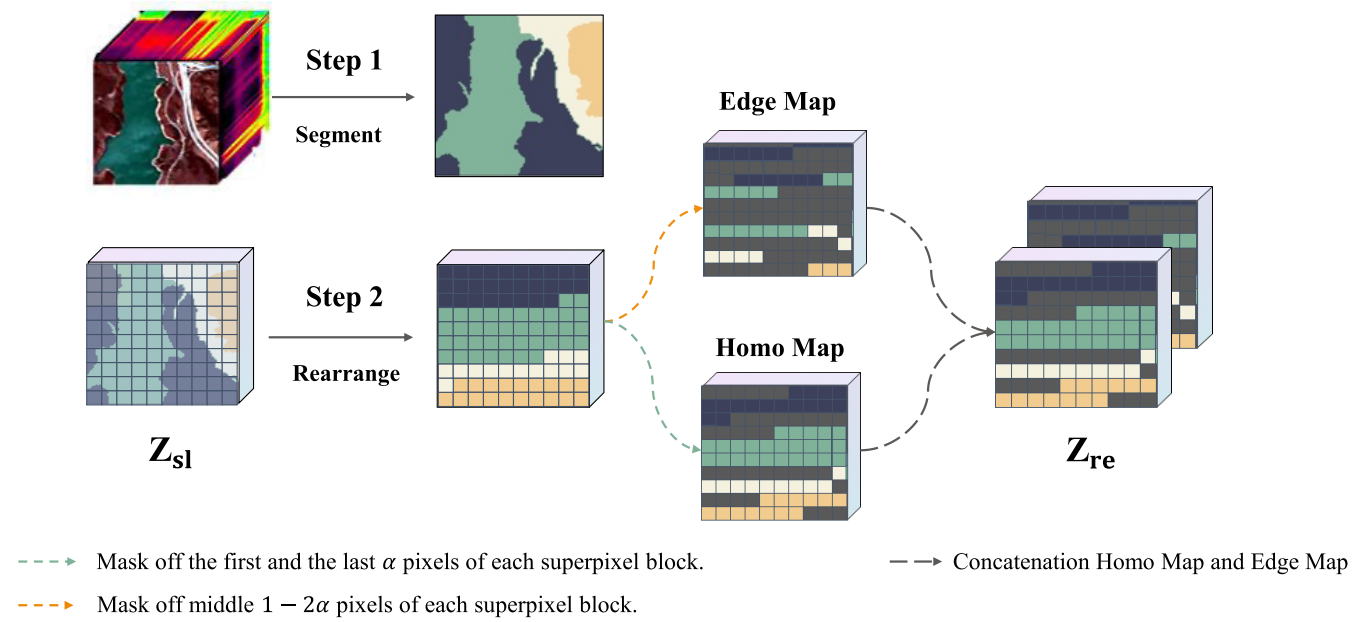


Fig. 2. Description of FMR.

2) ITS Strategy: Similar to the vision transformer, the ITS strategy transforms the sorted feature map Z_{re} into a sequence for subsequent processing. Specifically, the ITS strategy divides the feature map into nonoverlapping patches of a fixed size (e.g., 1×1) and projects each patch to a certain-dimensional C_d feature space using a convolutional layer. The resulting feature vectors are then concatenated to form a sequence Z_{ITS}.

3) HRA Strategy: As shown in Fig. 3, the HRA strategy is a multilayered network, and each layer consists of two phases: the homogeneous region correlation calculation (HRCC) phase and the cross-window (CW) phase. Each phase comprises a window partition stage (WP), a window reverse stage (WR), and a window-based multihead self-attention (W-MSA) stage. The input feature map $\mathbf{Z}_{\text{ITS}} \in \mathbb{R}^{B\times H\times W\times C_d}$ is divided into S nonoverlapping windows $[\mathbf{Z}_{w_1}, \mathbf{Z}_{w_2}, \ldots, \mathbf{Z}_{w_S}]$ inspired by the SwinTransformer [46], where $\mathbf{Z}_{w_i} \in \mathbb{R}^{b_1\times h_1\times w_1\times c_1}$, B, H, W, and P denote the batch size, height, width, and channel number of the feature map, respectively. b_1, h_1, w_1, and c_1, respectively, denote the batch size, height, width, and channel of each window.

To maintain the homogeneity of the pixels within each window, we employ the following method to compute the height of each window. The height of each window is computed as $h_1 = (N_{\text{spb}} \bmod H) \times 0.2$, where N_{spb} is the total number of pixels in the smallest homogeneous block. The value of 0.2 has been adopted to balance the tradeoff between window size and pixel homogeneity, ensuring that each window is small enough to capture homogeneity while still maintaining sufficient size for effective analysis. The width and the channel remain consistent with the input feature map, while the batch size of each window is computed as $b1 = B \times (H/h_1) \times (W/w_1)$

In the HRCC stage, the W-MSA module computes the self-attention for each window and reverses the partitioned windows back to the input feature map of HRCC by concatenating

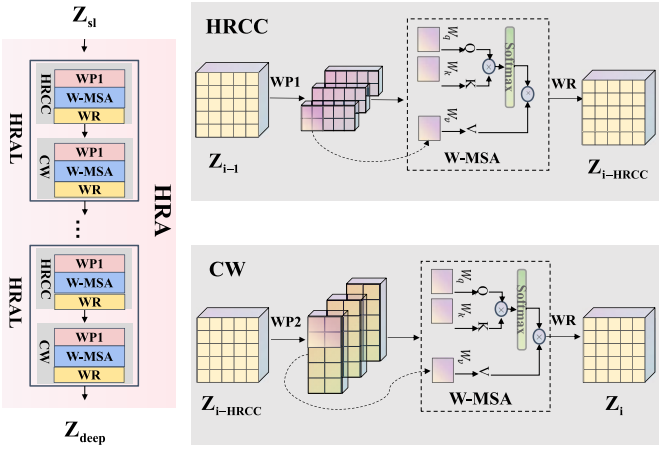


Fig. 3. Details of HRA.

the processed windows along the spatial dimensions

$$\mathbf{Z}_{i-\text{HRCC}} = \text{WR}(\text{W-MSA}(\text{WP1}(\mathbf{Z}_{i-1}))) \tag{8}$$

where WP denotes the WP module, and $\mathbf{Z}_{i-1}$ is the output feature map of the $(i-1)$th HRA layer (HRAL).

In the CW stage, the WP module sets the height of each window to w_1 and the width to h_1 to enable information exchange across windows. The W-MSA module then computes the self-attention for each window

$$\mathbf{Z}_i = \text{WR}(\text{W-MSA}(\text{WP2}(\mathbf{Z}_{i-\text{HRCC}}))) \tag{9}$$

where $\mathbf{Z}_i$ is the output feature map of the ith layer.

D. Spectral Group-Aware

The proposed method introduces an SGA module shown in Fig. 1 that exploits the similarities among adjacent spectra in

hyperspectral images to extract salient spectral information. The module consists of two main strategies: spectral group partition and intragroup spectral self-attention calculation.

1) Spectral Group Partition Strategy: The spectral bands of the hyperspectral image inherently exhibit similar traits between neighboring bands, and wavelengths in close proximity typically correspond to analogous colors. Exploiting this feature, we first divide the hyperspectral image into distinct color groups depending on its wavelength range. In our experimental setup, we consider a total of nine color groups ($k = 9$) comprising red, orange, yellow, green, cyan, blue, purple, near-infrared, and far-infrared. Subsequently, the hyperspectral image is fractionated into these nine groups according to their respective wavelengths.

We employ the k-means clustering technique to accomplish this color grouping. The k-means algorithm operates by initializing "k" centroids in the dataset, where "k" is the number of desired clusters, in our case, $k = 9$. Each spectral band is then assigned to the cluster whose centroid is nearest, based on the Euclidean distance. Following this, new centroids are computed as the mean of all data points within each cluster. This iterative process continues until the positions of the centroids stabilize and no spectral bands change clusters. The procedure can be denoted as

$$Y = [Y_1, Y_2, \ldots, Y_k] \tag{10}$$

where $Y_i \in \mathbb{R}^{H\times W\times C_i}$ represents the ith spectral group and k denotes the number of spectral groups.

By leveraging the inherent similarities between neighboring spectral bands and utilizing the k-means clustering algorithm, we can efficiently partition the hyperspectral image into distinct color groups, facilitating further analysis.

2) Intragroup Spectral Self-Attention Calculation: In the subsequent stage, the self-attention of each spectral group is calculated using a self-attention mechanism. Specifically, the spectral information within each spectral group is projected into query, key, and value spaces using learnable weight matrices, respectively. The self-attention matrix $A(i)$ is then obtained by applying a softmax function to the dot product of the query and key matrices and multiplying it with the value matrix. This enables the module to emphasize the salient spectral information and suppress the irrelevant information within each spectral group. The procedure can be represented as

$$A(i) = \text{softmax}\left(Q_i K_i^T\right) V_i \tag{11}$$

where Q_i, K_i, and V_i are the query, key, and value matrices, respectively. The output of the SGA module $\overline{Y}$ is obtained by concatenating the self-attention matrices of all spectral groups

$$\overline{Y} = [A(1); A(2); \ldots; A(k)]. \tag{12}$$

III. Experiments

Our experimental evaluation comprises one synthetic dataset and three representative real datasets, namely Samson, JasperRidge, and Apex datasets. We compare our methods against existing approaches on all datasets. Further details are provided below.

A. Data Description

1) Samson: The Samson dataset, obtained by the SAMSON sensor, is depicted in Fig. 4(a). The dataset comprises a region of interest (ROI) with dimensions of 95 × 95, which includes three endmembers, namely *Soil*, *Water*, and *Tree*. The ROI is observed on 156 channels, covering wavelengths from 0.401 to 0.889 μm at each pixel.

2) JasperRidge: The AVIRIS sensor collected the dataset over Jasper Ridge in central California, USA, as shown in Fig. 4(b). The original image has 512 × 614 pixels, 224 wavelength bands ranging from 0.38 to 2.5 μm. Researchers often conduct experiments with subimages containing 100 × 100 pixels selected from the original image. The endmembers in this data are: *Soil*, *Water*, *Tree*, and *Road*.

3) Apex: To further verify the effectiveness of our model, we use the Apex dataset which covers the wavelength range from 413 to 2420 and is cropped into a shape of 110 × 110. There are four endmembers [i.e., *Water*, *Tree*, *Road*, and *Roof*, shown in Fig. 4(c)].

4) Synthetic Dataset: To evaluate the performance of hyperspectral unmixing algorithms, we generated a simulated dataset containing 224 spectral bands and a spatial resolution of 100 × 100 pixels. The dataset shown in Fig. 4(d) was created using five endmembers that represent a diverse range of materials commonly found on Earth's surface, including Muscovite, Kaolinite, Calcite, Smectite, and Alunite. To more accurately simulate the noise present in real hyperspectral image collection, we added 20, 30, and 40 dB of noise to the simulated data.

B. Experimental Setup

1) Methods for Comparison: To assess our method, we choose five classic and state-of-the-art methods for comparison, which include the geometry-based VCA [47], FCLSU [48], the statistical-based NMF-QMV [49], and the deep-learning-based GLA [41], CNNAEU [50], and DeepTrans-HsU [45].

2) Performance Metrics: We evaluate the obtained abundance maps $\hat{Z}$ with ground-truth Z using average root MSE (aRMSE), which can be written as

$$\text{aRMSE}\left(Z, \hat{Z}\right) = \sqrt{\frac{1}{pn}\sum_{i=1}^{P}\sum_{j=1}^{N}\left\|Z_{ij} - \hat{Z}_{ij}\right\|_2^2}. \tag{13}$$

We use SAD as the evaluation criterion for endmember estimation, which can be expressed as

$$\text{SAD}\left(\hat{M}, M\right) = \frac{1}{P}\sum_{i=1}^{P}\arccos\left(\frac{\hat{M}_i^T M_i}{\left\|\hat{M}_i\right\|_2\left\|M_i\right\|_2}\right) \tag{14}$$

where M denotes the reference endmember and $\hat{M}$ is the estimated endmember.

C. Experimental Results

1) Synthetic Data: The statistical results of the average root-mean-square error (RMSE), SAD, and standard deviation acquired by our proposed model are presented in

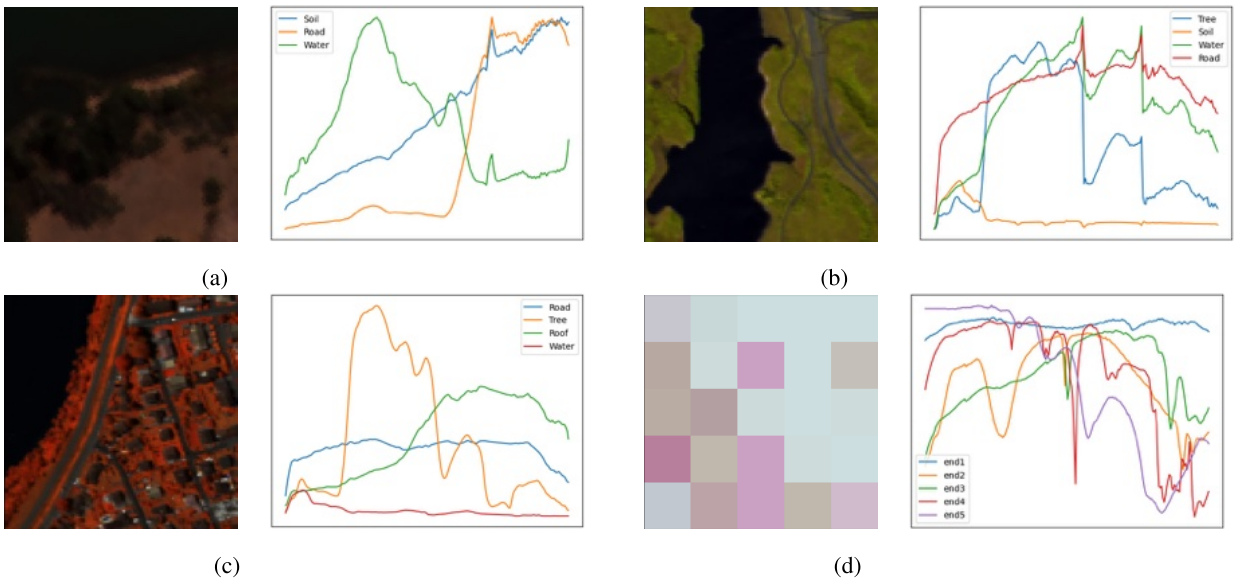


Fig. 4. Data description of all datasets: (left) Each subfigure displays a true-color image and (right) its corresponding endmembers. (a) Samson dataset. (b) JasperRidge dataset. (c) Apex dataset. (d) Synthetic dataset.

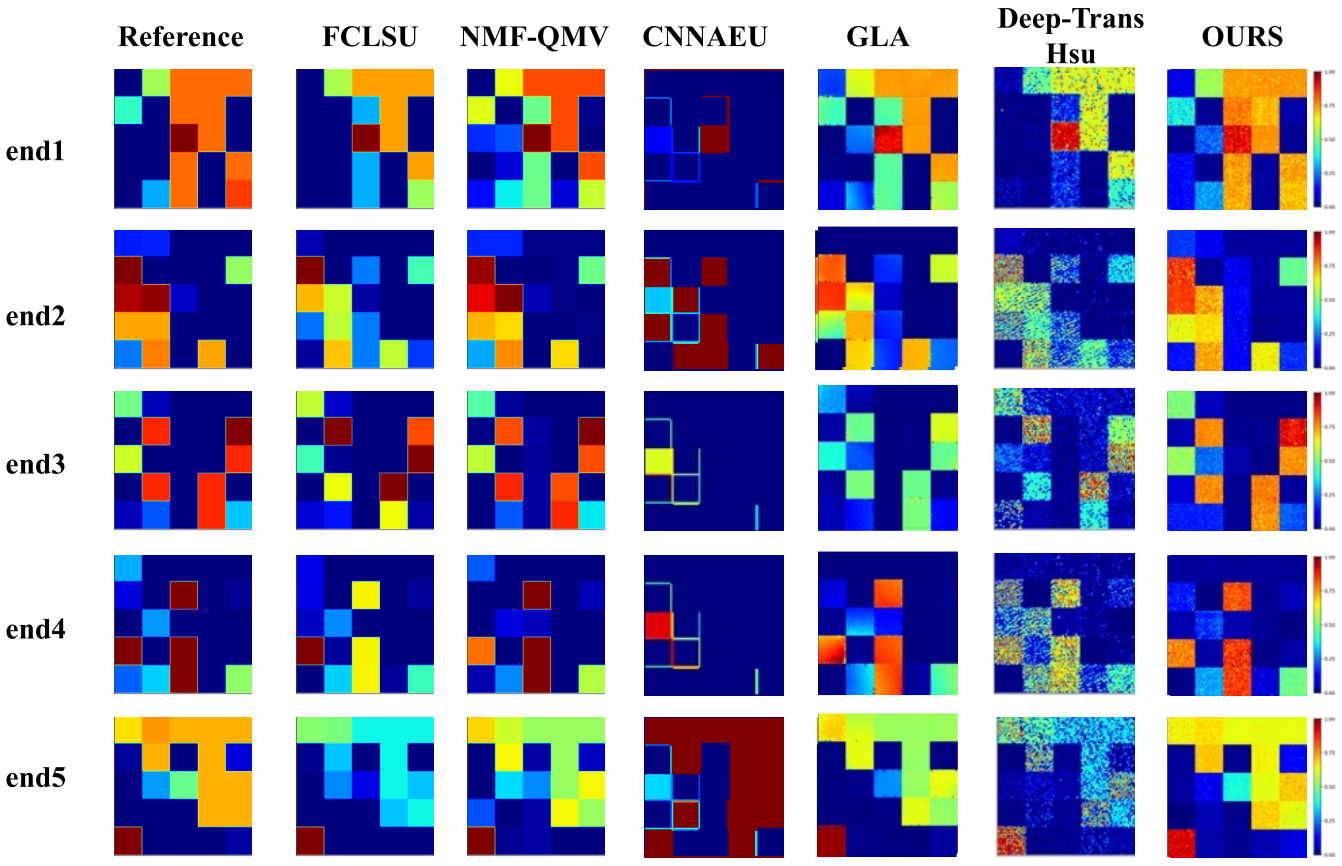


Fig. 5. Visual comparison of abundance maps obtained by different methods on synthetic data with a signal-to-noise ratio (SNR) of 40 dB.

Tables I and II. It is encouraging to observe that our model consistently demonstrates favorable performance and comparable outcomes, showcasing its robustness across varying levels of noise.

For a visual comparison of abundance maps obtained by different methods, we refer to Fig. 5. Notwithstanding the presence of some noise, our approach exhibits the generation of abundance maps that closely resemble the ground truth.

Furthermore, Fig. 6 reveals that our method achieves more precise estimations of the true values. We acknowledge that there might be areas for improvement, but these results provide promising evidence of the efficacy of our proposed approach.

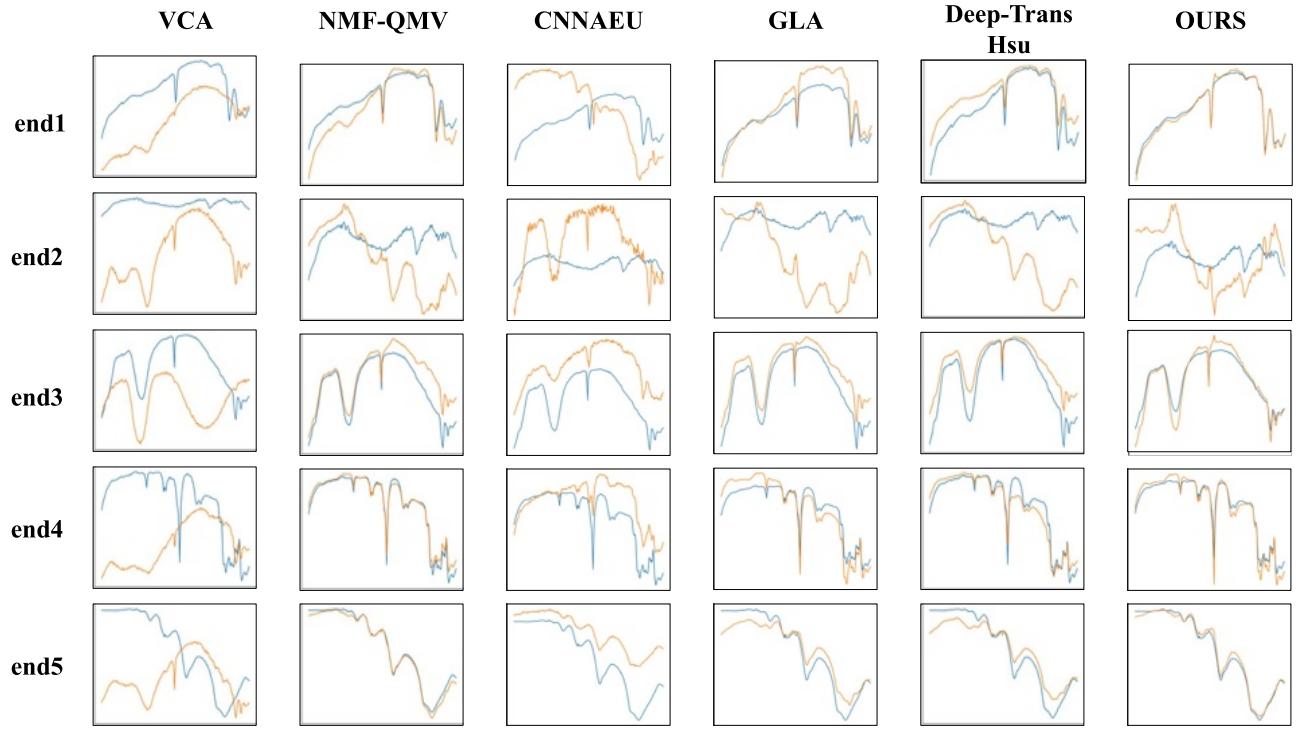


Fig. 6. Visual comparison of endmember maps obtained by different methods on synthetic data with an SNR of 40 dB. Blue line: ground-truth endmembers; orange line: estimated endmembers.

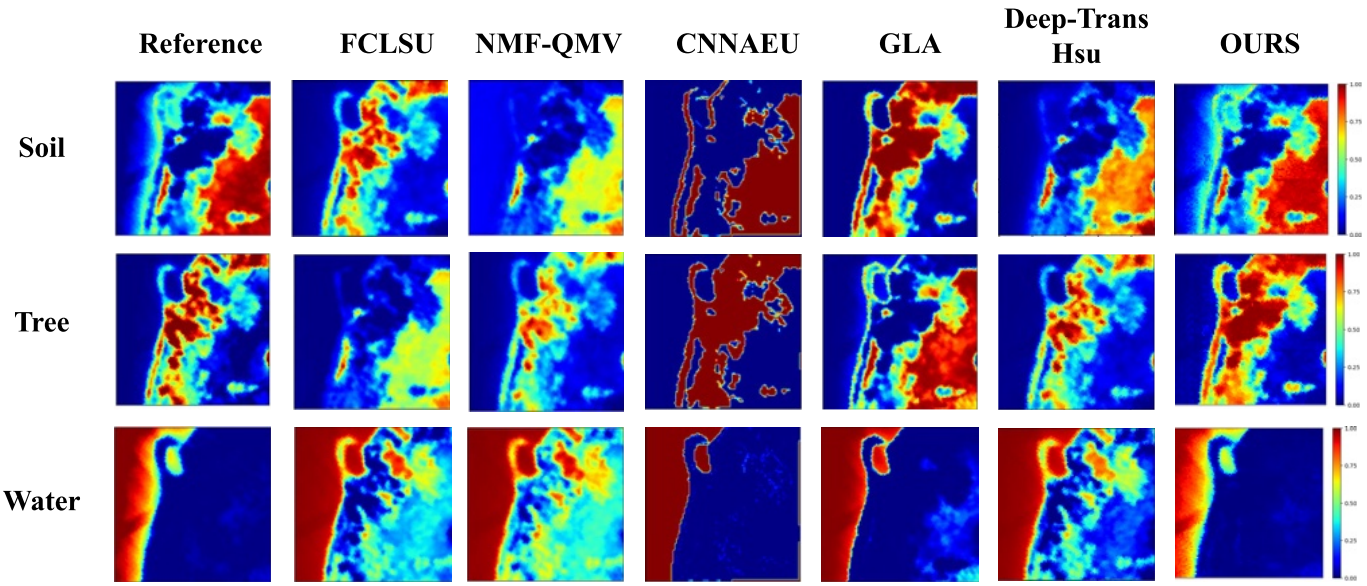


Fig. 7. Visual comparison of abundance maps obtained by different methods on synthetic data with an SNR of 40 dB.

TABLE I

MEAN RMSE RESULTS OF THE SYNTHETIC DATASET. THE BEST ONE IS SHOWN IN BOLD (×10)

	FCLSU	NMF-QMV	CNNAEU	GLA	DeepTrans-HsU	OURS
20dB	3.7407±0.01%	3.4831±0	3.8121±0.17%	2.9602±0.26%	2.9642±1.92%	**2.7052±0.25%**
30dB	3.5078±0.01%	3.6102±0	3.3923±0.58%	3.0312±0.03%	2.9321±1.36%	**2.0902±0.38%**
40dB	2.5145+0.14%	0.70963±0	4.4812±0.26%	0.7278±0.01%	2.3155±0.29%	**0.5462±0.41%**

2) Samson Dataset: Table III presents quantitative results on the Samson dataset. The actual values are listed in the accompanying table and have been multiplied by a factor of 10 for ease of comparison. Figs. 7 and 8 offer

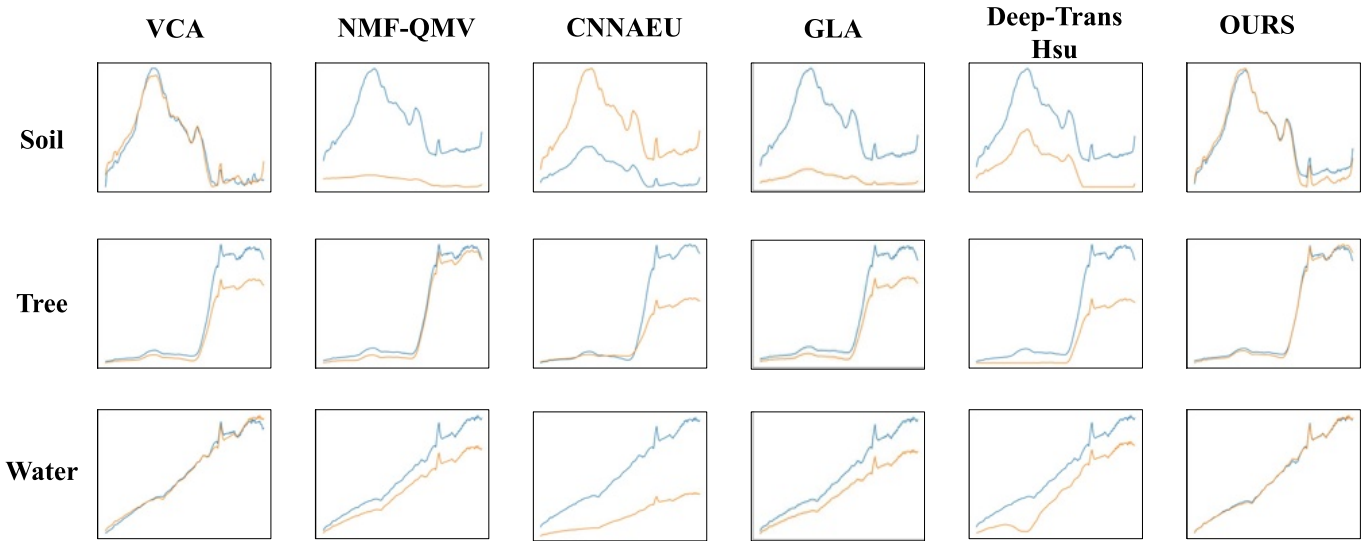


Fig. 8. Visual comparison of abundance maps obtained by different methods on synthetic data with an SNR of 40 dB.

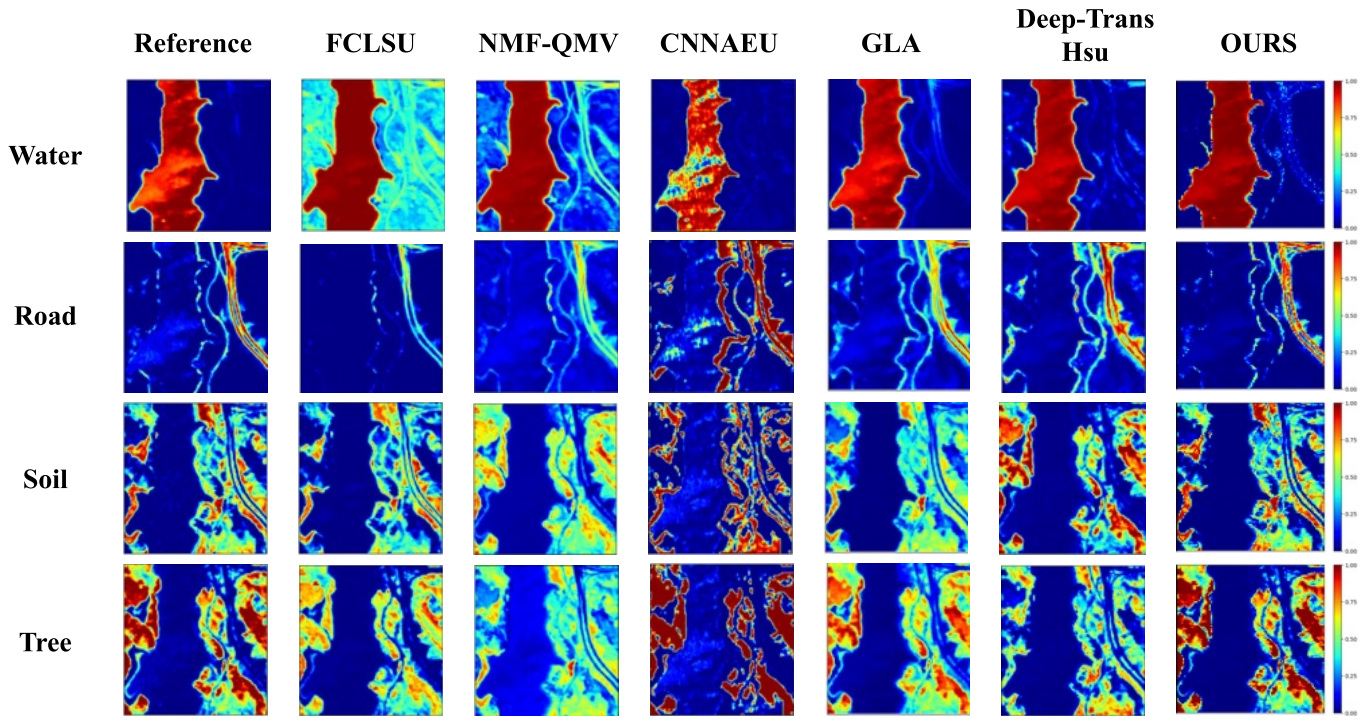


Fig. 9. Visual comparison of abundance maps obtained by different methods on the JasperRidge dataset.

TABLE II
MEAN SAD RESULTS ON THE SYNTHETIC DATASET. THE BEST ONE IS SHOWN IN BOLD (×10)

	VCA	NMF-QMV	CNNAEU	GLA	DeepTrans-HsU	OURS
20dB	2.6561±0.12%	0.4144±0	2.7549±0.15%	1.0512±0	1.0571±0.53%	**0.6893±0.02%**
30dB	2.403±0.05%	1.5707±0	2.4063±0.24%	1.7609±0.01%	1.8861±0.21%	**1.4766±0.11%**
40dB	2.2420±0.09%	0.6347±0	2.085±0.05%	0.7279±0.08%	0.9432±0.15%	**0.5077±0.01%**

visual comparisons of six representative methods. Our method yields results that are visually comparable to ground truths, as expected. A detailed examination of Fig. 7 reveals that our proposed scored-based HRA allows for better handling of boundary pixels between various objects compared to other methods. The proposed model outperforms other techniques,

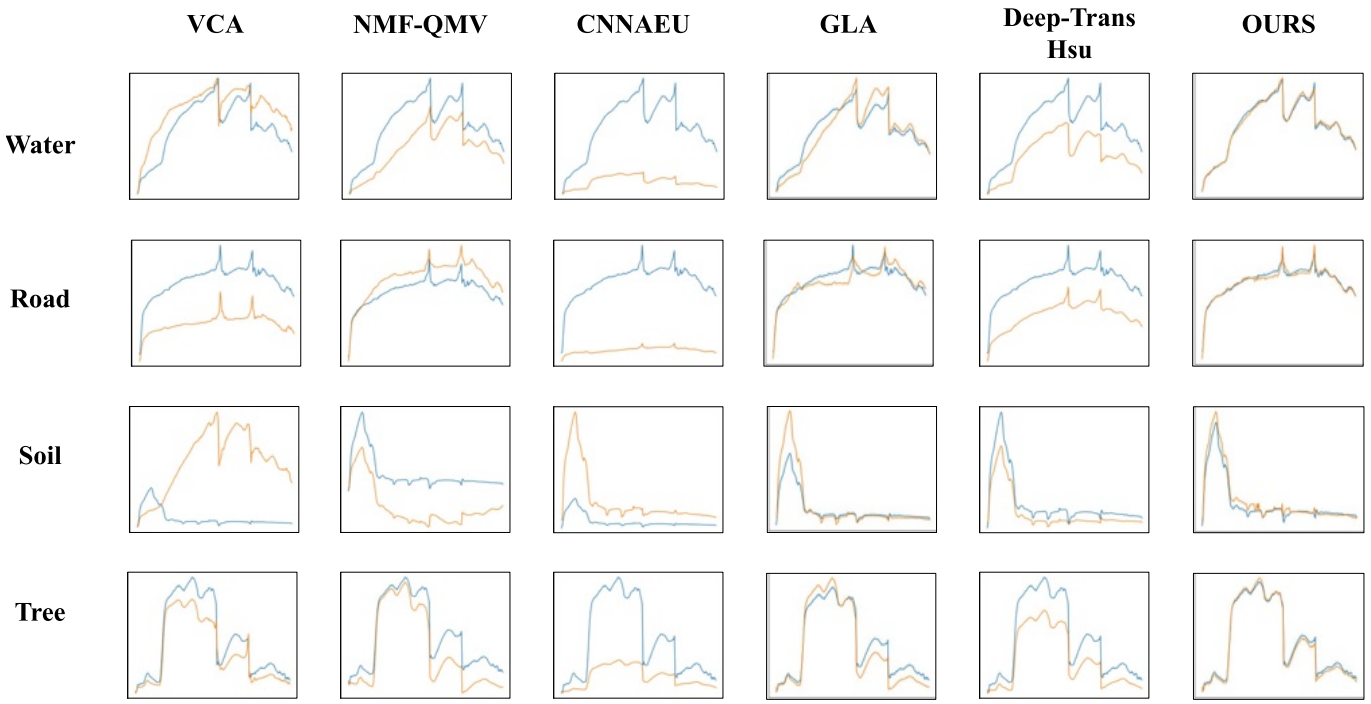


Fig. 10. Visual comparison of endmembers obtained by different methods on the JasperRidge dataset. Orange line: ground-truth endmembers; blue line: estimated endmembers.

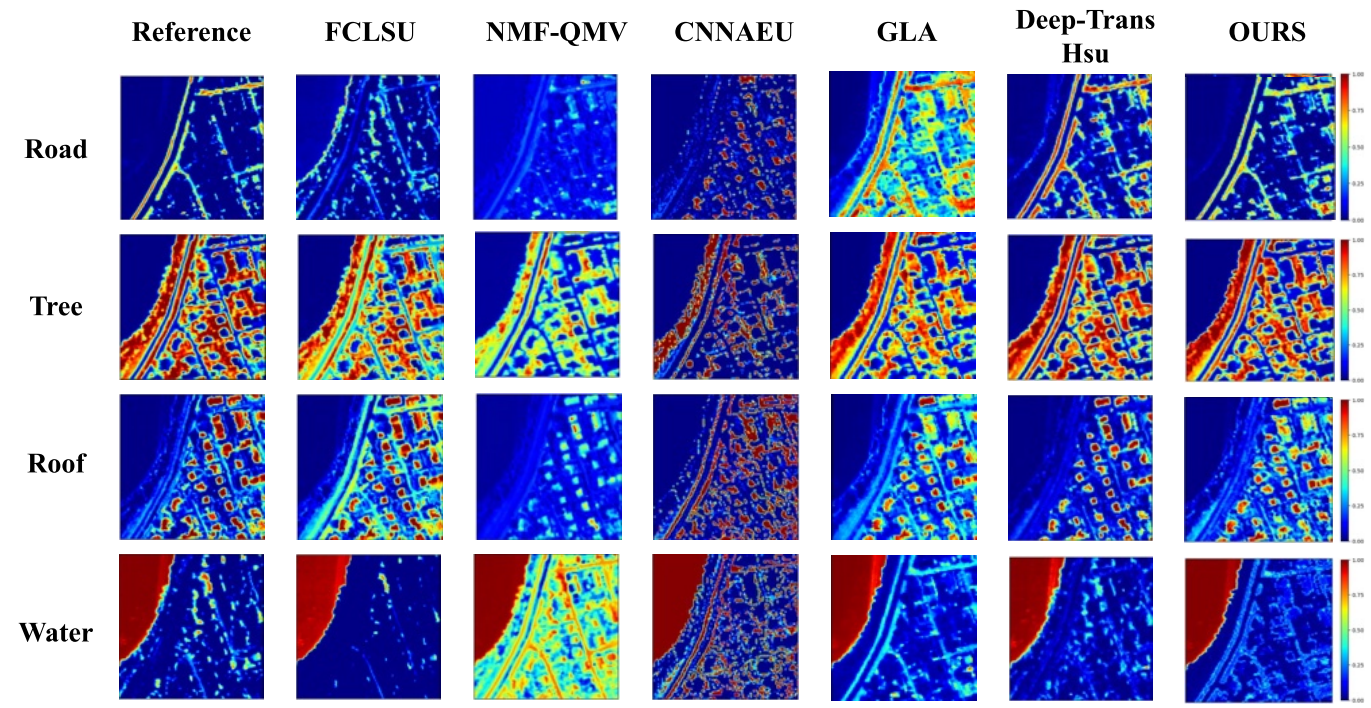


Fig. 11. Visual comparison of abundance maps obtained by different methods on the Apex dataset.

with a mean RMSE of 0.3429 and a mean SAD value of 0.1926. These results indicate the effectiveness of our proposed methods.

3) JasperRidge Dataset: The JasperRidge image dataset was used to evaluate the effectiveness of various methods, as presented in Table IV. Our proposed method performed the best in terms of abundance RMSE and spectral SAD metrics, while DeepTrans showed less proficiency in abundance estimation. On the other hand, FCLSU and NMF-QMV performed poorly in both abundance RMSE and spectral SAD metrics.

Fig. 9 depicts the poor results obtained by FCLSU, CNNAEU, and NMF-QMV methods, whereas DeepTrans failed to distinguish between the two endmembers "Tree" and "Road." In contrast, our proposed method produced abundances that exhibited a high degree of similarity to the ground truth. Moreover, Fig. 10 suggests that our

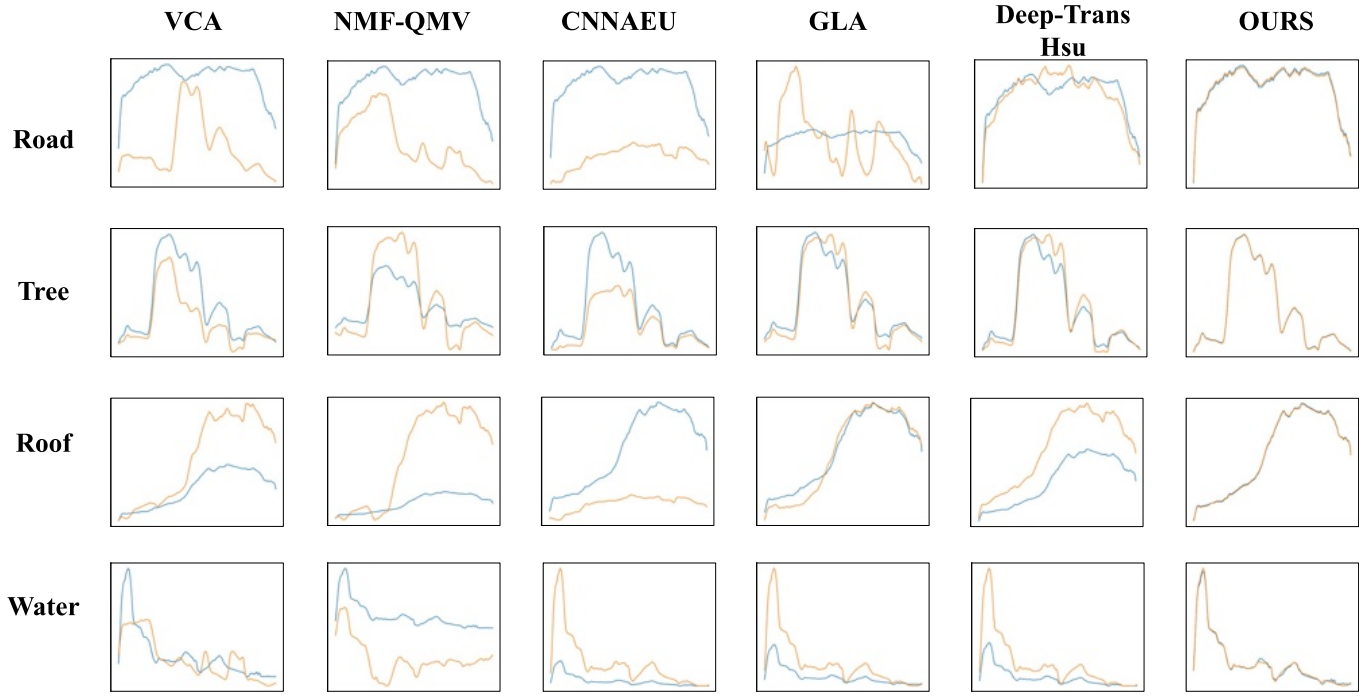


Fig. 12. Visual comparison of endmembers obtained by different methods on the Apex dataset. Orange line: ground-truth endmembers; blue line: estimated endmembers.

TABLE III
QUANTITATIVE RESULTS OF THE SAMSON DATASET. THE BEST ONE IS SHOWN IN BOLD (×10)

	FCLSU	VCA	NMF-QMV		CNNAEU		GLA		DeepTrans-HsU		OURS	
	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD
Soil	5.1791	0.9202	2.3298	0.2326	3.1571	0.7565	1.0612	1.5958	2.8261	1.5900	**0.4306**	**0.1191**
Tree	3.8002	0.7413	2.4432	0.6086	2.9114	0.5440	0.6451	0.5735	1.8093	1.1049	**0.2854**	**0.3775**
Water	3.3053	1.7792	3.7621	14.5643	1.5524	0.3642	1.1218	0.1915	1.9390	3.0116	**0.3128**	**0.0813**
Overall	4.0949	1.1469	2.8450	5.1352	2.5403	0.5549	0.9427	0.7869	2.1915	1.9022	**0.3429**	**0.1926**

TABLE IV
QUANTITATIVE RESULTS OF THE JASPERRIDGE DATASET. THE BEST ONE IS SHOWN IN BOLD (×10)

	FCLSU	VCA	NMF-QMV		CNNAEU		GLA		DeepTrans-HsU		OURS	
	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD
Water	2.9886	1.6316	2.0375	2.8387	2.1176	6.0821	0.4532	1.8562	1.0209	0.6448	**0.4211**	**0.2811**
Road	0.2006	0.6616	1.8446	14.6974	2.0429	0.5187	0.8464	1.0513	0.7961	1.1485	**0.7792**	**0.3306**
Soil	1.7473	11.1550	1.4647	1.7326	2.9783	3.3806	1.2752	1.2791	1.3829	1.0724	**0.7007**	**0.9074**
Tree	1.1199	1.1211	1.2528	0.5016	3.1690	3.1044	0.9283	1.5289	1.2560	0.6319	**0.6031**	**0.4519**
Overall	1.5141	3.6423	1.6499	4.9426	2.5770	3.2714	0.8758	1.4289	1.1140	0.8744	**0.6260**	**0.4927**

TABLE V
QUANTITATIVE RESULTS OF THE APEX DATASET. THE BEST ONE IS SHOWN IN BOLD (×10)

	FCLSU	VCA	NMF-QMV		CNNAEU		GLA		DeepTrans-HsU		OURS	
	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD	RMSE	SAD
Water	2.0575	3.2571	3.2571	4.1067	2.5669	0.4983	1.8914	0.4513	2.0245	1.8428	**1.6099**	**0.4450**
Road	1.4517	2.8580	2.8580	2.7636	3.1502	3.3235	1.6196	5.1296	1.3248	1.3698	**1.2311**	**0.4381**
Roof	1.3659	1.3661	1.3661	1.8906	2.5836	3.0619	1.2489	0.9504	**1.0908**	1.2562	1.3476	**0.4715**
Tree	1.3366	9.0444	3.7539	18.1938	3.8636	1.7377	1.5503	1.5410	**1.1067**	0.3898	1.3541	**0.5892**
Overall	1.5529	4.1314	2.8088	6.7387	3.0411	2.1554	1.5776	2.0181	1.3867	1.2147	**1.3857**	**0.4859**

method exhibits a degree of improved efficacy in extracting endmembers.

4) Apex Dataset: The quantitative findings garnered through various methodologies as applied to this dataset are

enumerated in Table V. The method we propose seems to display superior performance in relation to other strategies, specifically concerning endmember estimation across all categories. However, it should be noted that when it comes to the abundance estimation of notably challenging endmembers like "Roof" and "Tree," our method ranks second.

Furthermore, the visual resemblance between the abundance maps and endmember maps generated by our proposed method and the corresponding ground-truth maps, as depicted in Fig. 11 and Fig. 12, provides an indication that the proposed technique is potentially proficient at accurately determining the abundances and endmembers of each material.

Based on the experimental results conducted on three real datasets and one simulated dataset, our proposed method has demonstrated promising performance in endmember and abundance extraction compared to other existing methods.

IV. Conclusion

In this article, we propose a novel UnDAT, which aims at simultaneously exploiting the region homogeneity and spectral correlation of hyperspectral imagery. The proposed UnDAT comprises two important modules: the SHA module and the SGA module. The SGA module highlights the significant spectral information, enabling us to extract a shallow feature map from the processed hyperspectral image using an encoder. We then reorganize these features based on their distances from various centroids.

Next, we utilize the SHA module to capture global interactions within homogeneous areas. This module consists of three strategies: FMR, image to sequences, and HRA. The combination of three strategies can effectively model complex relationships between different parts of the hyperspectral image. Finally, we use a linear decoder to reconstruct the hyperspectral image.

To evaluate the performance of our proposed architecture, we conducted experiments on three real-scene datasets and a synthetic dataset. We compared our approach against geometrical and deep-learning-based methods, and the results demonstrate that our proposed architecture works better than all other approaches.

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 Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] Y. Su, J. Li, H. Qi, P. Gamba, A. Plaza, and J. Plaza, "Multi-task learning with low-rank matrix factorization for hyperspectral nonlinear unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 2127–2130.

[3] Q. Zhu et al., "S^3TRM: Spectral–spatial unmixing of hyperspectral imagery based on sparse topic relaxation-clustering model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5515613.

[4] Z. Zhang, Q. Wang, and Y. Yuan, "Hyperspectral unmixing VIA L1/4 sparsity-constrained multilayer NMF," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 2143–2146.

[5] X. Xu, Z. Shi, B. Pan, and X. Li, "A classification-based model for multi-objective hyperspectral sparse unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 9612–9625, Dec. 2019.

[6] X. Tao et al., "A new deep convolutional network for effective hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6999–7012, 2022.

[7] S. Jia, J. Liang, L. Deng, and Y. Qian, "Minimum variance block-based nonnegative matrix factorization algorithm for hyperspectral unmixing," in *Proc. 4th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2012, pp. 1–4.

[8] 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.

[9] R. Heylen, M. Parente, and P. Gader, "A review of nonlinear hyperspectral unmixing methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 1844–1868, Jun. 2014.

[10] J. M. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2011, pp. 1135–1138.

[11] M. Song, Y. Li, L. Zhang, and C.-I. Chang, "Recursive orthogonal vector projection algorithm for linear spectral unmixing," in *Proc. 6th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2014, pp. 1–4.

[12] H.-C. Li and C.-I. Chang, "Linear spectral unmixing using least squares error, orthogonal projection and simplex volume for hyperspectral images," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.

[13] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "Convex analysis based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Apr. 2009, pp. 1089–1092.

[14] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.

[15] F. Amiri and M. H. Kahaei, "New Bayesian approach for semi-supervised hyperspectral unmixing in linear mixing models," in *Proc. Iranian Conf. Electr. Eng. (ICEE)*, May 2017, pp. 1752–1756.

[16] K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semisupervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.

[17] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.

[18] R. Li, B. Pan, X. Xu, T. Li, and Z. Shi, "Toward convergence: A gradient-based multiobjective method with greedy hash for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509114.

[19] S. Zhang et al., "Superpixel-guided sparse unmixing for remotely sensed hyperspectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 2155–2158.

[20] S. Zhang, J. Li, H.-C. Li, C. Deng, and A. Plaza, "Spectral–spatial weighted sparse regression for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3265–3276, Jun. 2018.

[21] Y. Miao and B. Yang, "Multilevel reweighted sparse hyperspectral unmixing using superpixel segmentation and particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[22] X. Shen, H. Liu, X. Zhang, K. Qin, and X. Zhou, "Superpixel-guided local sparsity prior for hyperspectral sparse regression unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[23] Y. Wei, X. Xu, B. Pan, T. Li, and Z. Shi, "A multiobjective group sparse hyperspectral unmixing method with high correlation library," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7114–7127, 2022.

[24] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "DAEN: Deep autoencoder networks for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4309–4321, Jul. 2019.

[25] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and Y. Huang, "Deep auto-encoder network for hyperspectral image unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 6400–6403.

[26] A. Min, Z. Guo, H. Li, and J. Peng, "JMNet: Joint metric neural network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5505412.

[27] L. Qi, F. Gao, J. Dong, X. Gao, and Q. Du, "SSCU-Net: Spatial–spectral collaborative unmixing network for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5407515.

[28] Z. Han, D. Hong, L. Gao, B. Zhang, and J. Chanussot, "Deep half-siamese networks for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 1996–2000, Nov. 2021.

[29] F. Kong, M. Chen, Y. Li, and D. Li, "A global spectral–spatial feature learning network for semisupervised hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 3190–3203, 2022.

[30] J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2189–2192.

[31] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1755–1759, Nov. 2018.

[32] V. S. Deshpande and J. S. Bhatt, "A practical approach for hyperspectral unmixing using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[33] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, "CyCU-Net: Cycle-consistency unmixing network by learning cascaded autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503914.

[34] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1427–1431, Sep. 2018.

[35] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Blind hyperspectral unmixing using autoencoders: A critical comparison," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1340–1372, 2022.

[36] C. Zhou and M. R. D. Rodrigues, "ADMM-based hyperspectral unmixing networks for abundance and endmember estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5520018.

[37] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "Spectral unmixing using autoencoder with spatial and spectral regularizations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 3321–3324.

[38] Z. Zhao, H. Wang, Y. Liang, T. Huang, Y. Xiao, and X. Yu, "Sparsity constrained convolutional autoencoder network for hyperspectral image unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 3317–3320.

[39] Z. Hua, X. Li, Y. Feng, and L. Zhao, "Dual branch autoencoder network for spectral–spatial hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[40] Q. Jin, Y. Ma, X. Mei, and J. Ma, "TANet: An unsupervised two-stream autoencoder network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5506215.

[41] X. Xu, X. Song, T. Li, Z. Shi, and B. Pan, "Deep autoencoder for hyperspectral unmixing via global-local smoothing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5524216.

[42] Y. Yu, Y. Ma, X. Mei, F. Fan, J. Huang, and H. Li, "Multi-stage convolutional autoencoder network for hyperspectral unmixing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, Sep. 2022, Art. no. 102981. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843222001728

[43] Z. Hua, X. Li, Q. Qiu, and L. Zhao, "Autoencoder network for hyperspectral unmixing with adaptive abundance smoothing," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 9, pp. 1640–1644, Sep. 2021.

[44] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[45] P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, and P. Scheunders, "Hyperspectral unmixing using transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535116.

[46] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

[47] J. M. P. Nascimento and J. M. B. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.

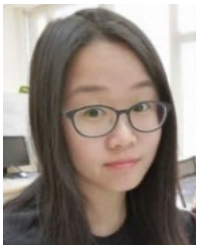
[48] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, Nov. 2011.

[49] M. Zhao, T. Gao, J. Chen, and W. Chen, "Hyperspectral unmixing via nonnegative matrix factorization with handcrafted and learned priors," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[50] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral–spatial hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 535–549, Jan. 2021.

Yuexin Duan received the B.S. degree from the School of Computer, Nankai University, Tianjin, China, in 2020, where she is currently pursuing the M.S. degree with the School of Cybersecurity.

Her research interests include machine learning and hyperspectral unmixing.



Xia Xu received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Astronautics, Beihang University, Beijing, China, in 2019.

She is currently an Assistant Professor with the College of Computer Science, Nankai University, Tianjin, China. Her research interests include hyperspectral unmixing, multiobjective optimization, and remote-sensing image processing.

Tao Li received the Ph.D. degree in computer science from Nankai University, Tianjin, China, in 2007.

He is currently a Professor with the College of Computer Science, Nankai University. His main research interests include heterogeneous computing, machine learning, and the Internet of Things.

Dr. Li is a Distinguished Member of the China Computer Federation (CCF).

Bin Pan (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote-sensing image processing, and multiobjective optimization.

Zhenwei Shi (Senior Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

He was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA, from 2013 to 2014. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored more than 200 scientific articles in refereed journals and proceedings, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), and the IEEE International Conference on Computer Vision (ICCV). His research interests include remote-sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi serves as an Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the *Pattern Recognition*, the ISPRS *Journal of Photogrammetry and Remote Sensing*, and the *Infrared Physics and Technology*. His personal website is http://levir.buaa.edu.cn/.