SSF-Net: A Spatial–Spectral Features Integrated Autoencoder Network for Hyperspectral Unmixing

Bin Wang, Huizheng Yao, Dongmei Song, Jie Zhang, and Han Gao, *Member, IEEE*

***Abstract*—In recent years, deep learning has received tremendous attention in the field of hyperspectral unmixing (HU) due to its powerful learning capabilities. Particularly, the unsupervised unmixing method based on an autoencoder (AE) has become a research hotspot. Most of the current AE unmixing networks mainly focus on information about pixels and their neighborhoods in images. However, they make insufficient use of information about spatial heterogeneity and spectral differences of endmembers in hyperspectral image (HSI) data. To this end, an AE HU network with the name of SSF-Net is proposed for fusing the spatial–spectral features. The network first extracts pseudoendmember information from the HSI using a regional vertex component analysis algorithm. Then, a dual-branch feature fusion module incorporating a spatial–spectral attention mechanism is constructed to make full use of the information in the HSI data, thereby improving the network's unmixing performance. It is worth stating that SSF-Net can fuse spatial–spectral information and utilize different attention maps to obtain more significant spectral difference information and more discriminative spatial difference information about the scene. The experimental results on synthetic and real datasets demonstrate that the proposed SSF-Net outperforms state-of-the-art unmixing algorithms.**

***Index Terms*—Attention, autoencoder (AE), deep learning (DL), feature fusion, hyperspectral unmixing (HU).**

I. Introduction

HYPERSPECTRAL image (HSI) can capture detailed spectral information of ground objects in hundreds of continuous bands from visible light to short-wave infrared and even wider spectral intervals while obtaining spatial distribution information of ground objects. Because of its rich spectral information, it has received a great deal of attention, especially in the fields of military investigation, target tracking, target identification, environmental monitoring, etc. [1], [2], [3], [4], [5]. However, due to the limitation of spatial resolution and the complex diversity of the natural land surfaces, the phenomenon of mixed pixels [6] is common in HSI. The mixed pixels are composed of a variety of pure material spectra, and their existence will have an enormous impact on the accuracy of hyperspectral remote sensing applications [7]. To better solve this problem, the hyperspectral unmixing (HU) technique is often used to decompose the mixed pixels into a series of different pure material spectra (endmembers) and the coverage ratio (abundances) of the endmembers [8]. Currently, HU has been widely used in mineral detection [9], [10] and agricultural detection [11], [12].

HU models can simply be classified into two categories: linear mixing model (LMM) [1] and nonlinear mixing model (NLMM) [13]. In this regard, the LMM is based on the assumption that the electromagnetic wave energy received by the sensor does not undergo secondary scattering during transmission, i.e., the spectrum of a mixed pixel is a linear combination of multiple pure spectra (endmembers) of the ground object according to certain proportions (abundances). Moreover, considering the physical mechanism in HU, the abundance needs to satisfy the nonnegative constraint (ANC) and abundance sum-to-one constraint (ASC) [14]. NLMM is often used to describe the intricate interactions between scattered light from multiple materials within a scene [1]. Although the NLMM is more in line with the actual transmission of electromagnetic waves, it requires consideration of numerous complex factors in implementation. Given the explicit physical mechanism and relatively straightforward solving process of the LMM and the relatively simple solution process, the simulation of mixed spectra can be efficiently achieved. Therefore, this study focuses on the LMM-based HU.

Traditional unmixing methods can be mainly categorized into geometric-based, statistical-based, and sparse regression-based unmixing methods. Among the geometry-based unmixing methods, the typical representatives include N-finder (N-FINDR) [15] and vertex component analysis (VCA) [16]. N-FINDR employs the projection of pixels into the feature space to form a simplex, where the endmembers are efficiently selected by identifying the pixel that constitutes the maximum volume simplex. The VCA does this by iteratively projecting the pixels into a direction orthogonal to the subspace formed by the already identified endmembers, where the new endmember corresponds to the extreme of the projection. Due to the complexity and variety of natural surfaces, it is a formidable challenge to identify the pure pixels in remote sensing images. Furthermore, geometry-based unmixing methods tend to fall into local optima in HSI with highly mixed ground objects. In contrast, statistical-based unmixing methods

Manuscript received 13 July 2023; revised 19 September 2023; accepted 17 October 2023. Date of publication 25 October 2023; date of current version 22 December 2023. This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR2022MD015, in part by the Key Program of Joint Fund of the National Natural Science Foundation of China and Shandong Province under Grant U1906217 and Grant U22A20586, in part by the National Natural Science Foundation of China under Grant 41701513, Grant 61371189, and Grant 41772350, and in part by the Key Research and Development Program of Shandong Province under Grant 2019GGX10133. *(Corresponding author: Dongmei Song.)*

The authors are with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao 266580, China (e-mail: wangbin007@upc.edu.cn; z21160110@s.upc.edu.cn; songdongmei@upc.edu.cn; zhangjie@upc.edu.cn; gaohangeo@upc.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3327549

are able to obtain the global optima, such as the unmixing methods with a Bayesian framework [17] and nonnegative matrix decomposition (NMF) [18]. Bayesian methods can effectively incorporate *a priori* information into the unmixing process, thus improving the accuracy of unmixing [19], [20]. Due to the advantages of learning part-based representations, NMF has become a prominent research focus in the field of HU. Notably, NMF can simultaneously capture the endmembers and abundances of HSI after completing the unmixing task [21]. Currently, many NMF-based HU methods have been proposed, mainly focusing on improving the unmixing accuracy through the incorporation of regularization constraints or the integration of spectral and spatial information [22], [23]. Besides, there is also unmixing by nonnegative tensor factorization [24], [25] to minimize the information loss in the unmixing process. Furthermore, there are model-inspired network-based unmixing approaches [26], [27], [28] that make the unmixing process more physically interpretable by combining it with a physical model. Finally, the sparse regression-based unmixing method [29] effectively estimates the endmembers and their corresponding abundances in HSI using *a priori* knowledge of a known spectral library. Although the sparse regression-based unmixing methods can mitigate the adverse effects of inaccurate endmember extraction, they cannot be widely applied to HSI images acquired in complex environments due to the poor mobility of the spectral library.

Deep learning has attracted much attention in the field of HU owing to its powerful feature representation ability. In particular, the autoencoder (AE) and its variants have been applied to HU with excellent unmixing results. Up to now, the self-encoder-based unmixing networks can be simply classified into two categories: pixel-level unmixing networks and spatial-level unmixing networks. And the typical representative of pixel-level unmixing networks includes EndNet [30], uDAS [31], DAEN [32], MUNet [33], CyCU-Net [34], and EGU-Net [35]. Among them, EndNet forms an unmixing network through two AEs and uses spectral angular distance (SAD) and K–L divergence as loss functions. uDAS reduces the effect of noise on unmixing by introducing denoising constraints into the AE and enhances the abundance estimation by introducing the so-called l_{21}-norm into the decoder [31]. To enhance the robustness of the model, DAEN first processes the outliers and noise in the data by using a stacked autoencoder (SAE) and then feeds the processed data into a variational AE to obtain more accurate unmixing results. MUNet constructs a multimodal unmixing network by additionally introducing LiDAR data, which improves the unmixing performance by integrating the elevation differences of the LiDAR data into the HSI. CyCU-Net uses a cyclic consistency network structure with two cycle-connected AEs to reduce the information loss during image processing, thus enhancing the unmixing ability of the network. EGU-Net reduces the influence of spectral variability (SV) on the unmixing results by introducing pseudoendmembers and utilizes the unmixing information derived from pseudoendmembers to guide the unmixing process to improve network performance. With the rise of CNNs in the field of computer vision, many spatial-level unmixing networks have emerged as well. Typically, CNNAEU [36] introduces CNNs into AE-based unmixing networks, omitting the use of any pooling or upsampling operations to maximize the retention of spatial information from HSI. SSCC-Net [37] utilizes both spectral and spatial information to train the spatial AE networks and spectral AE networks, respectively, in an end-to-end manner. Particularly, DeepTrans [38] was the first to use a transformer in combination with the convolutional AE for HU. Although the above methods perform remarkable performance in HU tasks, they still have some limitations. The pixel-level unmixing networks mainly utilize the spectral information of pixels for unmixing without considering the spectral differences between pseudoendmembers. The spatial-level unmixing networks enhance the receptive field by introducing convolutional operations to capture spatial contextual information but neglect the spatial differences between different ground objects. Although these joint spectral–spatial networks utilize the contextual information of the spectral and spatial features in HSI, they still do not take into account the effects of spectral differences between the pseudoendmembers as well as the spatial heterogeneities of distributed features. Therefore, how to make full use of the spatial discrepancy in HSI and the spectral difference between pseudoendmembers to maximize the accuracy of HU has become an urgent challenge to overcome.

To this end, this article proposes a novel spatial–spectral features integrated AE HU network, called SSF-Net, which integrates spectral attention mechanism and spatial attention mechanism within the AE unmixing framework to effectively extract the spatial discrepancy in HSI and the spectral difference between pseudoendmembers in an unsupervised manner. Compared with the current unmixing methods that only use spectral information or neighboring pixel information, SSF-Net can better extract the feature difference information between different ground objects, thus significantly improving the network unmixing accuracy. Specifically, the main contributions of this study can be summarized as follows.

1) SSF-Net improves the unmixing performance by exploiting the spatial difference information in HSI and the spectral difference information between pseudoendmembers. To the best of our knowledge, this study represents the first utilization of DL to investigate the unmixing task of multifeature fusion.
2) A two-branch feature fusion module incorporating a spatial–spectral attention mechanism is built into SSF-Net to address the problem of underutilization of spatial–spectral information. This module extracts the relevant spatial–spectral information by integrating spatial and channel attention, thus improving the accuracy of the unmixing.
3) An unsupervised learning approach is adopted, which makes the training process no longer dependent on labeled data. The network model is able to learn and extract features from the data itself autonomously, thus improving the generalization ability of the model.

The rest of this article is organized as follows. Section II briefly describes the principles of AE-based unmixing. Section III describes the network structure of SSF-Net in detail.

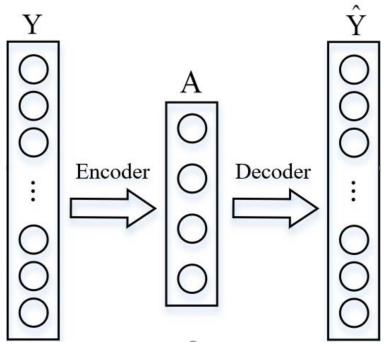

Fig. 1. Schematic diagram of an AE. The abstract features are first obtained by encoding the input data, and then the abstract features are decoded to reconstruct the input data.

Section IV gives the results of SSF-Net on several datasets. Finally, Section V concludes this article.

II. AE-Based Unmixing Model

This section mainly devotes to the introduction of an unmixing method based on LMM, which can usually be expressed as follows:

$$Y = EA + N \tag{1}$$

where $Y \in \mathbb{R}^{l\times n}$ is the HSI expressed in the form of a two-dimensional (2-D) matrix, l is the number of bands, and n is the number of pixels. Besides, $E \in \mathbb{R}^{l\times p}$ is an endmember matrix representing the p endmembers in the HSI, $A \in \mathbb{R}^{p\times n}$ is the abundance matrix corresponding to the p endmembers, and $N \in \mathbb{R}^{l\times n}$ refers to the added noise matrix.

Considering that abundance represents the proportion of different feature types in each mixed pixel, the abundance vector a_j also needs to satisfy the constraints of ANC and ASC

$$\begin{cases} a_i \geq 0 \\ \sum_{i=1}^{p} a_{i,j} = 1. \end{cases} \tag{2}$$

This study primarily focuses on HU based on the AE to simultaneously obtain the endmember matrix E and abundance matrix A in an unsupervised manner with the powerful learning and characterization capabilities of deep neural networks. As illustrated in Fig. 1, a comprehensive AE typically consists of an encoder and a decoder.

Encoder: The encoder is typically a multilayer network structure that converts the input original hyperspectral data $\{y_i\}_{i=1}^{n} \in \mathbb{R}^l$ into a hidden layer data denoted by h_i, as formulated in the following equation:

$$h_i = f_E(y_i) = f\left(W^{(e)T} y_i + b^{(e)}\right) \tag{3}$$

where $f(\cdot)$ represents the activation function of the encoder, $W^{(e)}$ represents the weight of the eth layer encoder, and $b^{(e)}$ represents the bias of the eth layer encoder.

Decoder: The decoder converts the hidden layer data h_i back to the original input data, denoted by $\{\hat{y}_i\}_{i=1}^{n} \in \mathbb{R}^l$, which can be formulated as follows:

$$\hat{y}_i = f_D(h_i) = W^{(d)T} h_i \tag{4}$$

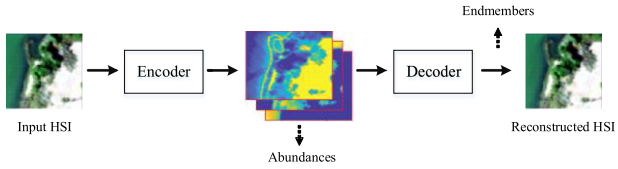

Fig. 2. Workflow of AE-based HU.

where $W^{(d)}$ represents the weight of the dth layer decoder.

Conventionally, the metric of mean square error (MSE) standard formulation (5) is always employed to quantify the reconstruction error of AE. However, when dealing with hyperspectral data, additional error metrics, such as the SAD, as depicted in (6), need to be taken into consideration for evaluating the reconstruction accuracy

$$L_{\text{AE-MSE}}(\hat{y}_i, y_i) = \frac{1}{n}\sum_{i=1}^{n} \|\hat{y}_i - y_i\|^2 \tag{5}$$

$$L_{\text{AE-SAD}}(\hat{y}_i, y_i) = \cos^{-1}\left(\frac{\hat{y}_i^{\mathrm{T}} y_i}{\|\hat{y}_i\|_2 \|y_i\|_2}\right). \tag{6}$$

Considering the inherent advantages of AE, such as a simple training process, flexible stacking of layers, and unsupervised learning paradigm, this study adopts the employment of AE for the HU. By leveraging the encoder to transform the input HSI data into abstract features and then convert them into abundance maps according to the ANC and ASC constraints, such operations ensure that the decoder is fully compliant with the LMM principle to reconstruct the abundance map back to HSI data. At this point, the decoder weights can be interpreted as the desired endmembers. The workflow of AE-based HU is illustrated in Fig. 2, wherein the network enables the simultaneous estimation of both abundances and endmembers. It is noteworthy that the entire AE unmixing process is conducted in an unsupervised manner, which effectively mitigates the issue of insufficient labeled samples in HSI data [39].

III. Proposed Method

In this study, a spatial–spectral features integrated AE network, abbreviated as SSF-Net, is proposed for the HU, with its overall structure, as shown in Fig. 3. The network consists of two parts: a spatial–spectral feature fusion encoder and a decoder. The former component fuses the deep-level features extracted from both HSI data and pseudoendmembers to fully exploit the spatial and spectral characteristics inherent in the original data. The latter component employs a commonly used decoder architecture to reconstruct the HSI by leveraging the extracted abundances. It is noteworthy that within the encoder, the in-depth integration of spectral features and spatial features is achieved by employing a dedicated module known as the spatial–spectral fusion module (SSFM). Through the SSFM, the fused high-level features encompass the intrinsic attributes from pseudoendmember spectra as well as the spatial global information from the HSI, which synergistically contributes to the enhancement of the unmixing accuracy of the network.

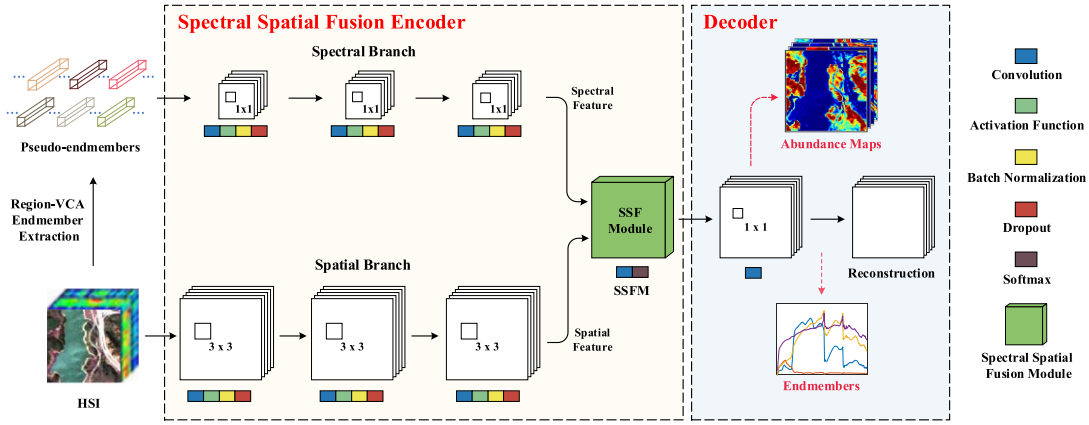

Fig. 3. Network structure of the proposed SSF-Net.

To endow the network with abundant spectral features, the regional VCA endmember extraction method is employed to acquire the pseudoendmember spectra from HSI. Specifically, the HSI data are first partitioned into several subpatches with a certain overlap rate, and the pseudoendmembers of each subpatches are extracted using the VCA algorithm. Subsequently, the K-means clustering algorithm is utilized to eliminate the duplicate pseudoendmembers, and all remaining pseudoendmembers are then aggregated into K clusters [35]. And the pseudoendmembers are then obtained by computing the centers of each cluster. Notably, the number of subpatches and K values can be determined by referring to the literature [40]. In this study, the K value is deliberatively set to about 20% of all hyperspectral pixels according to several trial experimental results. It should be noted that the pseudoendmembers obtained by the above process can reduce the influence of SV on the unmixing results because they contain rich spectral information of features, perturbation information, and a certain amount of noise. The following sections devote to a detailed description of the SSF-Net framework.

A. *Spatial–Spectral Feature Fusion Encoder*

To fully exploit the information contained in the HSI data, the spatial–spectral fusion encoder is ingeniously designed as a dual-branch structure. The encoder consists of a spectral branch and a spatial branch, which encode the input data from different views to capture the high-level spectral features and spatial features in the HSI. In the spectral branch, the spectral features in the pseudoendmembers are extracted by using three consecutive feature extraction blocks. Within each feature extraction block, a 1×1 convolution is employed to compress the spectral information of the pseudoendmembers, and an activation function (such as Sigmoid, ReLU, or LeakyReLU) is introduced after the convolution. Since the ReLU activation function may lead to the problem of neuron invalidation [41], the LeakyReLU activation function is employed in this network. Moreover, the batch normalization strategy is introduced to alleviate problems, such as gradient vanishing and gradient exploding, as well as to improve the overall computational speed of the network. To mitigate overfitting in the network, the dropout is added at the end of the blocks. The high-level spectral features, denoted as F_{spe}, are obtained from the pseudoendmembers through the three feature extraction blocks in the spectral branch. Moreover, considering the strong correlation between pixels and their surrounding scenes in HSI [37], [42], [43], the 3×3 convolutions are incorporated in the feature extraction blocks of the spatial branch to effectively capture the spatial information of neighboring pixels. Thus, the high-level spatial features obtained from the HSI data through these feature extraction blocks are denoted as F_{spa}.

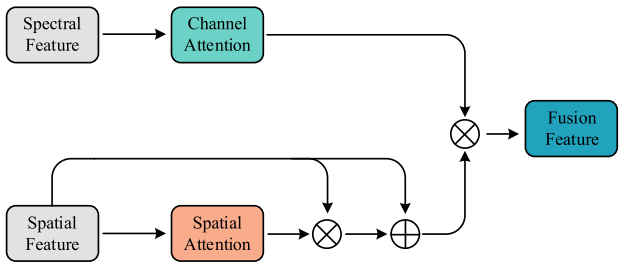

Fig. 4. Framework of SSFM.

To effectively combine the spectral features from the pseudoendmembers with the spatial features from the HSI, SSFM is constructed in this study. This module aims to enhance the network model's ability of utilizing and integrating spectral and spatial information to improve the unmixing accuracy. As shown in Fig. 4, the overall structure of SSFM contains the channel attention mechanism and the spatial attention mechanism, which are described in detail as follows.

1) *Channel Attention Mechanism:* During the process of HU, the pseudoendmembers are remarkably high resemblance to the pure endmembers. Therefore, in the absence of the pure endmembers, the spectral differences between different pseudoendmembers can be used as a substitute for the spectral differences between the pure endmembers. In view of this, a channel attention mechanism is introduced to enhance or suppress the channel features that are responsive to the differences between different

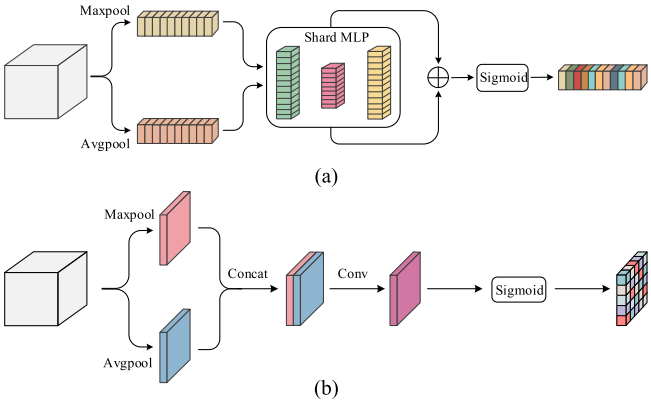

Fig. 5. Framework of attention mechanism. (a) Channels attention. (b) Spatial attention.

pseudoendmembers to improve the accuracy of abundance estimation during HU. The workflow of the channel attention module is shown in Fig. 5(a).

First, the input spectral feature $F_{\text{spe}} \in \mathbb{R}^{l\times H\times W}$ is processed using the global average pooling and global maximum pooling operations to obtain the average pooling feature $F^c_{\text{avg}} \in \mathbb{R}^{l\times 1\times 1}$ and the maximum pooling feature $F^c_{\text{max}} \in \mathbb{R}^{l\times 1\times 1}$, respectively. Then, to capture the association information between F^c_{avg} and F^c_{max}, they are separately put into a shared multilayer perceptron (MLP) and the outputs are summed. Finally, the channel attention map is output $M_C(F_{\text{spe}})$ by virtue of the sigmoid function. To reduce the number of parameters, the hidden layer size in the MLP is set to l/r, where r is the compression ratio. Through multiple sets of experiments, it is found that when r is 16, the weight operator corresponding to the entire channel attention module can significantly improve the representation ability of the network model for abundance.

The channel attention module can be expressed as follows:

$$\begin{aligned} &M_C\left(F_{\text{spe}}\right) \\ &\quad= \delta\left[\text{MLP}\left(\text{Maxpool}\left(F_{\text{spe}}\right)\right) \oplus \text{MLP}\left(\text{Avgpool}\left(F_{\text{spe}}\right)\right)\right] \\ &\quad= \delta\left[\text{MLP}\left(F^c_{\text{max}}\right) \oplus \text{MLP}\left(F^c_{\text{avg}}\right)\right] \end{aligned} \tag{7}$$

where δ represents the sigmoid function, and $\oplus$ refers to the operation of elementwise addition.

2) *Spatial Attention Mechanism:* The complex distribution of real-world environments and the susceptibility of the HSI imaging process to interference from external factors cause the spectral profiles of pixels in the HSI to be affected by SV [44], [45], resulting in significant differences in the contributions to the unmixing of HIS pixels from different regions. The SV can often lead to deviations between the spectral curves of some pixels in the HSI and the ideal spectral curves, thereby affecting the accuracy of HU. To this end, the spatial attention mechanism is introduced to enhance or suppress the importance of pixels in different regions during the HU process to improve the unmixing ability of the network. The workflow of the spatial attention mechanism is shown in Fig. 5(b).

First, the input spatial features $F_{\text{spa}} \in \mathbb{R}^{l\times H\times W}$ are subjected to the operations of global average pooling and global max pooling, which results in the average-pooled feature $F^s_{\text{avg}} \in \mathbb{R}^{1\times H\times W}$ and the max-pooled feature $F^s_{\text{max}} \in \mathbb{R}^{1\times H\times W}$, respectively. Then, these two feature maps are merged together by concatenation operation along the channel dimension, yielding the fused feature $F^s_{A-M} \in \mathbb{R}^{2\times H\times W}$, which serves for calculating the spatial weights. Subsequently, the fused feature F^s_{A-M} is further processed by a 2-D convolution with a kernel size of 7×7, yielding the spatial weight operator. Finally, this spatial weight operator is converted to a spatial attention map $M_S(F_{\text{spa}})$ using the sigmoid activation function.

The spatial attention module can be formulated as follows:

$$\begin{aligned} &M_S\left(F_{\text{spa}}\right) \\ &\quad= \delta\left[f^{7\times 7}\left(\text{Concat}\left(\text{Maxpool}\left(F_{\text{spa}}\right); \text{Avgpool}\left(F_{\text{spa}}\right)\right)\right)\right] \\ &\quad= \delta\left[f^{7\times 7}\left(\text{Concat}\left(F^s_{\text{max}}; F^s_{\text{avg}}\right)\right)\right] \end{aligned} \tag{8}$$

where $f^{7\times 7}$ denotes the 2-D convolution with a kernel size of 7×7. Concat represents the concatenation operation of vertically stacking the feature maps along the channel dimension, and δ refers to the sigmoid function.

3) *Fusion of Spectral–Spatial Features:* SSFM facilitates the comprehensive mining and utilization of spectral and spatial information by synchronizing the fusion of features using the channel attention mechanism and spatial attention mechanism, thus significantly improving the unmixing accuracy. The former mechanism evaluates the importance of channels based on the spectral features to improve the abundance estimation accuracy, while the latter mechanism focuses on elevating the significance of different pixels in space to obtain better endmember extraction results. The fusion process of SSFM can be formulated as follows:

$$F_{\text{fused}} = M_C\left(F_{\text{spe}}\right) \otimes \left[F_{\text{spa}} \oplus \left(F_{\text{spa}} \otimes M_S\left(F_{\text{spa}}\right)\right)\right] \tag{9}$$

where $\otimes$ and $\oplus$ denote the operations of element-by-element multiplication and addition, respectively; and F_{fused} refers to the fused features.

Furthermore, the abundance maps are obtained by employing a 3×3 convolutional operation upon the F_{fused}, where the number of abundance maps is consistent with the number of endmembers. In the end, a softmax function is used immediately after the convolution layer to ensure that the abundance results satisfy the ANC and ASC constraints.

B. Decoder

The decoder reconstructs the input pixels by integrating the estimated abundances and the corresponding endmembers. Such a process can be expressed as follows:

$$\hat{y}_i = f\left(W^{(d)}\hat{a}_i\right) = \hat{E}\hat{a}_i. \tag{10}$$

In the equation above, $\{\hat{y}_i\}_{i=1}^n \in \mathbb{R}^l$ is the reconstructed pixels, and $\hat{E} \in \mathbb{R}^{l\times p}$ represents the estimated endmember matrix. $\{\hat{a}_i\}_{i=1}^n \in \mathbb{R}^p$ denotes the generated abundance vector.

Notably, to reduce the training time, the method of VCA is employed to initialize the weights $W^{(d)}$ of the decoder.

C. Objective Function

To achieve the best possible training results, the loss function of the SSF-Net model is designed to consist of SAD and MSE. The SAD exhibits spectral scale invariance as it evaluates the similarity between two spectral curves by calculating the angle between the target spectrum and the reference spectrum. And a smaller angle between the two spectral curves indicates more similarity. The calculation formula of SAD is given as follows:

$$J_{\text{SAD}} = \frac{1}{n}\sum_{i=1}^{n} \arccos \frac{\langle \hat{y}_i, y_i \rangle}{\|\hat{y}_i\| \cdot \|y_i\|} \tag{11}$$

where y_i and $\hat{y}_i$ denote the input and reconstructed pixel data, respectively. n denotes the total number of pixels.

Although SAD can improve the accuracy of endmember extraction, it is prone to larger errors in abundance estimation as it only considers the scale invariance of endmembers. To this end, the MSE is also introduced into the objective function to ensure that the network can obtain more accurate abundance

$$J_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n} \|\hat{y}_i - y_i\|^2. \tag{12}$$

To strive for better unmixing results, the overall loss function of the network in this study is defined as a weighted combination of SAD error and MSE error, as shown in the following equation:

$$L = \alpha J_{\text{SAD}} + \beta J_{\text{MSE}}. \tag{13}$$

Here, α and β represent the hyperparameters of the loss function.

IV. Experiments

In this section, a comparatively experimental analysis is carried out with the state-of-the-art HSI unmixing methods to demonstrate the superiority of the SSF-Net network. The algorithms selected for comparison include three classical methods: fully constrained least-squares unmixing (FCLSU) [14], multilayer nonnegative matrix factorization (MLNMF) [46], spatial group sparsity regularized nonnegative matrix factorization (SGSNMF) [47], SNMF-Net [48], and four deep-learning-based methods: uDAS [31], DAEU [49], CNNAEU [36], and CyCU-Net [34]. These methods are widely recognized and highly represented in the field of HU. To ensure fairness in the experiments, the VCA [16] is first adopted for generating the initial endmember for all the comparison algorithms.

A. Data Description

1) Synthetic Dataset: The synthetic dataset is composed of five randomly selected spectral curves from the ASTER spectral library, as curated by Jin et al. [50]. This dataset consists of 60×60 pixels, with a total of 200 spectral bands spanning from 0.4 to 14 μm. The abundance maps follow a Dirichlet distribution. To simulate the endmember variability in real HSI data, this dataset is made by using asphalt as the background color and the remaining four endmembers are randomly scattered in the corners. Moreover, to enhance the realism of the synthetic hyperspectral data, Gaussian noise with different signal-to-noise ratios (SNRs) was introduced to the synthetic dataset. The data contain a total of five endmembers: limestone, conifer, basalt, concrete, and asphalt. Fig. 6(a) shows the RGB true color image corresponding to this data area.

(a) (b) (c) (d)

Fig. 6. RGB image of four datasets. (a) Synthetic. (b) Samson. (c) Jasper. (d) Urban.

2) Samson Dataset: The Samson dataset is acquired using the SAMSON sensor. The image consists of 952×952 pixels with a total of 156 spectral bands ranging from 0.401 to 0.889 μm. Considering that the size of the original image is too large, an area of 95×95 pixels is cropped out starting from the position of (252,332) pixels in the original image as the experimental data. Specifically, the cropped data contain three endmembers: Soil, Tree, and Water. Fig. 6(b) presents the corresponding RGB true color image of these data.

3) Jasper Dataset: The Jasper data were collected by the airborne visible infrared imaging spectrometer sensor. The image is 521×614 pixels and contains 224 bands with a spectral range from 0.38 to 2.50 μm. Since the original image was too large, only a cropped subimage containing 100×100 pixels is used in this experiment, with its first pixel starting from the position of (252,332) pixels in the original image. After removing some of the bands affected by high water vapor concentration and atmospheric effects, only 198 channels are retained in these data, which contain four endmembers: Road, Soil, Water, and Tree. Fig. 6(c) shows the RGB true color image corresponding to this data area.

4) Urban Dataset: The Urban data were obtained from the hyperspectral digital image collection experimental sensor for the urban area of Copoas, Texas, USA. The image consists of 307×307 pixels and contains 210 bands with a spectral range from 0.4 to 2.5 μm. Only 162 bands are retained after removing bands affected by high water vapor concentrations and atmospheric effects. In these HIS data, there are five endmembers: Asphalt, Grass, Tree, Roof, and Dirt. Fig. 6(d) presents the corresponding RGB true color image of these data.

B. Experimental Settings

1) Hyperparameter Settings: The implementation of the SSF-Net model in this study is based on the PyTorch framework with an i9-9900K CPU and an NVIDIA 2080 8GB GPU as the hardware platform. During the training process, the Adam optimizer is used to update the network parameters, where the learning rate is set to 1×10^{-3}. To further improve the network

TABLE I
QUANTITATIVE RESULTS FOR THE SYNTHETIC DATASET

Methods		FCLSU	MLNMF	SGSNMF	SNMF	uDAS	DAEU	CNNAEU	CyCU-Net	SSF-Net
RMSE	Asphalt	0.1276	0.0798	0.0663	0.1443	0.0883	0.1093	0.2641	0.4226	0.0686
	Conifer	0.0448	0.0281	0.0264	0.0299	0.0172	0.0229	0.2054	0.1130	0.0274
	Concrete	0.0383	0.0427	0.0439	0.1525	0.0361	0.2514	0.1123	0.2620	0.0369
	Basalt	0.0259	0.0224	0.0175	0.0756	0.0144	0.0549	0.1137	0.1075	0.0152
	Limestone	0.0450	0.0591	0.0224	0.1267	0.0402	0.0442	0.1686	0.1609	0.0371
Mean_RMSE		0.0566	0.0464	0.0353	0.1058	0.0393	0.1067	0.1826	0.2235	0.0370
SAD	Asphalt	0.0162	0.0190	0.0092	0.0339	0.0126	0.0231	0.0580	0.0540	0.0156
	Conifer	0.0265	0.0260	0.0105	0.1248	0.0139	0.0904	0.1197	0.0661	0.0199
	Concrete	0.0620	0.0697	0.0120	0.2637	0.0422	0.0288	0.1238	0.1804	0.0138
	Basalt	0.0212	0.0228	0.0097	0.1418	0.0241	0.0339	0.1116	0.1275	0.0177
	Limestone	0.0408	0.0444	0.0153	0.1912	0.0254	0.0232	0.0398	0.2877	0.0455
Mean_SAD		0.0333	0.0364	0.0113	0.1511	0.0236	0.0398	0.0906	0.1431	0.0225

Best results are shown in red, followed by green.

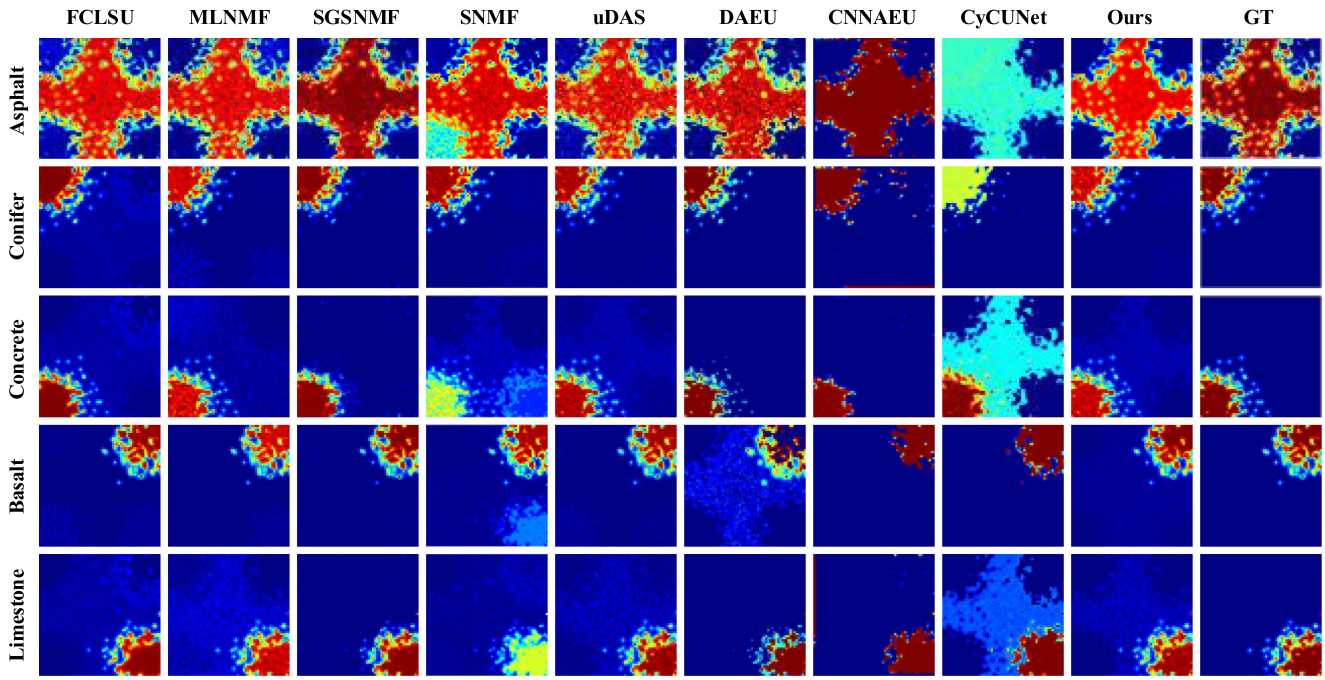

Fig. 7. Abundance maps of five materials from the synthetic data obtained by different algorithms.

accuracy, the learning rate decay strategy is adopted, that is, the learning rate is decayed once every 40 epochs of training, and the maximum number of iterations is set to 800.

2) Evaluation Metrics: To evaluate the network unmixing performance, the following two metrics are introduced into the experiments: root-mean-square error (RMSE) and SAD, which are defined as follows:

$$L_{\text{RMSE}}(a_i, \hat{a}_i) = \sqrt{\frac{1}{N}\sum_1^N \|\hat{a}_i - a_i\|_2^2} \tag{14}$$

$$L_{\text{SAD}}(m_i, \hat{m}_i) = \arccos\left(\frac{\langle m_i, \hat{m}_i\rangle}{\|m_i\|_2\|\hat{m}_i\|_2}\right) \tag{15}$$

where a_i and $\hat{a}_i$ represent the true abundance vector and generated abundance vector, respectively. m_i and $\hat{m}_i$ denote the true endmember and extracted endmember, respectively.

C. Experimental Result and Analysis

1) Synthetic Dataset: The quantitative results of the RMSE as well as the SAD for each endmember on the synthetic dataset by the different algorithms are presented in Table I. Meanwhile, Figs. 7 and 8 show the abundance maps and corresponding endmember results extracted by different algorithms on the synthetic dataset. In the synthetic dataset, SGSNMF achieves better unmixing results than FCLSU and MLNMF. SGSNMF uses spatial information to divide the HSI into spatial groups and incorporates a sparsity constraint of spatial group into nonnegative matrix factorization, which results in a superior decomposition structure. Although SNMF is constructed based on a nonnegative matrix model with L–p sparse constraints, its failure to take the spatial information into account results in unmixing performance in synthetic data without significant advantages. The uDAS algorithm achieves promising results in

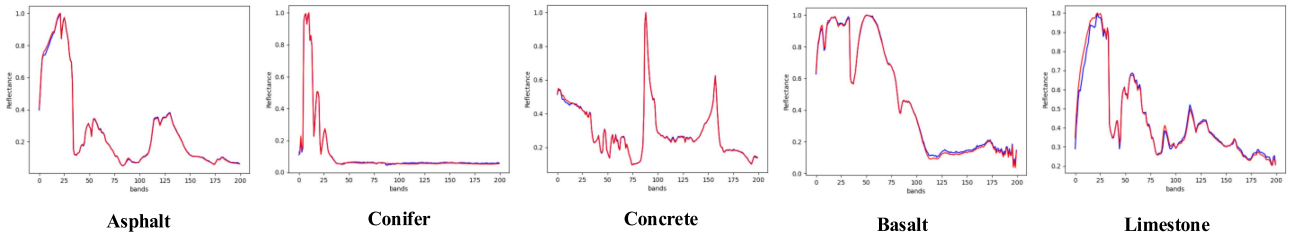

Fig. 8. Comparison of endmembers between SSF-Net (blue curves) and the corresponding GT (red curves) on the synthetic dataset.

TABLE II
QUANTITATIVE RESULTS OF MEAN_RMSE AND MEAN_SAD FOR THE SYNTHETIC DATASET UNDER DIFFERENT NOISES

Methods		FCLSU	MLNMF	SGSNMF	SNMF	uDAS	DAEU	CNNAEU	CyCU-Net	SSF-Net
Mean_RMSE	20db	0.0566	0.0464	0.0353	0.1058	0.0393	0.1067	0.1826	0.2235	0.0370
	30db	0.0485	0.0453	0.0267	0.0954	0.0395	0.0906	0.1583	0.1950	0.0297
	40db	0.0458	0.0404	0.0209	0.0782	0.0312	0.0873	0.1510	0.2020	0.0223
Mean_SAD	20db	0.0333	0.0364	0.0113	0.1511	0.0236	0.0398	0.0906	0.1431	0.0225
	30db	0.0256	0.0241	0.0104	0.1456	0.0227	0.0266	0.0936	0.1392	0.0161
	40db	0.0224	0.0199	0.0081	0.1279	0.0212	0.0232	0.0930	0.1405	0.0109

Best results are shown in red, followed by green.

TABLE III
QUANTITATIVE RESULTS FOR THE SAMSON DATASET

Methods		FCLSU	MLNMF	SGSNMF	SNMF	uDAS	DAEU	CNNAEU	CyCU-Net	SSF-Net
RMSE	Soil	0.2656	0.1882	0.1887	0.2487	0.2598	0.1146	0.1688	0.1980	0.0511
	Tree	0.2507	0.2035	0.2531	0.2603	0.2553	0.1065	0.1099	0.1720	0.0502
	Water	0.4236	0.3216	0.3700	0.3507	0.4108	0.0407	0.1925	0.2020	0.0272
Mean_RMSE		0.3133	0.2378	0.2706	0.2866	0.3086	0.0873	0.1571	0.1910	0.0428
SAD	Soil	0.0433	0.0457	0.0495	0.0103	0.0534	0.0345	0.0803	0.0417	0.0092
	Tree	0.1567	0.1244	0.2523	0.0277	0.0307	0.0172	0.2567	0.0478	0.0314
	Water	0.0236	0.0264	0.0143	0.1752	0.1379	0.0582	0.1141	0.1889	0.0373
Mean_SAD		0.0745	0.0655	0.1054	0.0711	0.0740	0.0366	0.1504	0.0928	0.0260

Best results are shown in red, followed by green.

abundance estimation by incorporating denoising constraints and attaching certain physical constraints. The DAEU, on the other hand, only focuses on spectral information and ignores spatial information. Although CNNAEU and CyCU-Net consider spatial information, they ignore the inter-SV, which leads to their unsatisfactory performance on synthetic datasets. The experimental results of these algorithms demonstrate the importance of taking both spectral and spatial information into consideration to obtain accurate HU results. In contrast, the proposed SSF-Net achieves superior results on synthetic data, which proves its superiority in the unmixing task.

To verify the robustness of the proposed SSF-Net network, the varying SNR values from 20 to 40 dB are added to the synthetic dataset. Correspondingly, the quantitative experimental results are presented in Table II. In general, the unmixing accuracy of these algorithms tends to decrease as the noise increases. The classical algorithm SGSNMF exerts good performance on the synthetic dataset. Benefiting from denoising processing, the uDAS network exhibits minimal variation in unmixing accuracy under different noise conditions. In contrast, the models of SNMF, CNNAEU, and CyCU-Net perform poorly in high-noise situations. Particularly important, the proposed SSF-Net has obtained remarkable accuracy in both abundance estimation and endmember extraction under varying noise levels, which fully demonstrates its effectiveness and robustness.

2) Samson Dataset: The results of the RMSE and SAD quantification for each endmember on the Samson dataset by the different algorithms are presented in Table III. Figs. 9 and 10 show the abundance maps and corresponding endmember results extracted by different algorithms on the Samson dataset. The Samson dataset is characterized by a relatively uniform spatial distribution of different materials, making it widely regarded as a relatively simple unmixing dataset. In general, the algorithms have all achieved relatively excellent results. However, despite the promising results obtained by the algorithm of SGSNMF on synthetic datasets, its performance on the Samson dataset is subpar. The reason for this discrepancy can be attributed to the distribution complexity of ground objects and the non-Gaussian nature of noise distribution in real scenes. Moreover, the unmixing accuracy of all classical algorithms is lower than that of deep learning models. Specifically, the proposed SSF-Net algorithm achieves the best results in terms of RMSE for each land cover category, and it also outperforms all other methods in terms of the overall mean endmember accuracy of SAD.

3) Jasper Dataset: The results of the RMSE and SAD quantification for each endmember on the dataset of Jasper by the

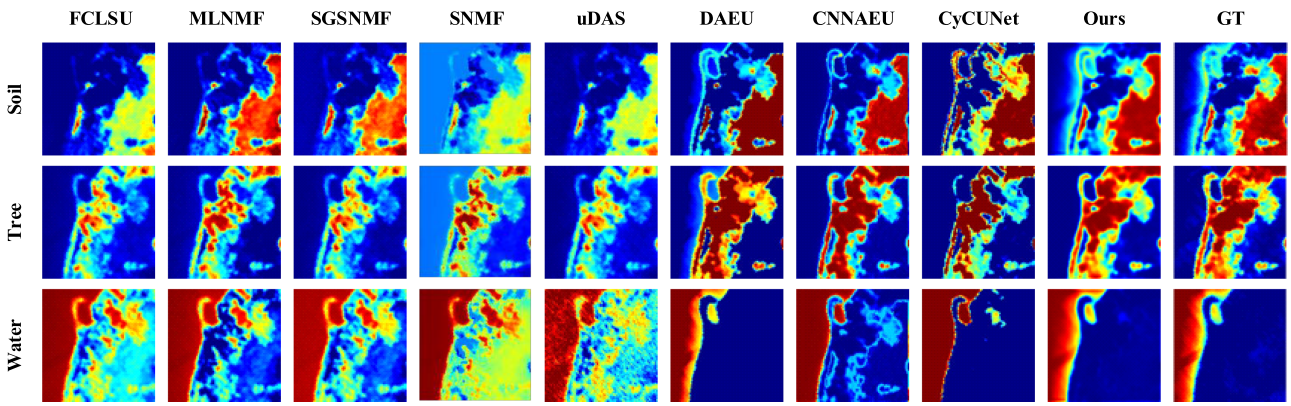

Fig. 9. Abundance maps of three materials from the Samson data obtained by different algorithms.

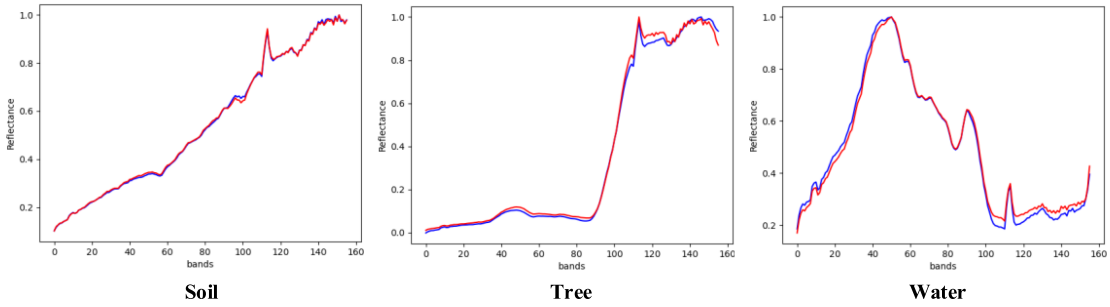

Fig. 10. Comparison of endmembers between SSF-Net (blue curves) and the corresponding GT (red curves) on the Samson dataset.

TABLE IV
QUANTITATIVE RESULTS FOR THE JASPER DATASET

Methods		FCLSU	MLNMF	SGSNMF	SNMF	uDAS	DAEU	CNNAEU	CyCU-Net	SSF-Net
RMSE	Tree	0.1563	0.2284	0.1385	0.2093	0.1327	0.1686	0.2489	0.1110	0.0721
	Water	0.1880	0.1205	0.1756	0.2093	0.1770	0.0906	0.1201	0.0894	0.0761
	Soil	0.1272	0.2497	0.1854	0.2133	0.1252	0.1129	0.3408	0.1457	0.0930
	Road	0.1450	0.1500	0.1213	0.1319	0.1228	0.1499	0.2718	0.1370	0.0782
Mean_RMSE		0.1541	0.1872	0.1552	0.1910	0.1394	0.1305	0.2454	0.1208	0.0798
SAD	Tree	0.1481	0.3288	0.2521	0.1332	0.1300	0.0577	0.1047	0.1142	0.0781
	Water	0.2234	0.0945	0.1394	0.1469	0.1843	0.0434	0.1903	0.1481	0.0293
	Soil	0.0901	0.1398	0.0487	0.0755	0.0558	0.0906	0.1136	0.0889	0.0484
	Road	0.3750	0.3933	0.2400	0.0957	0.1390	0.0569	0.1164	0.1133	0.0238
Mean_SAD		0.2092	0.2391	0.1700	0.1128	0.1273	0.0622	0.1312	0.1161	0.0449

Best results are shown in red, followed by green.

different algorithms are presented in Table IV. Figs. 11 and 12 show the abundance maps and the corresponding endmember results extracted by different algorithms on this dataset. The distribution of ground objects in the dataset of Jasper exhibits a higher level of complexity compared with the Samson dataset. As shown in Fig. 11, the unmixing algorithms, such as FCLSU, MLNMF, SGSNMF, SNMF, and uDAS, have deficiencies in accurately extracting all the roads, which is manifested in the fact that some of the roads are misclassified as water bodies, which affects the final unmixing accuracy. In contrast, deep learning algorithms are better able to identify roads and, thus, generate more accurate abundance maps. As can be seen in Table IV, the proposed SSF-Net network achieves excellent results in the Jasper dataset. The quantification results show that all accuracy evaluation metrics, except for SAD_Tree, achieve optimal accuracy levels, indicating that SSF-Net has excellent performance in the unmixing task.

4) Urban Dataset: The results of the RMSE and SAD quantification for each endmember on the Urban dataset by the different algorithms are presented in Table V. Figs. 13 and 14 show the abundance maps and corresponding endmember results extracted by different algorithms on the Urban dataset. Among four experimental datasets, the Urban dataset is the most heavily mixed dataset in terms of the mixing degree of ground objects. Visually, the abundance maps generated by the proposed SSF-Net network exhibit the highest degree of similarity to the real abundance maps. From the quantitative results, SSF-Net achieves the best accuracy in terms of the Mean_SAD

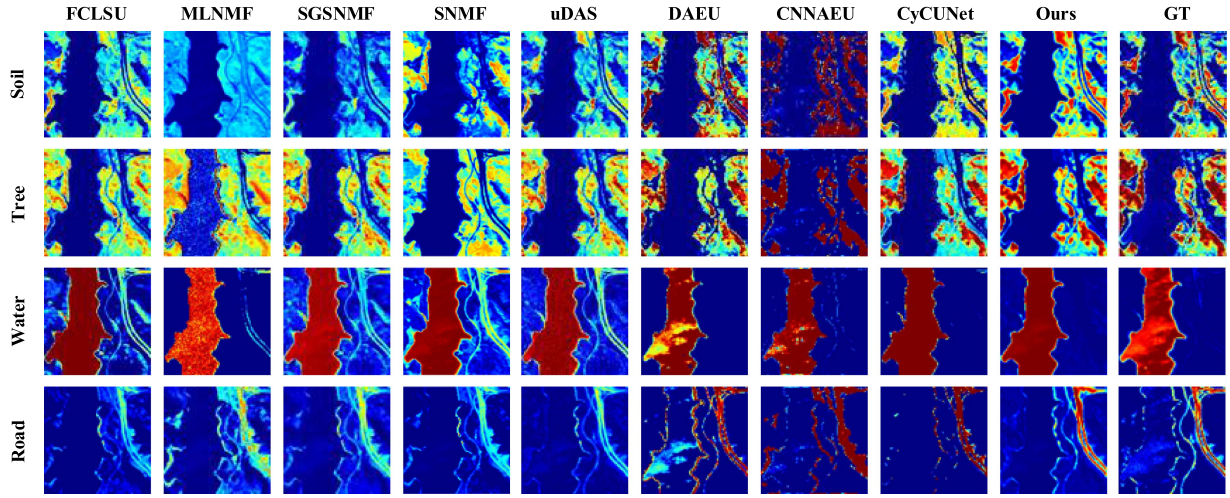

Fig. 11. Abundance maps of four materials from the Jasper ridge data obtained by different algorithms.

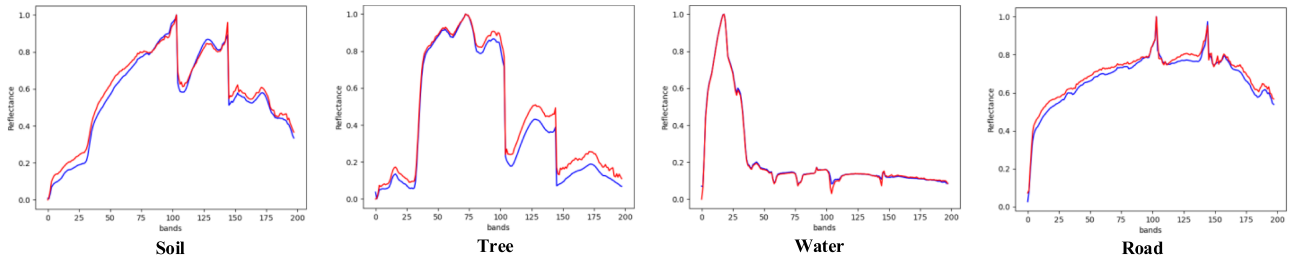

Fig. 12. Comparison of endmembers between SSF-Net (blue curves) and the corresponding GT (red curves) on the Jasper dataset.

TABLE V
QUANTITATIVE RESULTS FOR THE URBAN DATASET

Methods		FCLSU	MLNMF	SGSNMF	SNMF	uDAS	DAEU	CNNAEU	CyCU-Net	SSF-Net
RMSE	Asphalt	0.3496	0.2728	0.3482	0.2152	0.3249	0.1223	0.2707	0.3078	0.1587
	Grass	0.3153	0.2105	0.1407	0.3424	0.4738	0.1824	0.3061	0.4745	0.1416
	Tree	0.2673	0.1048	0.2428	0.2370	0.2693	0.1371	0.2215	0.3879	0.1179
	Roof	0.2258	0.2138	0.2620	0.2154	0.1976	0.0756	0.1253	0.1543	0.0909
	Dirt	0.2215	0.1876	0.2565	0.2001	0.2483	0.1593	0.2754	0.1554	0.1192
Mean_RMSE		0.2759	0.1979	0.2500	0.2420	0.3159	0.1354	0.2398	0.2960	0.1256
SAD	Asphalt	0.1416	0.6032	0.1662	0.1713	0.2022	0.1454	0.0907	0.1717	0.0629
	Grass	0.1103	0.2027	0.6543	0.1662	0.7363	0.1086	0.1460	0.1503	0.0411
	Tree	0.1388	0.2489	0.1505	0.2554	0.1078	0.0349	0.0379	0.4377	0.0850
	Roof	0.7798	0.4915	0.1485	0.1501	0.1216	0.0365	0.0578	0.1205	0.0487
	Dirt	0.1127	0.3611	0.1172	0.0601	0.1473	0.1712	0.0395	0.2067	0.1065
Mean_SAD		0.2566	0.3815	0.2474	0.1606	0.2630	0.0993	0.0744	0.2174	0.0688

Best results are shown in red, followed by green.

and Mean_RMSE. By comparing the experimental results on multiple datasets, the proposed SSF-Net network can consistently outperform other methods in generating more accurate and reliable estimates of endmembers and abundances, which fully demonstrates the remarkable superiority of the SSF-Net network in HU tasks.

D. Computational Cost

Table VI presents the run time by all comparison methods on different datasets. It can be seen that the classical methods, FCLS, MLNMF, and SGSNMF, have a relatively simple computational process and, thus, have a low time overhead. In contrast, deep learning models usually run significantly longer than classical methods due to their complex network structure and inclusion of a large number of parameters. Among these deep learning models, SNMF, uDAS, and DAEU belong to the pixel-level unmixing networks. Since they are unmixed pixel-by-pixel, their processing time increases with the number of pixels. However, CNNAEU and CyCU-Net are spatial-level unmixing networks, which capture spatial context information by introducing a receptive field. However, it is also the inclusion of the receptive field that causes their time overhead to significantly exceed that of pixel-level unmixing networks. Overall, the time expenditure of the proposed method in this study is between the pixel-level

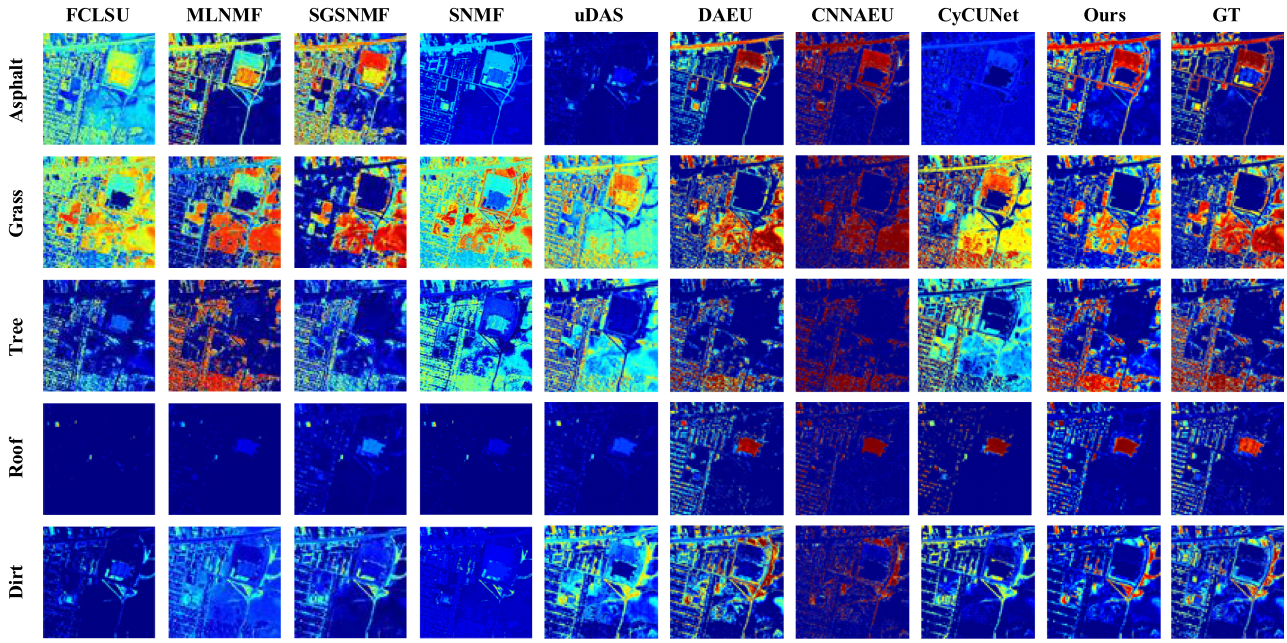

Fig. 13. Abundance maps of five materials from the urban data obtained by different algorithms.

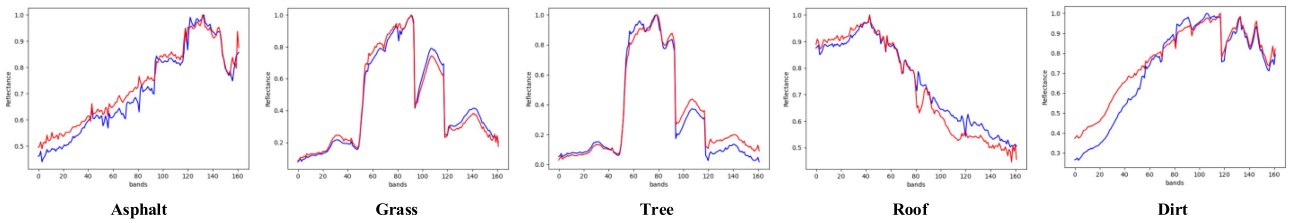

Fig. 14. Comparison of endmembers between SSF-Net (blue curves) and the corresponding GT (red curves) on the urban dataset.

TABLE VI
COMPUTATIONAL COST OF ALL COMPARISON METHODS ON DIFFERENT DATASETS IN TERMS OF SECONDS (S)

Methods	Synthetic	Samson	Jasper	Urban
FCLSU	1.7	5.1	10.3	105.7
MLNMF	1.9	4.6	8.5	146.4
SGSNMF	8.9	15.9	19.1	288.5
SNMF	37.6	37.5	39.9	43.1
uDAS	4.3	7.1	147.2	1257.8
DAEU	18.4	18.1	19.7	22.4
CNNAEU	182.4	160.0	227.3	280.9
CyCU-Net	91.5	128.0	179.3	250.6
Ours	14.3	15.7	19.0	64.0

TABLE VII
ABLATION ANALYSIS OF SSF-NET ON URBAN DATASET COMBINED WITH DIFFERENT NETWORK MODULES

Module		Mean_SAD	Mean_RMSE
Spectral Module	Spatial Module		
×	×	0.0961	0.2906
✓	×	0.0994	0.1469
×	✓	0.0701	0.1860
✓	✓	0.0688	0.1256

The best results are shown in red.

and spatial-level unmixing network models as mentioned above. The network model proposed in this study also constructs an SSFM module in the encoder, which contains spatial attention and spectral attention mechanisms, which helps to take into account the spatial contextual information in the HSI data and the spectral disparity information between the pseudoendmembers. Despite the increase in model runtime overhead caused by the introduction of the attention mechanism, the proposed method is able to significantly improve the unmixing accuracy.

E. Ablation Analysis

To verify the effectiveness of each module in the SSF-Net network, ablation experiments are conducted in this section for the spectral attention module and the spatial attention module on the Urban dataset. As can be seen from Table VII, when the SSF-Net network removes both the spectral attention module in spectral branch and the spatial attention module in spatial branch, the performance of the network becomes the worst, which also indicates to a certain extent that the potential information contained in the HSI data is not fully exploited.

In the SSF-Net network, the addition of the spectral attention module can enhance the network's ability of abundance estimation. The spatial attention module, on the other hand, is able to improve the endmember extraction accuracy of the network. It is worth emphasizing that the spectral attention module plays a crucial role in enhancing the abundance estimation capability of the network, mainly by focusing on the distinct disparities between different spectral features. And that, the spatial attention module exhibits higher sensitivity to the difference between different spatial regions, which significantly enhances the endmember extraction ability of the network. The results of the ablation experiments show that the joint use of both spectral attention module and spatial attention module in SSF-Net can be more effective in mining high-dimensional features from HSI, thus obtaining better unmixing results.

F. Discussion

By conducting a quantitative analysis of the experimental results on the four datasets, it is found that SSF-Net behaves prominently superior unmixing performance to the other comparative methods. Meanwhile, the complexity of the spatial distribution of features in real images far exceeds that of synthetic datasets, resulting in the poor performance of some NMF-based unmixing methods in real datasets. Besides, SNMF-Net is of high physical interpretability as it is built by unrolling L–p sparsity constrained NMF model, so it achieves higher accuracy than other NMF methods in the real datasets. However, because it only performs unmixing at the pixel level without introducing spatial information, it does not achieve particularly good unmixing accuracy. Similarly, the pixel-level-based unmixing method also includes DAEU, which pays more attention to endmember extraction, and a good endmember extraction result will further enhance the abundance estimation result, thus it also achieves relatively excellent unmixing accuracy in these comparative experiments. Although CNNAEU introduces spatial information, its loss function only considers SAD, which makes its unmixing results not as good as other DL unmixing networks. The accuracy of CyCU-Net is poor because its receptive field does not cover the complete image, resulting in its lack of extensive information and remote dependencies. Most importantly, the proposed method in this study, however, achieves the optimal accuracy in real datasets and the unmixing results can be perceived as the closest to the ground truth in terms of visualization. This further confirms the excellence of the proposed network model in the unmixing task.

V. Conclusion

In this study, a convolutional AE HU network called SSF-Net is proposed ingeniously integrating both spatial and spectral features. The architecture of this network is conceived in such a manner that it initiates its operation by employing a regional VCA algorithm to extract the pseudoendmembers from HSI data. Then, the network utilizes the spatial attention module along with the spectral attention module to learn the spatial difference information contained within the HSI data and spectral difference information amongst the pseudoendmembers, respectively, in such a way that the network makes the best use of the information inherent in the HSI data, resulting in more reasonable and superior unmixing results. Experiments confirm the effectiveness of the SSF-Net network proposed in this article on synthetic and real hyperspectral datasets with higher unmixing accuracy compared with other state-of-the-art HU methods. The proposed SSF-Net network is built based on LMM. However, considering the complexity of the hyperspectral imaging process, the NLMM is more suitable for elaborating its imaging principles. Thus, our future research aims to develop more general and powerful NLMM-based unmixing networks that integrate the spatial–spectral features. Meanwhile, the introduction of LiDAR data to aid in HU has been shown to be feasible. Thus, designing a network architecture that fuses the spatial–spectral features of HSI with those extracted from the LiDAR point cloud can help to address the unmixing issue more effectively.

References

[1] J. M. Bioucas-Dias et al., "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012, doi: 10.1109/JSTARS.2012.2194696.

[2] X. Mei, Y. Ma, C. Li, F. Fan, J. Huang, and J. Ma, "Robust GBM hyperspectral image unmixing with superpixel segmentation based low rank and sparse representation," *Neurocomputing*, vol. 275, pp. 2783–2797, Jan. 2018, doi: 10.1016/j.neucom.2017.11.052.

[3] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2019, doi: 10.1109/TGRS.2018.2861992.

[4] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, Jul. 2001, doi: 10.1109/36.934072.

[5] J. Ma, H. Zhou, J. Zhao, Y. Gao, J. Jiang, and J. Tian, "Robust feature matching for remote sensing image registration via locally linear transforming," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6469–6481, Dec. 2015, doi: 10.1109/TGRS.2015.2441954.

[6] Q. Jin et al., "Gaussian mixture model for hyperspectral unmixing with low-rank representation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Yokohama, Japan, 2019, pp. 294–297, doi: 10.1109/IGARSS.2019.8898410.

[7] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial–spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021, doi: 10.1109/TCYB.2020.3028931.

[8] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002, doi: 10.1109/79.974727.

[9] F. Poulet, B. L. Ehlmann, J. F. Mustard, M. Vincendon, and Y. Langevin, "Modal mineralogy of planetary surfaces from visible and near-infrared spectral data," in *Proc. 2nd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Reykjavik, Iceland, 2010, pp. 1–4, doi: 10.1109/WHISPERS.2010.5594898.

[10] D. A. Roberts, M. Gardner, R. Church, S. Ustin, G. Scheer, and R. O. Green, "Mapping chaparral in the Santa Monica mountains using multiple endmember spectral mixture models," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 267–279, Sep. 1998, doi: 10.1016/S0034-4257(98)00037-6.

[11] C. Yang, J. H. Everitt, Q. Du, B. Luo, and J. Chanussot, "Using high-resolution airborne and satellite imagery to assess crop growth and yield variability for precision agriculture," *Proc. IEEE*, vol. 101, no. 3, pp. 582–592, Mar. 2013, doi: 10.1109/JPROC.2012.2196249.

[12] M. Teke, H. S. Deveci, O. Haliloglu, S. Z. Gurbuz, and U. Sakarya, "A short survey of hyperspectral remote sensing applications in agriculture," in *Proc. 6th Int. Conf. Recent Adv. Space Technol.*, Istanbul, Turkey, 2013, pp. 171–176, doi: 10.1109/RAST.2013.6581194.

[13] 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, doi: 10.1109/JSTARS.2014.2320576.

[14] D. C. Heinz and C.-I- Chang, "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, doi: 10.1109/36.911111.

[15] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, 1999, doi: 10.1117/12.366289.
[16] 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, doi: 10.1109/TGRS.2005.844293.
[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, doi: 10.1109/TSP.2009.2025797.
[18] 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, doi: 10.1109/TGRS.2006.888466.
[19] O. Eches, N. Dobigeon, C. Mailhes, and J.-Y. Tourneret, "Bayesian estimation of linear mixtures using the normal compositional model: Application to hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1403–1413, Jun. 2010, doi: 10.1109/TIP.2010.2042993.
[20] J. M. P. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 863–878, Mar. 2012, doi: 10.1109/TGRS.2011.2163941.
[21] S. Jia and Y. Qian, "Constrained nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 161–173, Jan. 2009, doi: 10.1109/TGRS.2008.2002882.
[22] F. Zhu, Y. Wang, B. Fan, S. Xiang, G. Meng, and C. Pan, "Spectral unmixing via data-guided sparsity," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5412–5427, Dec. 2014, doi: 10.1109/TIP.2014.2363423.
[23] R. Huang, X. Li, and L. Zhao, "Spectral–spatial robust nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8235–8254, Oct. 2019, doi: 10.1109/TGRS.2019.2919166.
[24] Y. Qian, F. Xiong, S. Zeng, J. Zhou, and Y. Y. Tang, "Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1776–1792, Mar. 2017, doi: 10.1109/TGRS.2016.2633279.
[25] F. Xiong, Y. Qian, J. Zhou, and Y. Y. Tang, "Hyperspectral unmixing via total variation regularized nonnegative tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2341–2357, Apr. 2019, doi: 10.1109/TGRS.2018.2872888.
[26] Y. Qian, F. Xiong, Q. Qian, and J. Zhou, "Spectral mixture model inspired network architectures for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7418–7434, Oct. 2020, doi: 10.1109/TGRS.2020.2982490.
[27] F. Xiong, J. Zhou, M. Ye, J. Lu, and Y. Qian, "NMF-SAE: An interpretable sparse autoencoder for hyperspectral unmixing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Toronto, ON, Canada, 2021, pp. 1865–1869, doi: 10.1109/ICASSP39728.2021.9414084.
[28] Y. Qian, F. Xiong, M. Ye, and J. Zhou, "Model-inspired deep neural networks for hyperspectral unmixing," in *Advances in Hyperspectral Image Processing Techniques*, C.-I. Chang, Ed., 1st ed. Hoboken, NJ, USA: Wiley, 2022, pp. 363–403, doi: 10.1002/9781119687788.ch13.
[29] W. He, H. Zhang, and L. Zhang, "Hyperspectral unmixing using total variation regularized reweighted sparse non-negative matrix factorization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Beijing, China, 2016, pp. 7034–7037, doi: 10.1109/IGARSS.2016.7730834.
[30] S. Ozkan, B. Kaya, and G. Bozdagi Akar, "EndNet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 482–496, Jan. 2019, doi: 10.1109/TGRS.2018.2856929.
[31] 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, doi: 10.1109/TGRS.2018.2868690.
[32] 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, doi: 10.1109/TGRS.2018.2890633.
[33] Z. Han, D. Hong, L. Gao, J. Yao, B. Zhang, and J. Chanussot, "Multimodal hyperspectral unmixing: Insights from attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5524913, doi: 10.1109/TGRS.2022.3155794.
[34] 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, doi: 10.1109/TGRS.2021.3064958.
[35] 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, doi: 10.1109/TNNLS.2021.3082289.
[36] 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, doi: 10.1109/TGRS.2020.2992743.
[37] 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, doi: 10.1109/TGRS.2022.3150970.
[38] 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, doi: 10.1109/TGRS.2022.3196057.
[39] C. Cui, Y. Zhong, X. Wang, and L. Zhang, "Realistic mixing miniature scene hyperspectral unmixing: From benchmark datasets to autonomous unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5502515, doi: 10.1109/TGRS.2023.3236677.
[40] B. Somers, M. Zortea, A. Plaza, and G. P. Asner, "Automated extraction of image-based endmember bundles for improved spectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 396–408, Apr. 2012, doi: 10.1109/JSTARS.2011.2181340.
[41] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," 2015, *arXiv:1505.00853*.
[42] O. Eches, N. Dobigeon, and J.-Y. Tourneret, "Enhancing hyperspectral image unmixing with spatial correlations," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4239–4247, Nov. 2011, doi: 10.1109/TGRS.2011.2140119.
[43] P. V. Giampouras, K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "Simultaneously sparse and low-rank abundance matrix estimation for hyperspectral image unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4775–4789, Aug. 2016, doi: 10.1109/TGRS.2016.2551327.
[44] J. Theiler, A. Ziemann, S. Matteoli, and M. Diani, "Spectral variability of remotely sensed target materials: Causes, models, and strategies for mitigation and robust exploitation," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 8–30, Jun. 2019, doi: 10.1109/MGRS.2019.2890997.
[45] R. A. Borsoi et al., "Spectral variability in hyperspectral data unmixing: A comprehensive review," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 4, pp. 223–270, Dec. 2021, doi: 10.1109/MGRS.2021.3071158.
[46] R. Rajabi and H. Ghassemian, "Spectral unmixing of hyperspectral imagery using multilayer NMF," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 38–42, Jan. 2015, doi: 10.1109/LGRS.2014.2325874.
[47] X. Wang, Y. Zhong, L. Zhang, and Y. Xu, "Spatial group sparsity regularized nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6287–6304, Nov. 2017, doi: 10.1109/TGRS.2017.2724944.
[48] F. Xiong, J. Zhou, S. Tao, J. Lu, and Y. Qian, "SNMF-Net: Learning a deep alternating neural network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5510816, doi: 10.1109/TGRS.2021.3081177.
[49] 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, doi: 10.1109/ACCESS.2018.2818280.
[50] 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, doi: 10.1109/TGRS.2021.3094884.

Bin Wang received the B.S. degree in information and computational science from the China University of Mining and Technology, Xuzhou, China, in 2008, and the M.Sc. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2011, and the Ph.D. degree in geographic information science from The Hong Kong Polytechnic University, Hong Kong, China, in 2015.

He is currently with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. His major research interests include spatial analyses and mathematical modeling and intelligence algorithms for remote sensing image processing.

Huizheng Yao received the B.S. degree in remote sensing science and technology from Shandong Agricultural University, Tai'an, China, in 2021. He is currently working toward the master's degree in surveying and mapping engineering with the School of Marine and Spatial Information, China University of Petroleum (East China), Qingdao, China.

Dongmei Song received the B.S. and M.Sc. degrees in ecology from Shenyang Agricultural University, Shenyang, China, in 1997 and 2000, respectively, and the Ph.D. degree in landscape ecology from the Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, in 2003.

She then studied as a Postdoctoral Researcher with the Institute of Geographic Sciences and Natural Resources Research, CAS, China, in 2005. She is currently a Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. Her major research interest focuses on remote sensing image processing.

Jie Zhang received the B.S. and M.S. degrees in mathematics from Inner Mongolia University, Hohhot, China, in 1984 and 1987, respectively, and the Ph.D. degree in applied mathematics from Tsinghua University, Beijing, China, in 1993.

He is a Professor with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China, and the Laboratory of Marine Physics and Remote Sensing, Ministry of Natural Resources, First Institute of Oceanography, Qingdao. He has a broad interest in marine physics and remote sensing applications. His research mainly focuses on the following: the SAR retrieval of ocean dynamics' processes and the SAR detection of marine targets, ocean hyperspectral remote sensing, high-frequency surface-wave radar ocean detection techniques, and the integration of marine remote sensing application systems. He has served as a member of multiple domestic/international committees and a principal investigator/coinvestigator of many projects from the National Science Foundation of China, the State High-Tech Development Plan (863), and other funding agencies. He has been the supervisor of nearly 40 Ph.D. degree students and has authored or coauthored more than 200 articles.

Han Gao (Member, IEEE) received the B.S. and M.S. degrees in geodesy and surveying engineering and the Ph.D. degree in photogrammetry and remote sensing from Central South University, Changsha, China, in 2015, 2018, and 2022, respectively.

Since 2022, he has been a Lecturer with the College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao, China. His research interests include crop and ocean remote sensing, time-series polarimetric SAR image processing, and pattern recognition.

Dr. Gao is a Reviewer for *Remote Sensing of Environment*, *IEEE Geoscience and Remote Sensing Letters*, and several other international journals in the remote sensing and image processing field.