

# A New Dual-Feature Fusion Network for Enhanced Hyperspectral Unmixing

Xuanwen Tao<sup>1</sup>, Member, IEEE, Bikram Koirala<sup>2</sup>, Member, IEEE, Antonio Plaza<sup>3</sup>, Fellow, IEEE, and Paul Scheunders<sup>4</sup>, Senior Member, IEEE

**Abstract**—Hyperspectral unmixing is a crucial technique in remote sensing data processing that aims to estimate component information from mixed pixels in hyperspectral images. Most existing deep learning-based hyperspectral unmixing models employ autoencoder (AE) networks to reconstruct hyperspectral images and estimate abundance maps. Here, the weight between the reconstructed and the softmax layers is used to extract/estimate endmember signatures. However, AEs are heavily dependent on initial weights, which introduces inherent randomness, potentially compromising unmixing accuracy. To address this issue, in this article, we present a new dual-feature fusion network (DFFN) for enhanced hyperspectral unmixing. Our DFFN mainly consists of four modules: 1) a feature fusion module (FFM); 2) an abundance estimation module (AEM); 3) an endmember estimation module (EEM); and 4) a reconstruction module (RM). First, FFM calculates spectral and spatial similarities and then enhances the hyperspectral image by matrix multiplications with similarity matrices. Second, AEM takes the enhanced hyperspectral image as input and uses convolutional layers to estimate abundances and reconstruct the image. Next, the reconstructed image is fed into EEM to automatically estimate endmembers. RE performs the final reconstruction through matrix multiplication of the estimated endmembers and abundances. Experiments on synthetic and real hyperspectral datasets, together with a comparison with state-of-the-art techniques, demonstrate the superiority of our newly proposed DFFN. The full code is released at <https://github.com/xuanwentao> for public evaluation.

**Index Terms**—Dual-feature fusion, hyperspectral unmixing, similarity, unmixing module.

## I. INTRODUCTION

**H**YPERSPECTRAL images collect hundreds of wavelength bands per pixel spanning the entire electromagnetic spectrum, thus enabling accurate identification or detection of materials with exceptional detail. This offers enormous potential and is highly sought after in a variety

of applications, including classification [1], [2], [3], anomaly detection [4], [5], super-resolution [6], [7], [8], and change detection [9], [10].

However, due to limitations inherent in the imaging acquisition equipment, there is an inevitable tradeoff between the acquired spectral and spatial information, limiting the spatial resolution of hyperspectral sensors. Accordingly, hyperspectral images may contain many mixed pixels, which significantly affects the accuracy of subsequent image processing tasks. To address this problem, a large number of hyperspectral unmixing techniques have been developed in the literature to decompose mixed pixels into a collection of material spectra (endmember signatures) and their corresponding per-pixel fractions (abundance maps).

### A. Traditional Unmixing Methods

Conventional hyperspectral unmixing approaches are mainly divided into linear [11], [12], [13], [14], [15] and nonlinear [16], [17], [18], [19]. The linear mixture model (LMM) assumes that incident light interacts with a single material, leading to the expression of each pixel spectrum as a linear combination of endmembers and their respective abundances. In this context, the fully constrained least squares unmixing algorithm (FCLS) [20] estimates the fractional abundances by minimizing the least squared errors between the actual reflectance spectra and the ones reconstructed by the LMM and obeys both the abundance nonnegativity constraint (ANC) and the abundance sum-to-one constraint (ASC).

Many endmember extraction algorithms have been proposed in the literature by maximizing the volume of the enclosing simplex in the hyperspectral dataset [21], [22], [23], [24], e.g., vertex component analysis (VCA) and N-FINDR. VCA [22] iteratively projects data in directions orthogonal to the subspace defined by the endmembers already identified. The signature of each new endmember is found by locating the extreme point of this projection. The process repeats until all endmembers are identified. The famous N-FINDR algorithm [24] is based on the fact that the simplex formed by the purest pixels has a larger volume than any other combination of pixels and iteratively inflates a simplex to find those purest pixels. When the endmembers are not available in the hyperspectral image (no pure pixel scenario), these can be estimated by finding the minimum volume linear simplex, which encloses all the data points [25], [26]. Unsupervised

Received 31 May 2024; revised 20 September 2024 and 24 October 2024; accepted 21 November 2024. Date of publication 25 November 2024; date of current version 9 December 2024. This work was supported by the Research Foundation Flanders under Project G031921N. The work of Bikram Koirala was supported by the Postdoctoral Fellow of the Research Foundation Flanders, Belgium [Fonds Wetenschappelijk Onderzoek (FWO)], under Grant 1250824N-7028. (Corresponding author: Xuanwen Tao.)

Xuanwen Tao, Bikram Koirala, and Paul Scheunders are with the Imec-VisionLab, Department of Physics, University of Antwerp, 2000 Antwerp, Belgium (e-mail: taoxuanwenupc@gmail.com; bikram.koirala@uantwerpen.be; paul.scheunders@uantwerpen.be).

Antonio Plaza is with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, University of Extremadura, 10071 Cáceres, Spain (e-mail: aplaza@unex.es).

Digital Object Identifier 10.1109/TGRS.2024.3505292

1558-0644 © 2024 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See <https://www.ieee.org/publications/rights/index.html> for more information.

unmixing techniques simultaneously estimate endmembers and abundances. Li et al. [27] proposed minimum volume simplex analysis to fit a minimum volume simplex to hyperspectral data while enforcing the constraint that the abundance fractions lie within the probability simplex. Miao and Qi [30] presented a minimum volume-constrained nonnegative matrix factorization method that incorporates a volume constraint into the nonnegative matrix factorization framework to perform the unmixing task. Li et al. [14] developed a gradient-based multiobjective method with greedy hash (GMOGH), which simultaneously addresses the problem of convergence and the discrete constraints imposed on endmembers to improve the accuracy of unmixing. These methods formulate the unmixing problem as a nonconvex optimization problem with respect to both endmembers and abundances. To deal with highly mixed scenarios, sparse unmixing techniques have also been proposed [32], [33], [34]. These methods are often described as semisupervised unmixing methods. Sparse unmixing utilizes a rich and well-designed library of pure spectra and applies sparse regression for abundance estimation. A major challenge is how to correct mismatches between the real reflectance spectra and the library spectra, caused by differences in the acquisition conditions of the two data types. Under the overly simple and unrealistic assumption that photons detected by the sensor interact with only one material at a macroscopic level, unmixing methods based on the LMM might not be appropriate for all applications.

Nonlinear spectral unmixing is based on the assumption that materials undergo multiple scattering. To address complex nonlinear higher order scattering effects and improve the accuracy of unmixing results, many nonlinear unmixing methods have been proposed in the literature. Nonlinear models include bilinear ones, such as the generalized bilinear model (GBM) [35], which assumes that the incident light interacts with a maximum of two pure materials before reaching the sensor. Various nonlinear mixing models exist, which aim to capture higher order interactions of incident light before reaching the sensor, including the multilinear mixing model (MLM) [36] and the  $p$ -linear ( $p > 2$ ) mixture model (pLMM) [37], [38], [39]). Other approaches include seminonnegative matrix factorization [40] and robust NMF (rNMF) [41]. However, if the selected model does not match the characteristics of the real scene, it will significantly reduce the accuracy of the endmember and abundance estimation. Hence, there is an urgent need for a more universally applicable method for hyperspectral unmixing purposes.

### B. Deep Learning-Based Unmixing Methods

Due to the success of deep learning-based networks [42], [43] in machine learning and computer vision, deep neural networks have recently been proposed for hyperspectral unmixing [44], [45], [46], [47], [48], [49], [50], [51]. Most available techniques are based on autoencoder (AE) networks, which can automatically learn low-dimensional embeddings and reconstruct the original data. An AE contains an encoder and a decoder. The encoder transforms the input spectra into

fractional abundances, and the decoder transforms the abundances into reconstructed spectra using linear layers and the endmembers as weights. When building a network for hyperspectral unmixing, researchers initially used fully connected (FC) layers to capture spectral information. Guo et al. [52] first used deep learning techniques for hyperspectral unmixing. Their approach combined a marginalized denoising AE (Mdae) with a nonnegative sparse AE to implicitly denoise the observed data and apply adaptive sparsity constraints, respectively. To consider physically meaningful endmember information, Han et al. [53] proposed a deep half-Siamese network that uses sharing weights to introduce the prior knowledge of endmembers extracted from hyperspectral images to further guide the unmixing network for improved abundance estimation. Su et al. [54] introduced a multitask bilinear unmixing framework that used a deep AE to linearly estimate endmember signatures and their respective abundance fractions. In addition, they used another AE to model second-order scattering interactions via a bilinear model.

Nevertheless, the aforementioned methods operate on a pixel-by-pixel basis and fail to fully learn the spatial and spectral features inherent in hyperspectral images. Considering the promising performance of generative adversarial networks (GANs) and convolutional neural networks (CNNs) and in image processing tasks, several unmixing methods have been developed using CNNs to effectively capture spectral-spatial information, thereby improving unmixing performance. Gao et al. [55] proposed a reversible generative network (Rev-Net), which includes a flow-based endmember learning module, and provided a theoretical proof for the reversibility of the endmember generation process to address the spectral variability challenge. The motivation behind Rev-Net is that if the endmember distribution can be represented by an explicit mathematical expression and that expression is reversible, the generation process will become more stable and reliable. Xu et al. [50] proposed a multiscale convolutional mask network (MsCM-Net) for hyperspectral unmixing. MsCM-Net introduced a mixed region mask strategy specifically designed for hyperspectral unmixing tasks. It utilized a multiscale convolutional AE as the baseline network to implement this mask strategy, enhancing the method's robustness in addressing ill-posed unmixing problems. Gao et al. [56] trained two cascaded AEs in an end-to-end manner to create a cycle-consistency unmixing network (Cycu-Net) for hyperspectral unmixing. By switching from the initial pixel-level reconstruction assumption to cycle consistency dominated by the cascaded AEs, Cycu-Net limited the loss of specific and precise information during reconstruction. Yu et al. [57] devised a multistage convolutional AE network (MSNet) aimed at capturing broad contextual information while preserving detailed features through a gradual multistage unmixing network. Zhao et al. [58] introduced an AE network with regularization by denoising (AE-RED), which divided the unmixing task into two subtasks. One task involved the use of deep AEs to implicitly regularize the estimates and model the mixing mechanism, while the other task used a denoiser to incorporate explicit information. This

approach combined the strengths of deep AE-based unmixing methods with additional priors offered by denoisers, thereby enhancing the overall performance. However, both of the above methods require endmember signatures acquired from other techniques, such as the well-known VCA algorithm [22], to serve as initial weights for optimizing the proposed network. As a consequence, the randomness inherent in VCA is inevitably transmitted to the network, thereby affecting the accuracy of the unmixing process. Recently, the minimum simplex convolutional network (MiSiCNet) [59] was proposed to incorporate both the spatial correlation between adjacent pixels and the geometrical properties of the linear simplex in order to tackle the randomness inherent to endmember initialization techniques such as VCA.

Very recently, the transformer was introduced as a new model [60] that employs a self-attention mechanism to capture global contextual information and shows promising results in the image processing domain. The transformer model has been applied to hyperspectral image processing [61], [62], [63]. In [64], a convolutional AE was combined with a transformer for unmixing using a multihead self-attention mechanism to identify nonlocal contextual feature dependencies by exploring long-range relations between image patches. Tao et al. [65] designed an abundance-guided spectral and spatial attention network (A2SAN) that aims at directly exploiting the attention mechanism rather than using it solely as a processing module for capturing information within other networks. Although A2SAN solves the issue encountered by most unmixing methods that rely on VCA initialization (that introduces randomness in the network), it still is not able to directly estimate endmember signatures via the network itself.

### C. Motivations

Most existing unmixing networks exploit AE frameworks to estimate abundances and reconstruct the images. The weights between the reconstructed and the softmax layers are extracted as endmember signatures and are often initialized using VCA to enhance the performance of the unmixing. In this sense, they mainly face two challenges: 1) they do not allow a direct estimation of the endmembers from the network, making it difficult to extend the method to nonlinear unmixing and 2) VCA produces different results on different runs. This randomness is passed on to the corresponding network and affects the final performance of the unmixing. Nevertheless, the deep learning approach is a workable and feasible way to address the latent hyperspectral unmixing challenges. This motivated us to develop a universal deep learning-based framework that provides direction for addressing nonlinear unmixing problems and estimates endmembers directly instead of weights while avoiding the randomness typically introduced by existing methods.

### D. Contributions

To achieve the abovementioned goals, we introduce a novel dual-feature fusion network (DFFN), called DFFN, to reconstruct hyperspectral images and estimate endmembers and abundances effectively and directly. More specifically, the

main contributions of the proposed DFFN can be summarized as follows.

- 1) We introduce a novel DFFN network that directly estimates endmembers rather than weights, distinguishing itself from other deep learning-based unmixing frameworks. Our innovative approach not only avoids introducing randomness from other methods into the network but is also the first one to estimate endmember signatures directly within the network.
- 2) In our FFM, we introduce a new strategy for computing both spectral and spatial similarities to enhance hyperspectral images. This operation not only performs denoising by using the neighboring pixels to weight the noise pixels but also improves the final unmixing accuracy.
- 3) We devise a novel loss function that contains a reconstruction error, a correction error, an abundance sum-to-one error, and an abundance nonnegativity error to train the proposed network in an end-to-end manner. These components guarantee improved image reconstruction and better endmember and abundance estimation performance.

The remaining of this article is organized as follows. In Section II, the proposed DFFN is described in detail. Section III presents the experimental results obtained using synthetic and real datasets. Section IV discusses the impact of VCA and FFM as initialization on the unmixing performance. Section V concludes this article with some remarks and outlines potential avenues for future research.

## II. PROPOSED METHOD

In this section, we provide a detailed explanation of the proposed DFFN, outlining how it estimates abundance maps and directly estimates endmembers through the network itself. The overall architecture of DFFN is presented in Fig. 1.

### A. Feature Fusion Module

Many existing unmixing methods often encounter challenges in extracting spectral-spatial information, especially when they rely solely on FC layers to design the network. To address this problem, we design FFM, a feature fusion module (FEM) that captures spectral-spatial information and enhances hyperspectral images for more accurate spectral unmixing.

Let  $\mathbf{Y} \in \mathbb{R}^{Nb \times Np}$  be the hyperspectral image with  $Nb$  bands and  $Np$  pixels. To capture spectral information, the band distance matrix  $\mathbf{D}_B \in \mathbb{R}^{Nb \times Nb}$  is defined as

$$(\mathbf{D}_B)_{ik} = \sum_{j=1}^{Np} (y_{ij} - y_{kj})^2 \quad (1)$$

where  $(\mathbf{D}_B)_{ik}$  denotes the squared Euclidean distance between band  $i$  and band  $k$ , summed over all pixels of the image.

Similarly, to capture spatial information, the pixel distance matrix  $\mathbf{D}_P \in \mathbb{R}^{Np \times Np}$  is defined as

$$(\mathbf{D}_P)_{jr} = \sum_{i=1}^{Nb} (y_{ij} - y_{ir})^2 \quad (2)$$

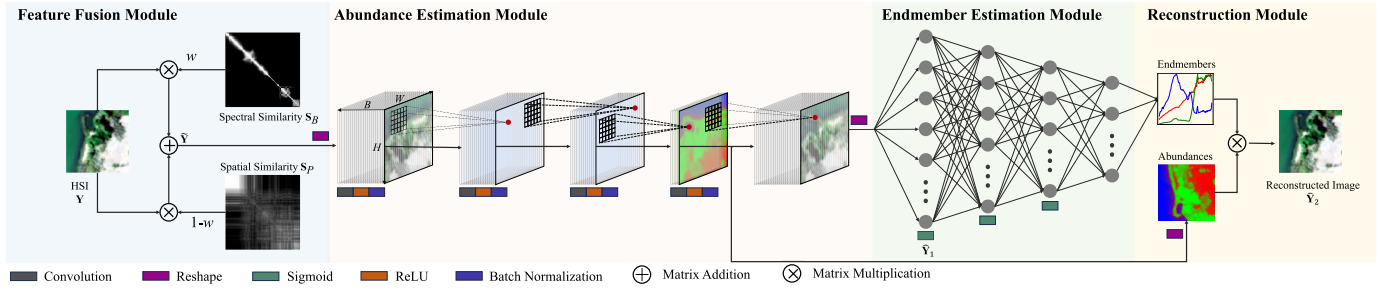


Fig. 1. Architecture of the proposed DFFN for hyperspectral unmixing and reconstruction.

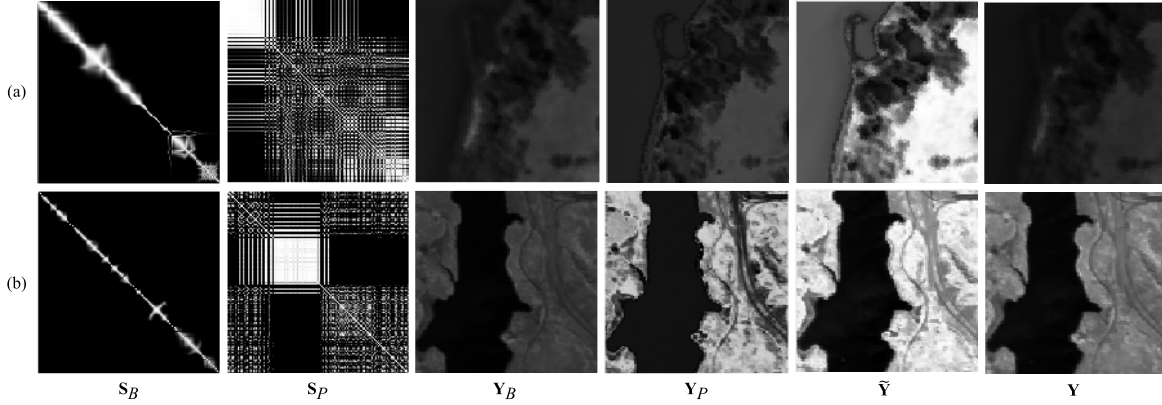


Fig. 2. Visualization of the FFM on (a) Samson and (b) Jasper Ridge datasets.  $\mathbf{S}_B$ : spectral similarity matrix;  $\mathbf{S}_P$ : spatial similarity matrix;  $\mathbf{Y}_B$ : spectrally enhanced image (band 40);  $\mathbf{Y}_P$ : spatially enhanced image (band 40);  $\tilde{\mathbf{Y}}$ : FFM output (band 40); and  $\mathbf{Y}$ : original image (band 40).

where  $(\mathbf{D}_P)_{jr}$  denotes the squared Euclidean distance between pixels  $j$  and  $r$ , summed over all bands.

In both matrices, a larger distance corresponds to a lower similarity. Therefore, a spectral similarity matrix  $\mathbf{S}_B \in \mathbb{R}^{Nb \times Nb}$  and a spatial similarity matrix  $\mathbf{S}_P \in \mathbb{R}^{Np \times Np}$  are defined by

$$\begin{aligned} \mathbf{S}_B &= e^{-\mathbf{D}_B} \\ \mathbf{S}_P &= e^{-\mathbf{D}_P} \end{aligned} \quad (3)$$

where  $\mathbf{S}_B$  and  $\mathbf{S}_P$  are used to enhance the hyperspectral image with spectral and spatial information, respectively,

$$\begin{aligned} \mathbf{Y}_B &= \mathbf{S}_B \mathbf{Y} \\ \mathbf{Y}_P &= \mathbf{Y} \mathbf{S}_P. \end{aligned} \quad (4)$$

$\mathbf{Y}_B$  and  $\mathbf{Y}_P$  are normalized by

$$\begin{aligned} \mathbf{Y}_B &= \frac{\mathbf{Y}_B - \min(\mathbf{Y}_B)}{\max(\mathbf{Y}_B) - \min(\mathbf{Y}_B)} \\ \mathbf{Y}_P &= \frac{\mathbf{Y}_P - \min(\mathbf{Y}_P)}{\max(\mathbf{Y}_P) - \min(\mathbf{Y}_P)}. \end{aligned} \quad (5)$$

Both spectral and spatial information sources are combined by a weighted fusion strategy

$$\tilde{\mathbf{Y}} = w\mathbf{Y}_B + (1 - w)\mathbf{Y}_P \quad (6)$$

where  $w$  is a weight parameter. After final normalization, we have

$$\tilde{\mathbf{Y}} = \frac{\tilde{\mathbf{Y}} - \min(\tilde{\mathbf{Y}})}{\max(\tilde{\mathbf{Y}}) - \min(\tilde{\mathbf{Y}})} \quad (7)$$

where  $\tilde{\mathbf{Y}}$ , a spectrally and spatially enhanced hyperspectral image, is obtained.

Fig. 2 visually shows the image enhancement capabilities of FFM on the well-known Samson and Jasper Ridge datasets. It shows both the spectral similarity matrix  $\mathbf{S}_B$  and the spatial similarity matrix  $\mathbf{S}_P$ , the enhanced hyperspectral images, namely,  $\mathbf{Y}_B$  with enhanced spectral information and  $\mathbf{Y}_P$  with enhanced spatial information, and the final enhanced image  $\tilde{\mathbf{Y}}$  that is clearer than the original hyperspectral image  $\mathbf{Y}$ . To obtain the final enhanced image, the optimal choice of  $w$  is image dependent. For the Samson dataset, one can observe that the information contained in  $\mathbf{Y}_B$  and  $\mathbf{Y}_P$  is similar in terms of clarity. In this sense, we choose 0.5 as the weight parameter. On the other hand, in the Jasper Ridge dataset, the spatial information is dominant, and we choose 0.9 as the weight parameter to enhance the spectral information.

The matrix multiplications with the spectral and spatial similarity matrices in (4) imply a local weighting of neighboring pixels and act as a denoising step. Since the enhanced hyperspectral images are used as the input of AEM, they make the final unmixing results less sensitive to noise.

### B. Abundance Estimation Module

Some deep learning-based hyperspectral unmixing methods have used convolution (Conv) layers in their networks, yielding outstanding results. Therefore, Conv layers are employed in this work to design a straightforward AEM. The proposed AEM includes four Conv layers with a kernel size of  $5 \times 5$ . The number of nodes in these layers is set to 128, 64,  $c$  (the number



TABLE I

PERFORMANCE EVALUATION OF DIFFERENT UNMIXING METHODS ON THE SYNTHETIC DATASET. BEST RESULTS ARE SHOWN IN BOLD, AND SECOND-BEST RESULTS ARE UNDERLINED

SNR	Metrics	CAE [66]	DAEU [45]	MTAEU [67]	SIDAEU [68]	mDAE [52]	A2SN [65]	A2SAN [65]	MSNet [57]	DFFN
10dB	Mean SAD	0.4095	0.1277	0.1332	0.1179	0.1441	0.0932	0.0747	0.0654	<b>0.0504</b>
	Mean RMSE	0.3053	0.2513	0.2154	0.1965	0.3204	0.1681	0.1530	<b>0.1194</b>	<u>0.1364</u>
	RE	0.3802	0.3592	0.3577	0.3956	0.4258	<b>0.3543</b>	<u>0.3547</u>	0.3577	0.3589
20dB	Mean SAD	0.3271	0.1430	0.1094	0.2630	0.1275	<u>0.0529</u>	0.0912	0.1579	<b>0.0228</b>
	Mean RMSE	0.2836	0.2174	0.1372	0.2211	0.3235	0.1313	0.1051	<u>0.1020</u>	<b>0.0681</b>
	RE	0.1785	0.1286	0.1219	0.1276	0.1765	0.1239	0.1256	0.1200	<b>0.1188</b>
30dB	Mean SAD	0.2510	0.0796	0.0938	0.1223	0.0574	0.0585	0.0418	<u>0.0264</u>	<b>0.0041</b>
	Mean RMSE	0.2657	0.2170	0.2320	0.2340	0.3265	0.1255	<u>0.1098</u>	0.1154	<b>0.0617</b>
	RE	0.1831	0.0419	0.0412	0.0470	0.1869	0.0407	<u>0.0346</u>	0.0377	<b>0.0344</b>

TABLE II

PERFORMANCE EVALUATION OF DIFFERENT UNMIXING METHODS ON THE SAMSON DATASET. BEST RESULTS ARE SHOWN IN BOLD, AND SECOND-BEST RESULTS ARE UNDERLINED

Methods		CAE [66]	DAEU [45]	MTAEU [67]	SIDAEU [68]	mDAE [52]	A2SN [65]	A2SAN [65]	MSNet [57]	DFFN
SAD	Soil	1.0500	0.0271	0.0393	0.0236	0.3958	0.0243	0.0208	<u>0.0163</u>	<b>0.0147</b>
	Tree	0.1614	0.0322	0.0430	0.0243	0.1263	<u>0.0224</u>	0.0259	<b>0.0218</b>	0.0236
	Water	0.3227	0.0467	<u>0.0439</u>	0.0542	0.1906	0.0551	0.0625	<b>0.0377</b>	0.0479
Mean SAD		0.5114	0.0353	0.0421	0.0340	0.2376	0.0339	0.0364	<b>0.0253</b>	<u>0.0287</u>
Mean RMSE		0.2462	0.0700	0.0786	0.0614	0.3390	0.0469	<u>0.0443</u>	0.0457	<b>0.0274</b>
RE		0.0965	0.0501	0.0544	0.0590	0.1053	<u>0.0370</u>	0.0390	0.0406	<b>0.0359</b>

of endmembers), and  $Nb$ , respectively. To improve the training process and stability, a batch normalization (BN) layer follows each Conv layer. In addition, the ReLU activation function is employed to introduce nonlinearity into the network. The input of the first Conv layer is given by

$$\mathbf{A}_0 = \text{reshape}(\tilde{\mathbf{Y}}) \quad (8)$$

where  $\tilde{\mathbf{Y}} \in \mathbb{R}^{Nb \times Np}$  is the output of the FFM. The operation “reshape” rearranges the array elements to shape the data as a tensor of size  $\mathbb{R}^{Nb \times W \times H}$ , where  $W$  and  $H$  are the width and height of the hyperspectral image, respectively. The output  $\mathbf{A}_l$  of the  $l$ th hidden layer is given by

$$\mathbf{A}_l = \text{ReLU}(\text{BN}(\text{Conv}(\mathbf{A}_{l-1}))) \quad (9)$$

where  $\mathbf{A}_{l-1}$  is the input of the  $l$ th hidden layer. The output  $\hat{\mathbf{A}}_3 \in \mathbb{R}^{c \times W \times H}$  of layer 3 (with  $c$  nodes) is an estimation of the abundance maps, and the output  $\hat{\mathbf{A}}_4 \in \mathbb{R}^{Nb \times W \times H}$  of layer 4 (with  $Nb$  nodes) is the reconstructed image.

### C. Endmember Estimation Module

The predominant deep learning-based unmixing methods in the literature involve the use of FC layers or Conv layers in their networks. Unfortunately, this tendency often results in an inability to effectively learn spectral-spatial information or in their performance being compromised by external initialization methods. Furthermore, these methods typically estimate endmember signatures from the weights between a softmax

and a reconstruction layer, rather than allowing the network to perform this task autonomously.

For an automatic estimation of the endmembers, we use FC layers to construct the EEM. The module contains three FC layers followed by a Sigmoid activation function, with 1000, 30, and  $c$  nodes, respectively. To include spectral-spatial information, the proposed EEM uses the reconstructed image after the AEM module as the input

$$\mathbf{E}_0 = \hat{\mathbf{Y}}_1 \quad (10)$$

where  $\hat{\mathbf{Y}}_1 = \text{reshape}(\hat{\mathbf{A}}_4) \in \mathbb{R}^{Nb \times Np}$ . The output  $\mathbf{E}_l$  of the  $l$ th hidden layer is obtained as

$$\mathbf{E}_l = \text{Sigmoid}(\text{FC}(\mathbf{E}_{l-1})). \quad (11)$$

The output of the last hidden layer  $\mathbf{E}_{\text{end}}$  provides the endmember signatures.

### D. Reconstruction Module (RM)

The final reconstructed image is obtained by

$$\hat{\mathbf{Y}}_2 = \hat{\mathbf{E}}\hat{\mathbf{A}} \quad (12)$$

where  $\hat{\mathbf{E}} = \hat{\mathbf{E}}_{\text{end}} \in \mathbb{R}^{Nb \times c}$  and  $\hat{\mathbf{A}} = \text{reshape}(\hat{\mathbf{A}}_3) \in \mathbb{R}^{c \times Np}$  are the estimated endmember signatures and abundance maps by the EEM and AEM, respectively.

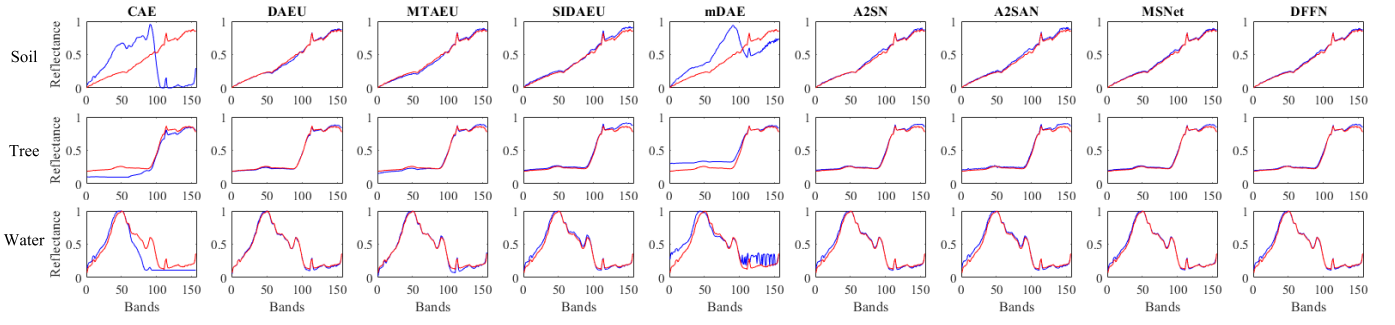


Fig. 3. Reference signatures (red) and estimated endmember signatures (blue) by different unmixing methods on the Samson dataset.

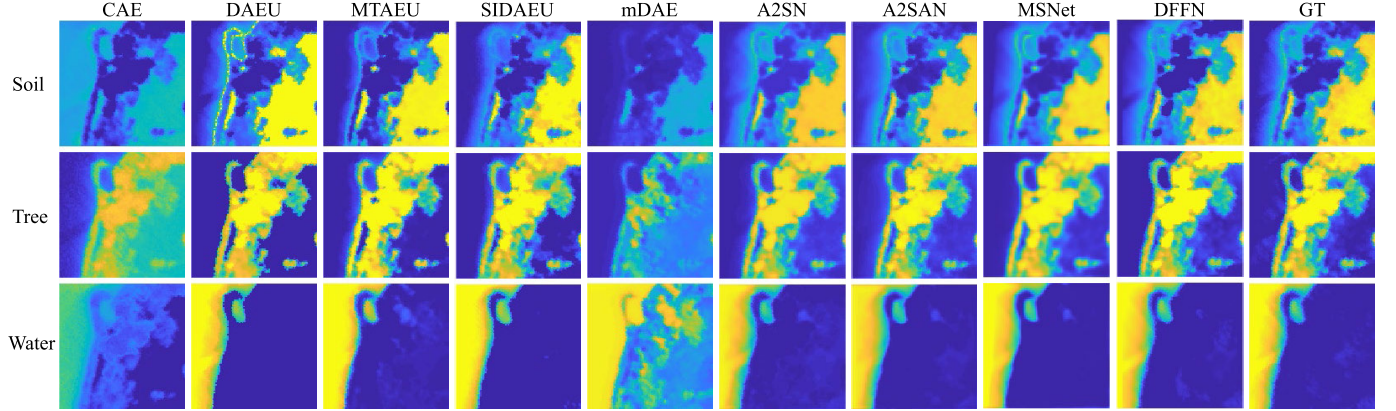


Fig. 4. Abundance maps estimated by different unmixing methods on the Samson dataset. The GT column shows the GT abundance maps.

### E. Loss Function

Our proposed DFFN is an end-to-end model that automatically reconstructs hyperspectral images and estimates abundances and endmembers. To optimize our network, the reconstruction loss and the abundance loss are used. The reconstruction loss is given by

$$L_R = \frac{1}{N_p} \sum_{j=1}^{N_p} \arccos \frac{\mathbf{y}_j \cdot \hat{\mathbf{y}}_{2j}}{\|\mathbf{y}_j\|_2 \|\hat{\mathbf{y}}_{2j}\|_2} \quad (13)$$

where  $\mathbf{y}_j \in \mathbb{R}^{N_b \times 1}$  is the original hyperspectral pixel,  $\hat{\mathbf{y}}_{2j} \in \mathbb{R}^{N_b \times 1}$  represents the reconstructed pixel from the RM module, and  $\|\cdot\|_2$  denotes the  $L_2$  norm.

To construct the abundance loss, both ASC and ANC should be met. To meet the ASC constraint, the ASC loss is defined as

$$L_{ASC} = \frac{1}{N_p} \sum_{j=1}^{N_p} \left[ \left( \sum_{m=1}^c \hat{a}_{mj} \right) - 1 \right]^2 \quad (14)$$

where  $\hat{a}_{mj}$  is the abundance of the  $m$ th endmember at the  $j$ th pixel and  $c$  is the number of endmembers. The ANC loss is defined as

$$L_{ANC} = \frac{1}{c \times N_p} \sum_{i=1}^c \sum_{j=1}^{N_p} \max(0, -\hat{a}_{mj}). \quad (15)$$

Finally, the entire abundance loss is defined as

$$L_A = L_{ASC} + L_{ANC}. \quad (16)$$

In addition, a correction error loss function is constructed to optimize the endmember estimation, which enhances the robustness of the proposed DFFN and minimizes the difference between the input of the EEM  $\hat{\mathbf{Y}}_1$  and the final reconstructed image  $\hat{\mathbf{Y}}_2$

$$L_C = \frac{1}{N_p} \sum_{j=1}^{N_p} \arccos \frac{\hat{\mathbf{y}}_{1j} \cdot \hat{\mathbf{y}}_{2j}}{\|\hat{\mathbf{y}}_{1j}\|_2 \|\hat{\mathbf{y}}_{2j}\|_2}. \quad (17)$$

Finally, the total loss function for optimizing the proposed DFFN is defined as

$$L = L_R + bL_A + cL_C \quad (18)$$

where  $b$  and  $c$  are regularization parameters to balance between image reconstruction, abundance estimation, and endmember estimation.

## III. EXPERIMENTAL RESULTS

In this section, we assess the performance of the proposed DFFN on one synthetic and three widely used real hyperspectral datasets: Samson, Jasper Ridge, and Urban. To ensure a fair comparison, we evaluate the proposed DFFN against several classic deep learning-based unmixing methods that address mixed pixels in the supervised scenario: the convolutional AE (CAE) [66], deep AE unmixing (DAEU) [45], multitask AE unmixing (MTAEU) [67], spectral information divergence AE unmixing (SIDAEU) [68], mDAE [52], abundance-guided spectral and spatial network (A2SN) [65], abundance-guided attention spectral and spatial attention network (A2SAN) [65], and MSNet [57].

TABLE III

PERFORMANCE EVALUATION OF DIFFERENT UNMIXING METHODS ON THE JASPER RIDGE DATASET. BEST RESULTS ARE SHOWN IN BOLD, AND SECOND-BEST RESULTS ARE UNDERLINED

Methods		CAE [66]	DAEU [45]	MTAEU [67]	SIDAEU [68]	mDAE [52]	A2SN [65]	A2SAN [65]	MSNet [57]	DFFN
SAD	Tree	0.2314	0.0185	0.0961	0.0300	0.1322	<b>0.0240</b>	0.0297	0.0442	<u>0.0270</u>
	Water	0.2400	0.0098	0.0955	0.0261	0.1810	<b>0.0120</b>	0.0201	0.0427	<u>0.0162</u>
	Soil	0.2434	0.0506	0.1048	0.0796	0.4673	<u>0.0390</u>	0.0502	0.0479	<b>0.0274</b>
	Road	0.4382	0.2056	0.2172	0.2433	0.1587	0.0934	<u>0.0428</u>	0.1014	<b>0.0524</b>
Mean SAD		0.2883	0.0711	0.1284	0.0948	0.2348	0.0421	<u>0.0357</u>	0.0591	<b>0.0308</b>
Mean RMSE		0.2468	0.1319	0.2700	0.1309	0.2748	<u>0.0956</u>	0.1164	0.0977	<b>0.0951</b>
RE		0.0957	0.0768	0.0888	0.0718	0.1048	0.0700	<b>0.0652</b>	0.0728	<u>0.0697</u>

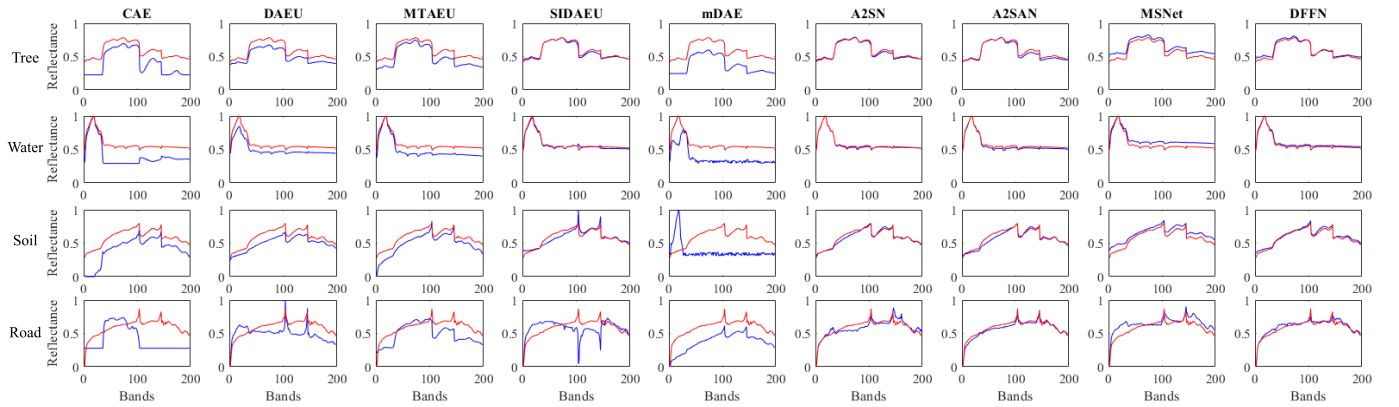


Fig. 5. Reference signatures (red) and estimated endmember signatures (blue) by different unmixing methods on the Jasper Ridge dataset.

### A. Data Description and Evaluation Metrics

1) *Synthetic Dataset*: The developed method assumes that the spectral reflectance of the mixtures can be reconstructed by linearly combining the spectra of pure materials. Thus, to demonstrate the potential of the proposed method, a dataset that perfectly follows the linear mixing model is first generated. For this purpose, we generated a synthetic dataset of  $60 \times 60$  pixels by combining five endmembers linearly. The endmembers were (randomly) chosen from the United States Geological Survey (USGS).<sup>1</sup> The spectral data have 188 bands with a spectral resolution of 10 nm and a wavelength range of  $0.38\text{--}2.5\ \mu\text{m}$ . We start from an image ( $64 \times 64$  pixels), divided into  $8 \times 8$  blocks. For each block, one of the five endmembers is randomly selected and fills the entire block. The abundance maps are generated by the convolution of that image with a low-pass filter, i.e., the  $5 \times 5$  mean filter. Then, the pixels at the edge of the image ( $2 \times 2$  rows and columns) are removed. In order to investigate the robustness of the proposed method in the presence of noise, Gaussian noise at different signal-to-noise ratio (SNR) levels was added to the spectra. For each noise level, different sets of endmembers were used to generate the synthetic dataset, but the spatial ordering remained the same.

2) *Samson Dataset*: The Samson dataset is composed of  $952 \times 952$  pixels with 156 bands, covering the spectral range from 401 to 889 nm. To alleviate computational costs, a region

of  $95 \times 95$  pixels is extracted, starting from the (252, 332)th pixel. The dataset contains three endmembers: soil, tree, and water.

3) *Jasper Ridge Dataset*: The Jasper Ridge dataset has 224 bands with wavelengths ranging from 380 to 2500 nm and is composed of  $512 \times 614$  pixels. A subset of  $100 \times 100$  pixels, starting from the (105, 269)th pixel, is extracted. Bands 1–3, 108–112, 154–166, and 220–224 are removed due to atmospheric interferers and dense water vapor, and the remaining 198 bands are retained. The final Jasper Ridge dataset includes four endmembers: water, soil, tree, and road.

4) *Urban Dataset*: The Urban dataset consists of 210 bands with a wavelength range of 400–2500 nm and  $307 \times 307$  pixels. To mitigate the effects of atmospheric interferers and dense water vapor, bands 1–4, 76, 87, 101–111, 136–153, and 198–210 are excluded, resulting in 162 remaining bands. The final Urban dataset comprises four endmembers: asphalt, grass, tree, and roof.

5) *Evaluation Metrics*: In our experiments, we introduce three commonly used evaluation metrics to assess the performance of algorithms: the spectral angle distance (SAD) between the estimated and ground-truth (GT) endmember signatures, the root mean square error (RMSE) between the estimated and GT abundances, and the reconstruction error (RE) between the reconstructed and the original spectra. These metrics are defined as follows:

$$\text{SAD}(\mathbf{e}_m, \hat{\mathbf{e}}_m) = \arccos \frac{\mathbf{e}_m \cdot \hat{\mathbf{e}}_m}{\|\mathbf{e}_m\|_2 \|\hat{\mathbf{e}}_m\|_2} \quad (19)$$

<sup>1</sup><https://www.usgs.gov/labs/spectroscopy-lab/science/spectral-library>

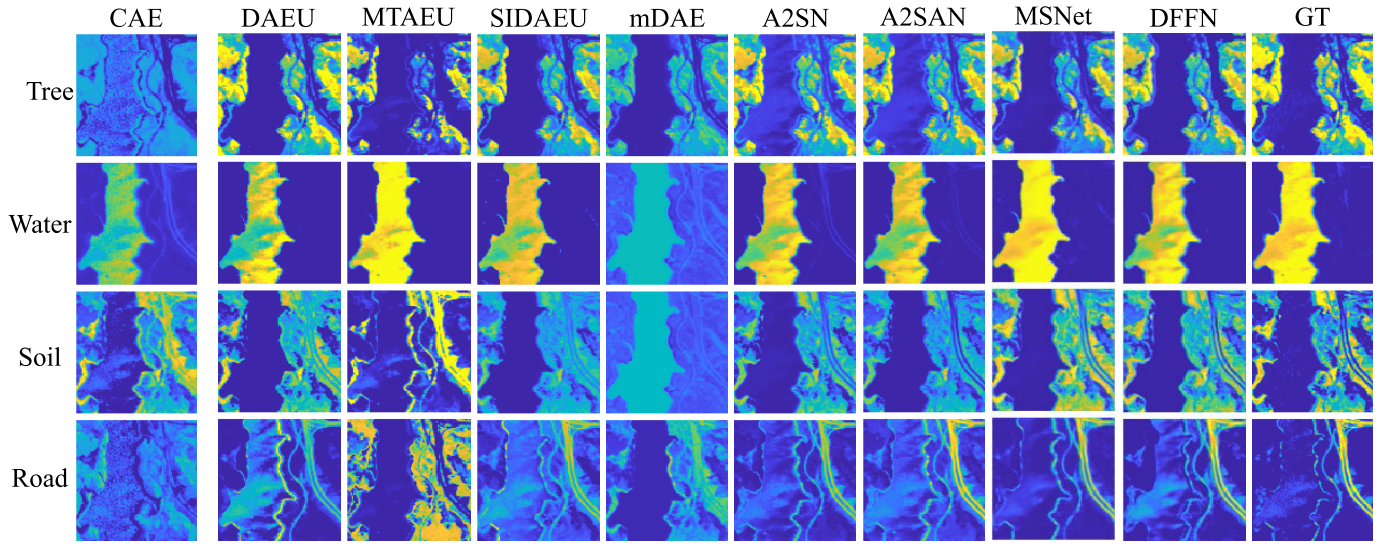


Fig. 6. Abundance maps estimated by different unmixing methods on the Jasper Ridge dataset. The GT column shows the GT abundance maps.

TABLE IV

PERFORMANCE EVALUATION OF DIFFERENT UNMIXING METHODS ON THE URBAN DATASET. BEST RESULTS ARE SHOWN IN BOLD, AND SECOND-BEST RESULTS ARE UNDERLINED

Methods		CAE [66]	DAEU [45]	MTAEU [67]	SIDAUE [68]	mDAE [52]	A2SN [65]	A2SAN [65]	MSNet [57]	DFFN
SAD	Asphalt	0.3523	0.2106	0.1411	0.2901	0.2163	<u>0.0735</u>	0.1113	0.1511	<b>0.0728</b>
	Grass	0.9602	0.2171	0.3427	0.3471	0.4843	<b>0.0520</b>	<u>0.0494</u>	0.1022	0.0654
	Tree	0.1629	0.0806	0.0773	0.1338	0.1567	<b>0.0199</b>	0.0557	0.0948	<u>0.0358</u>
	Roof	0.2044	0.2436	0.1473	0.2555	0.2105	<b>0.0945</b>	0.1337	0.1736	<u>0.1296</u>
Mean SAD		0.4199	0.1880	0.1771	0.2566	0.2670	<b>0.0600</b>	0.0875	0.1304	<u>0.0759</u>
Mean RMSE		0.2679	0.1454	0.1560	0.1670	0.2704	<b>0.1013</b>	0.1065	0.1259	<u>0.1049</u>
RE		0.1696	0.0548	0.0713	<b>0.0509</b>	0.1151	0.0650	<u>0.0540</u>	0.0837	0.0707

$$\text{RMSE}(\hat{\mathbf{a}}_j, \mathbf{a}_j) = \sqrt{\frac{1}{N_p} \sum_{j=1}^{N_p} (\hat{\mathbf{a}}_j - \mathbf{a}_j)^2} \quad (20)$$

$$\begin{aligned} \text{RE}(\mathbf{Y}, \hat{\mathbf{Y}}_2) &= \frac{1}{N_p} \sum_{j=1}^{N_p} \text{RE}(\mathbf{y}_j, \hat{\mathbf{y}}_{2j}) \\ &= \frac{1}{N_p} \sum_{j=1}^{N_p} \arccos \frac{\mathbf{y}_j \cdot \hat{\mathbf{y}}_{2j}}{\|\mathbf{y}_j\|_2 \|\hat{\mathbf{y}}_{2j}\|_2} \end{aligned} \quad (21)$$

where  $\hat{\mathbf{e}}_m$ ,  $\hat{\mathbf{a}}_j$ , and  $\hat{\mathbf{y}}_{2j}$  stand for the estimated endmember signatures, abundance maps, and hyperspectral images, respectively. Similarly,  $\mathbf{e}_m$ ,  $\mathbf{a}_j$ , and  $\mathbf{y}_j$  are the real endmember signatures, abundance maps, and hyperspectral images.

### B. Experiments on the Synthetic Dataset

Table I lists the quantitative results obtained by different unmixing methods on the synthetic data with different noise levels. It can be seen that the proposed DFFN obtained the best results in terms of mean SAD, mean RMSE, and RE on the synthetic data with 20- and 30-dB noises. Although DFFN did not obtain the best results on synthetic data with 10-dB noise in terms of mean RMSE and RE, its performance was very close to the best performers (i.e., MSNet in terms of mean RMSE and A2SN in terms of RE). Consequently, our

proposed DFFN consistently showed exceptional capability in estimating endmember signatures, estimating abundance maps, and reconstructing images.

### C. Experiments on the Samson Dataset

Table II reports the performance of different unmixing methods in terms of endmember estimation, abundance estimation, and image reconstruction on the Samson dataset. From the table, it can be observed that both DFFN and MSNet performed well in the task of estimating endmembers. MSNet accurately estimated tree and water, while DFFN obtained the best result in estimating soil. In terms of RMSE and RE, DFFN outperformed the other methods and produced the best results in the task of estimating abundance maps. Figs. 3 and 4 visually show the endmember signatures and abundance maps obtained by all the unmixing methods on the Samson dataset, respectively. From Fig. 3, it can be seen that CAE and mDAE could not accurately estimate endmembers compared with the other methods, while the endmember signatures obtained by MSNet and DFFN agree well with the GT. From Fig. 4, A2SN, A2SAN, MSNet, and DFFN performed well in estimating abundance maps, and the results of DFFN are closest to the GT, which further shows the effectiveness and superiority of the proposed approach in abundance estimation.



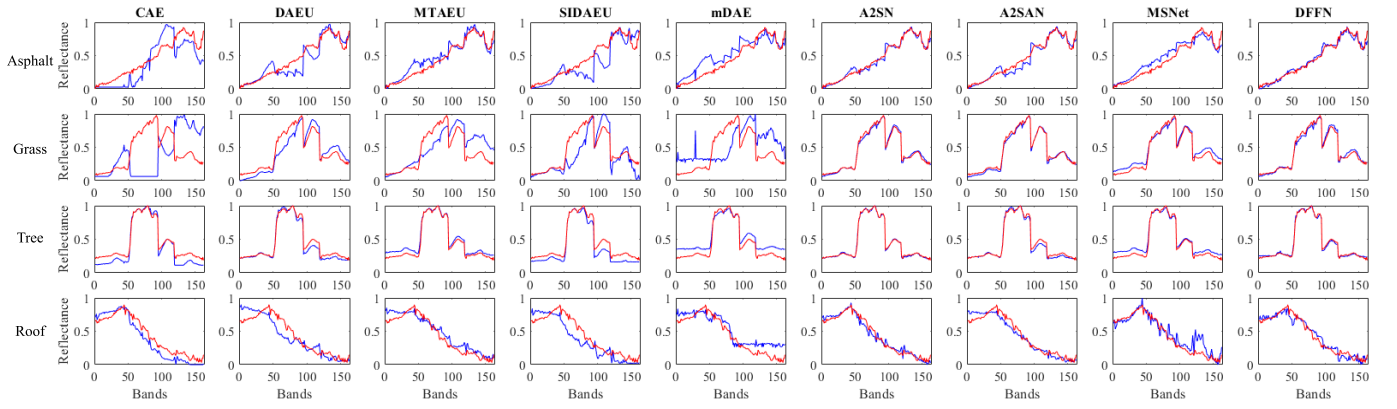


Fig. 7. Reference signatures (red) and estimated endmember signatures (blue) by different unmixing methods on the Urban dataset.

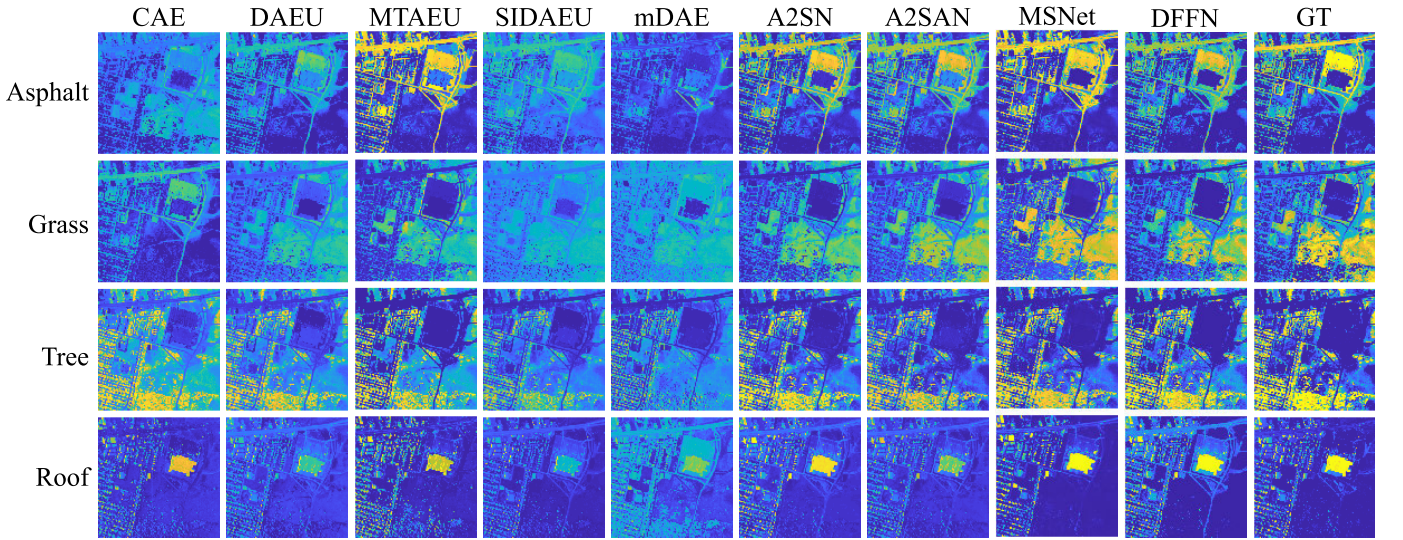


Fig. 8. Abundance maps estimated by different unmixing methods on the Urban dataset. The GT column shows the GT abundance maps.

#### D. Experiments on the Jasper Ridge Dataset

Table III shows the quantitative results obtained by the different unmixing methods on the Jasper dataset in terms of mean SAD for estimating endmember signatures, mean RMSE for estimating abundance maps, and RE for reconstructing images. From the table, we observe that A2SN could accurately estimate the two endmembers tree and water, while DFFN performed well in estimating soil and road. The proposed DFFN performed the best in terms of overall endmember estimation. Moreover, DFFN obtained the best result and second-best result in abundance estimation and image reconstruction. Fig. 5 shows the estimated endmembers by the different unmixing methods on the Jasper dataset. One can observe that the endmember signatures obtained by the last four methods, i.e., A2SN, A2SAN, MSNet, and DFFN, were close to the reference signatures, while the proposed DFFN performed the most consistently. Fig. 6 shows the abundance maps estimated by the different unmixing methods on the Jasper dataset. A2SN, MSNet, and DFFN obtained the best results.

#### E. Experiments on the Urban Dataset

Table IV summarizes the mean SAD, mean RMSE, and RE for all competitive methods on the Urban dataset. From the

table, it can be observed that DFFN obtained the best results in estimating the endmember of asphalt, while A2SN estimated the grass, tree, and roof endmembers well. Generally, in terms of overall endmember estimation and abundance estimation, A2SN performed the best and DFFN performed second best, while the gap with A2SN was very small. Moreover, SIDAUE performed the best in image reconstruction. Fig. 7 visually shows the endmember signatures obtained by the different unmixing methods on the Urban dataset. It is obvious that the spectral curves of A2SN, A2SAN, and DFFN were closer to the reference signatures compared with other methods. Fig. 8 depicts the abundance maps estimated by all compared methods for the Urban dataset. It can be observed that CAE, DAEU, MTAEU, SIDAUE, and mDAE failed to recover the asphalt and grass abundances, while DFFN successfully separated these materials and provided accurate abundance estimation.

#### F. Parameter Setting and Parameter Analysis

Figs. 9–11 report the results of hyperparameter analysis on the different datasets. It can be observed that the proposed DFFN produced the best results on the three datasets for a learning rate of  $1e^{-3}$ . Fig. 9(b) and (c) shows that DFFN performed best on the Samson dataset when the regularization

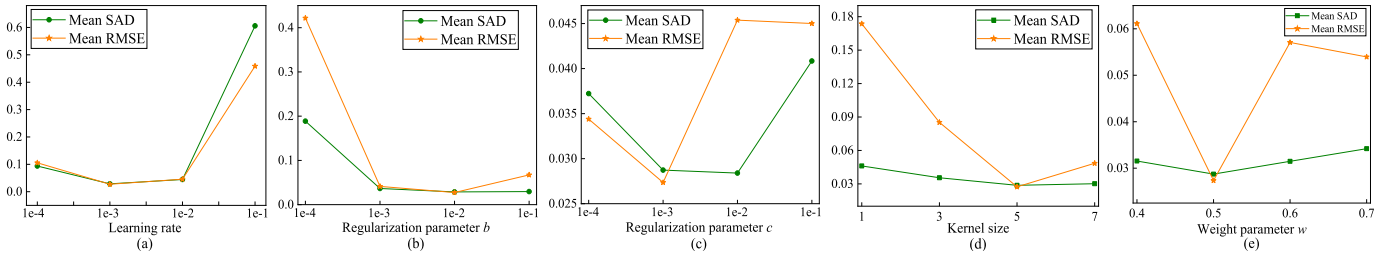


Fig. 9. Hyperparameter analysis on the Samson dataset by varying, (a) Learning rate, (b) regularization parameter  $b$ , (c) regularization parameter  $c$ , (d) kernel size, and (e) weight parameter  $w$ .

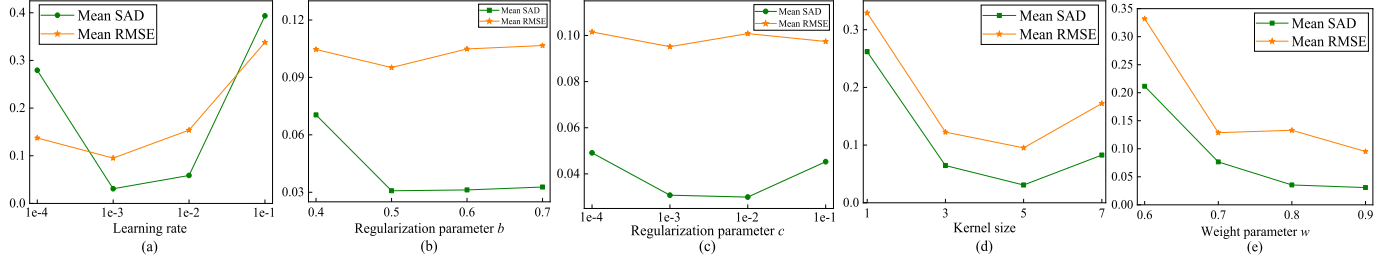


Fig. 10. Hyperparameter analysis on the Jasper Ridge dataset by varying, (a) Learning rate, (b) regularization parameter  $b$ , (c) regularization parameter  $c$ , (d) kernel size, and (e) weight parameter  $w$ .

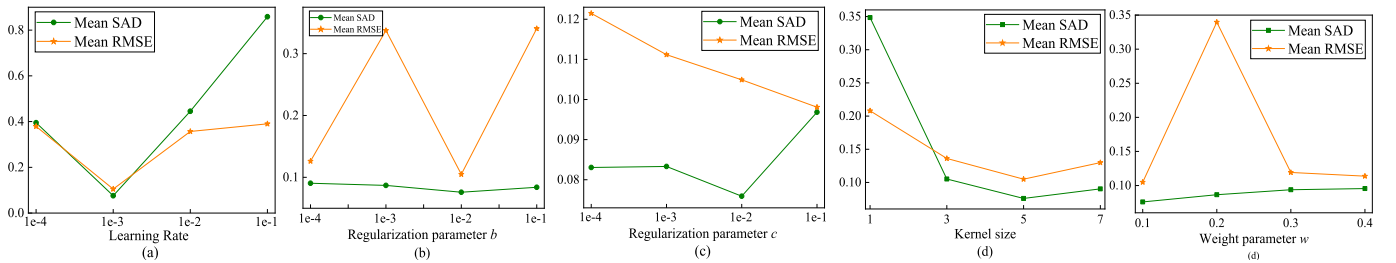


Fig. 11. Hyperparameter analysis on the Urban dataset by varying, (a) Learning rate, (b) regularization parameter  $b$ , (c) regularization parameter  $c$ , (d) kernel size, and (e) weight parameter  $w$ .

TABLE V

RUNNING TIME (S) OF DIFFERENT UNMIXING METHODS ON THE CONSIDERED DATASETS

Methods	Synthetic (20dB)	Samson	Jasper Ridge	Urban
CAE	17.8016	37.4707	53.3654	441.3218
DAEU	18.4865	43.8835	47.9196	52.8121
MTAEU	35.5518	85.9920	99.3320	107.5883
SIDAEU	4.9552	41.2642	10.7187	48.4249
mDAE	13.5610	11.2525	12.4647	103.4156
A2SN	24.0497	41.8415	24.9551	180.3824
A2SAN	23.9539	32.6343	27.5029	214.4112
MSNet	10.7441	11.2745	11.6601	49.6279
DFFN	27.6473	43.2153	64.1353	277.9177

TABLE VI

ABLATION STUDY ON OUR PROPOSED DFFN WITH VARYING INFORMATION ENHANCEMENT ON DIFFERENT DATASETS. BEST RESULTS ARE SHOWN IN BOLD

Datasets	Spectral enhancement	Spatial enhancement	Mean SAD	Mean RMSE
Synthetic data (20dB)	✓	✗	<b>0.0213</b>	<b>0.0650</b>
	✗	✓	0.1705	0.1264
	✓	✓	0.0228	0.0681
Samson	✓	✗	0.0378	0.0400
	✗	✓	0.0328	0.0755
	✓	✓	<b>0.0287</b>	<b>0.0274</b>
Jasper Ridge	✓	✗	<b>0.0275</b>	0.1017
	✗	✓	0.1613	0.2847
	✓	✓	0.0308	<b>0.0951</b>
Urban	✓	✗	0.1221	0.3381
	✗	✓	0.0826	0.3416
	✓	✓	<b>0.0759</b>	<b>0.1049</b>

parameters  $b$  and  $c$  are set to  $1e^{-1}$  and  $1e^{-3}$ , respectively. Fig. 10(b) and (c) shows that DFFN performed best on the Jasper Ridge dataset when the regularization parameters  $b$  and  $c$  are set to 0.5 and  $1e^{-2}$ . Fig. 11(b) and (c) reveals that DFFN performed best on the Urban dataset when the regularization parameters  $b$  and  $c$  both are set to  $1e^{-2}$ . Figs. 9–11 show that the optimal kernel size of the convolutional layers in DFFN is  $5 \times 5$ . The weight parameter  $w$  that balances between spectral

information and spatial information in the FFM is optimally set to 0.5, 0.9, and 0.1 for the Samson, Jasper Ridge, and Urban datasets, respectively.

### G. Computational Cost

All experiments were conducted on a computer with a 2.6-GHz Intel Core i7 CPU and 16 GB of memory (NVIDIA GeForce RTX 2060 GPU). The running times of the different

TABLE VII

IMPACT OF INITIALIZATION BY VCA ON THE UNMIXING PERFORMANCE OF MSNET ALONG WITH RESULTS OF THE PROPOSED DFFN. BEST RESULTS ARE SHOWN IN BOLD

Methods	Synthetic (20dB)		Samson		Jasper		Urban	
	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE
MSNet without VCA	0.5403	0.3243	0.2055	0.2125	0.4810	0.3424	0.3946	0.2693
MSNet with VCA-1	0.1579	0.1020	<b>0.0253</b>	0.0457	0.0591	0.0977	0.1304	0.1259
MSNet with VCA-2	0.0694	0.1123	0.0398	0.0612	0.0729	0.0971	0.0802	0.1259
MSNet with VCA-3	0.1656	0.1127	0.0349	0.0453	0.0432	0.1035	0.1269	0.1567
DFFN	<b>0.0228</b>	<b>0.0681</b>	0.0287	<b>0.0274</b>	<b>0.0308</b>	<b>0.0951</b>	<b>0.0759</b>	<b>0.1049</b>

TABLE VIII

UNMIXING PERFORMANCE OF MSNET WITH AND WITHOUT FFM ALONG WITH THE RESULTS OF THE PROPOSED DFFN. BEST RESULTS ARE SHOWN IN BOLD

Methods	Synthetic (20dB)		Samson		Jasper		Urban	
	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE	Mean SAD	Mean RMSE
MSNet without FFM	0.1579	0.1020	0.0253	0.0457	0.0591	0.0977	0.1304	0.1259
MSNet with FFM	0.0381	0.0803	<b>0.0248</b>	0.0432	0.0546	<b>0.0929</b>	0.0826	0.1229
DFFN	<b>0.0228</b>	<b>0.0681</b>	0.0287	<b>0.0274</b>	<b>0.0308</b>	0.0951	<b>0.0759</b>	<b>0.1049</b>

unmixing methods on the different datasets are reported in Table V. It can be observed that SIDAEU, mDAE, and MSNet are more efficient than other unmixing methods. Considering the results of endmember estimation, abundance estimation, and image reconstruction, although the proposed DFFN is not the fastest unmixing method, its computational cost is acceptable.

#### IV. DISCUSSION

##### A. Ablation Study

To validate the FFM of the proposed approach, an ablation study was performed in which either the spectral or spatial enhancement step (or both) has been considered. Results on all datasets are shown in Table VI. From the table, it can be observed that the proposed method achieved the best results on the synthetic data when spatial enhancement was removed. This may be due to nonrealistic spatial patterns in the synthetic dataset. From the table, it can also be observed that considering both spectral enhancement and spatial enhancement leads to the best results in endmember estimation and abundance estimation on the Samson and Urban datasets. The ablation study reflects that the proposed fusion strategy FFM is very effective for the unmixing task.

##### B. Initialization Analysis of VCA

Most existing unmixing methods based on deep learning use VCA as the initial weight to improve the performance of endmember extraction. However, the randomness of VCA will inevitably be propagated to the corresponding network and greatly affects the final unmixing performance.

Given that MSNet generates superior unmixing results compared to other methods, we consider MSNet as a representative algorithm of the SOTA and make use of it to assess the

impact of initialization by VCA. For this, MSNet is run without any initialization and with three different runs of VCA as initialization. Results are reported in Table VII. To compare, the results of the proposed DFFN are shown as well. As can be observed, the performance of MSNet reduces significantly when endmembers are not initialized properly. Moreover, although VCA is an effective initialization, it produces different endmembers on different runs. This randomness is passed to the network, which subsequently affects the unmixing performance. To address the randomness problem caused by initialization, our proposed DFFN directly estimates endmember signatures and abundance maps without relying on VCA and demonstrates superior performance compared to other methods.

##### C. Evaluation of the FFM Module

The proposed approach includes a feature fusion strategy, FFM, to enhance hyperspectral images. In theory, FFM should also be effective for other unmixing methods. Considering the superiority of MSNet, we again use it as a representative algorithm of the SOTA to test the effectiveness of FFM. Table VIII shows the mean SAD and mean RMSE of MSNet with and without using FFM on all datasets. When including FFM, the performance of MSNet improves drastically. Remark, however, that the proposed approach remains superior. Therefore, the proposed FFM serves not only as a denoising tool to improve the robustness of the proposed DFFN but also as a preprocessing step of other methods to enhance hyperspectral images for more effective unmixing.

#### V. CONCLUSION

In this article, we have developed a new dual-feature feature fusion network (referred to as DFFN) for hyperspectral



unmixing. Our newly proposed DFFN contains an FFM that calculates pixel and band similarities to enhance the hyperspectral image using spatial and spectral information. The enhanced hyperspectral image is used as the input of an abundance estimation module (AEM) and an endmember estimation module (EEM). The reconstructed image is obtained by the matrix multiplication between the endmember signatures and abundance maps obtained by these modules. Experimental results show that our newly proposed DFFN can effectively estimate endmember signatures, estimate abundance maps, and reconstruct images. Its performance is compared to state-of-the-art methods and is found to be highly competitive. In the future, we will devote our efforts to designing a general model to automatically determine the number of endmembers and estimate endmember signatures and abundance maps.

## REFERENCES

- [1] N. Audebert, B. Le Saux, and S. Lefevre, "Deep learning for classification of hyperspectral data," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.
- [2] P. Ghamisi et al., "New frontiers in spectral–spatial hyperspectral image classification: The latest advances based on mathematical morphology, Markov random fields, segmentation, sparse representation, and deep learning," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 3, pp. 10–43, Sep. 2018.
- [3] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.
- [4] H. Su, Z. Wu, H. Zhang, and Q. Du, "Hyperspectral anomaly detection: A survey," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 1, pp. 64–90, Mar. 2022.
- [5] Z. Wu et al., "Background-guided deformable convolutional autoencoder for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5531816.
- [6] W. Dong et al., "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [7] Y. Xu et al., "Hyperspectral image super-resolution with ConvLSTM skip-connections," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5519016.
- [8] Q. Qu, B. Pan, X. Xu, T. Li, and Z. Shi, "Unmixing guided unsupervised network for RGB spectral super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 4856–4867, 2023.
- [9] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 140–158, Jun. 2019.
- [10] G. Wang, G. Cheng, P. Zhou, and J. Han, "Cross-level attentive feature aggregation for change detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 7, pp. 6051–6062, Jul. 2024.
- [11] W. K. Ma et al., "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, Jan. 2014.
- [12] J. Wei and X. Wang, "An overview on linear unmixing of hyperspectral data," *Math. Problems Eng.*, vol. 2020, pp. 1–12, Aug. 2020.
- [13] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tournet, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," *IEEE Trans. Signal Process.*, vol. 64, no. 2, pp. 525–538, Jan. 2016.
- [14] R. Li, B. Pan, X. Xu, T. Li, and Z. Shi, "Toward convergence: A gradient-based multiobjective method with greedy hash for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5509114.
- [15] X. Tao et al., "Fast orthogonal projection for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5523313.
- [16] J. M. Nascimento and J. M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," *Proc. SPIE*, vol. 7477, pp. 157–164, Sep. 2009.
- [17] J. Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 480–492, Jan. 2013.
- [18] 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.
- [19] B. Yang, B. Wang, and Z. Wu, "Nonlinear hyperspectral unmixing based on geometric characteristics of bilinear mixture models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 694–714, Feb. 2018.
- [20] D. C. Heinz, "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.
- [21] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4112–4122, Nov. 2011.
- [22] J. M. P. Nascimento and J. M. Bioucas-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.
- [23] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4418–4432, Nov. 2009.
- [24] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, vol. 3753, pp. 266–275, Jul. 1999, doi: 10.1117/12.366289.
- [25] W. E. Full, R. Ehrlich, and J. E. Klovian, "EXTENDED QMODEL—Objective definition of external end members in the analysis of mixtures," *J. Int. Assoc. Math. Geol.*, vol. 13, no. 4, pp. 331–344, Aug. 1981.
- [26] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542–552, May 1994.
- [27] J. Li, A. Agathos, D. Zaharie, J. M. Bioucas-Dias, A. Plaza, and X. Li, "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067–5082, Sep. 2015.
- [28] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in *Proc. 1st Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Aug. 2009, pp. 1–4.
- [29] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tournet, and A. O. Hero, "Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Trans. Signal Process.*, vol. 57, no. 11, pp. 4355–4368, Nov. 2009.
- [30] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765–777, Mar. 2007.
- [31] 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.
- [32] M.-D. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2014–2039, Jun. 2011.
- [33] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484–4502, Nov. 2012.
- [34] B. Rasti and B. Koirala, "SUNCNN: Sparse unmixing using unsupervised convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [35] A. Halimi, Y. Altmann, N. Dobigeon, and J. Tournet, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.
- [36] R. Heylen and P. Scheunders, "A multilinear mixing model for nonlinear spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 240–251, Jan. 2016.
- [37] A. Marinoni and P. Gamba, "A novel approach for efficient  $p$ -linear hyperspectral unmixing," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1156–1168, Sep. 2015.
- [38] 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.
- [39] A. Marinoni, A. Plaza, and P. Gamba, "Harmonic mixture modeling for efficient nonlinear hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4247–4256, Sep. 2016.



- [40] N. Yokoya, J. Chanussot, and A. Iwasaki, "Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 2, pp. 1430–1437, Feb. 2014.
- [41] C. Févotte and N. Dobigeon, "Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4810–4819, Dec. 2015.
- [42] M. F. Guerri, C. Distant, P. Spagnolo, F. Bougourzi, and A. Taleb-Ahmed, "Deep learning techniques for hyperspectral image analysis in agriculture: A review," *ISPRS Open J. Photogramm. Remote Sens.*, vol. 12, Apr. 2024, Art. no. 100062.
- [43] Y. Fu, Z. Liu, and J. Lyu, "Transferable adversarial attacks for remote sensing object recognition via spatial-frequency co-transformation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5636812.
- [44] J. S. Bhatt and M. V. Joshi, "Deep learning in hyperspectral unmixing: A review," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Sep. 2020, pp. 2189–2192.
- [45] 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.
- [46] 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.
- [47] X. Tao et al., "A new deep convolutional network for effective hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6999–7012, 2022.
- [48] P. Jia, M. Zhang, and Y. Shen, "Deep spectral unmixing framework via 3D denoising convolutional autoencoder," *IET Image Process.*, vol. 15, no. 7, pp. 1399–1409, May 2021.
- [49] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "UnDIP: Hyperspectral unmixing using deep image prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5504615.
- [50] M. Xu, J. Xu, S. Liu, H. Sheng, and Z. Yang, "Multiscale convolutional mask network for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 3687–3700, 2024.
- [51] M. Li, B. Yang, and B. Wang, "EMLM-net: An extended multi-linear mixing model-inspired dual-stream network for unsupervised nonlinear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5509116.
- [52] 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.
- [53] Z. Han, D. Hong, L. Gao, B. Zhang, and J. Chanussot, "Deep half-siamese networks for hyperspectral unmixing," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 11, pp. 1996–2000, Nov. 2021.
- [54] Y. Su, X. Xu, J. Li, H. Qi, P. Gamba, and A. Plaza, "Deep autoencoders with multitask learning for bilinear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8615–8629, Oct. 2021.
- [55] Y. Gao, B. Pan, X. Xu, X. Song, and Z. Shi, "A reversible generative network for hyperspectral unmixing with spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5519115.
- [56] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, "CyCU-net: Cycle-consistency unmixing network by learning cascaded autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
- [57] Y. Yu, Y. Ma, X. Mei, F. Fan, J. Huang, and H. Li, "Multi-stage convolutional autoencoder network for hyperspectral unmixing," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 113, Sep. 2022, Art. no. 552981.
- [58] M. Zhao, J. Chen, and N. Dobigeon, "AE-RED: A hyperspectral unmixing framework powered by deep autoencoder and regularization by denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5512115.
- [59] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "MiSiC-Net: Minimum simplex convolutional network for deep hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 552981.
- [60] K. Han et al., "A survey on vision transformer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2022.
- [61] L. Scheibenreif, M. Mommert, and D. Borth, "Masked vision transformers for hyperspectral image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2023, pp. 2165–2175.
- [62] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral-spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5522214.
- [63] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
- [64] P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, and P. Scheunders, "Hyperspectral unmixing using transformer network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5535116.
- [65] X. Tao, M. E. Paoletti, Z. Wu, J. M. Haut, P. Ren, and A. Plaza, "An abundance-guided attention network for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5505414.
- [66] Y. Ranasinghe et al., "Convolutional autoencoder for blind hyperspectral image unmixing," in *Proc. IEEE 15th Int. Conf. Ind. Inf. Syst. (ICIS)*, Nov. 2020, pp. 174–179.
- [67] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, 2019.
- [68] F. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, "Neural network hyperspectral unmixing with spectral information divergence objective," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 755–758.