

# CyCU-Net: Cycle-Consistency Unmixing Network by Learning Cascaded Autoencoders

Lianru Gao<sup>1</sup>, Senior Member, IEEE, Zhu Han<sup>2</sup>, Student Member, IEEE, Danfeng Hong<sup>3</sup>, Member, IEEE, Bing Zhang<sup>4</sup>, Fellow, IEEE, and Jocelyn Chanussot<sup>5</sup>, Fellow, IEEE

**Abstract**—In recent years, deep learning (DL) has attracted increasing attention in hyperspectral unmixing (HU) applications due to its powerful learning and data fitting ability. The autoencoder (AE) framework, as an unmixing baseline network, achieves good performance in HU by automatically learning low-dimensional embeddings and reconstructing data. Nevertheless, the conventional AE-based architecture, which focuses more on the pixel-level reconstruction loss, tends to lose some significant detailed information of certain materials (e.g., material-related properties) in the reconstruction process. Therefore, inspired by the perception mechanism, we propose a cycle-consistency unmixing network, called CyCU-Net, by learning two cascaded AEs in an end-to-end fashion, to enhance the unmixing performance more effectively. CyCU-Net is capable of reducing the detailed and material-related information loss in the process of reconstruction by relaxing the original pixel-level reconstruction assumption to cycle consistency dominated by the cascaded AEs. More specifically, cycle consistency can be achieved by a newly proposed self-perception loss, which consists of two spectral reconstruction terms and one abundance reconstruction term. By taking advantage of the self-perception loss in the network, the high-level semantic information can be well preserved in the unmixing process. Moreover, we investigate the performance gain of CyCU-Net with extensive ablation studies. Experimental results on one synthetic and three real hyperspectral data sets demonstrate the effectiveness and competitiveness of the proposed CyCU-Net in comparison with several state-of-the-art unmixing algorithms.

**Index Terms**—Cascaded autoencoders (AEs), cycle consistency, deep learning (DL), hyperspectral unmixing (HU), remote sensing (RS), self-perception.

Manuscript received December 3, 2020; revised January 20, 2021 and February 24, 2021; accepted March 5, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 42030111 and Grant 41722108, by MIAI@Grenoble Alpes, (ANR-19-P31A-0003) and the AXA Research Fund. (Corresponding author: Danfeng Hong.)

Lianru Gao is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: gaolr@aircas.ac.cn).

Zhu Han and Bing Zhang are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: hanzhu19@mails.ucas.ac.cn; zb@radi.ac.cn).

Danfeng Hong is with the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), 82234 Wessling, Germany, and also with Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (e-mail: danfeng.hong@dlr.de)

Jocelyn Chanussot is with Université Grenoble Alpes, INRIA, CNRS, Grenoble INP, LJK, 38000 Grenoble, France, and also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China (e-mail: jocelyn@hi.is).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TGRS.2021.3064958>.

Digital Object Identifier 10.1109/TGRS.2021.3064958

## I. INTRODUCTION

**H**YPERSPECTRAL imagery (HSI) has received an increasing attention in remote sensing (RS) applications, such as land cover classification [1], [2], data fusion [3]–[5], and anomaly/target detection [6]–[8], due to its high spectral resolution, which enables varieties of ground objects to be identified and detected [9]. However, due to the relatively low spatial resolution of sensors and the complex distribution of materials, many mixed pixels exist in the HSI and inevitably degrade the performance of high-level data processing [10], [11]. To reveal the intrinsic material interaction of mixed pixels, hyperspectral unmixing (HU) has become an emerging strategy to address this issue. HU can be regarded as a source separation problem whose goal is to separate the measured spectrum as a combination of spectral signatures, termed endmembers, and a set of fractional abundances. In the RS community, HU techniques have been widely used in a variety of applications, such as mineral exploration [12], [13] and agriculture monitoring [14], [15].

The linear mixing model (LMM) is a simple but effective model for spectral unmixing [16], which assumes the single interaction between the incident light and one material. Nevertheless, the assumption of an LMM is untenable due to the existence of spectral variability (SV) [17] and nonlinear interactions [18]. SV refers to a certain degree of endmember deformation because of lighting or atmospheric conditions [19], [20]. Most LMM-based approaches usually adopt additional parameters to model SVs. For example, Hong *et al.* [21] proposed a novel spectral mixing model, called the augmented LMM (ALMM), to model the principle scaling factors and other SVs simultaneously, yielding a desirable unmixing performance. To address more complex SV problems, nonlinear mixing models (NLMMs) have been proposed to consider physical interactions between multiple materials in a scene, which may occur at the intimate or multilayered level [22], [23]. Although numerous NLMMs have achieved good unmixing results in some specific scenarios [24]–[26], they usually require some prior knowledge about the characteristics of nonlinear interactions to establish the unmixing model in practice. Therefore, it is still necessary to find a more generalized way to unmix HSI.

In recent years, deep learning (DL) has attracted increasing attention in HU due to its powerful learning and data fitting ability [27]–[29]. The autoencoder (AE) framework, as an unmixing baseline network, has been proven to be

successful for HU by automatically learning low-dimensional embeddings (e.g., the abundance) and reconstructing the data with the corresponding base (e.g., the endmember) [30]. In general, AEs consist of two parts: an encoder and a decoder. The encoder extracts the representations of the data, and the decoder reconstructs the data from representations. By enforcing the nonnegative and sum-to-one constraints for representations, an AE can be trained quite efficiently in the process of minimizing reconstruction errors (REs) [31]. Moreover, to meet various functional requirements, many improvements have been made on the basis of AE architecture. For instance, nonnegative sparse and denoising AEs have been applied in HU by introducing superior denoising and self-adaption abilities, which achieved excellent unmixing performance, especially in highly noisy environments [32]–[34]. What is more, to fully synthesize the spatial information of HSI [35], convolutional neural networks (CNNs) [36]–[38] and multitask learning technologies [39] have been introduced into the traditional AE architecture to further enhance the unmixing accuracy. In addition, several other AE-based methods considering the second-order nonlinear interactions have been developed [40], [41], demonstrating the potential of AEs in the task of unmixing. Despite being able to be successfully applied in HU, these aforementioned AE-based unmixing methods, which focus more on the pixel-level reconstruction loss, tend to lose some detailed physical information of certain materials in the reconstruction process. As a result, the extracted endmembers and their abundances have certain errors compared with the ground truth (GT).

Inspired by the success of the perception mechanism, significant progress has been made in the area of image transformation by utilizing perceptual loss functions that depend on the extracted high-level feature to improve the network performance [42]–[45]. The reason is because the traditional pixel-level reconstruction similarity cannot guarantee that the extracted deep features are consistent [46], thereby resulting in detailed information loss during network training. For this reason, we propose a cycle-consistency unmixing network, called CyCU-Net, in which two cascaded AEs are trained in an end-to-end fashion, to effectively enhance the unmixing performance. CyCU-Net is capable of utilizing the learned high-level information dominated by cascaded AEs to further guide the unmixing network toward a more accurate estimation direction. More specifically, the main contributions of this article can be summarized as follows.

- 1) We propose an end-to-end deep network by cascading two AEs for HU tasks, called CyCU-Net. CyCU-Net is capable of estimating the abundance information of materials more efficiently and effectively than previous methods, thereby yielding a significant performance improvement.
- 2) A novel self-perception loss is introduced into the proposed CyCU-Net by designing a cycle-consistency strategy to further refine the detailed information in the unmixing process. More specifically, the proposed loss consists of two spectral reconstruction terms and one abundance reconstruction term. To the best of our

knowledge, this is the first study to transfer the high-level semantic information into an AE-based unmixing network by relaxing the original pixel-level reconstruction assumption and imposing a cycle-consistency constraint.

- 3) Due to the structural diversity of current AE-based unmixing methods, it is difficult to choose a suitable structure to meet unmixing needs. To this end, we investigate the performance gain of CyCU-Net with extensive ablation studies, involving different objective function types, abundance constraints, convolution filter sizes, and more complex unmixing scenes.

The remainder of this article is organized as follows. Section II describes the proposed CyCU-Net architecture. Section III reports the experimental results in synthetic and real hyperspectral data sets and gives a specific analysis. Finally, the conclusion is summarized in Section IV.

## II. PROBLEM FORMULATION AND METHOD

In this section, we start with a review of the existing AE approaches and then provide a detailed description of the proposed CyCU-Net method.

### A. Brief Review of AE-Based Unmixing Networks

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

$$\mathbf{X} = \mathbf{M}\mathbf{A} + \mathbf{N} \quad (1)$$

where  $\mathbf{X} \in \mathbb{R}^{L \times N}$  is the observed hyperspectral image with  $L$  bands and  $N$  pixels.  $\mathbf{N} \in \mathbb{R}^{L \times N}$  is the residual matrix, including the additive noise and other errors.  $\mathbf{M} \in \mathbb{R}^{L \times p}$  represents the endmember matrix with  $p$  endmember categories, and  $\mathbf{A} \in \mathbb{R}^{p \times N}$  denotes the corresponding abundance matrix.

Moreover, three physical constraints need to be satisfied in the unmixing problem. More precisely, the endmember matrix  $\mathbf{M} \geq 0$ . The abundance vectors  $\mathbf{a}_j$  should satisfy the abundance nonnegative constraint (ANC) and the abundance sum-to-one constraint (ASC) by the following equations:

$$\mathbf{a}_j \geq 0 \quad (2)$$

$$\sum_{i=1}^p \mathbf{a}_{ij} = 1. \quad (3)$$

Due to their powerful learning and reconstruction capabilities, AEs have become a typical representative in the field of HU. In general, AEs are simply composed of two parts: an encoder and a decoder.

1) *Encoder*: The encoder part transforms the input pixel  $\{\mathbf{x}_i\}_{i=1}^N \in \mathbb{R}^L$  into a hidden representation  $\mathbf{v}_i$  by utilizing some trainable network parameters, which can be expressed as

$$\mathbf{v}_i = f_E(\mathbf{x}_i) = f(\mathbf{W}^{(e)T} \mathbf{x}_i + \mathbf{b}^{(e)}) \quad (4)$$

where  $f$  denotes a nonlinear activation function, such as the sigmoid, rectified linear unit (ReLU), or the leaky ReLU (LReLU) function.  $\mathbf{W}^{(e)}$  and  $\mathbf{b}^{(e)}$  represent the weight and bias, respectively, in the  $e$ th encoder layer.

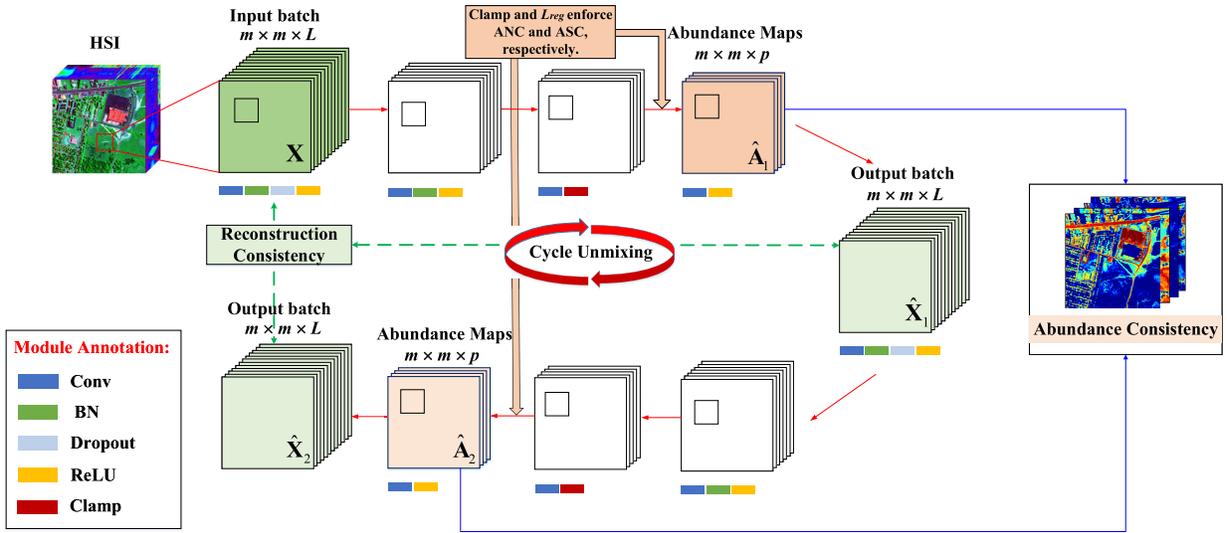


Fig. 1. Architecture of the proposed framework (CyCU-Net) that consists of two cascaded AEs that aim to utilize cycle-consistency constraints to enhance the unmixing performance. Specifically, the cycle-consistency constraints include the reconstruction and abundance consistency, respectively. BN stands for batch normalization.

2) *Decoder*: Based on the definition of the LMM, the decoder part adopts a hidden layer to reconstruct the input pixel from the representation  $\mathbf{v}_i$  and is denoted by

$$\hat{\mathbf{x}}_i = f_D(\mathbf{v}_i) = \mathbf{W}^{(d)T} \mathbf{v}_i \quad (5)$$

where  $\mathbf{W}^{(d)}$  is the weight matrix in the decoder part, which represents the connection between the hidden layer and the output layer, and  $\hat{\mathbf{x}}_i$  denotes the reconstructed pixel.

The training of the AE network is mainly realized by minimizing the mean squared error (MSE) objective function between  $\mathbf{x}_i$  and  $\hat{\mathbf{x}}_i$ , which is given by

$$L_{AE} = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{x}}_i - \mathbf{x}_i)^2. \quad (6)$$

When the entire network is well trained, the AE network can effectively encode and reconstruct HSI data. Due to the characteristic of (5), it can be understood that the output of the encoder is the estimated abundance vector  $\hat{\mathbf{a}}_i$ , and the weight matrix of the decoder represents the extracted endmember matrix  $\hat{\mathbf{M}}$ . Therefore, the process of solving endmembers and the corresponding abundances for most AE-based unmixing methods can be expressed as

$$\hat{\mathbf{a}}_i \leftarrow \mathbf{v}_i \quad (7)$$

$$\hat{\mathbf{M}} \leftarrow \mathbf{W}. \quad (8)$$

### B. Cycle-Consistency Unmixing Network

To reduce the detailed and material-related information loss in the process of reconstruction, the proposed CyCU-Net method is used to relax the original pixel-level reconstruction loss and introduce the cycle consistency to further explore the high-level semantic information in HSI. Fig. 1 illustrates the proposed CyCU-Net architecture.

The architecture of the proposed CyCU-Net is composed of two cascaded AEs, in which the output of one AE unmixing network is used as the input of the other same network.

This hybrid structure is designed to help the entire network make full use of the cycle-consistency constraints to improve the accuracy of unmixing. As shown in Fig. 1, the input batch  $\mathbf{X}$  extracted from HSI is compressed into the abundance  $\hat{\mathbf{A}}_1$  and then reconstructed into the first output batch  $\hat{\mathbf{X}}_1$  in the first AE network. Subsequently, after the second AE network, the abundance  $\hat{\mathbf{A}}_2$  and the second output batch  $\hat{\mathbf{X}}_2$  are generated again. Here,  $m$  represents the size of the input batch. Note that the cycle consistency is achieved by a newly proposed self-perception loss, including two reconstruction terms and one abundance term. The definition of self-perception loss is

$$L_S = L_R + \delta L_A \quad (9)$$

where  $\delta$  is the regularization parameter to control the balance between the reconstruction term and the abundance term. Specifically, the MSE losses of the network reconstruction and the abundance consistency are written as

$$L_R = \frac{\beta}{N} \sum_{i=1}^N (\hat{\mathbf{x}}_{1i} - \mathbf{x}_i)^2 + \frac{(1-\beta)}{N} \sum_{i=1}^N (\hat{\mathbf{x}}_{2i} - \mathbf{x}_i)^2 \quad (10)$$

$$L_A = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{a}}_{1i} - \hat{\mathbf{a}}_{2i})^2 \quad (11)$$

where  $\beta$  is the tradeoff parameter to balance the reconstruction consistency of the two cascaded AEs and the range of  $\beta$  is set to  $[0, 1]$ .  $\hat{\mathbf{a}}_{1i}$  and  $\hat{\mathbf{a}}_{2i}$  denote the  $i$ th abundance vector in  $\hat{\mathbf{A}}_1$  and  $\hat{\mathbf{A}}_2$ , respectively.

The network configuration for each AE in the proposed CyCU-Net architecture is shown in Table I. According to the characteristics of different layers, the architecture of each AE network is divided into six blocks. Among them, blocks 1–5 represent the encoder, and block 6 is the decoder. It is emphasized that the  $1 \times 1$  convolution (Conv) is used here to represent the operations of extracting features in the AE network since the input batch and the extracted abundance map are the same size. In fact, the  $1 \times 1$  Conv operation is

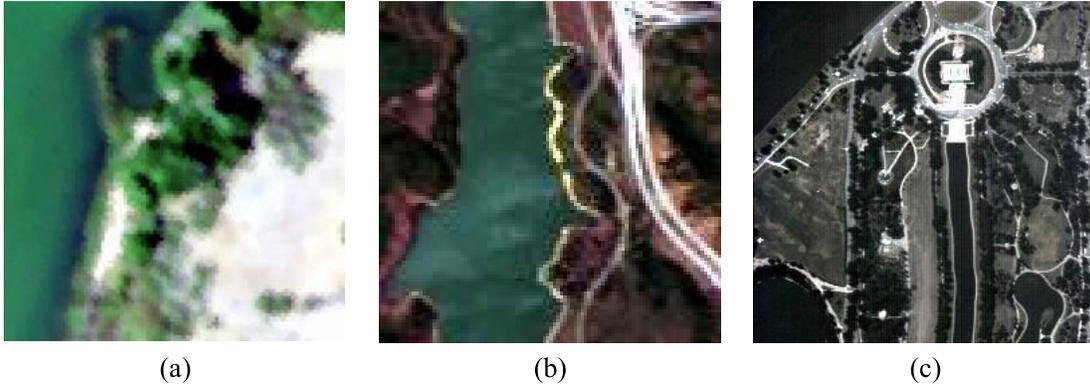


Fig. 2. RGB images from the three real hyperspectral data sets used in the experiment. (a) Samson. (b) Jasper Ridge. (c) Washington DC Mall.

TABLE I  
NETWORK CONFIGURATION FOR EACH AE IN THE  
PROPOSED CYCU-NET ARCHITECTURE

Pathway	Layer composition	Unit	Bias
Block 1	1 × 1 Conv BN Dropout ReLU	$L$	Yes
Block 2	1 × 1 Conv BN ReLU	$16p$	Yes
Block 3	1 × 1 Conv BN ReLU	$8p$	Yes
Block 4	1 × 1 Conv BN ReLU	$4p$	Yes
Block 5	1 × 1 Conv Clamp	$p$	Yes
Block 6	1 × 1 Conv ReLU	$L$	No

a dot product operation, so it is equivalent to the fully connected structure in the traditional AE network and there is no additional computation cost. Note that the proposed CyCU-Net can be extensible to spatial–spectral unmixing frameworks, such as  $3 \times 3$  and  $5 \times 5$  filters, to further improve the unmixing performance by combining the spatial information of HSIs. That is not to say, however, that the filter with a larger size must have a better performance. This is a tradeoff that will be further clarified by experiments in Section III-G3. The proposed cycle-consistency constraint can be seen as a network regularization. Since the single AE architecture is a severely nonconvex model, it makes it difficult to find the local optimal solution. Adding meaningful and important regularization will help shrink the solution space and get a better unmixing result. Before applying the activation function, batch normalization (BN) and dropout layers are applied to normalize the input data and prevent the SV effect to a certain extent.

To satisfy the ANC, the clamp function is applied to the output of the encoder to force all components of the abundance into the range of  $[0, 1]$ . Many researchers have introduced the clamp function to control the output of a specific network layer, which shows excellent performance in the

HSI compared to the traditional activation functions [47], [48]. In Section II-B, we will focus on analyzing the effect of the clamp function.

Moreover, to guarantee the ASC in (3), we limit the sum of each abundance vector to meet the sum-to-one property, which is given by

$$L_{\text{reg}} = \sum_{i=1}^N \left\| 1 - \sum_{j=1}^p \hat{\mathbf{a}}_{1ji} \right\|_1 + \sum_{i=1}^N \left\| 1 - \sum_{j=1}^p \hat{\mathbf{a}}_{2ji} \right\|_1 \quad (12)$$

where  $\hat{\mathbf{a}}_{1ji}$  and  $\hat{\mathbf{a}}_{2ji}$  represent the element in the  $j$ th row and the  $i$ th column abundance matrices  $\hat{\mathbf{A}}_1$  and  $\hat{\mathbf{A}}_2$ , respectively. Although the softmax function can directly force the output of the network to satisfy the ANC and ASC, it cannot produce sparse output and achieve a certain convergence accuracy since the softmax function considers only the relative probability between different categories, leading to the lower convergence rate. Compared with the softmax method, the proposed strategy in (12) can obtain more separate and accurate abundance results. Section III will explain this result in detail.

Finally, the overall loss of CyCU-Net can be formulated as

$$L = L_S + \gamma L_{\text{reg}} \quad (13)$$

where  $\gamma$  is the tradeoff parameter used to balance the cycle-consistency constraint and abundance regularization. It is obvious that three parameters ( $\beta$ ,  $\delta$ , and  $\gamma$ ) together affect the performance of the entire network. To be specific, the parameters  $\beta$  and  $\delta$  jointly determine the cycle-consistency constraint of the network, while the parameter  $\gamma$  is adopted to control ASC. Here, due to the abundance consistency constraint, the extracted abundance results of  $\hat{\mathbf{A}}_1$  and  $\hat{\mathbf{A}}_2$  are basically similar. Therefore, in the subsequent analysis, we adopt  $\hat{\mathbf{A}}_1$  as the extracted abundance results in the proposed CyCU-Net method for the experimental analysis.

### III. EXPERIMENTS

To assess the performance of the proposed methods, one synthetic and three real hyperspectral data sets are adopted in the experiment. The RGB images of the three real hyperspectral data sets are shown in Fig. 2. Note that, instead of the Washington DC Mall data set, the GT endmembers and the corresponding abundances are referenced in [49].

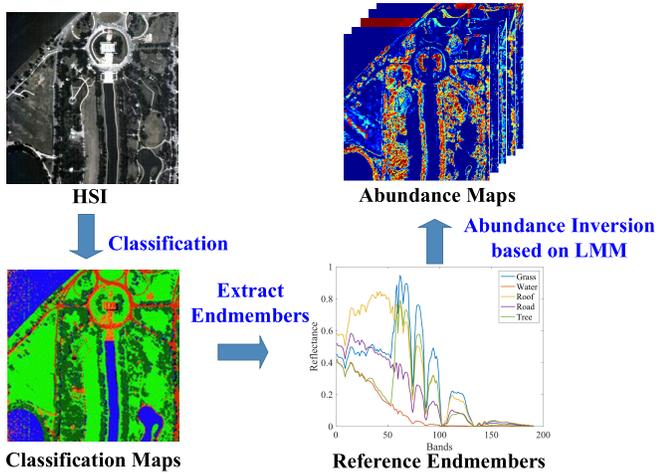


Fig. 3. Illustration of the processing chain for generating the reference GT in the Washington DC Mall data set.

Due to the lack of the GT in the Washington DC Mall data set, we adopt a simple but feasible way to generate the reference endmember and abundance results. For illustrative purposes, the flowchart depicting the reference GT acquisition is given in Fig. 3. First, we adopt the powerful classifiers for HSI to obtain the classification maps. According to the classification results, some spectral sets of different materials in small regions are manually selected, and then, the average spectral value is calculated as the reference endmember. Finally, the abundance maps are obtained by LMM-based methods, such as the scaled constrained least-squares (SCLS) method [17].

In addition, six classical and state-of-the-art unmixing methods are adopted for comparison: the fully constrained least-squares unmixing (FCLSU) [50], sparse unmixing by variable splitting and augmented Lagrangian (SUNSAL) [51], ALMM [21], spatial group sparsity regularized nonnegative matrix factorization (SGSNMF) [52], deep autoencoder unmixing (DAEU) [31], and CNN AE unmixing (CNNAEU) [37] techniques. For all the comparison methods, the initial endmembers are extracted by vertex component analysis (VCA) [53].

#### A. Data Description

1) *Synthetic Data Set*: The synthetic data set is simulated using six endmember references with 224 bands selected from the United States Geological Survey (USGS) spectral library in the range of 0.39–2.56  $\mu\text{m}$ , and the corresponding abundance maps satisfy the ANC and ASC with a size of  $100 \times 100$  pixels. To validate the effectiveness of the proposed method, two different mixing models are used to generate the synthetic data set: the LMM and the generalized bilinear model (GBM). It should be noted that the abundance maps follow a Dirichlet distribution, and the maximum purity of the synthetic data set is set to 0.9. Fig. 4 shows the illustration of the synthetic data set, including six endmembers and the data distribution result. Since the abundance satisfies the Dirichlet

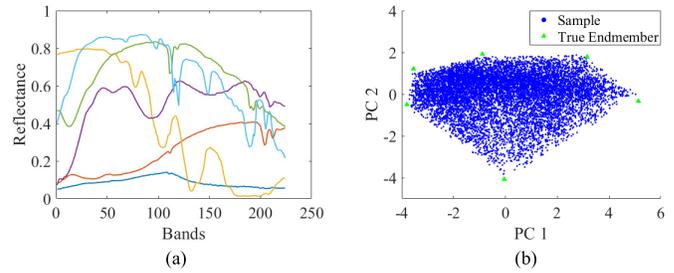


Fig. 4. Illustration of the synthetic data set, including six endmembers and the data distribution result. (a) Endmembers. (b) Results of data samples projected to two PCs.

distribution, the synthetic data samples are distributed in the  $p - 1$  dimensional simplex, as shown in Fig. 4(b).

2) *Samson Data Set*: The Samson data set obtained by the SAMSON sensor [54] is one of the most widely used hyperspectral data sets for HU. The original image contains  $952 \times 952$  pixels and 156 bands ranging from 0.401 to 0.889  $\mu\text{m}$ . The adopted scene is a  $95 \times 95$  pixel image cropped from the original image. In this studied scene, three main endmembers are investigated, that is, #1 *Soil*, #2 *Tree*, and #3 *Water*.

3) *Jasper Ridge Data Set*: The Jasper Ridge data set was captured via the airborne visible/infrared imaging spectrometer (AVIRIS) of the Jet Propulsion Laboratory (JPL). The original image is  $512 \times 614$  pixels recorded in 224 bands ranging from 0.38 to 2.50  $\mu\text{m}$ . We select a popular region of interest (ROI) with a size of  $100 \times 100$  pixels and retain 198 bands after removing the channels affected by the water vapor and atmospheric effects. Four dominant materials in this scene are investigated: #1 *Water*, #2 *Soil*, #3 *Tree*, and #4 *Road*.

4) *Washington DC Mall Data Set*: The Washington DC Mall data set was collected by the hyperspectral digital image collection experiment (HYDICE) sensor [55]. The original image is  $1208 \times 307$  pixels and has 210 bands, covering the spectra channels from 0.40 to 2.40  $\mu\text{m}$ . We investigate a  $256 \times 256$  pixel subimage cropped from the top-left corner of the original image. Due to the effect of water vapor and noise, only 191 bands remain. The five endmembers in this scene are #1 *Grass*, #2 *Water*, #3 *Roof*, #4 *Road*, and #5 *Tree*.

#### B. Experimental Setup

1) *Hyperparameter Settings*: The hyperparameters in the proposed CyCU-Net are listed as follows: the number of layers and neurons are determined empirically, and the size of the input batch  $m$  is 50. The Adam optimizer is utilized to update the network parameters, and the learning rate is set to  $10^{-3}$ . Moreover, the momentum is parameterized by 0.9, and the dropout rate is 0.9. The minibatch size is set to 20, and the maximum number of iterations is set to 500. When the maximum number of iterations is reached, CyCU-Net stops training and outputs the unmixing result.

2) *Evaluation Metrics*: In the experiment, three evaluation metrics are introduced to assess the performance of algorithms: spectral angle distance (SAD), root mean square

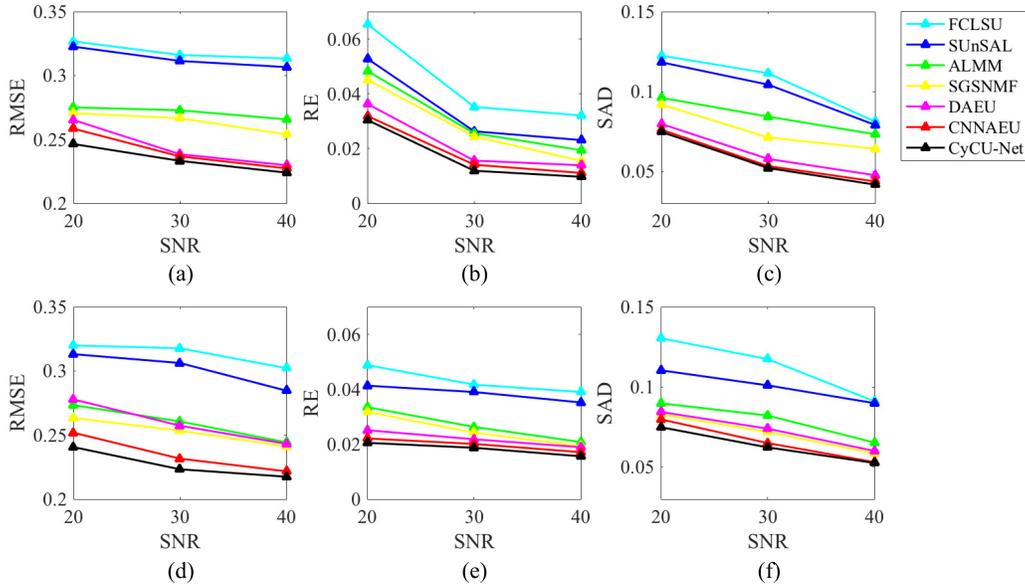


Fig. 5. Robustness evaluation of the synthetic data set with different SNR values, where the first and second rows represent the unmixing result of LMM and GBM, respectively. (a) and (d) RMSE. (b) and (e) RE. (c) and (f) SAD.

error (RMSE), and RE that are defined as

$$\text{SAD}(\hat{\mathbf{m}}_i, \mathbf{m}_i) = \frac{1}{p} \sum_{i=1}^p \arccos\left(\frac{\hat{\mathbf{m}}_i^T \mathbf{m}_i}{\|\hat{\mathbf{m}}_i\| \|\mathbf{m}_i\|}\right) \quad (14)$$

$$\text{RMSE}(\hat{\mathbf{a}}_j, \mathbf{a}_j) = \sqrt{\frac{1}{N} \sum_{j=1}^N \|\hat{\mathbf{a}}_j - \mathbf{a}_j\|_2^2} \quad (15)$$

$$\text{RE}(\hat{\mathbf{x}}_j, \mathbf{x}_j) = \sqrt{\frac{1}{LN} \sum_{j=1}^N \|\hat{\mathbf{x}}_j - \mathbf{x}_j\|_2^2} \quad (16)$$

where  $\hat{\mathbf{m}}_i$  and  $\mathbf{m}_i$  denote the extracted endmember and the reference endmember, respectively.  $\hat{\mathbf{a}}_j$  and  $\mathbf{a}_j$  are the estimated abundance and the actual abundance, respectively.  $\hat{\mathbf{x}}_j$  and  $\mathbf{x}_j$  represent the reconstructed and original pixels, respectively.

### C. Experiment With Synthetic Data Set

1) *Noise Robustness Analysis*: To investigate the robustness of the proposed method, different signal-to-noise ratio (SNR) values from 20 to 40 dB are added in the synthetic experiment. Fig. 5 illustrates the quantitative results on the synthetic data set with different SNR values in terms of the RMSE, RE, and SAD. As expected, the proposed CyCU-Net method achieves the optimal RMSE, RE, and SAD results compared with those of other state-of-the-art approaches, indicating that it can extract more accurate endmember and abundance results. In the LMM mixing data set, the DL-based unmixing methods, such as the DAEU and CNNAEU techniques, generally perform better than the traditional methods. It can be explained that the powerful fitting and learning capabilities of DL-based methods enable the network to be better reconstructed, resulting in more accurate unmixing results. However, for the GBM mixing data set, the RMSE and SAD results of SGSNMF and ALMM are superior to those of the DAEU technique in some cases. The reason may be that these

two methods are able to handle specific mixing problems by considering the spatial group sparsity information and modeling endmember variabilities, respectively. Regarding the RE and SAD results, the performance of CyCU-Net is similar to that of the CNNAEU technique. In contrast, the RMSE of CyCU-Net is obviously lower. The main reason might be that CyCU-Net can avoid the loss of detailed information and guide the network toward extracting more accurate abundance results due to the introduction of cycle-consistency constraints. Overall, CyCU-Net can obtain lower RMSE, RE, and SAD values compared with other methods, which demonstrates the robustness and superiority of the proposed method.

2) *Parameter Analysis*: The performance of the proposed CyCU-Net method is, to some extent, sensitive to the setting of three regularization parameters ( $\beta$ ,  $\delta$ , and  $\gamma$ ). For this reason, the corresponding experiments are conducted to investigate the effects of parameter setting, as shown in Fig. 6. It can be observed from Fig. 6 that, for the low SNRs circumstances, such as 20 and 30 dB, the best RMSE result in CyCU-Net comes from  $\beta = 0.5$  and  $\gamma = 1e - 6$ , indicating that the cycle-consistency constraint is effective for improving the accuracy of unmixing compared to the RMSE result when  $\beta = 0$  or  $\beta = 1$ . However, for the higher SNR circumstance, such as 40 dB, the value of  $\beta$  tends to be larger, and the value of  $\gamma$  is suitable to take  $1e - 7$ . It can be explained that the weight of the cycle-consistency constraint can be appropriately reduced in the case of low noise. For the parameter  $\delta$ , the optimal value is always  $\delta = 1e - 2$  under different SNR circumstances. Therefore, we can empirically summarize a general trend for the parameter setting. It is expected that, when the SNR is low,  $\beta$  and  $\gamma$  are suggested to be around 0.5 and  $1e - 6$  for noise suppression. When the SNR is high,  $\beta$  is suggested to set larger than 0.7, and  $\gamma$  tends to be around  $1e - 7$ . Moreover, it is clear that the optimal choice for  $\delta$  is  $1e - 2$ .

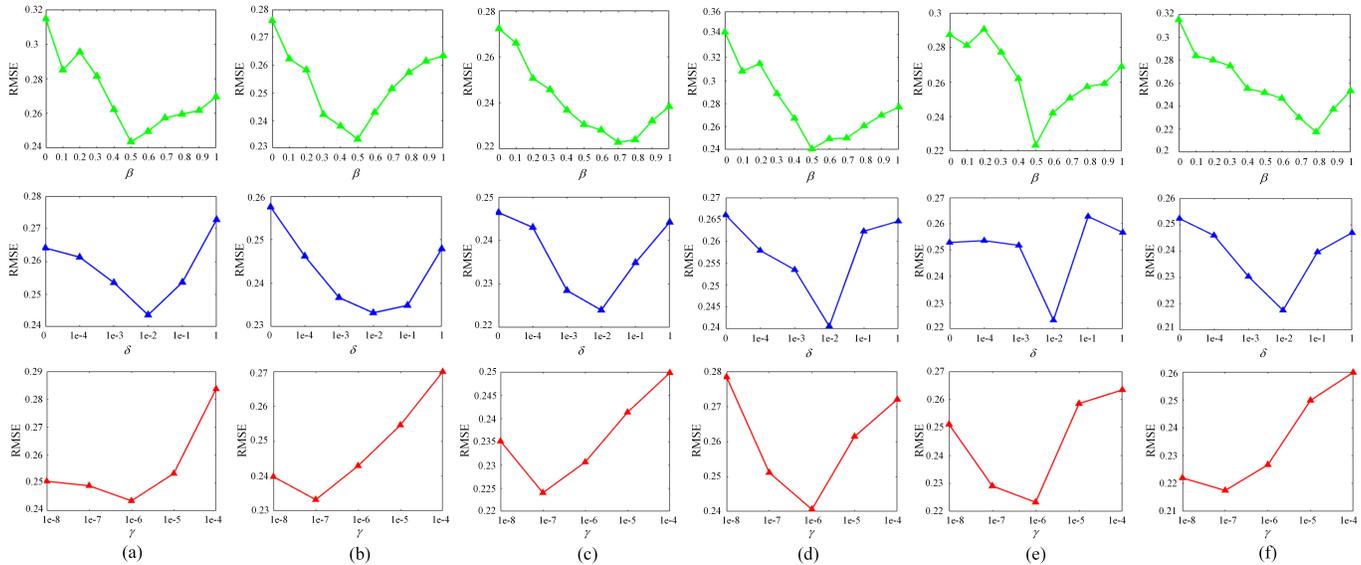


Fig. 6. Parameter sensitivity analysis of the proposed CyCU-Net method in the synthetic data set. (a) LMM with SNR = 20 dB. (b) LMM with SNR = 30 dB. (c) LMM with SNR = 40 dB. (d) GBM with SNR = 20 dB. (e) GBM with SNR = 30 dB. (f) GBM with SNR = 40 dB.

TABLE II  
QUANTITATIVE RESULTS FOR THE SAMSON DATA SET, WHERE THE SAD FOR EACH MATERIAL AND THE MEAN SAD, RMSE, AND RE ARE REPORTED. THE BEST RESULTS ARE SHOWN IN BOLD

Methods		FCLSU	SUnSAL	ALMM	SGSNMF	DAEU	CNNAEU	CyCU-Net
SAD	Soil	0.0280	0.0445	0.0296	0.0132	0.0229	<b>0.0117</b>	0.0188
	Tree	0.0530	0.0674	0.0395	0.0542	0.0419	<b>0.0377</b>	0.0380
	Water	0.1739	0.1302	0.1277	0.1805	0.0952	0.0916	<b>0.0406</b>
Mean SAD		0.0850	0.0807	0.0656	0.0826	0.0533	0.0470	<b>0.0325</b>
RMSE		0.6044	0.5581	0.5015	0.5427	0.4503	0.4650	<b>0.4164</b>
RE		0.0199	0.0156	0.0138	0.0151	0.0113	0.0116	<b>0.0107</b>

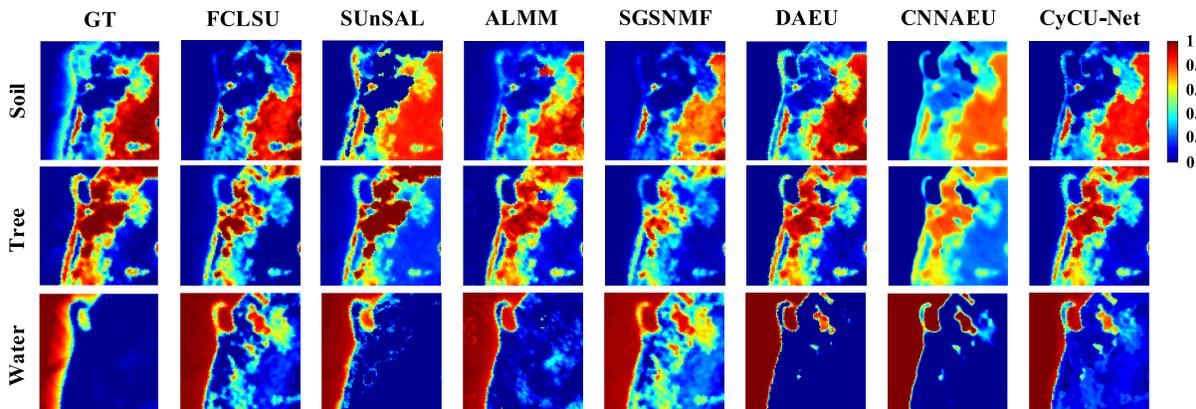


Fig. 7. Abundance maps of the soil, tree, and water from the Samson data set obtained by the different algorithms.

#### D. Experiment With Samson Data Set

Table II and Fig. 7 present the quantitative results and the corresponding abundance maps, respectively, in the Samson data set. It can be seen that the proposed CyCU-Net method achieves the best performance in terms of the SAD, RMSE, and RE, which validates the effectiveness of the proposed method. In addition, the abundance maps obtained by CyCU-Net are similar to the GT in Fig. 7. For illustrative

purposes, the extracted endmember comparison between the different algorithms and the corresponding GTs is shown in Fig. 8. It is obvious that the DL-based unmixing methods can obtain relatively smaller SAD results than the traditional methods. Moreover, although the CNNAEU method achieves the best results in terms of the extraction of soil and tree, the proposed CyCU-Net ranks first in terms of the overall endmember accuracy.

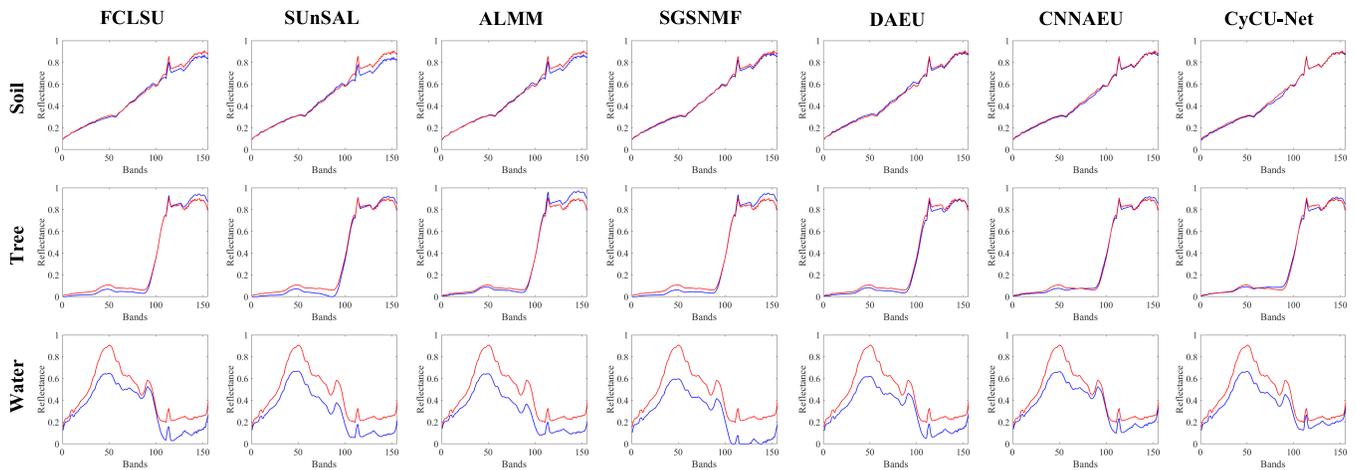


Fig. 8. Extracted endmember comparison between the different algorithms and the corresponding GTs in the Samson data set.

TABLE III  
QUANTITATIVE RESULTS FOR THE JASPER RIDGE DATA SET, WHERE THE SAD FOR EACH MATERIAL AND THE MEAN SAD, RMSE, AND RE ARE REPORTED. THE BEST RESULTS ARE SHOWN IN BOLD

Methods		FCLSU	SUnSAL	ALMM	SGSNMF	DAEU	CNNAEU	CyCU-Net
SAD	Water	0.0733	0.1043	0.0842	0.1108	0.0727	0.0607	<b>0.0595</b>
	Soil	0.1157	0.0829	0.0881	0.1402	0.1098	0.1161	<b>0.0377</b>
	Tree	0.1519	0.1543	0.1577	0.1591	0.1507	<b>0.1466</b>	0.1488
	Road	0.0996	0.0746	0.0750	0.0562	0.0615	0.0696	<b>0.0199</b>
Mean SAD		0.1101	0.1040	0.1013	0.1166	0.0987	0.0983	<b>0.0665</b>
RMSE		0.5028	0.4957	0.4502	0.3881	0.3428	0.3256	<b>0.2876</b>
RE		0.0109	0.0103	0.0097	0.0085	0.0081	0.0076	<b>0.0066</b>

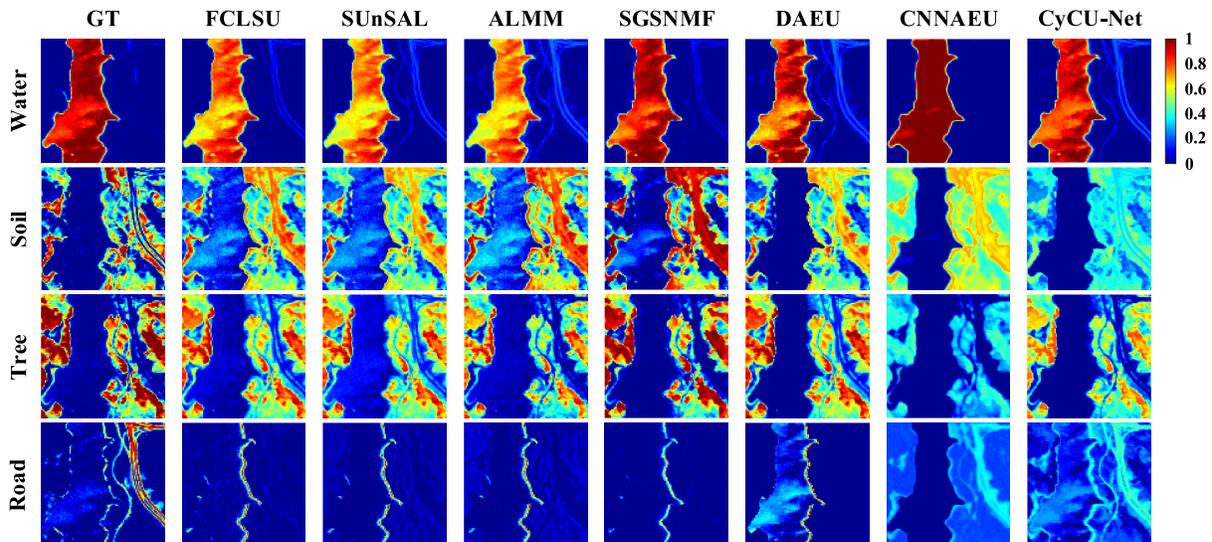


Fig. 9. Abundance maps of the water, soil, tree, and road from the Jasper Ridge data set obtained by the different algorithms.

### E. Experiment With Jasper Ridge Data Set

Table III quantitatively lists the performance assessment for all algorithms, and the corresponding abundance maps are displayed in Fig. 9. It can be clearly observed that most of the comparison methods cannot entirely separate the road endmember even though they have undergone the same VCA initialization. However, as shown in Fig. 9, the DL-based unmixing methods have less trouble with the road endmember due to their powerful learning abilities

and some reasonable model settings. Overall, CyCU-Net can obtain better competitive results than other comparison methods in terms of mean SAD, RMSE, and RE. For illustrative purposes, the extracted endmember comparison between the different algorithms and the corresponding GTs is depicted in Fig. 10. It is obvious that the extracted endmember signatures of CyCU-Net basically match well with GT, which further validates its effectiveness for HU in real scenarios.

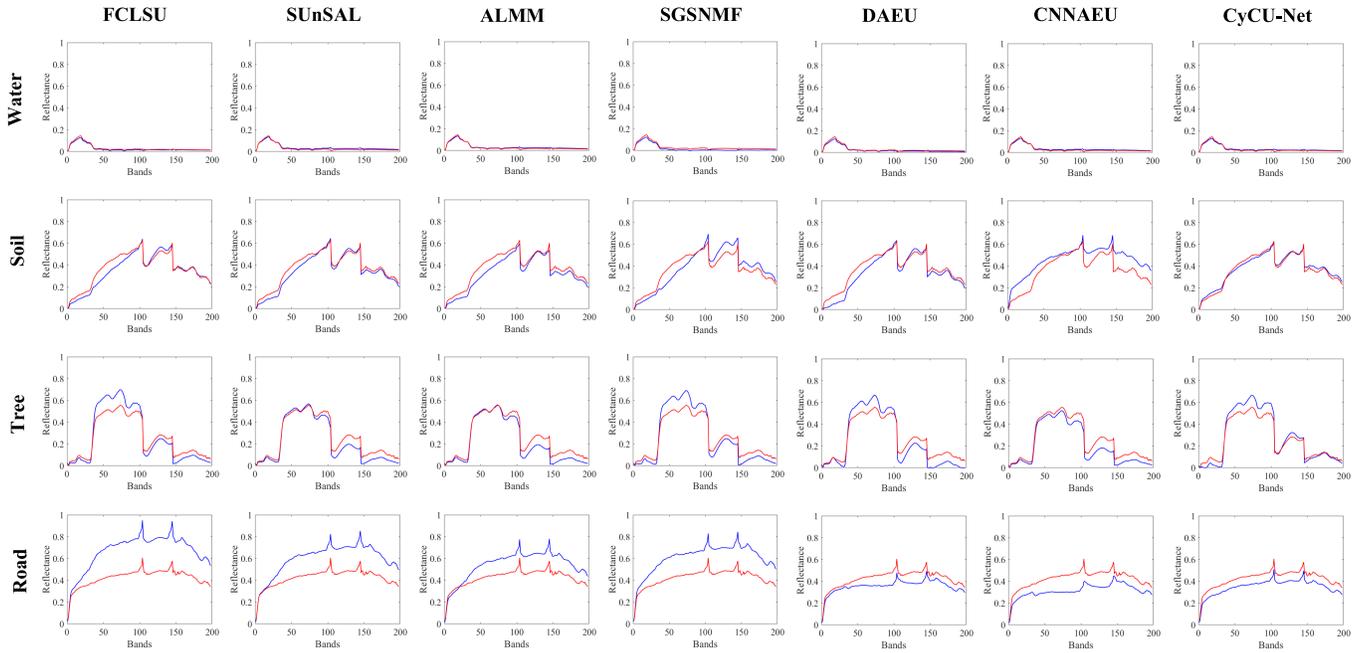


Fig. 10. Extracted endmember comparison between the different algorithms and the corresponding GTs in the Jasper Ridge data set.

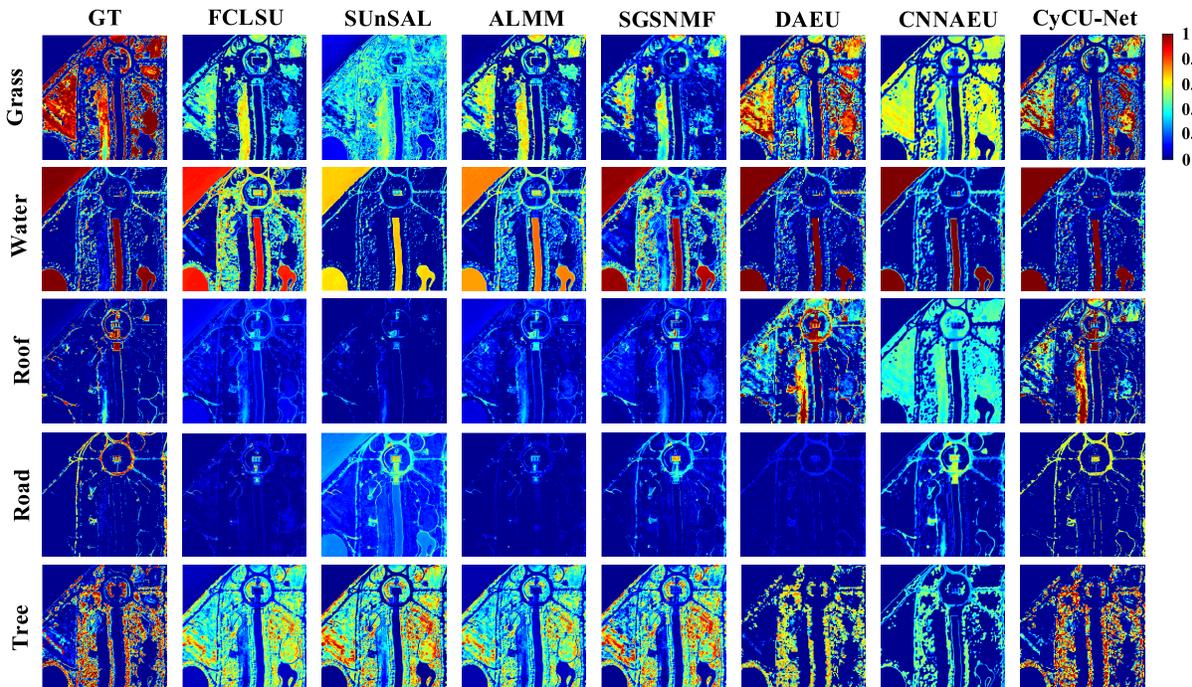


Fig. 11. Abundance maps of the grass, water, roof, road, and tree from the Washington DC Mall data set obtained by the different algorithms.

#### F. Experiment With Washington DC Mall Data Set

The cropped Washington DC Mall data set has a larger size than the other two data sets mentioned above. Table IV illustrates the quantitative results of different algorithms in the Washington DC Mall data set, and the corresponding abundance maps are depicted in Fig. 11. As seen from Table IV, the DL-based unmixing methods can obtain more accurate endmember and abundance results than traditional methods. Moreover, the corresponding abundance maps for

DL-based unmixing methods are more separable and close to the reference GTs. For example, although the endmember signatures for grass and tree are basically similar, DL-based unmixing methods can still separate them compared with the traditional methods, which proves the effectiveness of DL-based unmixing methods. Among the DL-based unmixing methods, CyCU-Net can explore the high-level features of the input HSI, thereby enhancing the extracted abundance results, especially for the tree and road, as shown in Fig. 11.

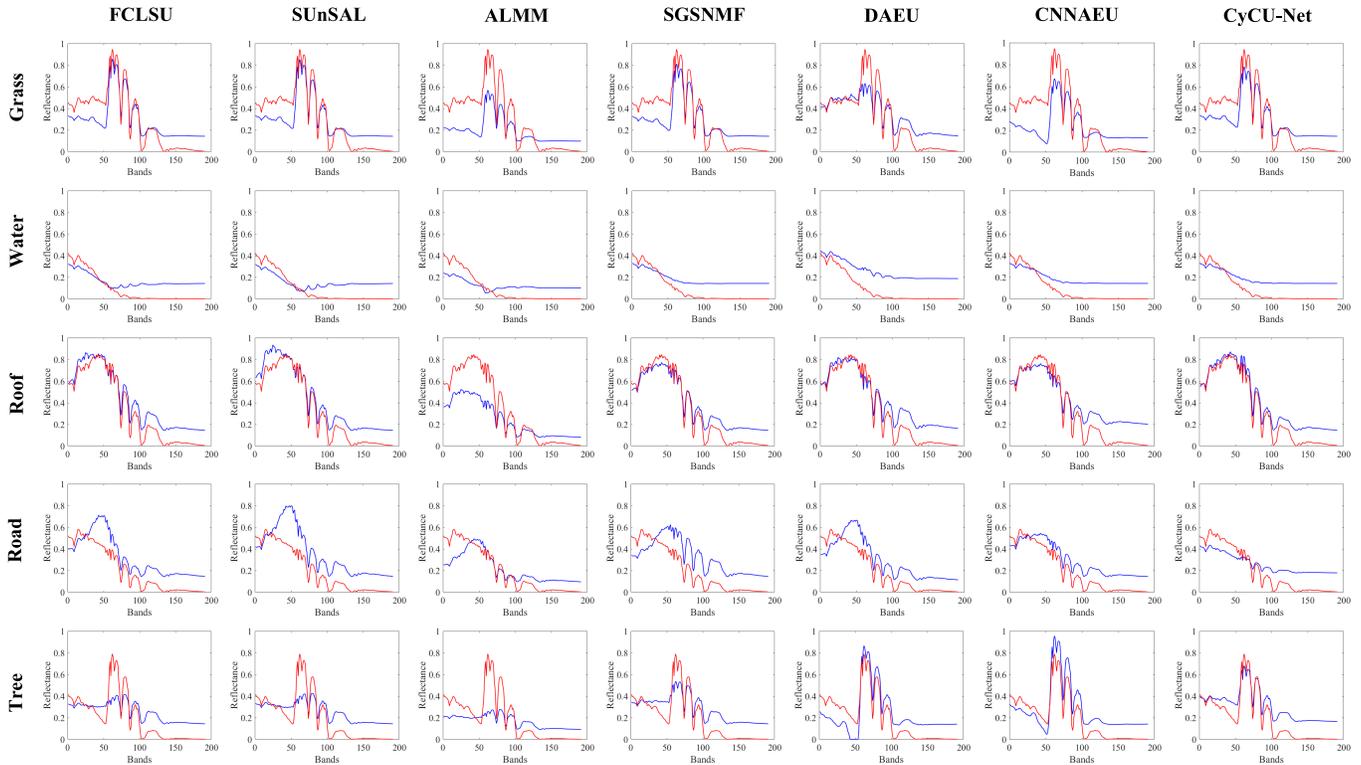


Fig. 12. Extracted endmember comparison between the different algorithms and the corresponding GTs on the Washington DC Mall data set.

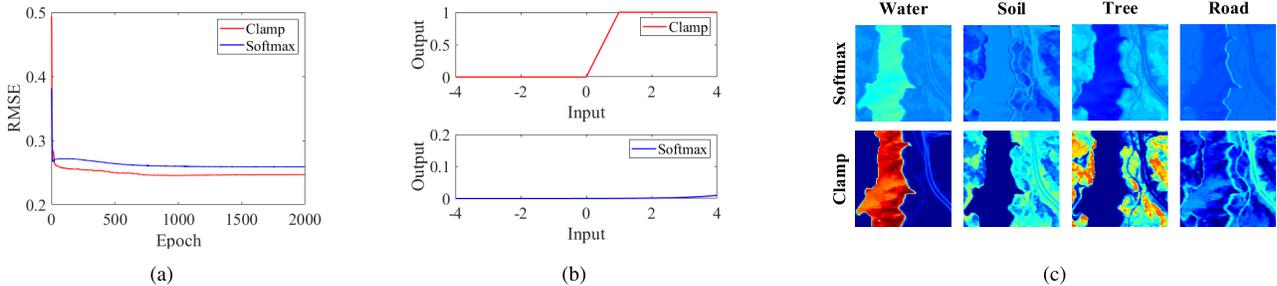


Fig. 13. Abundance comparison results when using different constraint functions in the Jasper Ridge data set. (a) RMSE convergence comparison. (b) Function curve comparison. (c) Corresponding abundance map comparison.

For illustrative purposes, the extracted endmember signatures are depicted in Fig. 12. It can be observed that the endmember signatures extracted by CyCU-Net are superior to those of other competitive methods, demonstrating the effectiveness and reliability of the proposed CyCU-Net method.

### G. Performance of Different Network Configurations

In this section, we will evaluate the performance of different network configurations (i.e., abundance constraints, objective functions, convolution filter sizes, and more complex unmixing scenes) in the proposed CyCU-Net.

1) *Abundance Constraints*: In this article, we introduce the clamp function and (12) to satisfy the ANC and ASC, respectively. To investigate the performance of different abundance constraints, the softmax function is adopted for comparison in this section. Fig. 13 shows the abundance comparison results when using these two strategies in the Jasper Ridge data set in terms of the RMSE convergence, function curve, and

corresponding abundance maps. The obtained RMSE value of the clamp function is smaller than that of the softmax function when converging, and the corresponding abundance maps for the clamp function tend to be more separate and remarkable. These results illustrate that the proposed clamp strategy can effectively enhance the performance of the abundance results compared with the softmax function. Fig. 13(b) can explain the results of this experiment. It can be seen that the output for the softmax function changes very little when the input is in the range of  $[-4, 4]$ , which causes the convergence speed to be slow and the loss function to not be efficiently updated, especially for the case in which the input value is small. In contrast, for the clamp function, the output can show a significant value change and have a faster update performance.

2) *Objective Functions*: Because the proposed CyCU-Net method is based on an AE-based network, the RE is the main factor influencing the unmixing performance in this process. According to the characteristics of the objective function, three typical objective functions namely, the SAD, MSE, and

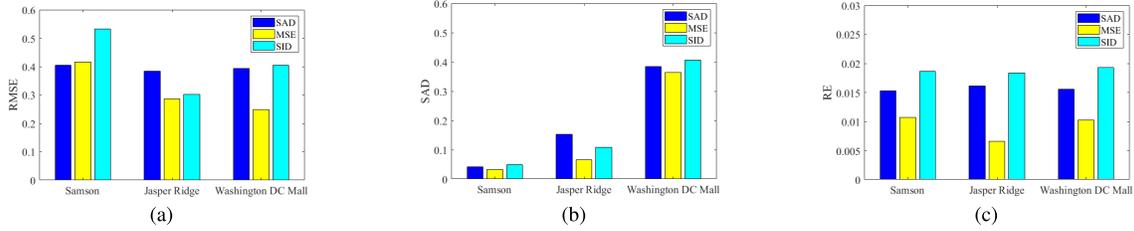


Fig. 14. Ablation study for the proposed CyCU-Net based on different objective functions. (a) RMSE. (b) SAD. (c) RE.

TABLE IV  
 QUANTITATIVE RESULTS FOR THE WASHINGTON DC MALL DATA SET, WHERE THE SAD FOR EACH MATERIAL AND THE MEAN SAD, RMSE, AND RE ARE REPORTED. THE BEST RESULTS ARE SHOWN IN BOLD

Methods	FCLSU	SUnSAL	ALMM	SGSNMF	DAEU	CNNAEU	CyCU-Net
SAD	Grass	0.3028	0.3016	0.3071	0.3151	0.3126	<b>0.2934</b>
	Water	0.6555	0.6748	0.6725	0.6506	0.6539	<b>0.6480</b>
	Roof	0.2015	0.1905	0.1825	0.2114	0.2477	<b>0.1809</b>
	Road	0.3679	0.3725	0.3765	0.4063	0.3598	<b>0.3370</b>
	Tree	0.5293	0.5158	0.5173	0.4495	0.4555	<b>0.3383</b>
Mean SAD	0.4114	0.4110	0.4112	0.4066	0.4059	0.3993	<b>0.3648</b>
RMSE	0.3853	0.3791	0.3476	0.2958	0.2721	0.2519	<b>0.2488</b>
RE	0.0194	0.0171	0.0164	0.0144	0.0137	0.0124	<b>0.0102</b>

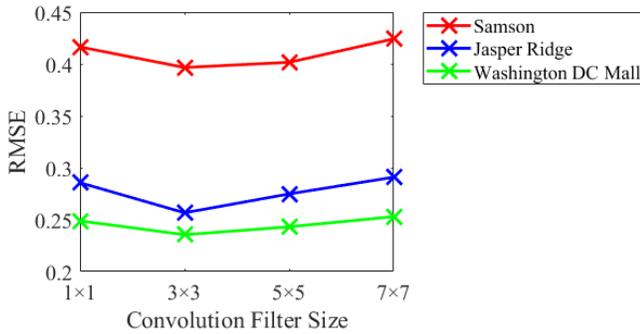


Fig. 15. Performance evaluation of the proposed CyCU-Net with different convolution filter sizes.

spectral information divergence (SID), are analyzed in the three real hyperspectral data sets. The specific definition of these objective functions can be found in [31]. Fig. 14 depicts the experimental results for the Samson, Jasper Ridge, and Washington DC Mall data sets, respectively. It can be seen that the results for the MSE objective function can obtain lower RMSE, SAD, and RE values than the other two objective functions. Although the SAD and SID objective functions focus on the proportional recovery of the input data and are scale-invariant, the experimental results reveal that the performance of the proposed CyCU-Net is more dependent on relatively accurate and identical reconstruction results. Therefore, the MSE objective function is more suitable for the proposed method.

3) *Convolution Filter Sizes*: By introducing different convolution filter sizes, the proposed CyCU-Net can be extended to a spatial-spectral framework. In order to find out the optimal size of convolution filter, the proposed method is evaluated with different filter sizes:  $1 \times 1$ ,  $3 \times 3$ ,  $5 \times 5$ , and  $7 \times 7$ . Fig. 15 reports the RMSE results of different convolution filter sizes in three real hyperspectral data sets. As can be seen, applying a certain convolution filter size can effectively improve the

TABLE V  
 COMPUTATIONAL COST OF ALL ALGORITHMS ON DIFFERENT DATA SETS IN TERMS OF SECONDS (s)

Method	Synthetic	Samson	Jasper Ridge	Washington DC
FCLSU	3.64	0.64	1.45	12.41
SUnSAL	0.12	0.03	0.06	0.32
ALMM	25.07	15.63	33.94	139.25
SGSNMF	45.76	25.30	27.17	107.32
DAEU	36.27	13.61	16.92	159.88
CNNAEU	37.93	26.72	48.26	227.56
CyCU-Net	43.19	30.65	48.61	293.75

unmixing accuracy, and the optimal filter size is  $3 \times 3$ . Thus, we recommend setting the convolution filter size to  $3 \times 3$  in the spatial CyCU-Net method.

4) *Abundance Visualization on Complex Cuprite Data*: In the section, we attempt to investigate a more complex spectral unmixing scene, e.g., mineral exploration, by using the proposed CyCU-Net. The used HSI is the well-known AVIRIS Cuprite with the size of  $250 \times 190 \times 188$  after removing the bands affected by the water vapor and low SNR [53]. The complex scene consists of multiple exposed minerals, where we select 12 principal endmembers to visualize and evaluate the unmixing performance, by following the works in [34] and [49]. The estimated endmembers and their corresponding abundance maps of the proposed CyCU-Net method are shown in Fig. 16. As can be seen, the proposed CyCU-Net can extract multiple endmembers in the Cuprite scene, and the result indicates the potential of the proposed method in real scenarios.

#### H. Computational Cost

In this section, we evaluate the computational cost of the proposed method, and the result of the running time is shown in Table V. The experiments are performed on a computer with a 2.9-GHz Intel Core i7 CPU and 16 GB of memory.

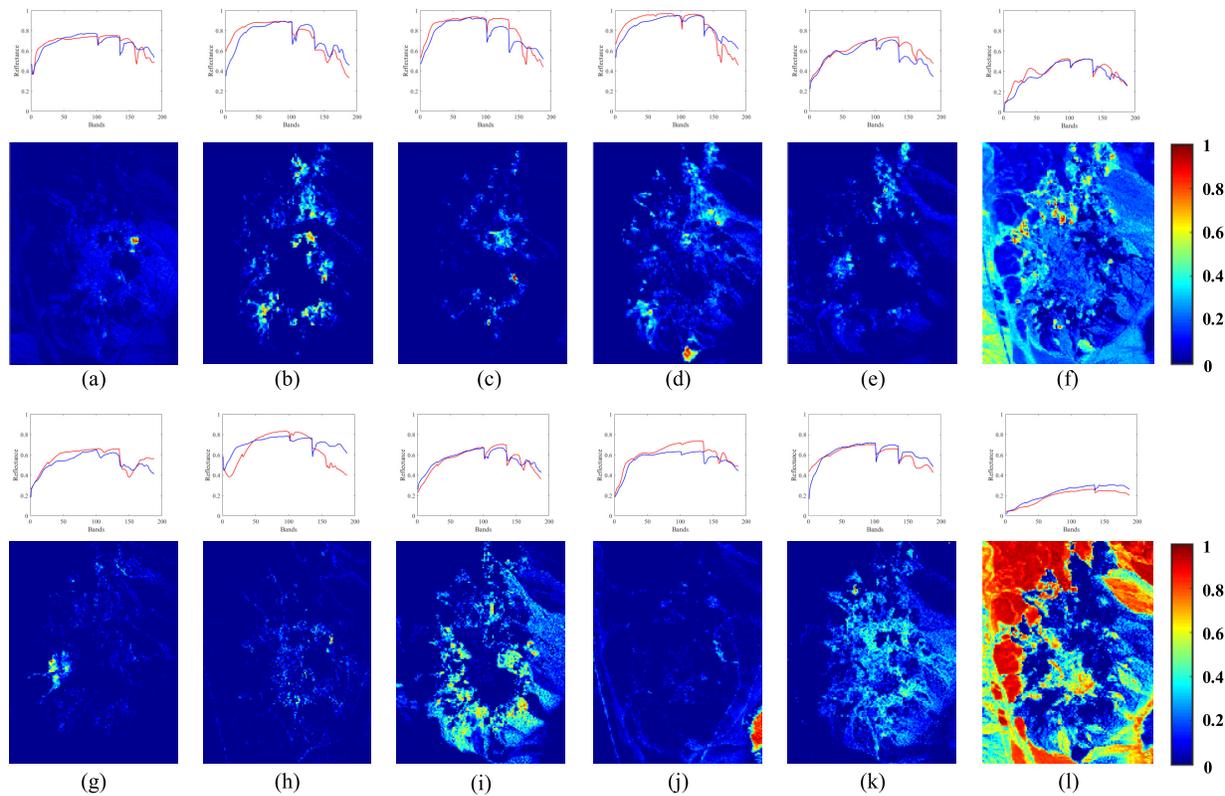


Fig. 16. Estimated endmembers (in blue), along with the references from the USGS library (in red), and their corresponding abundance maps obtained from the Cuprite mineral scene by the proposed CyCU-Net. (a) Muscovite. (b) Alunite. (c) Kaolinite KGa-2 (pxyl). (d) Kaolinite KGa-1 (wxyl). (e) Kaolin/Smect KLF506. (f) Nontronite. (g) Buddingtonite. (h) Dumortierite. (i) Kaolin/Smect H89-FR-5. (j) Montmorillonite. (k) Chalcedony. (l) Desert Varnish.

It can be seen that the proposed CyCU-Net can achieve similar computational efficiency as other comparison methods on relatively small data sets, such as the synthetic data set and the Jasper Ridge data set. However, our approach seems to spend more time on the large data set (due to two AEs involved in the training phase), yet the computational cost is acceptable.

#### IV. CONCLUSION

In this article, we propose a novel cycle-consistency unmixing network, called CyCU-Net, by training two cascaded AEs in the end-to-end architecture. Benefiting from the cycle consistency dominated by the cascaded AEs, the proposed CyCU-Net method can reduce the detailed and material-related information loss in the process of reconstruction more effectively than previous methods, which further yields more reasonable and superior unmixing results. Experiments with synthetic and real hyperspectral data sets validate the effectiveness and robustness of the proposed methods compared with other state-of-the-art unmixing approaches. In the future, we will aim to improve CyCU-Net by considering the spectral-spatial information with the help of DL.

#### REFERENCES

- [1] A. Di Gregorio, *Land Cover Classification System: LCCS—Classification Concepts and User Manual for Software*, vol. 2. Food & Agriculture Org., 2005.
- [2] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, “Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, Jan. 2019.
- [3] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson, “Challenges and opportunities of multimodality and data fusion in remote sensing,” *Proc. IEEE*, vol. 103, no. 9, pp. 1585–1601, Sep. 2015.
- [4] 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.
- [5] 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.
- [6] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, “Anomaly detection in hyperspectral images based on low-rank and sparse representation,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2016.
- [7] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, “ORSIm detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5146–5158, Jul. 2019.
- [8] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, “Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, Feb. 2020.
- [9] 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.
- [10] N. Keshava and J. F. Mustard, “Spectral unmixing,” *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.
- [11] 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.*, early access, Nov. 11, 2020, doi: 10.1109/TCYB.2020.3028931.

- [12] 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.
- [13] 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.*, vol. 1, Jun. 2010, pp. 1–4.
- [14] 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.
- [15] 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. (RAST)*, Jun. 2013, pp. 171–176.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] R. Close, P. Gader, and J. Wilson, "Hyperspectral unmixing using macroscopic and microscopic mixture models," *J. Appl. Remote Sens.*, vol. 8, no. 1, Apr. 2014, Art. no. 083642.
- [24] M. Tang, L. Gao, A. Marinoni, P. Gamba, and B. Zhang, "Integrating spatial information in the normalized P-linear algorithm for nonlinear hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1179–1190, Apr. 2018.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 18, 2020, doi: [10.1109/TGRS.2020.3015157](https://doi.org/10.1109/TGRS.2020.3015157).
- [29] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 24, 2020, doi: [10.1109/TGRS.2020.3016820](https://doi.org/10.1109/TGRS.2020.3016820).
- [30] G. A. Licciardi and F. Del 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.
- [31] 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, May 2018.
- [32] R. Guo, W. Wang, and H. Qi, "Hyperspectral image unmixing using autoencoder cascade," in *Proc. 7th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.
- [33] 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. (IGARSS)*, Jul. 2017, pp. 205–208.
- [34] 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.
- [35] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
- [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. (IGARSS)*, Jul. 2019, pp. 357–360.
- [37] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravorty, "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](https://doi.org/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, Feb. 2020.
- [39] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, Oct. 2019.
- [40] 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.
- [41] 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. (IGARSS)*, Jul. 2019, pp. 2127–2130.
- [42] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," 2015, *arXiv:1508.06576*. [Online]. Available: <http://arxiv.org/abs/1508.06576>
- [43] X. Hou, L. Shen, K. Sun, and G. Qiu, "Deep feature consistent variational autoencoder," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 1133–1141.
- [44] X. Hou, K. Sun, L. Shen, and G. Qiu, "Improving variational autoencoder with deep feature consistent and generative adversarial training," *Neurocomputing*, vol. 341, pp. 183–194, Mar. 2019.
- [45] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 11, pp. 3943–3956, Nov. 2020, doi: [10.1109/TCSVT.2019.2920407](https://doi.org/10.1109/TCSVT.2019.2920407).
- [46] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, Cham, Switzerland: Springer, 2016, pp. 694–711.
- [47] K. Zheng *et al.*, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2487–2502, Mar. 2021.
- [48] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. ECCV*, 2020, pp. 208–224.
- [49] 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.
- [50] D. C. Heinz and Chein-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.
- [51] J. M. Bioucas-Dias and M. A. T. Figueiredo, "Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing," in *Proc. 2nd Workshop Hyperspectral Image Signal Processing: Evol. Remote Sens.*, Jun. 2010, pp. 1–4.
- [52] 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.
- [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.
- [54] C. O. Davis, M. Kavanaugh, R. Letelier, W. P. Bissett, and D. Kohler, "Spatial and spectral resolution considerations for imaging coastal waters," *Proc. SPIE*, vol. 6680, Oct. 2007, Art. no. 66800P.

- [55] L. J. Rickard, R. W. Basedow, E. F. Zalewski, P. R. Silverglate, and M. Landers, "Hydice: An airborne system for hyperspectral imaging," in *Proc. ISTE*, vol. 1937, Sep. 1993, pp. 173–179.



**Lianru Gao** (Senior Member, IEEE) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, and the Ph.D. degree in cartography and geographic information system from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 2007.

He was a Visiting Scholar at the University of Extremadura, Cáceres, Spain, in 2014, and at Mississippi State University (MSU), Starkville, MS, USA, in 2016. He is a Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, CAS. He has published more than 160 peer-reviewed papers, and there are more than 80 journal articles included by SCI. He has coauthored an academic book entitled *Hyperspectral Image Classification and Target Detection*. He holds 28 National Invention Patents in China. His research focuses on hyperspectral image processing and information extraction.

Dr. Gao was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016 and was supported by the China National Science Fund for Excellent Young Scholars in 2017, and won the Second Prize of The State Scientific and Technological Progress Award in 2018. He received the recognition of the Best Reviewer of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2015 and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2017. In the last ten years, he was the PI of ten scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2010–2012, 2016–2019, 2018–2020) and by the Key Research Program of the CAS (2013–2015).



**Zhu Han** (Student Member, IEEE) received the B.S. degree in electrical engineering from the North China University of Technology, Beijing, China, in 2019. She is pursuing the Ph.D. degree in cartography and geographic information system with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS), Beijing.

Her research interests include hyperspectral image processing, deep learning, and artificial intelligence.



**Danfeng Hong** (Member, IEEE) received the M.Sc. degree (*summa cum laude*) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, and the Dr.-Ing. degree (*summa cum laude*) from Signal Processing in Earth Observation (SIPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

Since 2015, he has been a Research Associate at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. He is a Research Scientist and leads the Spectral Vision Working Group, IMF, DLR, and also an Adjunct Scientist in GIPSA-lab, Grenoble INP, CNRS, Université Grenoble Alpes, Grenoble, France. His research interests include signal/image processing and analysis, hyperspectral remote sensing machine/deep learning, and artificial intelligence and their applications in Earth vision.

Dr. Hong received the recognition of the Best Reviewer of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2020. He is a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and *Remote Sensing*. He has been serving as a Topical Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING since 2021.



**Bing Zhang** (Fellow, IEEE) received the B.S. degree in geography from Peking University, Beijing, China, in 1991, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, in 1994 and 2003, respectively.

He is a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 25 years. He has developed five software systems in image processing and applications. He has authored more than 300 publications, including more than 170 journal articles. He has edited six books/contributed book chapters on hyperspectral image processing and subsequent applications. His research interests include the development of mathematical and physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas.

Dr. Zhang has been serving as a Technical Committee Member of the IEEE Workshop on Hyperspectral Image and Signal Processing since 2011, the President of the Hyperspectral Remote Sensing Committee of the China National Committee of International Society for Digital Earth since 2012, and the Standing Director of the Chinese Society of Space Research (CSSR) since 2016. He is the Student Paper Competition Committee Member in IGARSS from 2015 to 2019. His creative achievements were rewarded ten important prizes from the Chinese government and special government allowances of the Chinese State Council. He was awarded the National Science Foundation for Distinguished Young Scholars of China in 2013 and the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS scholars. He is serving as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.



**Jocelyn Chanussot** (Fellow, IEEE) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998.

Since 1999, he has been with Grenoble INP, where he is a Professor of signal and image processing. He has been a Visiting Scholar at Stanford University, Stanford, CA, USA, KTH, Stockholm, Sweden, and NUS, Singapore. Since 2013, he has been an Adjunct Professor with the University of Iceland, Reykjavik, Iceland. From 2015 to 2017, he was a Visiting Professor at the University of California, Los Angeles (UCLA), Los Angeles, CA, USA. He holds the AXA Chair in remote sensing and is an Adjunct Professor at the Chinese Academy of Sciences, Aerospace Information Research Institute, Beijing, China. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning, and artificial intelligence.

Dr. Chanussot is the founding President of the IEEE Geoscience and Remote Sensing French Chapter from 2007 to 2010, which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia from 2017 to 2019. He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair from 2009 to 2011 and the Co-Chair of the GRS Data Fusion Technical Committee from 2005 to 2008. He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society from 2006 to 2008 and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing in 2009. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and the PROCEEDINGS OF THE IEEE. He was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING from 2011 to 2015. In 2014, he served as a Guest Editor for the *IEEE Signal Processing Magazine*. He is a member of the Institut Universitaire de France from 2012 to 2017 and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, 2018–2019).