An Improved Hyperspectral Unmixing Approach Based on a Spatial–Spectral Adaptive Nonlinear Unmixing Network

Xiao Chen, Xianfeng Zhang, Miao Ren, Bo Zhou, Ziyuan Feng, and Junyi Cheng

***Abstract*—The autoencoder (AE) framework is usually adopted as a baseline network for hyperspectral unmixing. Totally an AE performs well in hyperspectral unmixing through automatically learning low-dimensional embedding and reconstructing data. However, most available AE-based hyperspectral unmixing networks do not fully consider the spatial and spectral information of different ground features in hyperspectral images and output relatively fixed ratios of linear and nonlinear photon scattering effects under different scenarios. Therefore, these methods have poor generalization abilities across different ground features and scenarios. Here, inspired by the two-stream network structure, we propose a spatial–spectral adaptive nonlinear unmixing network (SSANU-Net) in which the spatial–spectral information of hyperspectral imagery is effectively learned using the two-stream encoder, followed by the simulation of the linear–nonlinear scattering component of photons using a two-stream decoder. Additionally, we adopt a combination of spatial–spectral and linear–nonlinear components using the optimized adaptive weighting strategy of learnable parameters. Experiments with several hyperspectral image datasets (i.e., Samson, Jasper Ridge, and Urban) showed that the proposed SSANU-Net network had higher unmixing accuracy and generalization performance compared with several conventional methods. This demonstrates that SSANU-Net represents a novel method for hyperspectral unmixing analysis.**

***Index Terms*—Adaptive weighting, autoencoder (AE), hyperspectral imagery, nonlinear mixing, spatial–spectral adaptive nonlinear unmixing network (SSANU-Net).**

I. Introduction

HYPERSPECTRAL imaging (HSI) technology collects information across the electromagnetic spectrum and has recently seen great advances and received increasing attention. Hyperspectral image data provide high spectral resolutions, and this rich spectral information can be used to identify ground features in monitored areas [1]. Therefore, HSI has become a widely used remote sensing technology for various applications, including land cover classification [2], data fusion [3], [4], anomaly detection [5], [6], and precision agriculture [7]. However, the usefulness of hyperspectral images has been limited by their low spatial resolution and the complicated distribution of ground features. Single pixels in a hyperspectral image often represent several ground features in the image scene, also known as mixed pixels, which seriously affect the interpretability of hyperspectral remote sensing data [8]. For this reason, hyperspectral unmixing (HU) has been developed and examined extensively. HU aims to separate the spectra of single pixels into a set of spectrally pure features, called endmembers, and subsequently, determine the abundance fraction associated with each endmember.

Among the various HU methods reported to date, the linear mixing model (LMM) is the most widely applied [9]. This type of model assumes that incident light is reflected only once by each ground feature in the scene before it is collected by the sensor, which would mean that the observed spectrum is a linear combination of the component endmembers in the pixel scene. This hypothesis makes the physical interpretation of pixels relatively clear and simple. Therefore, based on LMM, various HU methods have been proposed, including the N-finder algorithm (N-FINDR) [10], vertex component analysis (VCA) algorithm [11], non-negative matrix decomposition (NMF) [12], and sparse learning method [13].

However, the assumption of the LMM is not always valid due to the presence of spectral variability (SV) [14] and nonlinear interactions [15]. SV refers to a certain degree of difference among endmember spectra due to illumination or atmospheric conditions [16], [17]. To solve this problem, some researchers modified the LMM with additional parameters to simulate SV, such as the augmented LMM (ALMM) [18]. Notably, nonlinear mixing models (NLMM) have been developed and applied for SV enhancement. These models consider multiple interactions among photons of multiple ground features and have achieved good unmixing performance in specific cases [19], [20], [21]. In practical applications, however, these methods usually utilize *a priori* knowledge of nonlinear interactions to build HU models and lack sufficient generalizability [22].

On the other hand, due to its extraordinary learning and data fitting capabilities, deep learning approaches have become increasingly common in modeling studies, and indeed deep learning has been successfully applied in HU [23], [24]. Furthermore, unmixing of hyperspectral images using a classifier model, based

Manuscript received 14 May 2023; revised 13 August 2023; accepted 27 September 2023. Date of publication 11 October 2023; date of current version 25 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 42171327. *(Corresponding author: Xianfeng Zhang.)*

The authors are with the Institute of Remote Sensing and Geographic Information, Peking University, Beijing 100871, China (e-mail: williamchen-x@pku.edu.cn; xfzhang@pku.edu.cn; renmiao@pku.edu.cn; zhioubo@stu.pku.edu.cn; 2201110698@stu.pku.edu.cn; junyicheng@pku.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3323748

© 2023 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

on a supervised learning HU method, has been proposed [25], [26]. The drawback of this class of methods is that they require a training set with known abundances. Unfortunately, it is difficult to obtain accurate abundances in practical scenarios, and thus, some other HU methods have to be utilized to generate the abundance sample set. Consequently, this reduces the usefulness of this class of methods.

This problem can be solved by using an autoencoder (AE) framework as a baseline network for self-supervised HU [22]. An AE usually consists of an encoder and a decoder, where the encoder is responsible for feature extraction and representation from input data and the decoder is responsible for data reconstruction from feature representation. Typically, the nonnegative constraint (ANC) and sum-to-one constraint (ASC) for abundance retrieval are imposed on feature representations so that the AE can be effectively trained by minimizing the reconstruction errors [27]. As a result, the AE is widely applied in HU and many improvements have been made to the basic AE to suit different functional requirements. To address the problem of relatively low signal-to-noise ratio (SNR) in hyperspectral images, non-negative sparse, and denoising, AEs have been applied in HU, achieving excellent unmixing performance in high noise environments [28], [29], [30]. Su et al. [31] introduced a two-stage AE methodology that uses a stacked AE to eliminate outliers, followed by another AE designed with non-negative sparse constraints to perform robust hypothesis testing. Vijayashekhar et al. [32] proposed a two-stage fully connected self-supervised deep learning network to alleviate these practical issues when performing blind HU. Ozkan et al. [33] proposed EndNet, which replaced the inner product of the encoder with spectral angular distance, and improved the sparsity of the estimated abundance by adding an additional penalty term. Wesley et al. [34] combined a generative adversarial network (GAN) and AE, employing a GAN joint training objective to condition the decoder to generalize to previously unseen abundance mixtures. In [35], Markov random fields and AE were combined to extract endmembers and estimate their abundances, using an inhomogeneous Gaussian Markov random field (IGMRF) as a regularized prior to produce more accurate unmixing results. These methods achieved good unmixing accuracy on both real and synthetic datasets, demonstrating the great potential of AE in the field of HU. However, these methods have focused more on modeling the spectral information of hyperspectral images, but the spatial relationship between a target pixel and its neighbors in an image has not yet been fully utilized.

Furthermore, to make full use of the spatial information of hyperspectral images, convolutional neural networks (CNNs) have also been applied in HU, and AEs combined with CNNs have achieved even higher unmixing accuracies [8], [36], [37], [38]. Zhang et al. [26] used the CNN to achieve good HU results for the first time, demonstrating its great potential in the field of spectral unmixing, but the algorithm utilized requires training samples, which limits its application potential. Palsson et al. [39] proposed a CNN-based blind spectral unmixing method that combines the CNN and AE for the first time. Hong et al. [8] proposed the EGU-Net and showed that, with similar network structures, networks using convolutional operators have better HU performance. However, the simple 2-D-CNN method independently performs cross-band analyses of hyperspectral images and ignores the spectral properties of the data. To address this, some researchers have divided images into cubic patches and used 3-D-CNN to jointly learn spatial and spectral features [40], [41], but the size of the cube will affect the accuracy of the HU drastically.

In addition, by combining high-order nonlinear interaction of photons with AE, additional methods have been developed [41], [42], [43] to simulate the nonlinear mixture of photons when incident to ground features. For example, Dhaini et al. [44] replaced linear layers with convolutional layers to simulate the nonlinear effects of photons and achieved good results.

Despite the aforementioned advances, the characterization of the high-order nonlinear effect of photons is still relatively crude, and most existing methods simply superimpose linear and nonlinear components. However, the ratio of linear and nonlinear photon interactions may vary in different practical scenarios. Furthermore, the contributions of spatial and spectral features to HU will also be quite different for various ground features in practical scenarios. Hence, only the effective combination of spatial and spectral information from hyperspectral images can achieve accurate HU results. Conventional AE-based HU methods also tend to focus only on either the spatial or spectral features of hyperspectral images. Although a few of studies have considered the spatial–spectral features in the networks, they adopt 3-D-CNN networks to learn spectral and spatial information simultaneously. Unfortunately, the complex implementation and structure of 3-D-CNN leads to relatively shallow learning and confusion of the spatial–spectral features. These problems also limit the potential for improvement in the accuracy of HU.

In the two-stream networks for the classification of hyperspectral images [36], [37], [38], one stream network is used to learn spatial features and the other stream network learns spectral features; the learned spatial and spectral features are fused together [36]. Inspired by this, we propose an AE-based spatial–spectral adaptive nonlinear unmixing network (SSANU-Net) to resolve the aforementioned problems. The contributions of this study are as follows.

1) The encoder of the SSANU-Net comprises a spatial–spectral two-stream network that learns spatial and spectral features of the images separately, and effectively distinguishes among different ground features through the adaptive weighting of learned spatial and spectral features.
2) The decoder of the SSANU-Net comprises two-stream network that models the linear and nonlinear components of interphoton interactions separately. The network can be applied to different scenarios due to its adaptive weighting strategy, which uses learnable parameters to simulate linear and nonlinear components.
3) The rectified linear unit (ReLU) function and a newly designed loss function are used to satisfy the ANC and ASC to improve HU performance compared to using the original *Softmax* function.

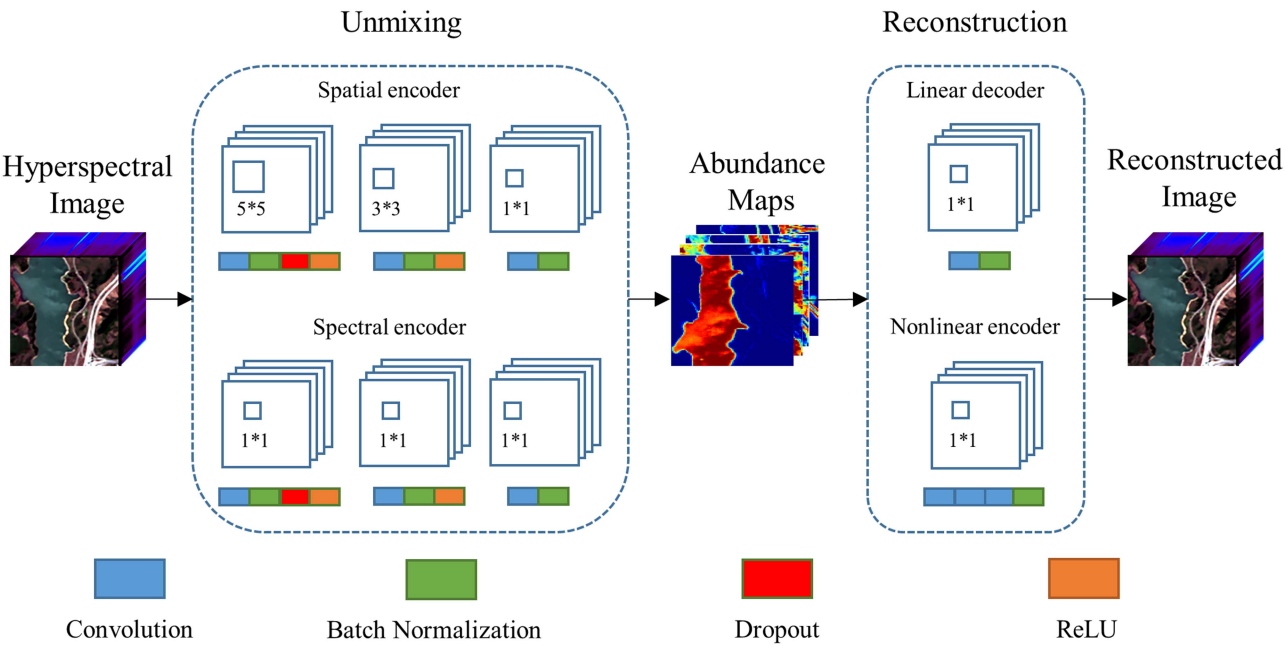

Fig. 1. Structure of the proposed SSANU-Net, including an encoder for unmixing and a decoder for reconstruction.

II. Framework and Methods

The mathematical expression of HU is introduced in this section followed by a novel solution that improves the unmixing performance.

A. Problem Formulation

Assume the observed hyperspectral image pixel $\mathrm{x} \in \mathbb{R}^B$, where B is the number of spectral bands. $\mathrm{M} = [m_1, m_2, \ldots, m_R]$ represents the $(\mathrm{B} * R)$ endmember matrix, where each column component $\mathrm{m_i}$ refers to the spectrum of the ith endmember, and R refers to the number of endmembers. $\mathbf{a} = [a_1, a_2, \ldots, a_R]^\top \in \mathbb{R}^R$ denotes the abundance component of each pixel.

A typical LMM model usually assumes that each observed pixel is obtained by the linear combination of abundance-weighted endmember spectra

$$\mathbf{x} = \mathbf{M} \times \mathbf{a} + \varepsilon \tag{1}$$

where $\varepsilon \in \mathbb{R}^B$ refers to the noise component.

Since the abundance represents the proportion of pixels occupied by each ground feature, a_i shall satisfy both ANC and ASC as

$$\forall i : a_i \geq 0 \tag{2}$$

$$\sum_{i}^{R} a_i = 1. \tag{3}$$

As mentioned previously, in complex scenes, there is nonlinear scattering of photons, and oftentimes, the more complex the scene the more severe the nonlinear scattering of photons. In such cases, the LMM assumption will not hold. Therefore, a nonlinear component and adaptive weights were introduced to (1) to obtain

$$\mathbf{x} = w_{d1} \times \mathbf{M} \times \mathbf{a} + w_{d2} \times \Psi(\mathbf{M}, \mathbf{a}) + \varepsilon \tag{4}$$

where ψ refers to the nonlinear component, and w_{d1} and w_{d2} are adaptive weights that vary with different scenarios.

B. General Framework for SSANU-Net

The proposed framework is an AE-based two-stream end-to-end network comprising one encoder and one decoder (see Fig. 1). The encoder f_E compresses the input hyperspectral image X into a low-dimensional representation of $\mathrm{H} \in \mathbb{R}^{R*N}$ as

$$\mathrm{H} = f_E(X) \tag{5}$$

$$f_E : \mathbb{R}^{B \times N} \rightarrow \mathbb{R}^{R \times N} \tag{6}$$

where H refers to the abundance matrix; B refers to the number of bands; N denotes the number of pixels; and R is the number of endmembers. Notably, R is known in this study and $R < B$. Decoder f_D is responsible for the decompression of H and raw data reconstruction

$$\hat{\mathrm{X}} = f_D(\mathrm{H}) \tag{7}$$

$$f_D : \mathbb{R}^{R \times N} \rightarrow \mathbb{R}^{B \times N}. \tag{8}$$

The detailed structure of the network is described in Sections II-C and D.

C. Encoder

A two-stream spatial–spectral adaptive weighted encoder is proposed in this study. The encoder contains spatial and spectral streams, each consisting of three blocks (see Table I).

TABLE I
NETWORK CONFIGURATION FOR EACH LAYER OF THE PROPOSED ENCODER STRUCTURE

Pathway	Spatial encoder	Spectral encoder	Unit
Block 1	5*5 Conv BN Dropout ReLU	1*1 Conv BN Dropout ReLU	128
Block 2	3*3 Conv BN ReLU	1*1 Conv BN ReLU	64
Block 3	1*1 Conv BN	1*1 Conv BN	R

TABLE II
NETWORK CONFIGURATION FOR EACH LAYER OF THE PROPOSED DECODER STRUCTURE

Pathway	Linear decoder	Nonlinear decoder	Unit
Block 1	1*1 Conv	1*1 Conv 1*1 Conv 1*1 Conv ReLU	B

In order to fully utilize the network capacity and reduce information loss, no specific constraint is imposed on the encoder and a stride strategy is used to compress the data instead of a pooling layer to reduce the size of the features. The same activation function is used for each layer, i.e., the ReLU function. The proposed encoder has spatial and spectral streams to learn spatial and spectral features, respectively, and compress the original hyperspectral image X into the abundance matrix H as

$$L_{\text{spa}} = f_{\text{Espa}}(X) \tag{9}$$

$$L_{\text{spe}} = f_{\text{Espe}}(X) \tag{10}$$

$$\mathrm{H} = w_{e1} \times L_{\text{spa}} + w_{e2} \times L_{\text{spe}} \tag{11}$$

where w_{e1} and $w_{e2} \in [0, 1]$ are learnable parameters that can achieve the adaptive weighted fusion of spatial and spectral features.

D. Decoder

A two-stream linear–nonlinear adaptive weighted decoder is proposed. The decoder contains linear and nonlinear streams, each consisting of one block, as shown in Table II.

The proposed decoder considers of both the linear component $\hat{\mathrm{X}}_{\text{lin}}$ and nonlinear component $\hat{\mathrm{X}}_{\text{nlin}}$ of photon interactions when reconstructing input signal $\hat{\mathrm{X}}$. Studies have shown that three-layer neural networks can represent arbitrary nonlinear relationships among inputs [9]. The proposed method uses one 1∗1 convolutional layer to simulate the linear interaction components of photons among different endmembers and three 1∗1 convolutional layers to simulate the higher order nonlinear interaction components of photons among different endmembers. To ensure acceptable unmixing performance, the endmember spectra are used to initialize the first convolutional layers of the linear and nonlinear decoders. Since endmember spectra are relatively easy to obtain (e.g., using algorithms such as VCA or field measurements), this does not affect the practicality of the proposed method.

It has been shown that, in simple scenarios, photons with high-order nonlinear interactions are typically exposed to weak interactions [9], while the nonlinear interactions induced by the multiple scattering of photons intensify as the complexity of the reaction environment increases [8]. Hence, reconstructed input images can be obtained by the adaptive weighted fusion of linear and nonlinear components

$$\hat{\mathrm{X}} = f_D(\mathrm{H}) = w_{d1} \times \hat{\mathrm{X}}_{\text{lin}} + w_{d2} \times \hat{\mathrm{X}}_{\text{nlin}} \tag{12}$$

where $\hat{\mathrm{X}}$ is the reconstructed input images, and $\mathrm{w_{d1}}$ and $\mathrm{w_{d2}} \in [0, 1]$ are learnable parameters for controlling the weights of linear and nonlinear components. The unmixing becomes a question to minimize the difference between the input imagery and the reconstructed imagery.

E. Loss Function

The loss function of the proposed AE network consists of multiple components. The proposed unmixing network is trained by minimizing the mean square error (MSE) between the input component x and the reconstruction component $\hat{\mathrm{x}}$:

$$L_{re} = \frac{1}{N} \times \sum_{i=1}^{N} (\hat{x}_i - x_i)^2 \tag{13}$$

where *N* refers to the number of pixels.

The unmixing process applies ANC and ASC to the abundance component. In some cases [8], [45], *Softmax* has been employed to directly force the network output to meet the requirements of ANC and ASC. However, this approach has a slow convergence rate because the *Softmax* function only considers the relative probabilities between different categories, which does not produce sparse outputs, reducing the convergence accuracy [22]. In our work, the ReLU activation function was employed to ensure the non-negativity of abundance component $\boldsymbol{a}$, and a loss function was employed to ensure the ASC

$$L_{\text{sum}} = \sum_{i=1}^{N} \|1 - \sum_{j=1}^{R} \alpha_{ji}\|_1 \tag{14}$$

where α_{ji} refers to the elements in row *j* and column *i* of the abundance matrix; and *R* refers to the number of endmembers.

The endmember and abundance in the self-supervised decomposition of mixing pixels are unknown, this makes unmixing difficult and generally requires reasonable a priori information to constrain the problem [9]. In our study, considering that unmixing is a sparsity problem, nuclear norm was used for the sparse constraint of abundance matrix H

$$L_{spa} = \|\mathrm{H}\|_*. \tag{15}$$

The final loss function of the network can be expressed as

$$\mathrm{L} = L_{\text{re}} + \lambda \times L_{\text{sum}} + \gamma \times L_{\text{spa}} \tag{16}$$

where λ and γ are two superparameters to control the intensities of the two loss functions.

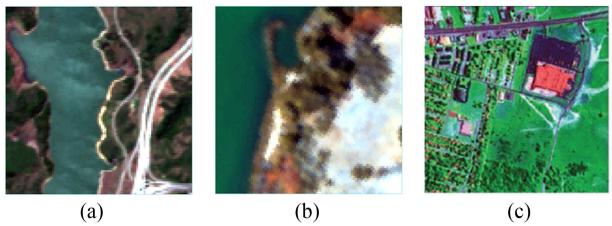

Fig. 2. True color synthesis images of the three hyperspectral datasets used in the experiment. (a) Jasper Ridge dataset. (b) Samson dataset. (c) Urban dataset.

F. Evaluation Metrics

The spectral angel distance (SAD) and root mean square error (RMSE) were employed in our work to assess the unmixing results

$$\mathrm{SAD} = \cos^{-1}\left(\frac{m^T \times \hat{m}}{\|m\| \times \|\hat{m}\|}\right) \tag{17}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \times \sum_{i=i}^{N} \|\alpha_i - \hat{\alpha}_i\|_2^2} \tag{18}$$

where m and $\hat{m}$ refer to the practical endmember and the endmember extracted by the model, respectively, which are obtained by averaging the spectra of pixels with abundances greater than 0.9; and α_i and $\hat{\alpha}_i$ represent the practical abundance and estimated abundance, respectively.

III. Experiments and Results

In order to evaluate the performance of the proposed network, one synthetic and three real hyperspectral image datasets were used in the experiments. The true color synthesis images of the three real hyperspectral image datasets are shown in Fig. 2. The random seed was set as 0. The experiments in this section were implemented on a Windows Server in the Python 3.6, Pytorch 1.8.1, and sklearn environments (CPU: Intel Xeon Silver 4116 at 2.1 GHz; RAM: 128 GB; GPU: GPU: NVIDIA RTX 1080TI).

Four typical and state-of-the-art HU methods were used for comparison, including the sequential maximum angle convex cone (SMACC) [46], the cycle-consistency unmixing network by learning cascaded autoencoders (CyCU-Net) [22], the sparse convolutional unmixing network (SUnCNN) [47], and the 3-D-CNN unmixing frame considering SV (3DCNN-var) [40]. The CyCU-net and SUnCNN methods were implemented using the source code demos provided by the authors, while the other methods were implemented using in-house-created code since the authors did not provide the source code.

A. Data Description

1) *Synthetic dataset:* The Matern Gaussian field synthetic dataset [48] comprises of five endmembers with 431 bands from the United States Geological Survey (USGS) spectral library as endmember spectra (see Fig. 3). These endmember spectra were used to generate a simulated dataset in the wavelength range of 0.39 $\sim$ 2.56 μm using the hyperspectral images synthesis (HYDRA) toolbox. The resulting abundances satisfy the ANC and ASC, and the image was 128 $\times$ 128 pixels in size. The dataset also incorporated SNR values ranging from 0 to 40 dB, which were used to verify the robustness of the model.

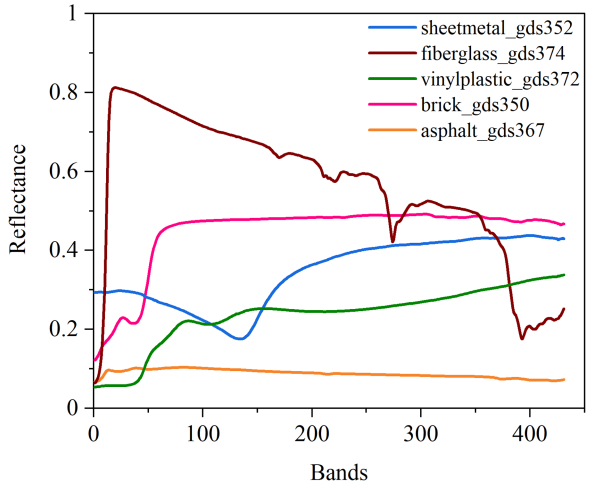

Fig. 3. Endmember spectra of the synthetic dataset.

2) *Samson dataset:* Obtained by the Samson sensor [43] and one of the most widely used hyperspectral datasets for HU. The original image was 952 $\times$ 952 pixels in size and had 156 bands and a wavelength range of 0.401 $\sim$ 0.889 μm. The adopted scene was an image cropped from the original image of 95 $\times$ 95 pixels. In this study, three main endmembers were studied, i.e., #1 soil, #2 tree, and #3 water. The dataset contains hyperspectral images, endmember spectra, and abundance reference data.
3) *Jasper Ridge dataset:* Acquired by the Jet Propulsion Laboratory (JPL) airborne visible/infrared imaging spectrometer (AVIRIS). The original image size was 512 $\times$ 614 pixels, distributed over 224 bands from 0.38 $\sim$ 2.50 μm [49]. The adopted scene was a popular region of interest (ROI) of 100 $\times$ 100 pixels, wherein 198 bands were retained after removing the water vapor and atmospheric influence channels. Four main materials were investigated in this scene: #1 water, #2 soil, #3 tree, and #4 roads. The dataset also contains hyperspectral images, endmember spectra, and abundance references.
4) *Urban dataset:* Obtained by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor [50] and one of the most widely used hyperspectral datasets used in the HU field [36], [51]. The image was 307 $\times$ 307 pixels in size and had 210 recorded bands covering the 400 $\sim$ 2500 nm wavelength range. However, several channels (i.e., 1–4, 76, 87, and 101–111, 136–153, and 198–210) were removed due to water-vapor absorption and atmospheric effects. The dataset comprises four constituent materials: asphalt (#1), grass (#2), tree (#3), and roof (#4). Additionally, the dataset includes hyperspectral images, endmember spectra, and abundance references.

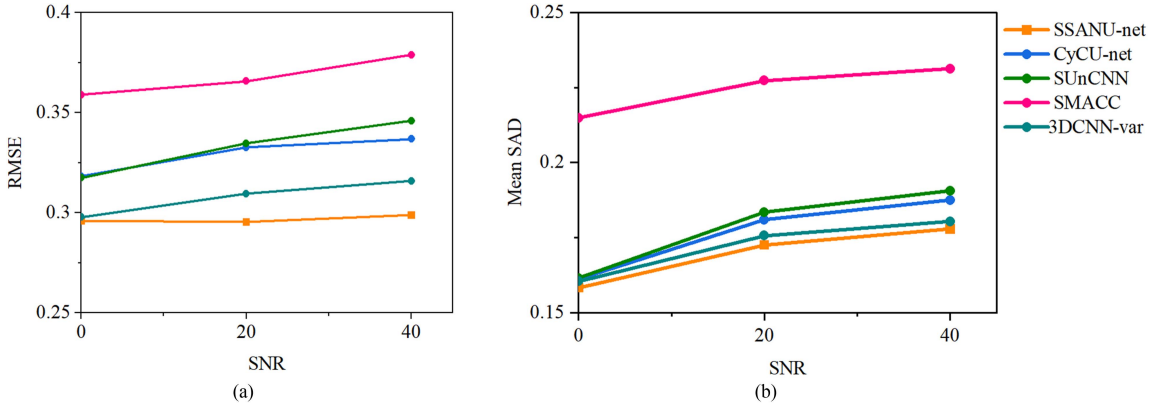

Fig. 4. Robustness assessment of the synthetic dataset under different SNRs. (a) RMSE; (b) Mean SAD.

TABLE III
UNMIXING OF THE SAMSON WITH THE BEST RESULTS HIGHLIGHTED IN BOLD

Methods		SMACC	SUnCNN	CyCU-net	3DCNN-var	SSANU-Net
S	Soil	00349	0.1361	0.033	0.0237	**0.0119**
A	Tree	0.0331	**0.0051**	0.0596	0.0434	0.0149
D	Water	0.0765	0.0371	0.0392	0.0314	**0.0114**
Mean SAD		0.0484	0.0594	0.0329	0.0985	**0.0127**
RMSE		0.4152	0.2118	0.2874	0.1750	**0.1668**

The bold values represent the best results by the models.

B. Noise Robustness Analysis With the Synthetic Dataset

To investigate the robustness of the proposed method, different SNRs from 20 to 40 dB were added into the synthetic experiment.

Experiments were conducted with the datasets of 0-, 20-, and 40-dB noise, respectively. The adaptive moment estimation (Adam) [52] optimizer was adopted, and the initial learning rate was set as 1×10^{-3}. The number of training epochs was set as 500. The initial values of w_{e1} and w_{e2} were set as 0.7 and 0.3, the initial values of w_{d1} and w_{d2} were set as 0.8 and 0.2, λ was set as 1×10^{-7}, and γ was set as 1×10^{-5}. Fig. 4 shows the quantitative results obtained using the synthetic dataset under different SNR values in terms of the RMSE and mean SAD. Deep learning-based unmixing methods typically achieve better unmixing results than traditional methods such as SMACC. This is because the powerful fitting and learning capabilities of deep learning methods enable the network to better learn image features and obtain more accurate unmixing results. Compared with the existing typical and state-of-the-art approaches, the proposed SSANU-Net achieved the best RMSE and mean SAD results, demonstrating its robustness and ability to more accurately extract endmember and abundance results.

C. Experiment With the Samson Dataset

For the dataset, the adaptive moment estimation (Adam) Adam optimizer was used. In extreme cases, a large learning rate may lead to fluctuations in the model accuracy during training, and a small learning rate may result in an inability of model fitting. In this study, the optimal initial learning rate was set as 1×10^{-2}. The number of training epochs was assigned as 500.

Appropriate initial values of w_{e1}, w_{e2}, w_{d1}, and w_{d2} will make the model converge more easily. Through experimentation, the optimal values in Samson dataset were determined to be 0.01, 0.99, 0.9, and 0.1, respectively.

The values of λ and γ can control the intensities of different loss functions. It is deduced that $\lambda = 1 \times 10^{-5}$ and $\gamma = 1 \times 10^{-5}$ lead to the best results in the Samson dataset.

Fig. 5 and Table III show the quantitative results and abundance plots of the Samson dataset using the proposed method and comparative methods. As observed, the proposed method obtained the best results, in terms of both RMSE and mean SAD, which demonstrated the effectiveness of the proposed network. Indeed, as shown in Table III, the unmixing method based on deep learning obtained the smallest RMSE results. Additionally, 3DCNN-var, which is similar to the proposed model in that it integrates the spatial–spectral features of hyperspectral images and the linear and nonlinear effects of photons, also obtained a high unmixing accuracy. However, the proposed method adopts a two-stream network structure that better learned the spatial–spectral features of the hyperspectral images and obtained the highest unmixing accuracy. Fig. 6 shows a comparison of the endmember spectra and ground truth values extracted with different methods. Although SUnCNN obtained the best results for the tree endmember spectra, the proposed network performed best overall in endmember spectra precision.

D. Experimental With the Jasper Ridge Dataset

In the experiment, the Adam optimizer was used, and the initial learning rate was set as 1×10^{-3}. The number of training epochs was as to 500. The initial values of w_{e1} and w_{e2} were

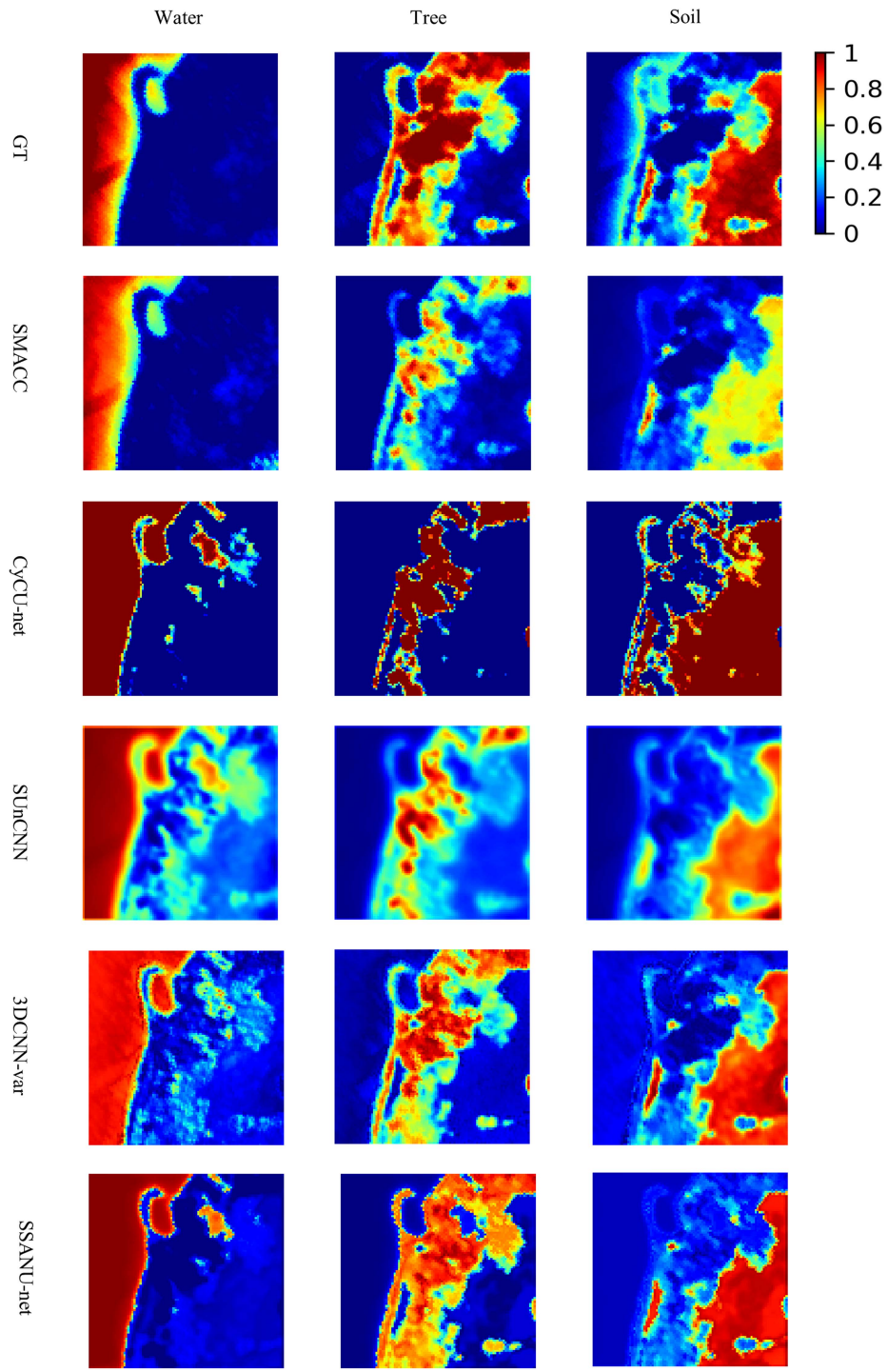

Fig. 5. Abundances of water, tree, and soil in the Samson dataset obtained by different methods.

assigned as 0.6 and 0.4, respectively, the initial values of w_{d1} and w_{d2} were assigned as 0.9 and 0.1, respectively, λ was set as 1×10^{-7}, and γ was set as 1×10^{-5}.

These hyperparameters were determined as the optimal values through experimentation. Table IV quantifies the experimental results of the proposed method and other comparative methods on Jasper Ridge dataset; the corresponding abundances are shown in Fig. 7. Theoretically, the differences in spectral features between trees and soils are relatively pronounced, so the unmixing of these two ground features should be relatively easy. However, the SADs of tree and soil obtained by most existing methods are relatively high (see Table IV). As shown in Fig. 7, the tree and soil abundances extracted by the existing methods also exhibited some confusion, possibly due to the complexity of the scene. The scene contains areas where soil and trees are mixed together, so there may be nonlinear scattering,

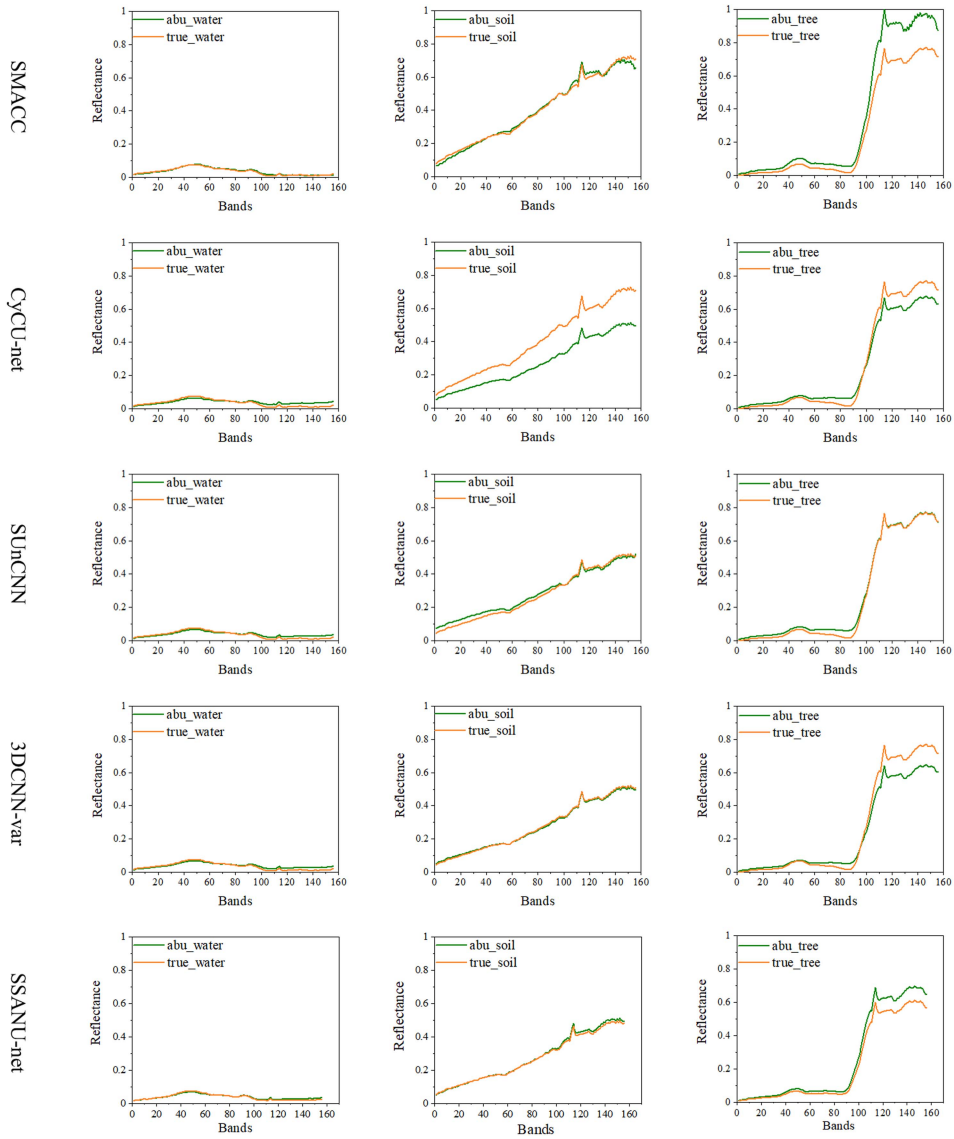

Fig. 6. Endmembers extracted by different methods and practical endmembers in the Samson dataset ("true" represents the measured endmember spectra and "abu" represents the modeled endmember spectra).

TABLE IV
UNMIXING OF THE JASPER RIDGE DATASET WITH THE BEST RESULTS HIGHLIGHTED IN BOLD

	Methods	SMACC	SUnCNN	CyCU-net	3DCNN-var	SSANU-Net
SAD	Water	**0.0323**	0.0595	0.1650	0.0682	0.0758
	Soil	0.0784	0.0377	0.0952	0.1274	**0.0226**
	Tree	0.1036	0.1488	**0.022**	0.0519	0.0322
	Road	0.0312	0.0317	0.1347	0.1174	**0.0312**
Mean SAD		0.0614	0.0694	0.1043	0.0912	**0.0404**
RMSE		0.5006	0.1794	0.1958	0.1997	**0.1570**

The bold values represent the best results by the models.

which results in the failure for most of the existing methods to obtain good unmixing performances. The proposed method, due to its comprehensive consideration of spatial–spectral and linear–nonlinear scattering factors, can distinguish well between mixed trees and soils, which is further demonstrated in the ablation study in Section III-D. Fig. 8 shows the endmember spectra extracted by different methods and the ground-truth values. Overall, the proposed network achieved the best results in terms of both RMSE and mean SAD, which demonstrated its effectiveness.

In summary, because the proposed SSANU-Net can better learn the nonlinear components in spectral mixing and fully exploit the spatial and spectral features of the image, it achieved the best unmixing performance in the aforementioned experiments.

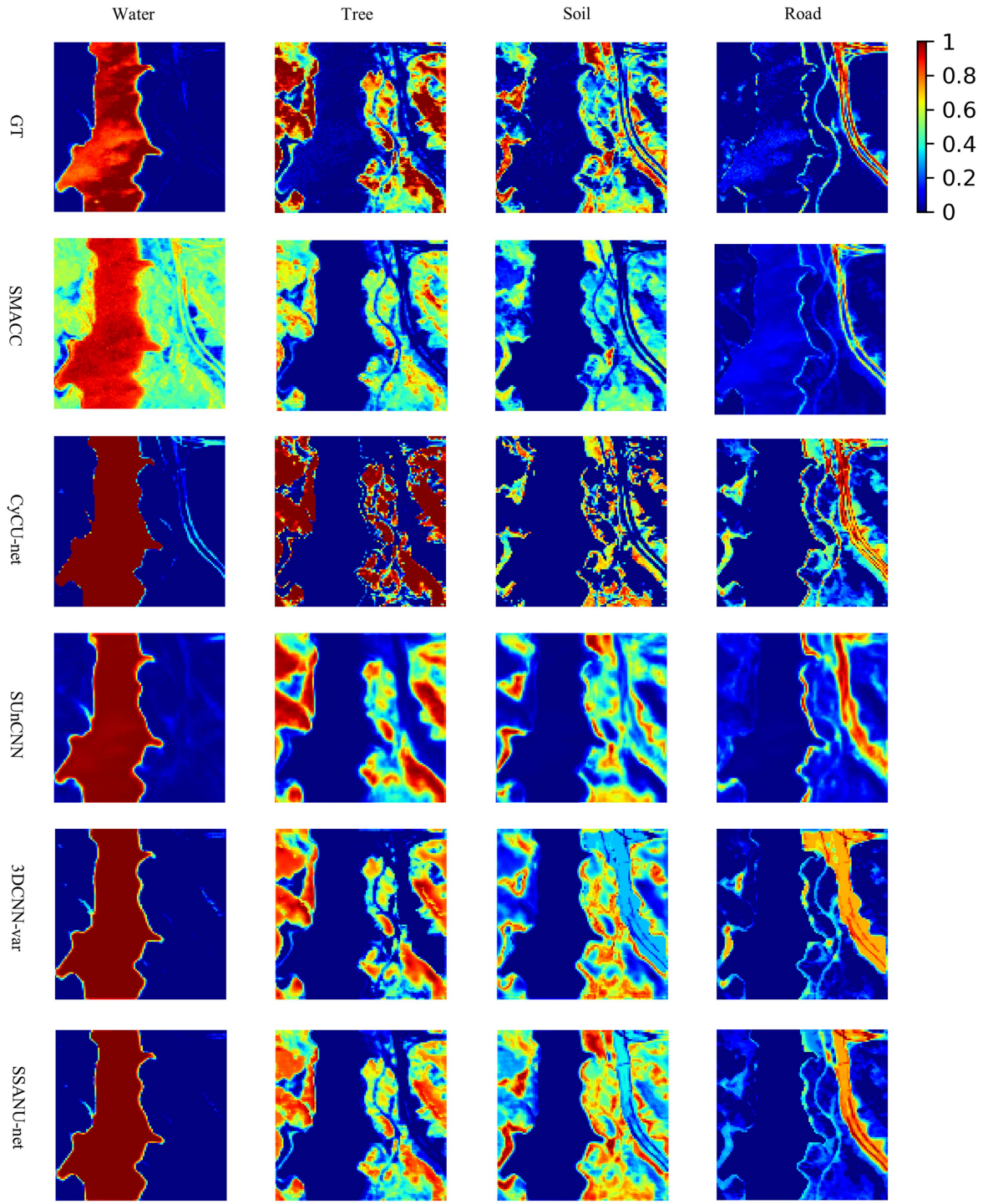

Fig. 7. Abundances of water, trees, soil, and roads in the Jasper Ridge dataset extracted by different methods.

E. Experiment With the Urban Dataset

In the experiment, the Adam optimizer was used, and the initial learning rate was set as 1×10^{-4}. The number of training epochs was set as 500. The initial values of w_{e1} and w_{e2} were assigned as 0.7 and 0.3, respectively, the initial values of w_{d1} and w_{d2} were assigned as 0.8 and 0.2, respectively, λ was set as 1×10^{-7}, and γ was set as 1×10^{-5}. These hyperparameters were determined as the optimal values through experimentation. The Urban dataset has a larger scale and more complex environment compared to the previous two real datasets. Table V quantifies the experimental results of the proposed method and the comparative methods on the Urban dataset; the corresponding abundances are shown in Fig. 9. The environment of the Urban dataset is complex and the spectral characteristics of different

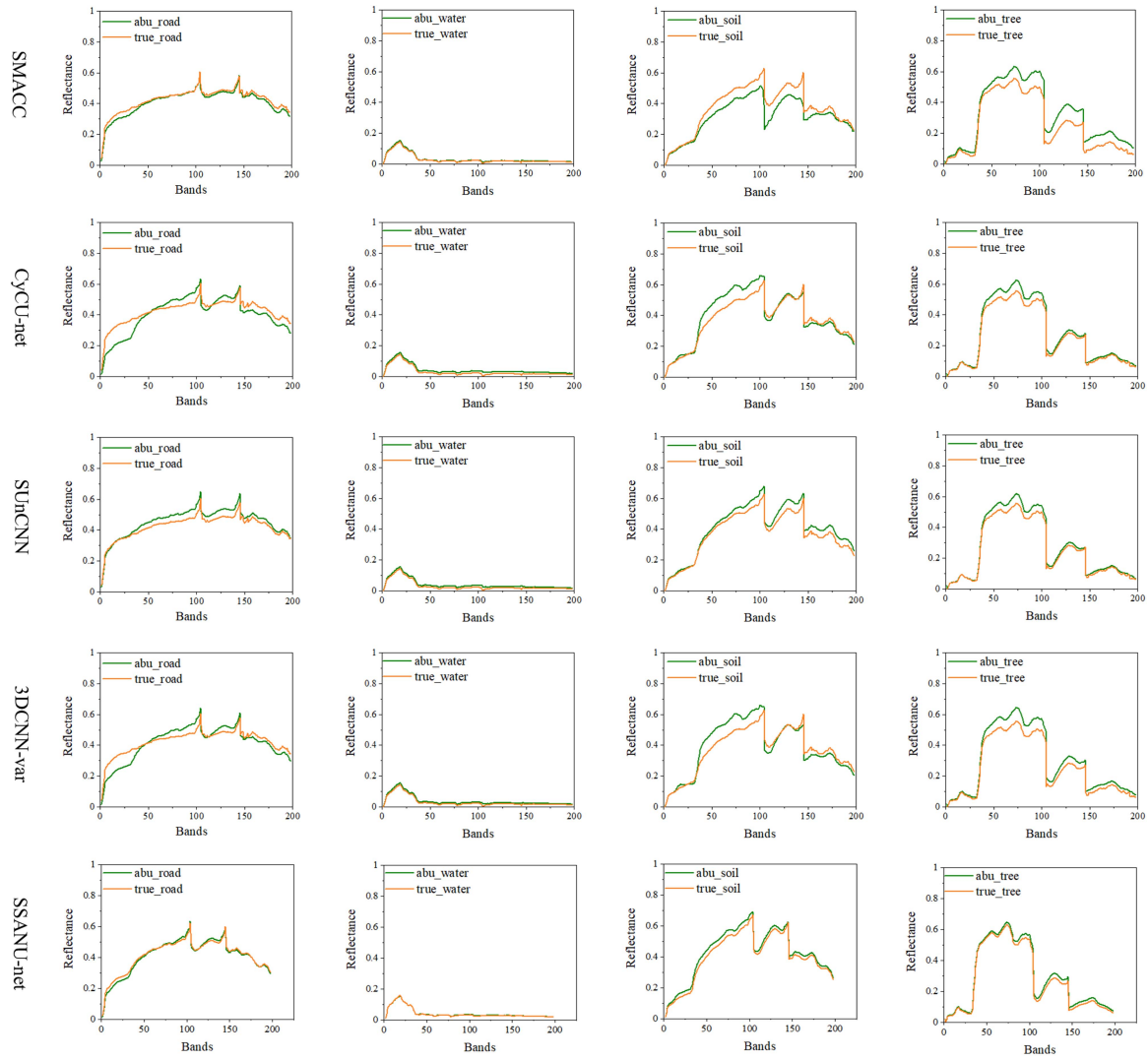

Fig. 8. Endmembers extracted by different methods and practical endmembers in the Jasper Ridge dataset ("true" denotes the measured endmember spectra and "abu" denotes the modeled endmember spectra).

TABLE V
UNMIXING OF THE URBAN DATASET WITH THE BEST RESULTS HIGHLIGHTED IN BOLD

Methods		SMACC	SUnCNN	CyCU-net	3DCNN-var	SSANU-Net
SAD	Asphalt	0.1316	0.0965	0.2065	0.0682	**0.0355**
	Grass	0.6530	0.0491	0.1895	0.0485	**0.0416**
	Tree	0.0743	0.1967	0.0706	0.0136	**0.0131**
	Roof	0.8670	0.2998	0.4704	0.2102	**0.1814**
Mean SAD		0.4314	0.1605	0.2343	0.0851	**0.0679**
RMSE		0.6291	0.2488	0.3365	0.2413	**0.2282**

The bold values represent the best results by the models.

endmembers can be quite similar, which makes the HU task more difficult. The traditional HU methods such as SMACC were almost completely unable to obtain correct abundance and endmember spectral results. Furthermore, HU methods based on traditional 2DCNNs, such as CyCU-net and SUnCNN, were unable to effectively learn the spatial–spectral features of the image, making it difficult to distinguish among easily confused endmembers, such as "tree" and "grass." The SSANU-Net method achieved good results for difficult-to-distinguish endmembers like "roof." This demonstrates the effectiveness of the proposed method in real complex scenes and also confirms its enormous potential for practical applications. Fig. 10 shows the endmember spectra extracted by different methods and the ground truth values. Overall, the proposed network achieved the best results in terms of both RMSE and mean SAD, which demonstrated its effectiveness.

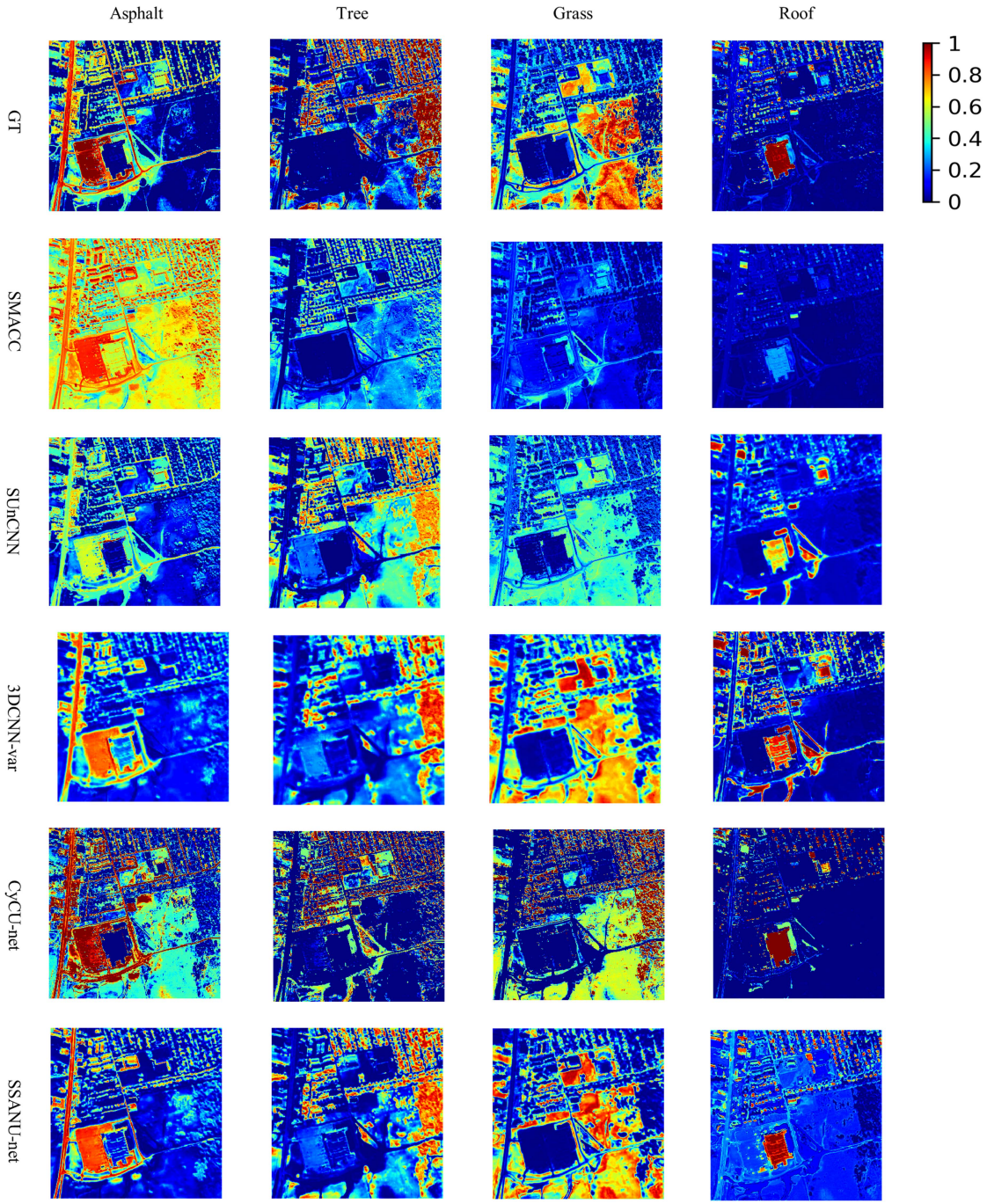

Fig. 9. Abundances of asphalt, tree, grass, and roof in the Urban dataset extracted by different methods.

F. Ablation Study

The proposed SSANU-Net model contains one spatial–spectral encoder and one linear–nonlinear decoder, which integrate the spatial–spectral features of hyperspectral images and the linear and nonlinear scattering of photons, thus it achieved better separating of the mixed pixels in the three real datasets and the synthetic dataset. To verify the effect of the network module on the unmixing results, we conducted ablation tests on the Samson dataset and Jasper Ridge dataset. The unmixing performances of the different modules in the SSANU-Net are summarized in Tables VI and VII.

As shown in Tables VI and VII, the best unmixing performance was achieved by the proposed SSANU-Net, which has all four modules. As shown in Table VI, the spectral module had the greatest impact on the unmixing performance of the Samson database, which was due to the simplicity of the data scene, the independent distribution of ground features, and the significant differences among the spectral features of water bodies, trees, and soils.

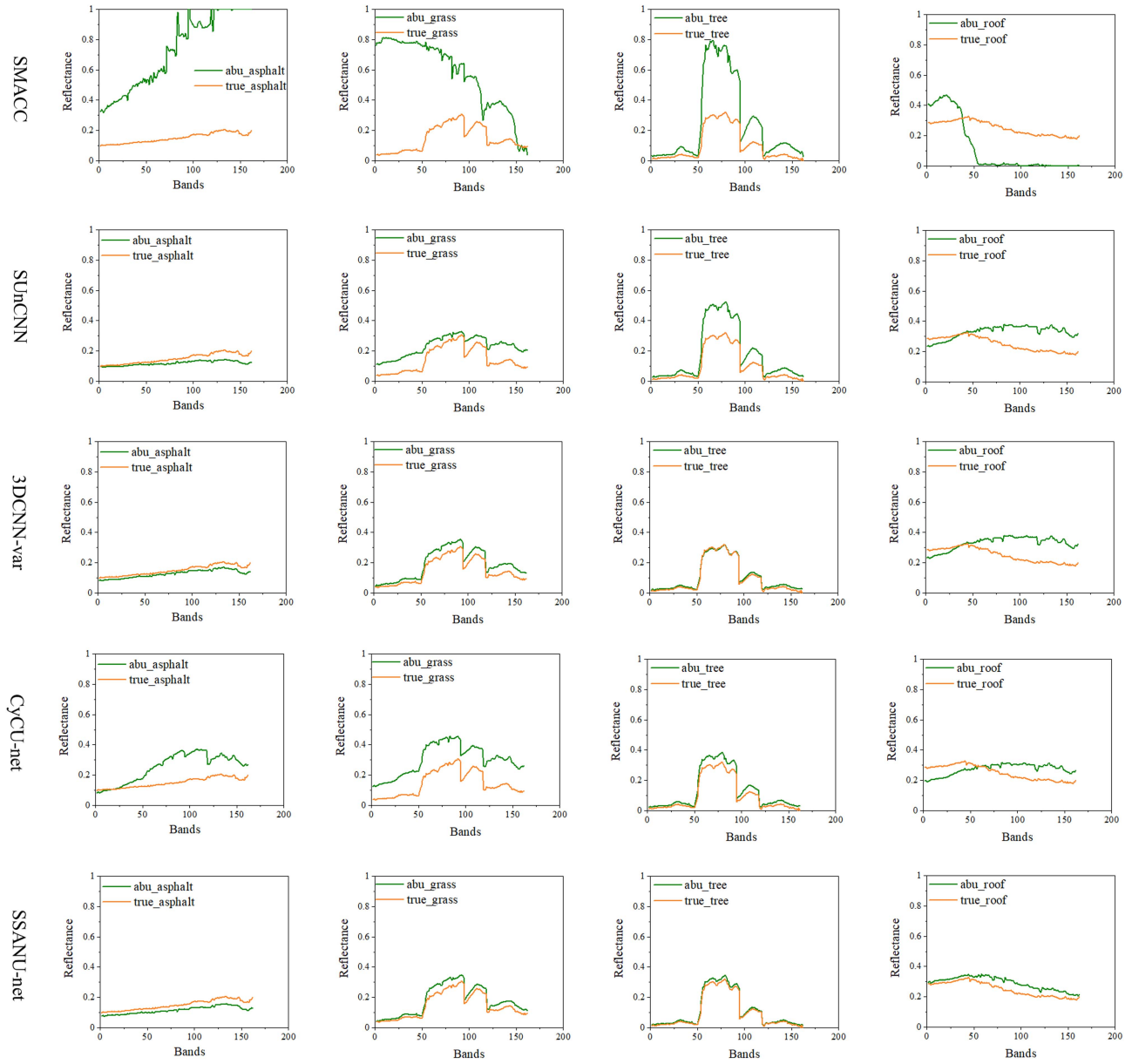

Fig. 10. Endmembers extracted by different methods and practical endmembers in the Urban dataset ("true" denotes the measured endmember spectra and "abu" denotes the modeled endmember spectra).

As shown in Table VII, the nonlinear module contributed much on the unmixing performance of the Jasper Ridge dataset. This was in line with our previous hypothesis that, when the scene is complex and trees and soil are heavily mixed, there is a certain degree of nonlinear scattering, which explained why the proposed method outperformed the compared methods. Moreover, the spatial module had greater contribution than the spectral module on the unmixing performance of the Jasper Ridge dataset. This is because the spectral features of land covers in the Jasper Ridge dataset present certain similarities, such as the soil and roads. Consequently, the unmixing performance is less effective when relying solely on spectral information.

The results of the ablation study on both datasets further demonstrated the effectiveness of the proposed SSANU-Net for mixed pixel decomposition problems in practical scenarios.

IV. Discussion

A. Assessment of the Unmixing Method

In this section, we aim to utilize the proposed SSANU-Net for spectral unmixing on some complex real-world scenarios, such as minerals, and evaluate the metrics of the unmixing methods. The HSI used was the well-known AVIRIS Cuprite with a size of 250 × 190 pixels in size and had 188 bands after removing the bands affected by the water vapor and low SNR [53]. This scene is quite complicated and contains multiple minerals. Based on the studies of [30] and [22], we selected 12 mineral end members. The endmember spectra and abundance maps extracted by the proposed SSANU-Net are shown in Fig. 11. The result demonstrated the potential application of the proposed SSANU-Net network in highly complex scenes.

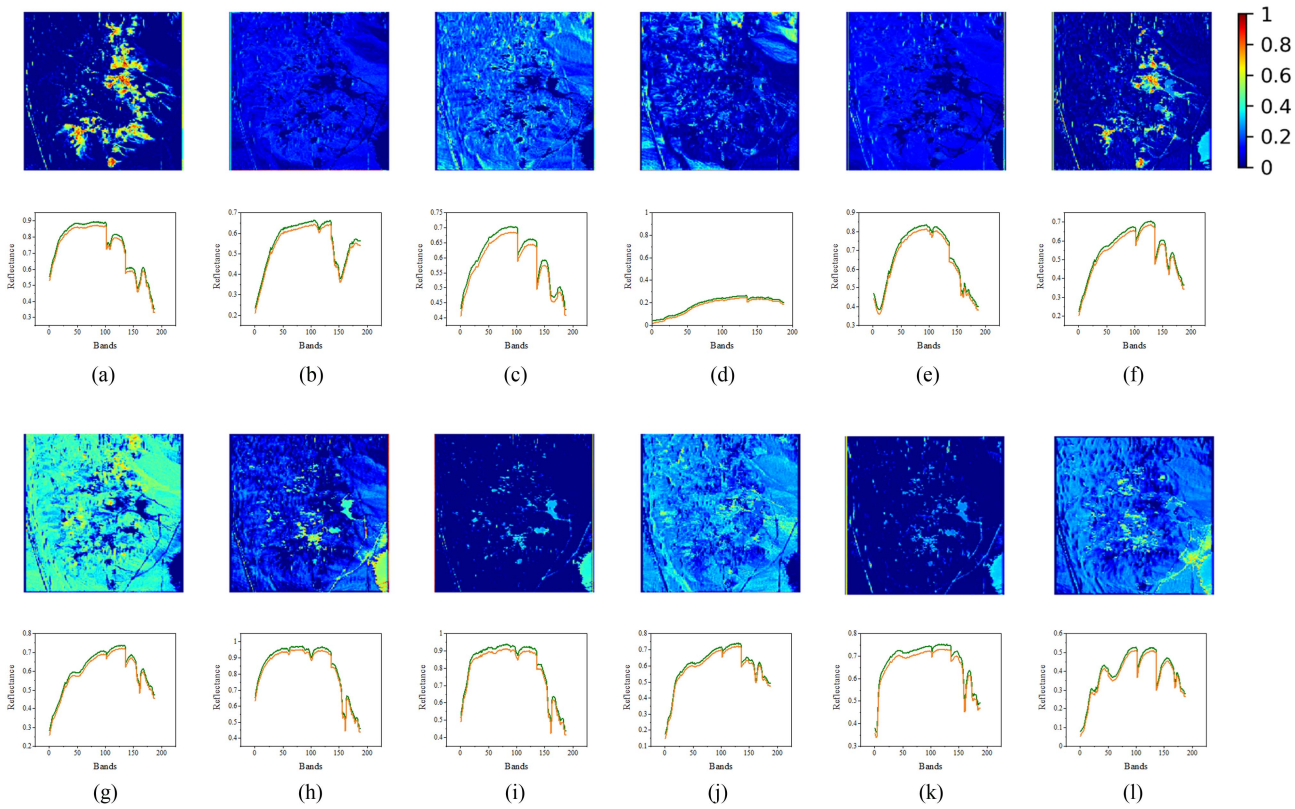

Fig. 11. Estimated endmembers (in orange) along with the references from the USGS library (in green) and their corresponding abundance maps obtained from the Cuprite mineral scene by the proposed SSANU-Net. The endmembers are labeled as follows. (a) Alunite. (b) Buddingtonite. (c) Chalcedony. (d) Desert Vanish. (e) Dumortierite. (f) Kaolin/SmectKLF506. (g) Kaolin/Smect H89-FR-5. (h) Kaolinite KGa-1 (wxyl). (i) Kaolinite KGa-2 (pxyl). (j) Montmorillonite. (k) Muscovite. (l) Nontronite.

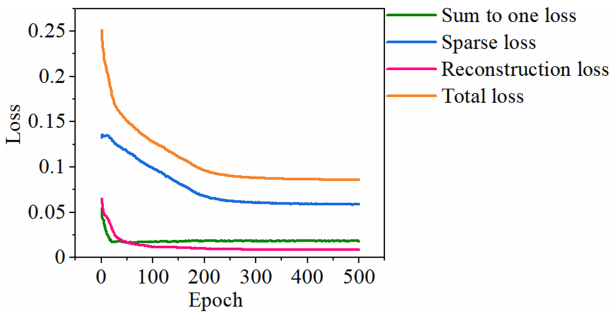

Fig. 12. Loss function curve for the Samson dataset. The sum to one loss and sparse loss were both weighted by hyperparameters λ and γ to balance their intensities.

TABLE VI
ABLATION TEST ON THE SAMSON DATASET

Spatial module	Spectral module	Linear module	Nonlinear module	RMSE	Mean SAD
✓	✓	✓	✓	**0.1668**	**0.0127**
✓		✓	✓	0.2784	0.0784
	✓	✓	✓	0.1736	0.0496
✓	✓	✓		0.1854	0.0646
✓	✓		✓	0.1879	0.0669

The bold values represent the best results by the models.

TABLE VII
ABLATION TEST ON THE JASPER RIDGE DATASET

Spatial module	Spectral module	Linear module	Nonlinear module	RMSE	Mean SAD
✓	✓	✓	✓	**0.1570**	**0.0404**
✓		✓	✓	0.1920	0.0815
	✓	✓	✓	0.2295	0.1078
✓	✓	✓		0.3182	0.2780
✓	✓		✓	0.2411	0.2681

The bold values represent the best results by the models.

Additionally, we have found that some methods excel in endmember spectral extraction but exhibit relatively poor performance in abundance extraction (see Figs. 5 and 6). We believe that the appearance of this phenomenon is due to the existence of spectral variability, which results in the ground-truth endmember spectra provided by the dataset may not be able to reflect the comprehensive spectral characteristics of the objects. Therefore, excessively pure endmembers may not necessarily yield better abundance results. Hence, a single metric cannot fully explain the effectiveness of a method.

Besides, we believe that in practical unmixing tasks, abundance is more crucial than endmember spectra. Endmember

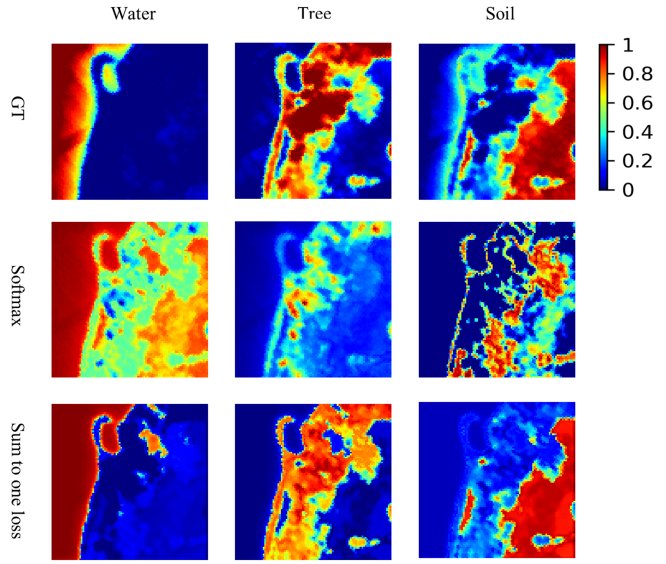

Fig. 13. Abundance maps of the Samson dataset using different abundance constraints ("*Softmax*" denotes the abundance with *Softmax* function and "Sum to one loss" denotes the abundance of the proposed method).

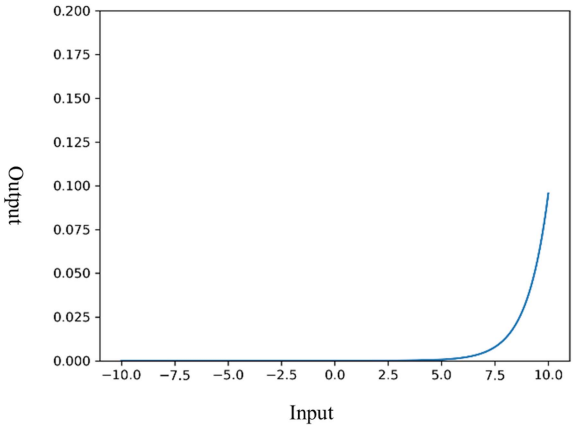

Fig. 14. Output of *Softmax* function with an input value range of [−10, 10].

TABLE VIII
COMPUTATIONAL COST OF ALL DEEP LEARNING METHODS ON DIFFERENT DATASETS IN TERMS OF FLOPS

Method	Samson	Urban
SUnCNN	8.05G	85.55G
CyCu-net	8.24G	67.71G
3DCNN-var	944.35G	14657.5G
SSANU-Net	**5.23G**	**63.8G**

The bold values represent the best results by the models.

spectra can be obtained through field measurements and similar means, whereas obtaining accurate abundance information is relatively more challenging. In consequence, we consider metrics like RMSE for assessing abundances and the visual representation of abundance maps to be more important than the SAD metric, which evaluates endmember spectra. And across multiple datasets, considering various metrics, the proposed method undoubtedly emerges as the optimal approach. Specifically, it outperforms all the compared methods in terms of extracting abundances.

B. Contribution of the Loss Function

As shown in (16), the loss function of the proposed SSAUN-net consists of three components: the reconstruction loss, the sparse loss, and the sum to one loss function. The learning curves of the multiple loss functions and the overall loss function on the Samson dataset are shown in Fig. 12. Compared to the sum-to-one loss, the sparse loss had more difficulty in converging. Therefore, it is necessary to appropriately increase the strength of the sparse loss.

In this study, the sum to one loss and the ReLU function were used instead of the *Softmax* function to satisfy the ANC and ASC of abundance. To investigate the performance of different abundance constraints, the *Softmax* function was adopted for comparison. Fig. 13 shows the abundances when these two strategies were used on the Samson dataset. Compared with the *Softmax* function, the proposed method yielded more discrete and significant abundance results. This was because when the input values were small, the variation in the *Softmax* function output was small (see Fig. 14) and the loss function could not be effectively updated, which led to a slower convergence in the model and increased the training difficulty. Additionally, the *Softmax* function greatly compresses the input value distribution, making it difficult to obtain a sparse abundance matrix using the *Softmax* function as the abundance constraint.

C. Computation Cost

We evaluated the computational cost of the proposed method based on floating point operations (FLOPs) on the Samson and Urban datasets. The results are shown in Table VIII. The computational cost of the proposed method mainly depends on the size of the dataset and the number of bands. Overall, the computational cost of the proposed method could be considered acceptable.

D. Applicability

Hyperspectral images have both spatial and spectral features, and because real land surfaces are often complex, there is both linear and nonlinear scattering of photons in the imaging process. The existing HU algorithms often rely on deep learning networks to learn in isolation, and do not take full advantage of the inherent properties of hyperspectral images. In this study, we developed a novel AE network that adopts a two-stream network to obtain the abundance learned from the original image features using a two-stream encoder for spatial–spectral features, and then, transforms the obtained abundance map into hyperspectral images using a two-stream decoder for linear–nonlinear interactions of photons to simulate the photon scattering process. Additionally, this study introduces learnable parameters as weights for the weighted fusion of spatial–spectral features and linear–nonlinear photon scattering components, respectively, which is

more suitable for spectral unmixing under practical scenarios. The aforementioned mechanism enables the proposed method to achieve good unmixing performance in complex scenes, such as the Urban dataset, where there is a high level of photon nonlinear scattering and different land cover categories with similar spectral characteristics. This undoubtedly contributes to the further practical applications of the proposed method.

V. Conclusion

This article presents an autoencoder-based self-supervised HU network SSANU-Net. The proposed model uses a two-stream spatial–spectral network as the encoder, a two-stream linear–nonlinear network as the decoder, and learnable parameters as weights for the adaptive weighted fusion of spatial–spectral features and linear–nonlinear scattering components. Compared with conventional self-supervised HU methods, the proposed method is superior in terms of network structure and can better learn spatial–spectral features from hyperspectral images and linear–nonlinear photon scattering component under simulated scenarios. Experiments were conducted on four real hyperspectral datasets and a comparative study was performed between the proposed SSANU-Net and four typical and state-of-the-art HU methods. As indicated, the proposed method achieved the best RMSE and mean SAD, especially for complex environments (e.g., where vegetation and soil were mixed). Additionally, the noise robustness of the proposed method was verified with the synthetic datasets under different SNRs.

To the best of our knowledge, this is the first study to apply the idea of a two-stream network to HU analysis. The results demonstrated the strong potential of using two-stream networks in learning hyperspectral spatial–spectral features and simulating the linear–nonlinear scattering of photons, which can be useful in further practical applications of HU analysis. To further improve the performance of the proposed study, future work will integrate the attention mechanism into the network.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments and questions, which help improve this manuscript.

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013, doi: 10.1109/MGRS.2013.2244672.

[2] D. Hong et al., "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019, doi: 10.1016/j.isprsjprs.2018.10.006.

[3] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019, doi: 10.1109/TGRS.2018.2890705.

[4] Y. Xu, Z. Wu, J. Chanussot, P. Comon, and Z. Wei, "Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 348–362, Jan. 2020, doi: 10.1109/TGRS.2019.2936486.

[5] M. S. Pagare and Y. R. Risodkar, "Low-rank and sparse representation for anomaly detection in hyperspectral images," in *Proc. Int. Conf. Adv. Commun. Comput. Technol.*, 2018, pp. 594–597.

[6] S. Feng, S. Tang, C. Zhao, and Y. Cui, "A hyperspectral anomaly detection method based on low-rank and sparse decomposition with density peak guided collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5501513, doi: 10.1109/TGRS.2021.3054736.

[7] J. A. J. Berni, P. J. Zarco-Tejada, L. Suarez, and E. Fereres, "Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 722–738, Mar. 2009, doi: 10.1109/TGRS.2008.2010457.

[8] D. Hong et al., "Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2022, doi: 10.1109/TNNLS.2021.3082289.

[9] 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, 2021, Art. no. 5509415.

[10] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Proc. SPIE*, 1999, pp. 266–275.

[11] J. M. P. Nascimento and J. M. B. Dias, "Does independent component analysis play a role in unmixing hyperspectral data?," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, pp. 175–187, Jan. 2005, doi: 10.1109/TGRS.2004.839806.

[12] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6076–6090, Oct. 2016, doi: 10.1109/TGRS.2016.2580702.

[13] A. Zare and P. Gader, "Sparsity promoting iterated constrained endmember detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 446–450, Jul. 2007, doi: 10.1109/LGRS.2007.895727.

[14] 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, doi: 10.1109/JSTSP.2018.2877497.

[15] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 82–94, Jan. 2014, doi: 10.1109/MSP.2013.2279274.

[16] 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, doi: 10.1109/MSP.2013.2279177.

[17] 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, doi: 10.1109/TIP.2016.2590324.

[18] 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, doi: 10.1109/TIP.2018.2878958.

[19] A. Marinoni and P. Gamba, "Improving reliability in nonlinear hyperspectral unmixing by multidimensional structural optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5211–5223, Aug. 2019, doi: 10.1109/TGRS.2019.2897430.

[20] A. Marinoni, J. Plaza, A. Plaza, and P. Gamba, "Nonlinear hyperspectral unmixing using nonlinearity order estimation and polytope decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2644–2654, Jun. 2015, doi: 10.1109/JSTARS.2015.2427517.

[21] F. Zhou et al., "Integrating convolutional neural network and gated recurrent unit for hyperspectral image spectral-spatial classification," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis.*, 2018, pp. 409–420.

[22] 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.

[23] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020, doi: 10.1109/MGRS.2020.2979764.

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

[25] G. A. Licciardi and F. D. Frate, "Pixel unmixing in hyperspectral data by means of neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4163–4172, Nov. 2011, doi: 10.1109/TGRS.2011.2160950.
[26] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional Neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1755–1759, Nov. 2018, doi: 10.1109/LGRS.2018.2857804.
[27] 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.
[28] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2015, pp. 1–4.
[29] Y. Su, A. Marinoni, J. Li, A. Plaza, and P. Gamba, "Nonnegative sparse autoencoder for robust endmember extraction from remotely sensed hyperspectral images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2017, pp. 205–208.
[30] 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.
[31] 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, doi: 10.1109/LGRS.2018.2841400.
[32] V. S. D. V. S. S and J. S. Bhatt, "A practical approach for hyperspectral unmixing using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5511505, doi: 10.1109/LGRS.2021.3127075.
[33] S. Ozkan, B. Kaya, and G. B. 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.
[34] J. H. Wesley and D. Qian, "Adversarially regularized autoencoder for hyperspectral image unmixing," *Proc. SPIE*, vol. 11533, 2020, Art. no. 115330U.
[35] J. R. Patel, M. V. Joshi, and J. S. Bhatt, "Spectral unmixing using autoencoder with spatial and Spectral regularizations," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 3321–3324.
[36] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spatial-spectral hyperspectral unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 357–360.
[37] 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.
[38] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 567–576, 2020, doi: 10.1109/JSTARS.2020.2966512.
[39] 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.
[40] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5521914, doi: 10.1109/TGRS.2022.3141387.
[41] 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, 2022, Art. no. 5509415, doi: 10.1109/TGRS.2021.3098745.
[42] M. Wang, M. Zhao, J. Chen, and S. Rahardja, "Nonlinear unmixing of hyperspectral data via deep autoencoder networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1467–1471, Sep. 2019, doi: 10.1109/LGRS.2019.2900733.
[43] Y. Su, J. Li, H. Qi, P. Gamba, A. Plaza, and J. Plaza, "Multi-task learning with low-rank matrix factorization for hyperspectral nonlinear unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2127–2130.
[44] M. Dhaini et al., "End-to-End convolutional autoencoder for nonlinear hyperspectral unmixing," *Remote Sens.*, vol. 14, 2022, Art. no. 3341.
[45] Z. Hua et al., "Gated autoencoder Network for Spectral–Spatial Hyperspectral unmixing," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3147.
[46] H. G. John, J. R. Anthony, and L. H. Michael, "The sequential maximum angle convex cone (SMACC) endmember model," *Proc. SPIE*, vol. 5425, pp. 1–14, 2004.
[47] B. Rasti and B. Koirala, "SUnCNN: Sparse unmixing using unsupervised convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 5508205, doi: 10.1109/LGRS.2021.3100992.
[48] *Hyperspectral Imagery Synthesis (EIAs) Toolbox*, U. d. P. V. E. H. U. U. E. Grupo de Inteligencia Computacional, Spain. Accessed: Oct. 18, 2023. [Online]. Available: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MATLAB
[49] 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, 2020, doi: 10.1109/TCI.2019.2948726.
[50] J. R. Lee et al., "HYDICE: An airborne system for hyperspectral imaging," *Proc. SPIE*, vol. 1937, pp. 173–179, 1993.
[51] 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.
[52] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, USA, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980
[53] 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.

Xiao Chen received the B.S. degree in geographic information science from China Agricultural University, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, China.

His main research interest includes remote sensing for disaster management, deep learning, and remotely sensed image classification.

Xianfeng Zhang received the Ph.D. degree in geography from the University of Western Ontario, London, Canada, in 2005.

He is currently a full Professor and Deputy Director with the Institute of Remote Sensing and GIS, Peking University, Beijing, China. His research interests include remote sensing for natural hazard management, deep learning, remote sensing Big Data, and remote sensing for transportation.

Dr. Zhang is an Associate Editor for *International Journal of Applied Earth Observation and Geoinformation*.

Miao Ren is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, China.

His main research interests include pavement condition measurement, deep learning, and remote sensing for transportation.

Bo Zhou received the B.Eng. degree in science and technology of remote sensing from Wuhan University, Wuhan, China, in 2021. He is currently working toward the M.S. degree in photogrammetry and remote sensing with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, China.

His research interests include hyperspectral image pan-sharpening, deep learning, and remotely sensed image segmentation.

Ziyuan Feng received the B.Eng. degree from the China University of Mining and Technology, Beijing, China, in 2022. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing.

His research interests include multispectral and hyperspectral image fusion and deep learning.

Junyi Cheng received the Ph.D. degree in photogrammetry and remote sensing from the Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, China, in 2023.

His main research interests include deep learning and space-time data mining.