Hyperspectral Unmixing With Multi-Scale Convolution Attention Network

Sheng Hu and Huali Li, *Member, IEEE*

***Abstract*—Hyperspectral unmixing is to decompose the mixed pixel into the spectral signatures (endmembers) with their corresponding abundances. However, the ignorance of endmember variability in hyperspectral unmixing results in low performance. To solve this problem, a multi-scale convolution attention network containing endmember unmixing network (EU-Net) and abundance unmixing network, which was called as-the hyperspectral unmixing with multi-scale convolution attention network (HUMSCAN) was proposed in this article. The EU-Net is composed of the variational autoencoder and the multi-scale effective convolution block attention module (MSECBAM), which is combined with the S-VCA pretraining to adaptively extract endmembers at the pixel and subpixel levels. The AU-Net is based on the MSECBAM frame jointed the spectral and spatial attention features. The proposed HUMSCAN method can simultaneously and unsupervisedly extract endmembers and their corresponding abundances, which can improve the accuracy and efficiency of spectral unmixing. The performance of the proposed method is evaluated both on synthetic and real datasets. Experimental results show its superiority in comparison with other state-of-the-art methods.**

***Index Terms*—Endmember variability, hyperspectral unmixing (HU), multi-scale effective convolution block attention module, variational autoencoder (VAE).**

I. Introduction

HYPERSPECTRA images (HSI) are widely used in remote sensing [1], environmental monitoring [2], [3], agriculture [4], [5], medical [6], and military purposes [7]. Due to atmospheric transmission and imaging device performance, mixed pixels are inevitable in HSI. However, the existence of mixed pixels seriously affects the accuracy of subsequent image processing. Therefore, hyperspectral unmixing (HU) is important, which decomposes the mixed pixel into the spectral signatures (endmembers) with their corresponding abundances [8].

HU models are mainly divided into linear mixture model (LMM) and nonlinear mixture model. In the LMM, pixel purity index [9], N-FINDR [10], and vertex component analysis [11] are classical pure pixel unmixing methods, while the minimum volume simplex analysis [12] and SUnSAL [13] can do unmixing without pure pixels. The traditional spectral unmixing methods are under the assumption that there is only one spectral curve for each type of ground object (endmember). Due to “same subject with different spectra” [14], [15], the spectral signature of a given material can not be expressed by a single signature. So, the endmember variability is considered into the LMM. The spectrum of each material are blindly decomposed into a reference spectrum and a perturbation term on perturbed linear mixing model [16]. The endmember variability is considered into spectral variation dictionary in augmented linear mixed model [17]. The reflectivity variability is simulated by multiplying the diagonal matrix with a fixed scaling factor is in extended linear mixing model [18]. To solve the problems of the extended linear model, low-rank attributes of subspaces for hyperspectral image unmixing has been proposed, such as subspace unmixing with low-rank attribute embedding for hyperspectral data analysis [19], low-rank subspace unmixing [20].

In recent years, deep autoencoder (AE) network and its variants [21], [22], [23] are developed in spectral unmixing. The HSI is mapped to the abundance coefficient by the encoder, with abundance nonnegative constraint (ANC) and abundance sum to one constraint (ASC) [24]. However, the spatial are ignored in these AE-based deep unmixing methods [25], [26], [27], [28], [29]. So, CNN-based AE model with the spatial information for spectral unmixing is developed, such as convolutional neural network autoencoder unmixing (CNNAEU) [30], 3D-CNNAEU [31]. cycle-consistency unmixing network [32], and endmember-guided unmixing network [33], etc. Transformers play an important role in HSI processing [34], [35], [36]. AE and transformer are combined into HU using transformer network (DeepTrans-HSU) [37] with multihead self-attention mechanism for spectral unmixing. Attention-based residual network with scattering transform [38] are based on attention mechanism residual network and scatter transform features with limited training samples.

To consider the impactness of endmember variability on the unmixing performance, the deep generative unmixing algorithm (DeepGUn) [39] is based on the deep generative model with endmember variability, while multiple pure pixels are extracted from the image with the variational autoencoder (VAE) [40]. However, the accuracy of spectral unmixing with DeepGUn relies on the number and quality of extracted pure pixels, which may lead to suboptimal solutions. Endmember variability, which is encoded by VAE, is combined with a probabilistic generative model in the method Probabilistic generative model for hyperspectral unmixing (PGMSU) [41]. However, this method ignored the spatial information. Deep generative model for

Manuscript received 19 July 2023; revised 4 October 2023; accepted 17 November 2023. Date of publication 23 November 2023; date of current version 8 January 2024. This work was supported by the National Natural Science Foundation of China under Grant 62071175. *(Corresponding author: Huali Li.)*

The authors are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: hushenghl@hnu.edu.cn; lihuali@hnu.edu.cn).

The source code for this work will be available at https://github.com/ma264/HUMSCAN.

Digital Object Identifier 10.1109/JSTARS.2023.3335907

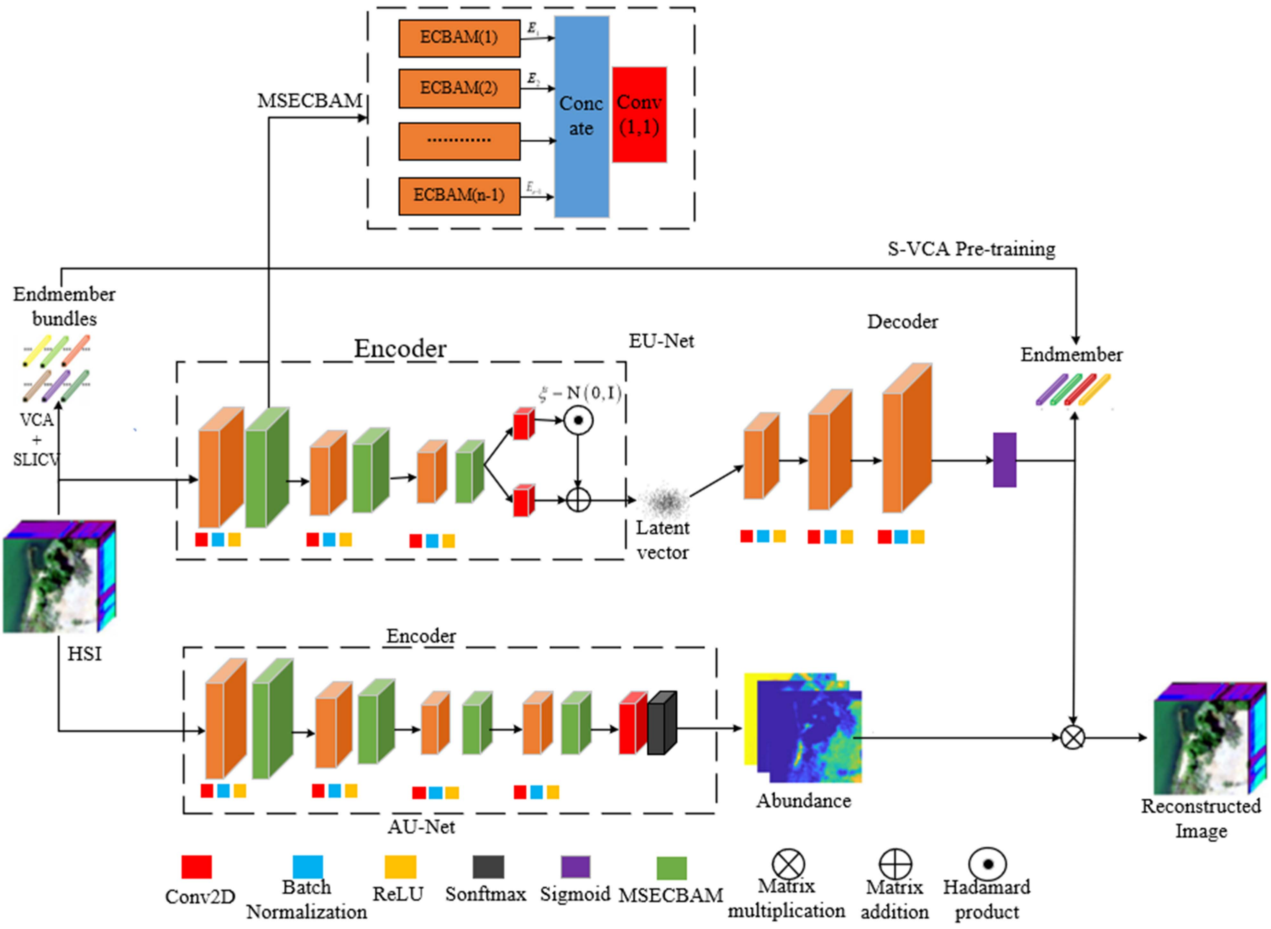


Fig. 1. Overview of the proposed HUMSCAN architecture. It is a two-stream network, including EU-Net and AU-Net.

spatial-spectral unmixing (DGMSSU) [42] is a deep generative model based on VAE with Bayesian inference methods, which is composed of CNN, graph neural networks (GNN), and self-attention mechanism. However, a lot of computing resources and training data are required in DGMSSU.

In order to solve these problems, considering VAE and the proposed multi-scale effective convolution block attention module (MSECBAM), a method named hyperspectral unmixing with multi-scale convolution attention network (HUMSCAN) is proposed for HU with spectral variability.

The improvement and main contributions of this proposed method are as follows:

1) The effective convolution block attention module (ECBAM) with different convolution kernel sizes is proposed in this article, which are used to adaptively extract multi-scale spectral-spatial attention features for the next multi-scale feature fusion.
2) A multi-scale convolution attention network called the HUMSCAN is proposed to accurately model endmember variability. MSECBAM and VAE are used to construct an endmember unmixing network (EU-Net). We use MSECBAM to construct an abundance unmixing network (AU-Net). Through joint learning of two networks, the proposed method can simultaneously perform endmember extraction and abundance estimation, more efficient and accurate.
3) The endmember extraction backbone network based on endmember beam is proposed, which adaptively extracts endmembers at pixel and subpixel levels under complex scenes.

The rest of this article is organized as follows. Section II introduces the proposed HUMSCAN framework in detail. Next, Section III discusses the experimental situation of the model in real HSI. Finally, Section IV concludes this article.

II. Proposed Framework

In this part, we mainly introduce MSECBAM, which combines ECBAM and multi-scale features, and propose the HUMSCAN framework as shown in Fig. 1. In the encoder part of the two-stream network, the global spatial and spectral informations are integrated with the MSECBAM of both EU-Net and AU-Net, respectively. AU-Net maps the input image directly to the abundance map. While EU-Net maps the input image into a potential representation of the endmember. After encoded with EU-Net and decoded, the endmember bundle are combined S-VCA pretraining to generate the final endmembers. So, the proposed HUMSCAN is forming an end-to-end unmixing network.

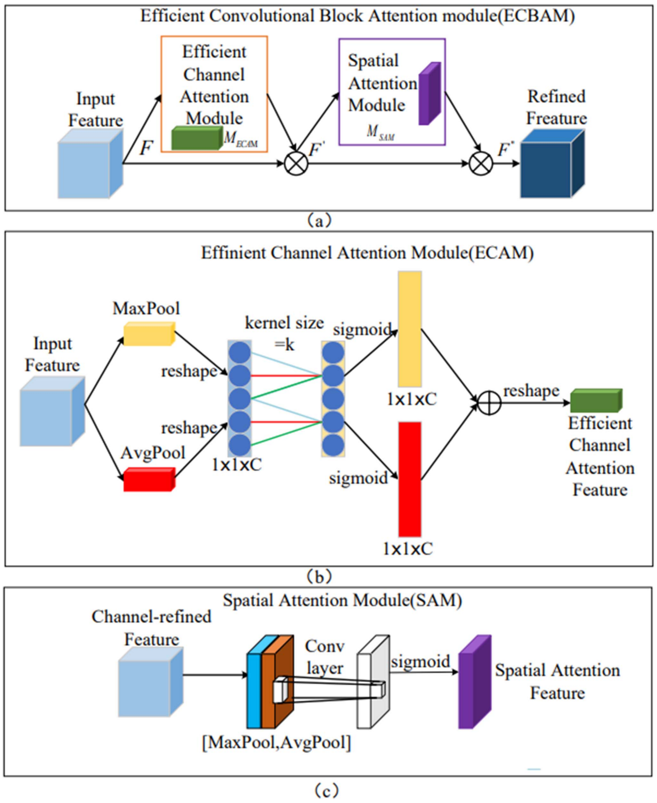


Fig. 2. Schematic diagram of the ECBAM.(a) ECBAM. (b) ECAM. (c) SAM.

A. Problem Formulation

According to the LMM, the observed spectral reflectance can be expressed as

$$\mathbf{Y} = \mathbf{MA} + \mathbf{Q} \tag{1}$$

$$\mathbf{A} \geq 0 \tag{2}$$

$$\mathbf{1}_P^T\mathbf{A} = \mathbf{1}_n^T. \tag{3}$$

Among $\mathbf{Y} \in \mathbb{R}^{L\times N}$ is the observed hyperspectral image, $\mathbf{M} \in \mathbb{R}^{L\times P}$ represents the endmember matrix of P endmember categories, $\mathbf{A} = \mathbb{R}^{P\times N}$ represents the corresponding abundance matrix, $\mathbf{Q} \in \mathbb{R}^{L\times N}$ represents the residual matrix, it contains additive noise and Gaussian noise, and $\mathbf{1}_n$ indicates an n-component column vector of ones, where L represents number of spectral bands, N represents number of pixels, P represents number of endmembers. In addition, three physical constraints should be satisfied in the unmixing process, namely the endmember matrix $\mathbf{M} \geq 0$, and the abundance vector should satisfy the constraints of ASC and ANC.

B. Efficient Convolutional Block Attention Module

As shown in Fig. 2(a), the ECBAM consists of two parts, namely the effective channel attention module (ECAM) and the spatial attention module (SAM). It replaces the channel attention module (CAM) of the convolutional block attention module (CBAM) [43] with the ECAM [44]. Through 1-D convolution with kernel function of k, the dependence between different channels can be captured more effectively, so as to improve the expression ability of image features. Compared with the traditional CABM, ECBAM has smaller computational complexity and faster calculation speed, which is suitable for scenarios with limited computing resources. ECBAM can also be integrated with other neural network structures to improve network performance. Therefore, the introduction of ECBAM makes the network give more weight to high-frequency effective information and ignore low-frequency invalid information.

As shown in Fig. 2(b), the input feature map $\mathbf{F} = \mathbb{R}^{H\times W\times L}$ of the ECAM, where *H*, *W*, and *L* are width, height, and number of bands, respectively. First, global average pooling and maximum pooling can be used to aggregate the spatial information of feature mapping, send it to the shared network, compress the spatial dimension of the input feature map, and sum the two to generate a channel attention map. Then, the local cross-channel spectral information is captured from HSI by a 1-D convolution with a convolution kernel size of *k*. Therefore, the weight calculation formula of $\boldsymbol{y}_i$ is

$$\boldsymbol{w}_i = \sigma\left(\sum_{j=1}^{k}\boldsymbol{\alpha}_i^j\boldsymbol{y}_i^j\right), \boldsymbol{y}_i^j \in \boldsymbol{\Omega}_i^k \tag{4}$$

where $\boldsymbol{\Omega}_i^k$ represents the set of k channels adjacent to $\boldsymbol{y}_i$. For this operation, the attention operation for each channel contains $k \times L$ parameters. To simplify the computation complexity, the learning parameters of all channels are the same, (4) can be written as

$$\boldsymbol{w}_i = \sigma\left(\sum_{j=1}^{k}\boldsymbol{\alpha}^j\boldsymbol{y}_i^j\right), \boldsymbol{y}_i^j \in \boldsymbol{\Omega}_i^k. \tag{5}$$

Since ECAM is usually implemented by using 1-D convolution with kernel function size k, formula (5) can be rewritten as

$$\boldsymbol{H}_{\text{ECAM}} = \sigma\left(C1D\left(\text{GAV}\left(\boldsymbol{F}\right)\right) + C1D\left(\text{MAX}\left(\boldsymbol{F}\right)\right)\right) \tag{6}$$

where $C1D$ represents 1-D convolution, GAV represents global average pooling, and MAX represents maximum pooling.

The kernel size k determines the captured interactive coverage, which is adaptive to channel dimension L. The mapping φ between kernel size k and channel dimension L is represented as follows:

$$L = \varphi(k). \tag{7}$$

The mapping relationship from [44] shows that k and L are nonlinearly proportional. The approximate mapping φ uses the following exponential function mapping:

$$\text{L} = \varphi(\text{k}) = 2^{(\varepsilon * k - b)}. \tag{8}$$

Finally, the value of k can be adaptively determined as

$$k = \lambda(L) = \left|\frac{\log_2(L)}{\varepsilon} + \frac{b}{\varepsilon}\right|_{\text{odd}} \tag{9}$$

where $|v|_{\text{odd}}$ represents the odd number closest to v.

As shown in Fig. 2(c), by applying average pooling and maximum pooling operations along the spectral dimension, the information regions are effectively highlighted [45] and

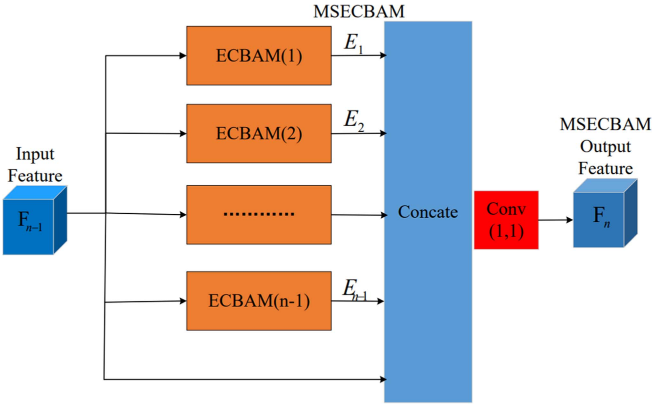


Fig. 3. Multi-Scale effective convolution block attention module.

connected to generate an effective feature descriptor. Then, 2-D spatial attention map is obtained by applying the 2-D convolution of a single convolution kernel. The calculation formula of SAM is as follows:

$$\boldsymbol{H}_{\text{SAM}} = \sigma\left(f^{n\times n}\left(\left[\text{AVG}\left(\boldsymbol{F}\right);\text{MAX}\left(\boldsymbol{F}\right)\right]\right)\right) \tag{10}$$

where σ represents the sigmoid function and $f^{n\times n}$ represents the convolution operation with a filter size of $n \times n$.

In general, for a given intermediate feature $\boldsymbol{F} = \mathbb{R}^{H\times W\times L}$, it can be generalized as

$$\boldsymbol{F}' = \boldsymbol{H}_{\text{ECAM}}\left(\boldsymbol{F}\right)\otimes\boldsymbol{F} \tag{11}$$

$$\boldsymbol{F}'' = \boldsymbol{H}_{\text{SAM}}\left(\boldsymbol{F}'\right)\otimes\boldsymbol{F}'. \tag{12}$$

C. Multi-Scale Effective Convolution Block Attention Module

As shown in Fig. 3, inspired by the multi-scale feature fusion in HSI super-resolution reconstruction [46], [47], [48], [49] and classification [50], [51], a multi-scale convolutional block attention module is proposed. By using ECBAM with $n-1$ different convolution kernel sizes, we can capture channel attention and spatial attention features of different sizes, and improve the receptive field of the network, so as to better solve the problem of HU. Then, the features obtained by the previous $n-1$ layer ECBAM and the input intermediate features are fused to retain and extract the initial feature image as much as possible, and enhance the reusability of the features. The fused images are combined by a 2-D convolution with a convolution kernel size of 1, so as to reduce dimensionality. It can not only fuse and filter the different feature information captured by ECBAM, but also improve the generalization ability of the model and avoid the occurrence of overfitting. The activation function ReLU is added after the convolution of the module to enhance the nonlinear expression ability of each module. Its specific representation is as follows:

$$\boldsymbol{F}_n = \text{ReLU}\left(\boldsymbol{W}^l_{1\times1}\times\left[\boldsymbol{F}_{n-1},\boldsymbol{E}_1,\boldsymbol{E}_2,\ldots\boldsymbol{E}_{n-1}\right]+\boldsymbol{b}^l\right) \tag{13}$$

where $[\boldsymbol{F}_{n-1},\boldsymbol{E}_1,\boldsymbol{E}_2,\ldots,\boldsymbol{E}_{n-1}]$ represents the fusion operation of features obtained by $n-1$ layer ECBAM and input intermediate features, variables $\boldsymbol{W}^l_{1\times1}$ and $\boldsymbol{b}^l$ represent the weight tensor and offset tensor of convolution operation in the same layer, respectively, ReLU represents the nonlinear activation function, $\boldsymbol{F}_{n-1}$ and $\boldsymbol{F}_n$ represent the input feature and the output feature of the current module, respectively.

D. Hyperspectral Unmixing With Multi-Scale Convolution Attention Network

VAE can model the spectral variability. Therefore, a HUMSCAN based on two-stream structure (EU-Net and AU-Net) is proposed. The previously proposed MSECBAM and VAE are combined to form EU-Net of HUMSCAN, which solves the problem of endmember spectral variability. The S-VCA pretraining depends on the endmember obtained by superpixel and VCA, which is combined with the EU-Net for the generation of endmembers. The AU-Net for abundance estimation is also based on the MSECBAM. In this article, the inference model is modeled by stochastic gradient variational Bayesian [40], which gives a deterministic approximation of the posterior distribution and is scalable for large datasets. That is, a neural network $\boldsymbol{q}_{\boldsymbol{\phi}_z}$ parameterized by $\boldsymbol{\phi}_z$ is used as an approximation of the true posterior distribution of the potential vector $\boldsymbol{z}_n$ under the condition of pixel $\boldsymbol{y}_n$. The posterior approximate Gaussian distribution is

$$\boldsymbol{q}_{\boldsymbol{\phi}_z}\left(\boldsymbol{z}_n \mid \boldsymbol{y}_n\right) = \mathcal{N}\left(\boldsymbol{\mu}_z\left(\boldsymbol{y}_n\right),\boldsymbol{\Sigma}_z\left(\boldsymbol{y}_n\right)\right). \tag{14}$$

Suppose the covariance matrix is a diagonal matrix, $\boldsymbol{\Sigma}_z = \text{diag}(\boldsymbol{\sigma}^2_z)$.

The decoder of the EU-Net (the generation model) maps latent variables to endmembers. According to LMM [52], considering the different endmembers in each pixel, we have

$$\boldsymbol{y}_n \mid (\boldsymbol{M}_n,\boldsymbol{a}_n) \sim \mathcal{N}\left(\boldsymbol{M}_n\boldsymbol{a}_n,\boldsymbol{\Lambda}\right) \tag{15}$$

where $n = 1,2,\ldots,N$ and $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_0\boldsymbol{I})^{-1}$ are the covariance matrices of the observed noise. Since we introduce a potential vector $\boldsymbol{z}_n$ through the encoder, it is a potential representation of the endmembers of the pixel and encodes their variability. For each pixel, we have

$$\text{vec}\left(\boldsymbol{M}_n\right) = f_\theta\left(\boldsymbol{z}_n\right) \tag{16}$$

where $\text{vec}(\boldsymbol{M}_n) = [\boldsymbol{m}^T_{1n},\boldsymbol{m}^T_{2n},\ldots,\boldsymbol{m}^T_{Bn}]^T$ and $f_\theta()$ are a nonlinear functions parameterized by θ, which connect the endmember matrix with its representation. The physical interpretation of $\boldsymbol{z}_n$ can be the quantitative value of the environmental conditions of the current pixel position, including the factors that cause the spectral variability such as illumination, topography, and atmospheric effects.

The AU-Net is mainly composed of inference models (encoders). The abundance is estimated by a nonlinear function in (17), as follows:

$$\boldsymbol{a}_n = f_{\boldsymbol{\phi}_a}\left(\boldsymbol{y}_n\right) \tag{17}$$

where $f_{\boldsymbol{\phi}_a}()$ is an $\boldsymbol{\phi}_a$-parameterized nonlinear function.

The softmax layer is used as the output of the AU-Net. With the ASC and ANC constraints, the expression of the softmax layer is as follows:

$$\text{softmax}:\boldsymbol{x}\leftarrow\frac{\exp\left(\boldsymbol{x}\right)}{\sum_i\exp\left(\boldsymbol{x}\right)}. \tag{18}$$

TABLE I
NETWORK CONFIGURATION OF HUMSCAN

EU-Net	
layers	neurons
Conv1-BN1-ReLU-MSECBAM1	32P
Conv2-BN2-ReLU-MSECBAM2	16P
Conv3-BN3-ReLU-MSECBAM3	4P
Conv4	J*
Conv5	J
Conv6-BN6-ReLU	16P
Conv7-BN7-ReLU	32P
Conv8-Sigmoid	LP
AU-Net	
Conv9-BN9-ReLU-MSECBAM4	32P
Conv10-BN10-ReLU-MSECBAM5	16P
Conv11-BN11-ReLU-MSECBAM6	4P
Conv12-BN12-ReLU-MSECBAM7	4P
Conv13-Softmax	P

* J represents number of dimensions of latent variable.
Conv represents the convolution layer.
BN represents the batch normalization layer.
ReLU represents the nonlinear activation function.

To normalize endmembers, the sigmoid function is added to the output layer of the EU-Net, and its expression is

$$\text{sigmoid} : \boldsymbol{x} \leftarrow \frac{1}{1+\exp(-\boldsymbol{x})}. \tag{19}$$

In short, the modeling process of endmember spectral variability is to infer the potential variables and abundances through (14) and (17), and then the final endmember can be derived from (16) based on the potential variables. Then, the parameters are optimized by (15) and the observation pixels are reconstructed. The network configuration is shown in Table I.

E. Loss Function and Training Process

Recent studies [39], [41], [42], [53] have shown that the use of VAE can better infer endmembers and abundances from mixed materials. HUMSCAN is optimized by using the loss function of VAE, which consists of reconstruction error and Kullback–Leibler (KL) divergence between variational distribution and Gaussian prior.

$$\begin{aligned} L_{\text{VAE}} &= L_{\text{rec}} + \lambda_{\text{KL}} L_{\text{KL}} \\ &= \frac{1}{N}\sum_{n=1}^{N}\frac{1}{N_n}\|\widehat{\boldsymbol{y}}_{\boldsymbol{n}} - \boldsymbol{y}_{\boldsymbol{n}}\|^2 \\ &\quad + \lambda_{\text{KL}}\frac{1}{N}\sum_{n=1}^{N} D_{\text{KL}}\left(\boldsymbol{q}_{\boldsymbol{\phi}_z}(\boldsymbol{z}_n \mid \boldsymbol{y}_n)\,\|\boldsymbol{p}(\boldsymbol{z})\right) \end{aligned} \tag{20}$$

where $\boldsymbol{q}_{\boldsymbol{\phi}_z}(\boldsymbol{z}_n \mid \boldsymbol{y}_n)$ is the variational distribution of VAE.

In addition, some expected properties can be introduced into HUMSCAN as regularization terms. First, the end elements of each material are expected to exhibit a similar shape, which is the main effect of the overall strength change. Therefore, we constrain the spectral angular distance (SAD) between the endmember and its mean, which can be expressed as

$$L_{\text{SAD}} = \frac{1}{NP}\sum_{j=1}^{P}\sum_{n=1}^{N}\arccos\frac{\overline{\boldsymbol{m}}_j^{\text{T}}\boldsymbol{m}_{nj}}{\|\overline{\boldsymbol{m}}_j\|\,\|\boldsymbol{m}_{nj}\|} \tag{21}$$

where $\bar{\boldsymbol{m}}_j = (1/N)\sum_{n=1}^{N}\boldsymbol{m}_{nj}$.

Second, the minimum volume constraint [54] applied to endmembers has shown the prospect of spectral unmixing. The endmember variance [55] is added as a simple regularization, which is constructed by

$$L_{\text{Vol}} = \frac{1}{NLP}\sum_{n=1}^{N}\sum_{j=1}^{P}\left\|\boldsymbol{m}_{nj} - \frac{1}{P}\sum_{j=1}^{P}\boldsymbol{m}_{nj}\right\|^2. \tag{22}$$

Finally, the objective function of HUMSCAN is

$$L(\boldsymbol{\theta}, \boldsymbol{\phi}_{\boldsymbol{a}}, \boldsymbol{\phi}_{\boldsymbol{z}}) = L_{\text{rec}} + \lambda_{\text{KL}}L_{\text{KL}} + \lambda_{\text{SAD}}L_{\text{SAD}} + \lambda_{\text{Vol}}L_{\text{Vol}} \tag{23}$$

where $\lambda_{\text{KL}}, \lambda_{\text{SAD}}$, and λ_{Vol} are hyperparameters.

In the training process, the pretraining stage is added to accelerate the training process and stabilize the unmixing results. Specifically, the endmember bundle extracted by superpixel segmentation and VCA is combined with EU-Net to constrain the generation of endmembers. The objective function before training are as follows:

$$L_{\text{Pre}}(\boldsymbol{\theta}, \boldsymbol{\phi}_{\boldsymbol{a}}, \boldsymbol{\phi}_{\boldsymbol{z}}) = L_{\text{rec}} + \lambda_{\text{KL}}L_{\text{KL}} + \lambda_{\text{S}-\text{VCA}}L_{\text{S-VCA}} \tag{24}$$

where $\lambda_{\text{S-VCA}}$ is hyperparameter.

Due to the small number of pure pixels in the actual HSI, there is a large error between the endmembers captured by VCA and the actual endmembers, which will affect the training process. Therefore, the HSI is segmented into superpixels by SLIC [56], and then the pseudopure endmembers are extracted by VCA, which can enhance the unmixing effect of the network

$$\text{SLIC}_{\text{feature}} = \sqrt{\frac{\Delta x_{ij}^2 + \Delta y_{ij}^2}{S^2} + \frac{\|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2}{m^2}} \tag{25}$$

where $\Delta x_{ij}^2 + \Delta y_{ij}^2$ is the square of the Euclidean distance between two pixels, S is the search size of SLIC, and m is a hyperparameter, which balances the influence of pixel distance and spectral similarity

$$L_{\text{S-VCA}} = \frac{1}{N}\sum_{n=1}^{N}\|\boldsymbol{M}_n - \boldsymbol{M}_{\text{S-VCA}}\|_F^2. \tag{26}$$

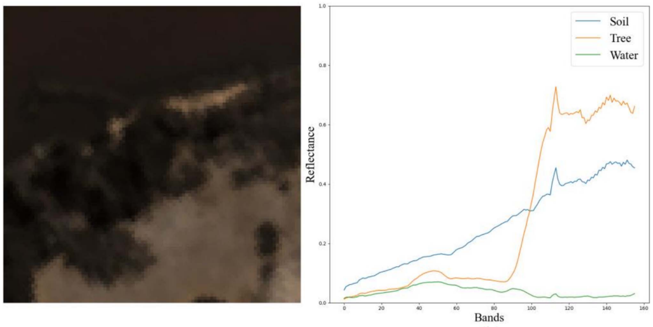
In summary, the learning strategy of HUMSCAN is shown in Algorithm 1.

III. EXPERIMENTS RESULTS

This section is devoted to illustrate the capabilities of the presented algorithm in different scenarios of HU. Several advanced HU algorithms are used, including traditional methods, such as FCLSU + VCA [11], [24], PLMM [16]. Neural network based models, such as DeepTrans-HSU [37], DeepGUn [39], PGMSU [41], DGMCNN [42]. Among them, the traditional method and DeepGUn are run on the MATLAB R2021a platform, and other deep learning-based algorithms are run on

Algorithm 1: HUMSCAN: Learning Strategy.

```
Input:HSI:Y $= \{\boldsymbol{y}_n, n = 1, 2, \ldots, N\}$
  Hyperparameters: $\lambda_{KL}, \lambda_{SAD}, \lambda_{Vol}, \lambda_{S-VCA}$
  Epochs: MaxIter
  Learning rate: LR
  Initialize the model parameters: $\boldsymbol{\Theta} = \{\boldsymbol{\theta}, \boldsymbol{\phi_a}, \boldsymbol{\phi_z}\}$
  for i=1,...,MaxIter do
  Training stage:
    Forward propagation:
    Encoder: $\boldsymbol{q}_{\boldsymbol{\phi}_z}(\boldsymbol{z}_n \mid \boldsymbol{y}_n), \boldsymbol{a}_n = f_{\boldsymbol{\phi}_a}(\boldsymbol{y}_n)$
    Decoder: $\mathrm{vec}(\boldsymbol{M}_n) = f_{\boldsymbol{\theta}}(\boldsymbol{z}_n)$
             $\boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}_n | \boldsymbol{M}_n, \boldsymbol{a}_n)$
    Calculate loss $L(\boldsymbol{\Theta}), L_{pre}(\boldsymbol{\Theta})$
  Back propagation:
    Calaulate gradient $\frac{\partial \boldsymbol{L}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}} or \frac{\partial \boldsymbol{L_{pre}}(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}}$
    Using the Adam optimizer update $\boldsymbol{\Theta}$
  End
Output:Abundance maps and Endmembers estimation:
 $\boldsymbol{a}_n, \boldsymbol{m}_n$
```



(a) (b)

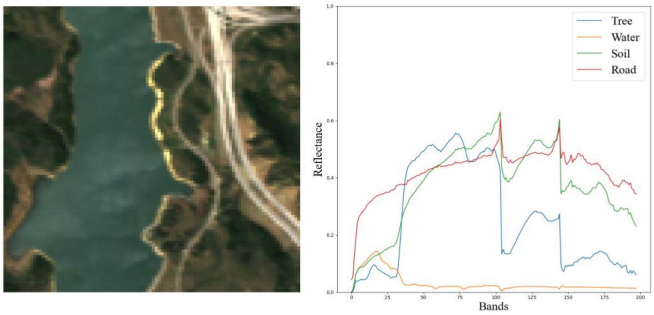
Fig. 4. Samson dataset. (a) False-color image. (b) Endmember spectrum.

Pycharm using the Pytorch module. All comparative experiments are run on the RTX3050 TI Laptop GPU.

A. Hyperspectral Image Data

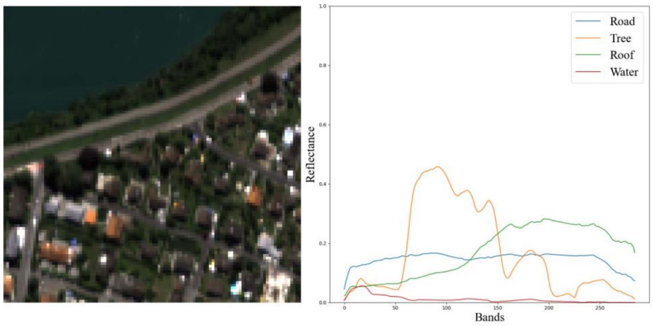
1) Samson dataset:[1] The original image [see Fig. 4(a)], with a size of 95×95, has 156 bands. There are three main substances, including soil, trees, water. There are many pure pixels in the image. The endmember spectral signatures are shown in Fig. 4(b).

2) Jasper Ridge dataset:[2] The original image size is 512×614, with 224 bands. After low signal-to-noise ratio bands removed, The subimage in Fig. 5(a) with 100×100 pixels and 198 bands were used in the experiments, containing vegetation, water, soil, and road. The endmember spectral signatures are shown in Fig. 5(b).



(a) (b)

Fig. 5. Jasper Ridge dataset. (a) False-color image. (b) Endmember spectrum.



(a) (b)

Fig. 6. Apex dataset. (a) False-color image. (b) Endmember spectrum.

3) Apex dataset:[3] The original image size is 111×122, with 285 bands. After the low signal-to-noise ratio bands removed, the subimage in Fig. 6(a) with 110×110 pixels and 285 bands were used in the experiments. Four endmembers were manually selected from the HSI, including water, tree, road, and roof. The endmember spectral signatures are shown in Fig. 6(b).

B. Experiment Setup

1) Hyperparameter Settings: The comparative experimental model uses the optimal parameter settings given in the reference. The structure of the HUMSCAN model is determined by experience, such as the number of layers to the neural network and the number of neurons. During the experiment, we use the Adam optimizer to update the network parameters, and the learning rate of the dataset is set to 5e-4. For the Samson and Jasper Ridge dataset epoch is 1000, the epoch of Apex dataset is 2000, and the n of MSECBAM is set to 4. In the unmixing model based on deep learning, the results usually depend largely on the hyperparameter settings. Choosing the appropriate value for the hyperparameters can significantly improve the results. The following will be discussed separately according to the hyperparameters of the three types of datasets. Fig. 7 shows the corresponding unmixing results of the proposed model under different hyperparameters. The red curve represents the average

[1] [Online]. Available: https://github.com/ma264/HUMSCAN/blob/main/samson_dataset.mat

[2] [Online]. Available: https://github.com/ma264/HUMSCAN/blob/main/jasper_dataset.mat

[3] [Online]. Available: https://github.com/ma264/HUMSCAN/blob/main/apex_dataset.mat

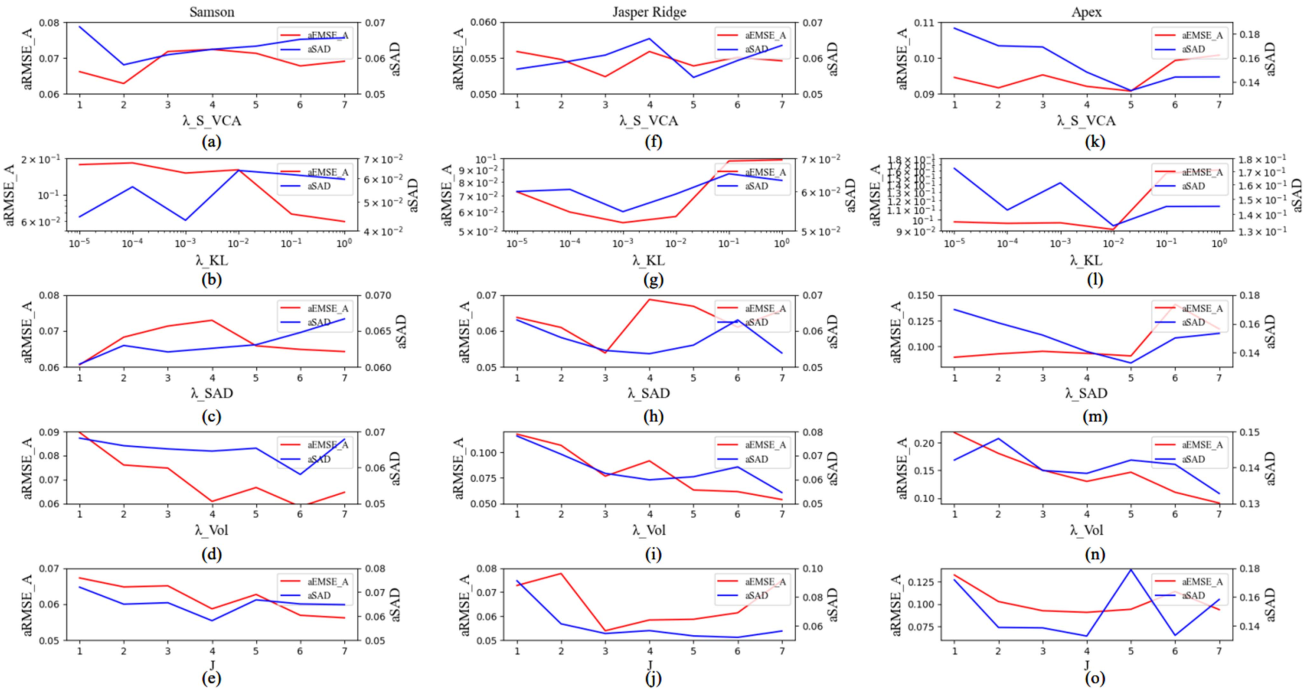


Fig. 7. HUMSCAN uses different hyperparameters and J for quality evaluation index diagram. (a)–(e) Samson dataset. (f)–(j) Jasper Ridge dataset. (k)–(o) Apex dataset.

RMSE of abundance ($a\mathrm{RMSE}_A$), and the blue curve represents the average spectral angle distance (SAD) of the endmember ($a\mathrm{SAD}_M$). The smaller the value of SAD is, the better the result of unmixing is. The following will be discussed further:

1) Samson Dataset: The optimal regularization parameters $\lambda_{S-\mathrm{VCA}}$, λ_{KL}, λ_{SAD}, λ_{Vol} in Fig. 7(a)–(d) should be set as 2, 1, 1, 6, respectively. The latent vector dimension J be set to 4 in Fig. 7(e).
2) Jasper Ridge Dataset: The optimal regularization parameters $\lambda_{\text{S-VCA}}$, λ_{KL}, λ_{SAD}, λ_{Vol} in Fig. 7(f)–(i) should be set as 5, 0.001, 3, 7, respectively. The latent vector dimension J be set to 3 in Fig. 7(j).
3) Apex Dataset: The optimal regularization parameters $\lambda_{\text{S-VCA}}$, λ_{KL}, λ_{SAD}, λ_{Vol} in Fig. 7(k)–(n) should be set as 5, 0.01, 5, 7, respectively. The latent vector dimension J be set to 4 in Fig. 7(o).

So, the hyperparameter λ_{KL} of the HUMSCAN is important. Because it directly affects the prior assumption of the unmixing model on the mixing ratio and the degree of regularization of the unmixing result. With the improper hyperparameters, stability and convergence affect the quality of the unmixing effect.

2) Image Segmentation: Because there are few or even no pure pixels in the real HSI, the endmembers obtained by VCA are not accurate enough, resulting in poor unmixing performance. Therefore, the endmember extraction framework composed of the endmember bundle obtained by superpixel segmentation and VCA and the EU-Net can adaptively extract endmembers at the pixel and subpixel levels [57], [58]. Superpixel segmentation of HSI is performed by using SLIC method. The results of superpixel segmentation with three datasets are shown in Fig. 8(a)–(c). The SAD of S-VCA and VCA in Table II represents the similarity between the endmember spectrum obtained after superpixel segmentation and the real endmember spectrum. It can be found that the overall accuracy of SAD with superpixel segmentation to all datasets has been improved, especially the SAD in roads and waters with the Apex dataset.

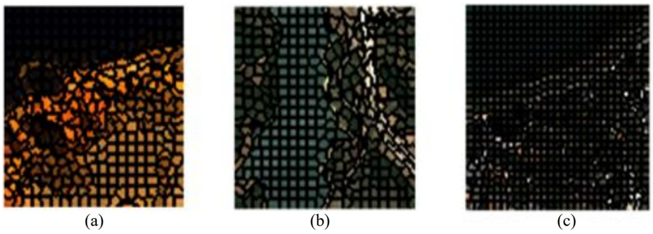

(a) (b) (c)

Fig. 8. (a) The Samson dataset is segmented into 340 superpixels. (b) The Jasper Ridge dataset is segmented into 374 superpixels. (c) The Apex dataset is segmented into 722 superpixels.

TABLE II
SAD OF ENDMEMBERS COMPARISON BETWEEN THE PSEUDOPURE ENDMEMBER OBTAINED BY VCA AFTER SUPERPIXEL SEGMENTATION AND THE ENDMEMBER OBTAINED BY VCA

		Soil	Tree	Water	
Samson	VCA	0.0259	0.0957	0.1554	
	S-VCA	0.0141	0.0775	0.1226	
		Vegetation	Water	Soil	Road
Jasper Ridge	VCA	0.1481	0.0892	0.1166	0.0901
	S-VCA	0.0492	0.1531	0.0179	0.0619
		Road	Tree	Roof	Water
Apex	VCA	0.6914	0.2490	0.1251	0.5176
	S-VCA	0.0703	0.1398	0.1626	0.1783

TABLE III
RMSE OF ABUNDANCE(SAMSON DATASET)

	Soil	Tree	Water	Overall
FCLSU+VCA [11], [24]	0.1757	0.0772	0.1587	0.1372
DeepTrans [37]	0.0843	0.0554	0.0654	0.0683
PLMM [16]	0.1739	0.0794	0.1552	0.1361
DeepGUn [39]	0.0852	0.0577	0.0489*	0.0639
PGMSU [41]	0.1240	0.0977	0.0756	0.0991
DGMCNN [42]	0.0784	0.0519	0.0565	0.0622
HUMSCAN	0.0639*	0.0493*	0.0562	0.0565*

* is the best performance result.

TABLE IV
SAD OF ENDMEMBERS(SAMSON DATASET)

	Soil	Tree	Water	Overall
FCLSU+VCA [11], [24]	0.0259	0.0957	0.1554	0.0923
DeepTrans [37]	0.0276	0.0810	0.1181	0.0755
PLMM [16]	0.0268	0.0965	0.1647	0.0960
DeepGUn [39]	0.0755	0.0745	0.0955	0.0818
PGMSU [41]	0.0428	0.0964	0.3587	0.1659
DGMCNN [42]	0.0772	0.1044	0.1585	0.1133
HUMSCAN	0.0234*	0.0710*	0.0948*	0.0630*

* is the best performance result.

3) Quantitative Performance Measures: In experiments, the root mean square error (RMSE) of abundance corresponding to different endmembers and the overall mean are often used to evaluate the unmixing performance of the comparison model. Defined as

$$\mathrm{RMSE}_{Ap} = \frac{1}{\mathrm{N}}\sqrt{\sum_{i=1}^{N}\left\|\boldsymbol{a}_{pi} - \hat{\boldsymbol{a}}_{pi}\right\|^2} \tag{27}$$

$$a\mathrm{RMSE}_{A} = \frac{1}{P}\sum_{p=1}^{P}\mathrm{RMSE}_{Ap} \tag{28}$$

where $\boldsymbol{a}_{pn}$ and $\hat{\boldsymbol{a}}_{pn}$ are estimated abundance and actual abundance, respectively.

The SAD of each endmember and the overall mean are introduced to compare the similarity between the ground truth (GT) endmember and the estimated endmember, which is defined as

$$\mathrm{SAD}_{Mp} = \frac{1}{N}\sum_{i=1}^{N}\arccos\left(\frac{\boldsymbol{m}_{ip}^{\mathrm{T}}\widehat{\boldsymbol{m}}_{ip}}{\|\boldsymbol{m}_{ip}\|\,\|\widehat{\boldsymbol{m}}_{ip}\|}\right) \tag{29}$$

$$a\mathrm{SAD}_{M} = \frac{1}{P}\sum_{p=1}^{P}\mathrm{SAD}_{Mp} \tag{30}$$

where $\boldsymbol{m}_{ip}$ and $\hat{\boldsymbol{m}}_{ip}$ represent the extracted endmember and the real endmember, respectively.

C. Experiment With Samson Dataset

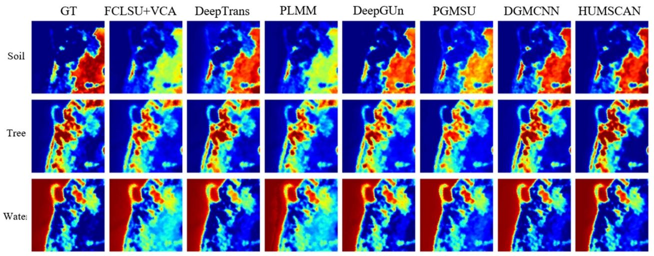


Fig. 9. Samson dataset—Visual comparison of the abundance maps obtained by the different unmixing techniques.

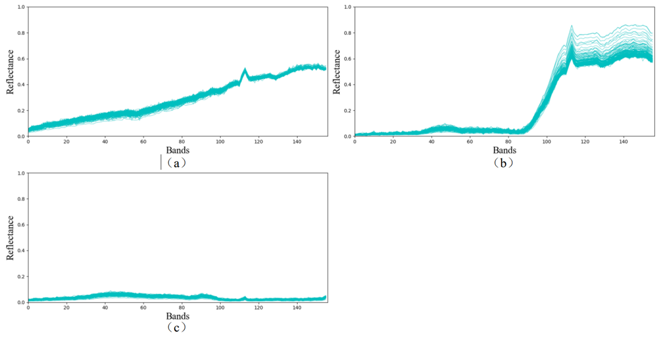


Fig. 10. Endmembers estimated by HUMSCAN for the Samson dataset. (a) Soil. (b) Tree. (c) Water.

TABLE V
RMSE OF ABUNDANCE(JASPER RIDGE DATASET)

	Vegetation	Water	Soil	Road	Overall
FCLSU+VCA [11], [24]	0.1588	0.2044	0.1301	0.1143	0.1519
DeepTrans [37]	0.0793	0.1118	0.1110	0.1350	0.1092
PLMM [16]	0.1570	0.2022	0.1349	0.1127	0.1517
DeepGUn [39]	0.0844	0.1078	0.0929	0.0808	0.0914
PGMSU [41]	0.0927	0.0872	0.1425	0.1013	0.1059
DGMCNN [42]	0.1136	0.1187	0.0986	0.1019	0.1082
HUMSCAN	0.0577*	0.0437*	0.0793*	0.0699*	0.0626*

* is the best performance result.

The quantitative performance comparison in terms of RMSE of abundance and SAD of endmembers on the Samson dataset are shown in Tables III and IV. Experiments show that the HUMSCAN model is superior to other techniques in terms of abundance RMSE and endmember SAD in most of the classes. Compared with the suboptimal results, the overall abundance RMSE and endmember SAD of the HUMSCAN were increased by 9.16% and 16.55%, respectively. That is because the effective spatial-spectral features are captured and fused into multiple attention feature with HUMSCAN, which significantly enhanced the accuracy of spectral unmixing. The corresponding abundance maps with various methods are shown in Fig. 9. Among these methods, the results with HUMSCAN are the most similar to the GT. In Fig. 10, the spectral variability of Tree endmember is more obvious than those of soil and water.

D. Experiment With Jasper Ridge Dataset

The quantitative performance comparison in terms of RMSE of abundance and SAD of endmembers on the Jasper Ridge dataset are shown in Tables V and VI. The endmembers with VCA are used for endmember initialization and constraint among these methods. The proposed HUMSCAN model demonstrated a good performance in RMSE of abundance and SAD of endmember, which had increased by 31.51% and 29.23%, respectively. That is because the HUMSCAN model does not only consider the combination of spatial information and spectral information, but also uses superpixel segmentation and VCA

TABLE VI
SAD OF ENDMEMBERS(JASPER RIDGE DATASET)

	Vegetation	Water	Soil	Road	Overall
FCLSU+VCA [11], [24]	0.1481	0.0892	0.1166	0.0901	0.1110
DeepTrans [37]	0.0604	0.0845	0.1438	0.0962	0.0962
PLMM [16]	0.1490	0.1145	0.1189	0.0865	0.1172
DeepGUn [39]	0.0840	0.1471	0.0721	0.0872	0.0976
PGMSU [41]	0.0348*	0.1330	0.0883	0.0559	0.0780
DGMCNN [42]	0.0973	0.1679	0.0762	0.0476	0.0972
HUMSCAN	0.0363	0.0807*	0.0589*	0.0452*	0.0552*

* is the best performance result.

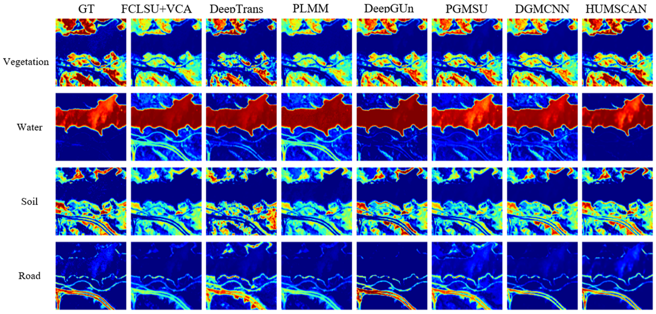
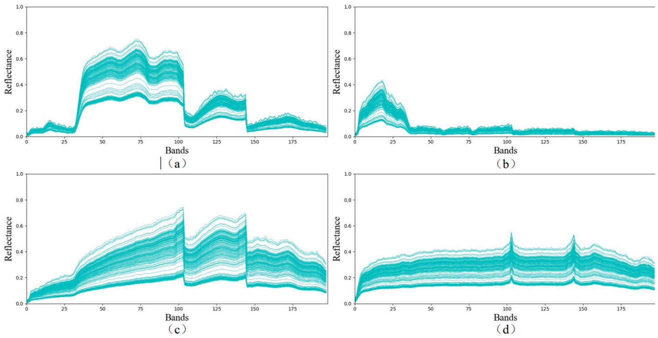


Fig. 11. Jasper Ridge dataset—Visual comparison of the abundance maps obtained by the different unmixing techniques.



（a） （b） （c） （d）

Fig. 12. Endmembers estimated by HUMSCAN for the Jasper Ridge dataset. (a) Vegetation. (b) Water. (c) Soil. (d) Road.

to extract pseudopure endmembers for endmember constraints, which help the results of endmember extraction and abundance estimation significantly improved. However, there are fewer pure pixels in the Jasper Ridge dataset, the accuracy of spectral unmixing with VCA initialization is low. But the methods based on deep learning had strong learning and data fitting capabilities, and can improve the unmixing accuracy. The corresponding abundance maps with various comparative methods in Fig. 11 reveal that Among these methods, the results with HUMSCAN are the most similar to the GT. In Fig. 12, the endmembers' spectrum with HUMSCAN are similar to the real endmember with spectral variabilities. The results in both Tables and Fig. 12 reveal that, the spectral variability are in the spectral intensity and the shape. And the deep learning unmixing model considering spectral variability is better than the method without spectral variability.

TABLE VII
RMSE OF ABUNDANCE(APEX DATASET)

	Road	Tree	Roof	Water	Overall
FCLSU+VCA [11], [24]	0.2057	0.1451	0.1365	0.1336	0.1552
DeepTrans [37]	0.1776	0.0972	0.1318	0.0943	0.1252
PLMM [16]	0.2733	0.0727	0.1306	0.1858	0.1656
DeepGUn [39]	0.2494	0.0946	0.1119	0.1249	0.1452
PGMSU [41]	0.1567	0.1436	0.1087	0.1649	0.1434
DGMCNN [42]	0.1692	0.1229	0.1624	0.2531	0.1769
HUMSCAN	0.1452*	0.0905*	0.0900*	0.0454*	0.0927*

* is the best performance result.

TABLE VIII
SAD OF ENDMEMBERS(APEX DATASET)

	Road	Tree	Roof	Water	Overall
FCLSU+VCA [11], [24]	0.6914	0.2490	0.1251	0.5176	0.3957
DeepTrans [37]	0.1176*	0.1311*	0.1166	0.0433*	0.1171*
PLMM [16]	0.3274	0.1976	0.1528	0.3508	0.2571
DeepGUn [39]	0.5441	0.1236	0.1486	0.2492	0.2663
PGMSU [41]	0.1369	0.1365	0.1033	0.5755	0.2380
DGMCNN [42]	0.1550	0.1727	0.1943	0.3050	0.2068
HUMSCAN	0.1413	0.1430	0.0876*	0.1524	0.1310

* is the best performance result.

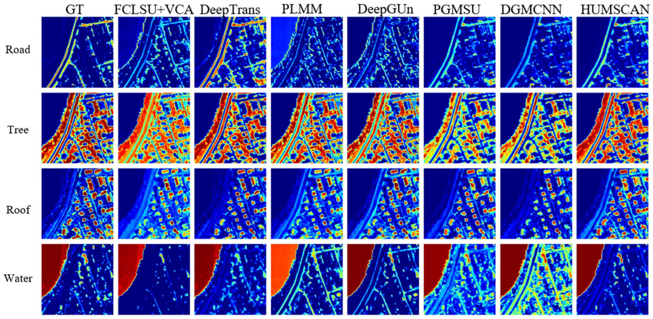


Fig. 13. Apex dataset—Visual comparison of the abundance maps obtained by the different unmixing techniques.

E. Experiment With Apex Dataset

The quantitative results on the Apex dataset are shown in Tables VII and VIII. Experiments show that there is a large SAD in the endmember road and water extracted by FCLSU + VCA, and those method with VCA initialization. The larger SAD is, the worse the result of spectral unmixing is. So, deep learning methods with VCA for endmember initialization have worse results. However, the proposed network model can obtain the optimal on the RMSE of abundance due to considering spectral variability. Because the spectral variability are not only the variablity of spectral intensity, but also the variability of spectral shape, which affects the performance of SAD. However, since the RMSE of abundance mainly considers the relative contribution degree, even the spectral variability has a relatively small impact on it. DeepTrans-HSU is without spectral variability, so its endmember SAD is not affected by spectral variability. However, DeepTrans-HSU did not consider spectral variability, even if the optimal endmember SAD value was obtained, the RMSE of abundance is still not optimal. Compared with the unmixing method with spectral variability, the HUMSCAN model demonstrates advantages superiorities in the SAD of endmembers and the RMSE of abundance. As shown in Fig. 13, the abundance

TABLE IX
ABLATION EXPERIMENT OF HUMSCAN, THE QUALITY EVALUATION INDEX OF DATASETS SAMSON, JASPER RIDGE AND APEX

	Samson		Jasper Ridge		Apex	
	$aRMSE_A$	$aSAD_M$	$aRMSE_A$	$aSAD_M$	$aRMSE_A$	$aSAD_M$
CBAM(no S-VCA)	0.0707	0.1025	0.1186	0.0881	0.1635	0.2287
ECBAM(no S-VCA)	0.0630	0.0850	0.1013	0.0799	0.1584	0.2162
MSECBAM(no S-VCA)	0.0620	0.0640	0.0915	0.0705	0.1466	0.1976
MSECBAM(no ECAM)	0.0680	0.0633	0.0795	0.0671	0.1155	0.1527
MSECBAM(no SAM)	0.0675	0.0632	0.0738	0.0622	0.1040	0.1521
MSECBAM(ECAM kernel=1,SAM kernel={3,5,7})	0.0618	0.0630	0.0660	0.0577	0.0987	0.1390
MSECBAM(ECAM kernel=3,SAM kernel={3,5,7})	0.0579	0.0631	0.0676	0.0555	0.0932	0.1333
MSECBAM(ECAM kernel={3,5,7},SAM kernel=1)	0.0599	0.0631	0.0659	0.0582	0.0927*	0.1310*
MSECBAM(ECAM kernel={3,5,7},SAM kernel=3)	0.0565*	0.0630*	0.0687	0.0557	0.0940	0.1360
MSECBAM(ECAM kernel={3,5,7},SAM kernel={3,5,7})	0.0577	0.0633	0.0626*	0.0552*	0.1007	0.1336

* is the best performance result.

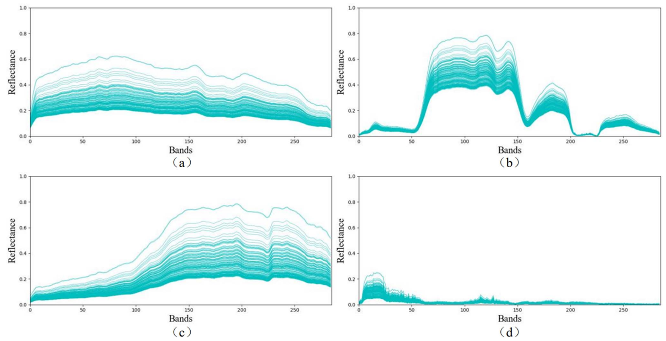


Fig. 14. Endmembers estimated by HUMSCAN for the Apex dataset. (a) Road. (b) Tree. (c) Roof. (d) Water.

map obtained by the HUMSCAN model is closer to the real feature abundance map. Fig. 14 is the endmember estimation of Apex dataset. It can be seen that the estimated endmembers are similar to the real endmembers, and demonstrate obvious spectral variability in both the spectral intensity and the spectral shape.

F. Ablation Experiment

Through comparative analysis of ablation experiments, the advantages of endmember bundle constraints on EU-Net and the importance of MSECBAM internal modules are verified, and the unmixing effects under different scale combinations are studied. In order to ensure the reliability of the experimental results, all experimental parameters are set to the same value. According to the quantitative results given in Table IX, after combining EU-Net with S-VCA pretraining, the unmixing accuracy is insignificantly improved for dataset (such as Samson dataset) with a large number of pure pixels. However, for Jasper Ridge and Apex datasets, the unmixing effect is significantly improved, especially on the Apex dataset. This is due to the large structural similarity difference between the endmembers generated by VCA and the real endmembers in Apex dataset. As a result, the error between the endmember spectrum generated by VCA preconstrained EU-Net and the actual spectrum is large, which makes the unmixing accuracy low. After adding the proposed MSECBAM to the Samson and Jasper Ridge datasets, the unmixing effect is still significantly improved compared with the first two modules, but the improvement for the Apex dataset is limited. However, endmembers generated by using superpixel segmentation and endmember bundle constraints obtained by VCA can solve this problem better.

In order to verify the importance of the internal module of MSECBAM, ECAM (spectral attention feature) and SAM (spatial attention feature) are removed, respectively. The experimental results show that the spectral information and spatial information of HSI are important in HU. By modifying the kernel function parameter k of ECAM and SAM, the experimental results of Table IX prove the influence of MSECBAM with different scales on the model unmixing efficiency. Therefore, through ablation experiments, not only the optimal MSECBAM parameter settings are selected for different dataset, but also the effectiveness of the proposed MSECBAM and the important role of the constraint strategy based on S-VCA pretraining in optimizing the HU of nonpure pixel HSI data are verified.

IV. CONCLUSION

In this article, a new convolution attention network for spectral unmixing with endmember variability (referred to as HUMSCAN) was proposed. The EU-Net was composed of the VAE and the MSECBAM, which was combined with the S-VCA pretraining to adaptively extract endmembers at the pixel and subpixel levels in complex scenes. The AU-Net was based on the MSECBAM frame jointed the spectral and spatial attention features. The proposed HUMSCAN method can simultaneously and unsupervisedly extracts endmembers and their corresponding abundances, which can improve the accuracy and efficiency of spectral unmixing. The experimental results showed that HUMSCAN considering endmember variability can outperform the competing methods. Future work will include developing more expressive network models and extending to nonlinear unmixing techniques with endmember variability.

REFERENCES

[1] M. Amani et al., “Google Earth engine cloud computing platform for remote sensing Big Data applications: A comprehensive review,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5326–5350, 2020.

[2] M. Liu, Z. Chai, H. Deng, and R. Liu, “A CNN-Transformer network with multiscale context aggregation for fine-grained cropland change detection,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4297–4306, May 2022, doi: 10.1109/JSTARS.2022.3177235.

[3] R. Chen et al., “Monitoring rainfall events in desert areas using the spectral response of biological soil crusts to hydration: Evidence from the Gurbantunggut desert, China,” *Remote Sens. Environ.*, vol. 286, no. 113448, pp. 1–19, 2023.

[4] H. Chen et al., “Stacked spectral feature space patch: An advanced spectral representation for precise crop classification based on convolutional neural network,” *Crop J.*, vol. 10, no. 5, pp. 1460–1469, 2022.

[5] N. Farmonov et al., “Crop type classification by DESIS hyperspectral imagery and machine learning algorithms,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1576–1588, Jan. 2023, doi: 10.1109/JSTARS.2023.3239756.

[6] A. Smith, B. Johnson, and C. Brown, “Hyperspectral imaging of microscopic medical preparations,” in *Proc. IEEE Int. Conf. Biomed. Imag.*, 2016, pp. 123–127.

[7] M. Shimoni, R. Haelterman, and C. Perneel, “Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques,” *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[8] J. Wang, Y. Li, and X. Zhang, “An algorithm for fully constrained abundance estimation in hyperspectral unmixing,” in *Proc. IEEE Int. Conf. Image Process.*, 2015, pp. 4567–4571.

[9] C. I. Chang and C. C. Wu, “Iterative pixel purity index,” in *Proc. Workshop Hyperspectral Image Signal Process.: Evol. Remote Sens.*, 2012, pp. 1–4.

[10] J. M. P. Nascimento and J. M. B. Dias, “N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 744–755, Mar. 2004.

[11] 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.

[12] 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.

[13] M. Parente and M. D. Iordache, “Sparse unmixing of hyperspectral data: The legacy of SUnSAL,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 21–24.

[14] A. Zare and K. C. Ho, “Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 95–104, Jan. 2014.

[15] A. Halimi, P. Honeine, and J. M. Bioucas-dias, “Hyperspectral unmixing in presence of endmember variability, nonlinearity, or mismodeling effects,” *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4565–4579, Oct. 2016.

[16] P. A. Thouvenin, N. Dobigeon, and J. Y. Tourneret, “Hyperspectral unmixing with spectral variability using a perturbed linear mixing model,” *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.

[17] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, “An augmented linear mixing model to address spectral variability for hyperspectral unmixing,” *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, Apr. 2019.

[18] L. Drumetz, S. Henrot, M. A. Veganzones, J. Chanussot, and C. Jutten, “Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability,” in *Proc. Workshop Hyperspectral Image Signal Processing, Evol. Remote Sens.*, 2015, pp. 1–4.

[19] D. Hong and X. X. Zhu, “SULoRA: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis,” *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1351–1363, Dec. 2018.

[20] Q. You, F. Li, S. Zhang, S. Wang, C. Deng, and C. Xu, “Low-rank subspace unmixing of remotely sensed hyperspectral image,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3849–3852.

[21] B. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, “Hyperspectral unmixing using a neural network autoencoder,” *IEEE Access*, vol. 6, pp. 25646–25656, 2018.

[22] Y. Qu, R. Guo, and H. Qi, “Spectral unmixing through part-based nonnegative constraint denoising autoencoder,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*,2017, pp. 209–212.

[23] Y. Qu and H. Qi, “uDAS: An untied denoising autoencoder with sparsity for spectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 3, pp. 1698–1712, Mar. 2019.

[24] D. C. Heinz and I. C. Chein, “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, Mar. 2001.

[25] 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.

[26] 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, pp. 1–15, Jul. 2021, doi: 10.1109/TGRS.2021.3094884.

[27] A. Min, Z. Guo, H. Li, and J. Peng, “JMnet: Joint metric neural network for hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Apr. 2021, doi: 10.1109/TGRS.2021.3069476.

[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] X. Zhang, W. Huang, Q. Wang, and X. Li, “SSR-NET: Spatial–spectral reconstruction network for hyperspectral and multispectral image fusion,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5953–5965, Jul. 2021.

[30] 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.

[31] M. Zhao, M. Wang, J. Chen, and S. Rahardja, “Hyperspectral unmixing for additive nonlinear models with a 3-D-CNN autoencoder network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Aug. 2021, doi: 10.1109/TGRS.2021.3098745.

[32] 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, pp. 1–14, Mar. 2021, doi: 10.1109/TGRS.2021.3064958.

[33] 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.

[34] C. Liu, J. Yang, Z. Qi, Z. Zou, and Z. Shi, “Progressive scale-aware network for remote sensing image change captioning,” in *Proc. IEEE Int. Geosci. Remote Sens. Sympos.*, Pasadena, CA, USA, 2023, pp. 6668–6671, doi: 10.1109/IGARSS52108.2023.10283451.

[35] C. Liu, R. Zhao, H. Chen, Z. Zou, and Z. Shi, “Remote sensing image change captioning with dual-branch transformers: A new method and a large scale dataset,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–20, Nov. 2022, doi: 10.1109/TGRS.2022.3218921.

[36] J. Chen et al., “DASNet: Dual attentive fully convolutional Siamese networks for change detection in high-resolution satellite images,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1194–1206, Nov. 2020, doi: 10.1109/JSTARS.2020.3037893.

[37] P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, and P. Scheunders, “Hyperspectral unmixing using transformer network,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Aug. 2022, doi: 10.1109/TGRS.2022.3196057.

[38] Y. Zeng, C. Ritz, J. Zhao, and J. Lan, “Attention-based residual network with scattering transform features for hyperspectral unmixing with limited training samples,” *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 400.

[39] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, “Deep Generative Endmember modeling: An application to unsupervised spectral unmixing,” *IEEE Trans. Comput. Imag.*, vol. 6, pp. 374–384, Oct. 2020, doi: 10.1109/TCI.2019.2948726.

[40] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” in *Proc. 2nd Int. Conf. Learn. Representations*, 2014, pp. 1–14.

[41] S. Shi, M. Zhao, L. Zhang, Y. Altmann, and J. Chen, “Probabilistic generative model for hyperspectral unmixing accounting for endmember variability,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Oct. 2021, doi: 10.1109/TGRS.2021.3121799.

[42] S. Shi, L. Zhang, Y. Altmann, and J. Chen, “Deep generative model for spatial–spectral unmixing with multiple endmember priors,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, Apr. 2022, doi: 10.1109/TGRS.2022.3168712.

[43] S. Woo, J. Park, J. Lee, and I. So Kweon, “Cbam: Convolutional block attention module,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[44] X. Wang, R. Girshick, A. Gupta, and K. He, “ECA-Net: Efficient channel attention for deep convolutional neural networks,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1234–1245.

[45] M. Zagoruyko and N. Komodakis, “Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1234–1245.

[46] J. Zhang, Z. Zhao, Y. Zhao, Q. Liu, J. Sun, and L. Zhang, “Multiscale residual network for image super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 1646–1654.

[47] L. Wang, H. Zhang, Y. Li, and Y. Zhang, “Hyperspectral image super-resolution based on multiscale residual block and multilevel feature fusion,” in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 2156–2160.

[48] Y. Sun, J. Qin, X. Gao, S. Chai, and B. Chen, “Attention-enhanced multiscale residual network for single image super-resolution,” *Signal Image Video Process.*, vol. 16, no. 5, pp. 1417–1424, 2022.

[49] D. Mishra and O. Hadar, “Self-FuseNet: Data free unsupervised remote sensing image super-resolution,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1710–1727, Jan. 2023, doi: 10.1109/JSTARS.2023.3239758.

[50] Y. Qing, Q. Huang, L. Feng, Y. Qi, and W. Liu, “MultiScale feature fusion network incorporating 3D self-attention for hyperspectral image classification,” *Remote Sens.*, vol. 14, no. 742, pp. 1–22, Feb. 2022.

[51] S. Liu, C. Wei, Y. Zhang, and X. Zhang, “Dynamic spectral-spatial multiscale feature extraction network for hyperspectral image classification,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4321–4324.

[52] L. Chen, J. Zhang, and Q. Du, “Variational autoencoders for hyperspectral unmixing with endmember variability,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1234–1238.

[53] 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.

[54] L. Miao and H. Qi, “Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.

[55] C. Li, D. Chen, J. Ma, and Y. Zhang, “Minimum distance constrained non-negative matrix factorization for the endmember extraction of hyperspectral images,” in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 5678–5682.

[56] J. Yin, T. Wang, Y. Du, X. Liu, L. Zhou, and J. Yang, “SLIC superpixel segmentation for polarimetric SAR images,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Jan. 2021, doi: 10.1109/TGRS.2020.3047126.

[57] X. Xu, X. Tong, A. Plaza, Y. Zhong, H. Xie, and L. Zhang, “A new spectral-spatial sub-pixel mapping model for remotely sensed hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6763–6778, Nov. 2018.

[58] X. Xu, X. Tong, A. Plaza, Y. Zhong, H. Xie, and L. Zhang, “Joint sparse sub-pixel mapping model with endmember variability for remotely sensed imagery,” *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 15.

Sheng Hu received the B.S. degree in electronic information engineering from the Hunan University, Changsha, China, in 2021. He is currently working toward the master’s degree in electronic information with Hunan University, Changsha, China.

His research interests include deep learning and remote sensing image processing.

Huali Li (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2007, and the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, in 2012.

She is currently an Associate Professor with the College of Electrical and Information Engineering, Hunan University, Changsha, China. Her research interests include pattern recognition, image processing, and remote sensing.