MiSiCNet : Minimum Simplex Convolutional Network for deep hyperspectral unmixing

MiSiCNet: Minimum Simplex Convolutional Network for Deep Hyperspectral Unmixing

Behnood Rasti, *Senior Member, IEEE,* Bikram Koirala, *Member, IEEE,* Paul Scheunders, *Senior Member, IEEE,* and Jocelyn Chanussot, *Fellow Member, IEEE*

***Abstract*—© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. In this paper, we propose minimum simplex convolutional network for deep hyperspectral unmixing (MiSiCNet). Unlike all the deep learning-based unmixing methods proposed in the literature, the proposed convolutional encoder-decoder architecture incorporates spatial and geometrical information of the hyperspectral data, in addition to the spectral information. The spatial information is incorporated using convolutional filters and implicitly applying a prior on the abundances. The geometrical information is exploited by incorporating a minimum simplex volume penalty term in the loss function for the endmember extraction. This term is beneficial when there are no pure material pixels in the data, which is often the case in real-world applications. We generated simulated datasets, where we consider two different no-pure pixel scenarios. In the first scenario, there are no pure pixels but at least two pixels on each facet of the data simplex (i.e., mixtures of 2 pure materials). The second scenario is a complex case with no pure pixels and only one pixel on each facet of the data simplex. Additionally, we evaluate the performance of MiSiCNet in three real datasets. The experimental results confirm the robustness of the proposed method to both noise and absence of pure pixels. Additionally, MiSiCNet considerably outperforms the state-of-the-art unmixing approaches. The results are given in terms of spectral angle distance in degree for the endmember estimation, and root mean square error in percentage for the abundance estimation. MiSiCNet was implemented in Python (3.8) using PyTorch as the platform for the deep network and is available online: https://github.com/BehnoodRasti/MiSiCNet.**

***Index Terms*—Hyperspectral image, unmixing, convolutional neural network, deep learning, deep prior, endmember extraction, minimum simplex volume, blind unmixing**

I. Introduction

A spectral pixel is generally a mixture of the pure spectra of the materials within the pixel, called endmembers [1], [2]. Estimating the fractional abundances of the endmembers within spectral pixels is called spectral unmixing. If the materials are intimately mixed, and multiple reflections are received at the sensor, then nonlinear unmixing is utilized. On the other hand, linear unmixing is found helpful in remote sensing applications, where hyperspectral images often consist of homogeneous regions of single materials [3], [4].

If there exist pure pixels for each material within the scene, then the endmembers can be easily extracted using a geometrical approach relying on the simplex of the data. After extracting the endmembers, the abundances can be estimated by minimizing the least squared errors between the actual spectra and reconstructed spectra from the endmembers, subjected to the physical constraints on the abundances, i.e., the abundance non-negativity constraint (ANC) and the abundance sum-to-one constraint (ASC) [5].

The endmembers are assumed to be located at the vertices of the data simplex. Therefore, they can be extracted by maximizing the data simplex e.g, by simplex volume maximization (SiVM) [6]. In [7], convex optimizations were used to solve the maximum volume inscribed simplex (MVIS) problem. MVIS searches for the data points whose simplex convex hull is inscribed in the data convex hull while the simplex volume is maximized. Alternatively, the vertices can be selected as the extreme points after iteratively projecting the data onto a particular direction. For instance, Vertex Component Analysis (VCA) [8] selects endmembers iteratively by projecting the data into an orthogonal direction to the subspace spanned by the already selected endmember. Pixel Purity Index (PPI) [9] scores the spectral vectors by projecting them onto a large set of random vectors (called skewers) and counting the number of times that each vector is an extreme point. N-FINDR [10] searches for pure pixels that form the largest simplex by gradually inflating a simplex inside the data.

In the absence of pure pixels, the geometrical approaches can still be successful when enough data points are available on the facets of the data simplex. Then, virtual endmembers located at the vertices can be estimated. For instance, the minimum volume enclosing simplex (MVES) problem seeks the minimum simplex which encloses the data points [11], [12] and therefore does not rely on the pure pixel assumption. In [13], a Maximum Volume Inscribed Ellipsoid (MVIE) method was proposed. MVIE seeks the maximum volume ellipsoid contained within the convex hull of the data points. This volume is maximal when touching the facets of the data simplex. The contact points provides clues on how to estimate the virtual endmembers.

In no pure pixel-scenarios, estimating the endmembers and the abundances simultaneously is called blind unmixing. A

Behnood Rasti (corresponding author) is with Helmholtz-Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Machine Learning Group, Chemnitzer Straße 40, 09599 Freiberg, Germany; b.rasti@hzdr.de

Bikram Koirala and Paul Scheunders are with Imec-Visionlab, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp, Belgium; Bikram.Koirala@uantwerpen.be; paul.scheunders@uantwerpen.be

J. Chanussot is with the Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, F-38000 Grenoble, France. (e-mail: jocelyn.chanussot@grenoble-inp.fr)

Manuscript received .....

common procedure is to use constrained (penalized) least squared error minimization (also referred to as non-negative matrix factorization in literature [14]) to form an optimization problem with respect to both endmembers and abundances [15], [16]. In those problems, the geometrical information can be induced by using a minimum volume-based penalty term. In [17], the square of the simplex volume was selected as the penalty term in a fully constrained penalized least squares estimation. The Euclidean distances between the endmembers and the center of the data points are considered in [14] as a geometrical penalty. In [18], the authors proposed a geometrical penalty defined by the Euclidean distance between the endmembers and the extremes of the data simplex extracted by a geometrical approach (i.e., VCA). In [19], the total variation (TV) of all the endmembers is proposed as a penalty term. In these optimization problems, the selection of the regularization parameter, which is the trade off between the fidelity term and the penalty term, is often a complex problem. In [20], a parameter selection technique is proposed for the fully constrained penalized least squares method with the three different geometrical penalties proposed in [14], [18] and [19].

Alternatively, the unmixing problem can be reformulated into a non-convex minimization (or a non-concave maximization) problem, which seeks the minimum volume data simplex, subjected to the ASC and ANC [21], [22]. Note that the initialization can considerably affect the performance of the aforementioned non-convex approaches.

Sparse unmixing is another group of unmixing techniques that estimate the fractional abundances using sparse regression techniques, relying on a rich and well-designed library of pure spectra. Each spectrum is assumed to be a sparse linear combination of the dictionary elements, i.e., the library. The optimization problem is often defined in the form of penalized least squares with a sparsity promoting penalty applied on the abundances [23], [24]. Recently, in [25], a deep learning approach was proposed for sparse unmixing, using a convolutional neural network (SUnCNN). The major disadvantage of sparse unmixing techniques is their dependency on the spectral library, which can considerably affect the abundance estimation performance. On the other hand, a specific group of sparse unmixing approaches considers the endmembers' spectral variability by using a dictionary of endmember bundles generated from the data itself. An example of such approach is the Collaborative LASSO (Least Absolute Shrinkage and Selection Operator [26]) [27] [28].

In recent years, deep learning techniques attracted increasing attention in remote sensing, including hyperspectral image analysis. A variety of architectures has been developed and proposed for different applications, such as hyperspectral image classification [29]. For deep hyperspectral unmixing, the most widely used architecture is based on autoencoders. The abundances are often generated by enforcing the constraints (i.e., ANC and ASC) in the final layer of the encoder. The decoder has one layer that reconstructs the signal, and its weights are the endmembers. In [30], a stack of nonnegative sparse autoencoders (SNSA) was proposed in which the last autoencoder is utilized for unmixing, and the rest of the network improves the robustness with respect to outliers. The Deep AutoEncoder Network (DAEN) [31] exploits a stacked autoencoder to initialize a variational autoencoder to estimate the endmembers and the abundances. In [32], a deep network was proposed that uses a variational autoencoder to generate the endmembers. An untied Denoising Autoencoder with Sparsity (uDAS) was proposed in [33] for spectral unmixing. uDAS benefits from an additional denoising constraint applied to the decoder and an $\ell_{2,1}$ sparsity constraint applied to the decoder. A sparse autoencoder was proposed in [34] which uses a loss function with a Kullback-Leibler divergence term, a SAD similarity, and a sparsity term. We should note that in all the above mentioned autoencoder-based techniques, the spatial information is ignored, as for training the network, the autoencoder receives a spectral pixel at a time [35]. In [36], an unmixing technique based on an autoencoder network is proposed that incorporates the spatial correlation between pixels by utilizing an adaptive abundance smoothing method. In [37], patch-wise or cube-wise convolutional autoencoders were proposed to incorporate the spatial information. In [38], parallel autoencoders were applied to spectral patches to exploit the spatial information. Recently, a 3D convolutional autoencoder was proposed for supervised hyperspectral unmixing (i.e., the endmembers are assumed to be known) in [39]. The patchwise and cubewise approaches that exploit the convolutional filters are beneficial for endmember estimation since they capture the variability of the spectra. On the other hand, they may blur the abundance maps [40]. Unmixing using deep image prior (UnDIP) [41] utilizes a convolutional encoder-decoder architecture and a deep image prior [42], [43]. UnDIP is robust to noise and provides sharp abundances. However, it relies on a geometrical approach (i.e., SiVM) to extract the endmembers, i.e., it only considers pure pixel-scenarios. The cycle-consistency unmixing network (CyCUNet) [44] utilizes two convolutional autoencoders, which are cascaded and performed cyclically. The proposed loss function contains two terms for spectral reconstruction and one for abundance reconstruction to incorporate high-level semantic information.

A major problem of the DL-based approaches is the absence of geometrical information. Therefore, as we will show in the experimental section, most of them fail to accurately estimate the endmembers when there are no pure pixels in the dataset. On the other hand, the advantage of simplex volume minimization for blind unmixing in no pure pixel-scenarios has been proven [45]. In this paper, we propose a convolutional encoder-decoder architecture for blind spectral unmixing called MiSiCNet (minimum simplex convolutional network). MiSiCNet utilizes a deep encoder-decoder network that incorporates both spatial and geometrical information. The spatial information is incorporated by using the convolutional operator and by implicitly applying a regularizer on the abundances. The geometrical information is exploited using a simplex volume penalized loss function. We show that MiSiCNet is superior in unmixing datasets which do not contain pure pixels. MiSiCNet was implemented in Python (3.8) using PyTorch as the platform for the deep network and is available online: https://github.com/BehnoodRasti/MiSiCNet.

A. Contributions and Novelties

The contribution of the proposed method is threefold.

1) MiSiCNet is the first deep blind unmixing network which considers the geometry of the data points by employing a simplex volume penalized loss function to incorporate the geometrical information, while the DL-based unmixing techniques from the literature directly optimize either reconstruction errors or spectral angle distance. Some conventional techniques, such as [17] and [20] incorporate the geometrical information using a simplex volume penalized least squares optimization. However, they do not incorporate spatial information, mostly because the optimization problem becomes too complex and the selection of the tuning parameter is complicated. In the experiments, we show that incorporating the geometrical information is crucial for the endmember estimation when there are no pure pixels available.
2) We propose a deep unsupervised convolutional network that solves blind unmixing problems with regularizers on both abundances and endmembers. In other words, we demonstrate how to turn a blind unmixing problem into the optimization of a deep unsupervised convolutional network. Additionally, we show how to implement all the physical constraints via applying constraints on the network parameters.
3) We propose a convolutional encoder-decoder with a skip connection that incorporates the spatial information using both convolutional filters and an implicitly induced regularizer on the abundances. Additionally, the skip connection augments the low-level features to high-level features, which sharpens the abundance maps.

The remaining of this paper is organized as follows. The proposed unmixing methodology is explained in detail in Section II. The experimental results are shown and discussed in Section IV. Section V concludes the paper.

II. Methodology

A. Hyperspectral Image Modeling

We assume that the mixing model for the observed HSI is given by:

$$\mathbf{Y} = \mathbf{EA} + \mathbf{N}, \quad \text{s.t.} \quad \mathbf{A} \geq 0, \mathbf{1}_r^T \mathbf{A} = \mathbf{1}_n^T, 0 \leq \mathbf{E} \leq 1 \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{p \times n}$ is the observed HSI, with n pixels and p bands, $\mathbf{N} \in \mathbb{R}^{p \times n}$ is noise, $\mathbf{E} \in \mathbb{R}^{p \times r}$, and $\mathbf{A} \in \mathbb{R}^{r \times n}$, $r \ll p$, contain the r endmembers and their fractional abundances, respectively. $\mathbf{1}_n$ indicates an n-component column vector of ones. In blind unmixing scenarios, the task is to estimate both $\mathbf{E}$ and $\mathbf{A}$ simultaneously.

A general optimization problem for estimating both $\mathbf{E}$ and $\mathbf{A}$ is given by:

$$\begin{aligned}(\hat{\mathbf{A}}, \hat{\mathbf{E}}) = &\arg\min_{\mathbf{A},\mathbf{E}} \frac{1}{2}||\mathbf{Y} - \mathbf{EA}||_F^2 + \lambda\phi(\mathbf{E}) + \beta R(\mathbf{A}) \\ &\text{s.t.} \mathbf{A} \geq 0, \mathbf{1}_r^T \mathbf{A} = \mathbf{1}_n^T, 0 \leq \mathbf{E} \leq 1\end{aligned} \tag{2}$$

where the first term is the fidelity term, ϕ and R are penalty terms on endmembers and abundances respectively, and λ and β control the trade-off between the penalty terms and the fidelity term. $\|.\|_F$ denotes the Frobenius norm. The problem is solved, subjected to the non-negativity and sum to one constraint. The penalty functions are often chosen based on the prior knowledge of the abundances and endmembers. The function R is often selected to capture the spatial correlation in the dataset. For instance, a total variation penalty or a sparsity enforcing penalty is often a common choice for R, since it induces piecewise smoothness on the abundance maps. The function ϕ is often chosen to minimize the volume of the data simplex. A reasonable choice is to enforce the endmembers towards the center of the data simplex [14]:

$$\phi(\mathbf{E}) = \left\|\mathbf{E} - \mathbf{m}\mathbf{1}_\mathbf{r}^\mathbf{T}\right\|_F^2 \tag{3}$$

where $\mathbf{m}$ is a vector that contains the mean values of the spectral pixels. A geometrical interpretation of this term is given in Fig. 1. As can be seen, by minimizing $\phi(\mathbf{E})$, the endmembers are pulled towards the vertices of the true data simplex from their initial values, i.e., the vertices of the initial data simplex.

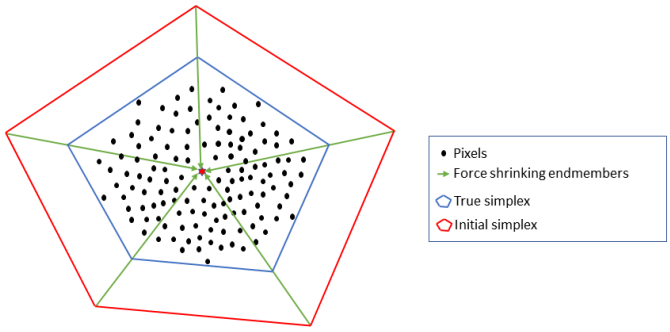


Fig. 1. A geometrical illustration of the minimum volume constraint, pulling the endmembers towards the center of the data simplex.

Therefore, the problem to solve becomes:

$$\begin{aligned}(\hat{\mathbf{A}}, \hat{\mathbf{E}}) = &\arg\min_{\mathbf{A},\mathbf{E}} \frac{1}{2}||\mathbf{Y} - \mathbf{EA}||_F^2 + \lambda \left\|\mathbf{E} - \mathbf{m}\mathbf{1}_\mathbf{r}^\mathbf{T}\right\|_F^2 \\ &+ \beta R(\mathbf{A}) \quad \text{s.t.} \quad \mathbf{A} \geq 0, \mathbf{1}_r^T \mathbf{A} = \mathbf{1}_n^T, 0 \leq \mathbf{E} \leq 1\end{aligned} \tag{4}$$

For more discussion on the selection of ϕ, we refer to [17]. In [14], function R is selected to promote sparsity on $\mathbf{A}$. It is worth mentioning that selecting λ and β is not a trivial task. This is partially addressed in [17] by neglecting the spatial dependency of the abundance maps, i.e., $\beta = 0$.

B. Deep Unmixing Using MiSiCNet

A general inverse image reconstruction task can be formulated as an optimization problem:

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2} \|\mathbf{Y} - \mathbf{X}\|_F^2 + \lambda R(\mathbf{X}) \tag{5}$$

where λ is the tuning parameter that acts as a trade-off between the fidelity term and the regularizer (R). The regularizer selection depends on the application and the available prior knowledge. On the other hand, deep image prior (DIP) [42],

[43] shows that the selection of a suitable regularizer can be substituted by optimizing a deep network's parameters (i.e., weights and biases):

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}||\mathbf{Y} - f_{\theta}(\mathbf{Z})||_F^2 \quad \text{s.t.} \quad \hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z}) \tag{6}$$

where $\mathbf{Z}$ is the network input that is fixed throughout the optimization, and is often selected as a random input image. Here, we explain how we solve equation (4) using a deep neural network and (6).

Deep unmixing techniques utilize deep encoder-decoder architectures to estimate abundances and endmembers. The encoder part often contains several layers to estimate the abundances at the bottleneck and the decoder contains only one linear layer. The reconstruction loss is often minimized to train the network. Let us denote the encoder network with f_{θ_1} with network parameters (i.e., weights and biases) denoted by θ_1. The decoder part is only a linear layer which we simply show by a set of weights i.e., θ_2. Momentarily, neglecting all the constraints, using (6) we can turn the optimization problem (4) into an optimization of the network parameters:

$$\begin{aligned}(\hat{\theta}_1, \hat{\mathbf{E}}) = \arg\min_{\theta_1, \mathbf{E}} \frac{1}{2}||\mathbf{Y} - \mathbf{E}f_{\theta_1}(\mathbf{Z})||_F^2 + \lambda \left\|\mathbf{E} - \mathbf{m1_r^T}\right\|_F^2 \\ \text{s.t.} \quad \hat{\mathbf{A}} = f_{\hat{\theta}_1}(\mathbf{Z})\end{aligned} \tag{7}$$

As $\mathbf{E}$ contains the weights of the final linear layer (i.e., the decoder), we rewrite (7) as:

$$\begin{aligned}(\hat{\theta}_1, \hat{\theta}_2) = \arg\min_{\theta_1, \theta_2} \frac{1}{2}||\mathbf{Y} - \theta_2 f_{\theta_1}(\mathbf{Z})||_F^2 + \lambda \left\|\theta_2 - \mathbf{m1}_r^T\right\|_F^2 \\ \text{s.t.} \quad \hat{\mathbf{E}}\hat{\mathbf{A}} = \hat{\theta}_2 f_{\hat{\theta}_1}(\mathbf{Z})\end{aligned} \tag{8}$$

where $\hat{\mathbf{Y}} = \hat{\mathbf{E}}\hat{\mathbf{A}}$. Hence, the optimization problem (4) can be solved using a deep network with a loss function given by:

$$\mathcal{L}(\mathbf{Y}, \hat{\mathbf{Y}}, \hat{\theta}_2, \mathbf{m}) = \frac{1}{2}||\mathbf{Y} - \hat{\mathbf{Y}}||_F^2 + \lambda \left\|\hat{\theta}_2 - \mathbf{m1}_r^T\right\|_F^2 \tag{9}$$

To enforce both the ASC and the ANC we use a softmax function in the final layer of the encoder:

$$softmax(\mathbf{A}) = \frac{e^{\mathbf{A}_{ij}}}{\sum_{i=1}^{r} e^{\mathbf{A}_{ij}}} \quad \forall i, j \tag{10}$$

and the weights of the decoder (i.e., θ_2) are constrained between 0 and 1 in every step of the optimization to enforce the endmember constraint i.e., $0 \leq \mathbf{E} \leq 1$.

III. The Architecture of MiSiCNet

The architecture of MiSiCNet is visualized in Fig. 2. The main architecture is based on the convolutional encoder-decoder with a skip connection. The skip connection is able to learn the identity function when the parameters become zero and avoids vanishing gradients in a deep network. Additionally, via the skip connection, the information is shifted from low level features to high level (deep) features, which considerably helps to keep details such as the structures in the deep feature representation. We use four convolutional layers (Conv), excluding the skip connection (ConvSkip). To preserve the abundances' spatial resolution, we do not apply any down-sampling and use reflection padding in each layer. The number of filters for each convolutional layer is given in Table I. Each convolutional layer is followed by a batch normalization (BN) layer which speeds up the learning process and provides more robustness for selecting the hyperparameters. To promote the nonlinearity, we use Leaky ReLU (rectified linear unit [46]) as the nonlinear activation function for all the convolutional layers, except for the last one where we use the softmax to enforce the ASC and ANC. We select random noise with the same size as the dataset as the input ($\mathbf{Z}$) and train the network iteratively to map the input to the dataset.

A. Hyperparameter Selection

The selection of hyperparameters is always challenging in a deep network. In this work, we fix all the hyperparameters used in the experiments, except the tuning parameter (λ). We set the negative slope of Leaky ReLU to 0.1. The convolutional layers use kernels of size 3×3, except for ConvSkip, which uses a kernel of size 1×1. We use four filters in the skip connection, 256 filters for Conv1, Conv2, and Conv3, and r (the number of pure materials) filters for Conv4. We should note that MiSiCNet is an unsupervised network, and we train the network over the entire dataset iteratively. Therefore, the number of iterations is a hyperparameter for the proposed network. We set the number of iterations to 8000. We also use exponentially weighted averaging over the outputs to make the technique robust to a possible jump in the loss function at the final iteration. Finally, we use the Adam optimizer and set the learning rate to 0.001. We use the PyTorch deep learning platform for the network implementation. Table I lists the hyperparameters used in the experiments.

TABLE I
HYPERPARAMETERS OF MISICNET USED IN THE EXPERIMENTS.

Hyperparameters				
	Input Ch.	Ouput Ch.	Filter Size	Stride
Conv1	p	256	3x3	1
Conv2	256	256	3x3	1
Conv3	260	256	3x3	1
Conv4	256	r	3x3	1
ConvSkip	p	4	1x1	1
		Negative Slope		
Leaky ReLU		0.1		
	Type	Learning Rate		Iterations
Optimizer	Adam	0.001		8000

IV. Experimental Results

We performed experiments on two simulated datasets and three real datasets. The description of the datasets is given below.

A. Hyperspectral Data Description

1) Simulated Dataset 1: A dataset of 105×105 pixels (see Fig. 3(a)) is simulated by generating linear mixtures of six endmembers. These endmembers are shown in Fig. 3(b) and contain 224 reflection values in the wavelength range [400-2500] nm. A PCA reduced data manifold is shown in Fig. 4(a). As can be observed, no pure pixels (red circles) are available in the dataset, but at least two mixed data points are available on each facet to geometrically reconstruct virtual endmembers.

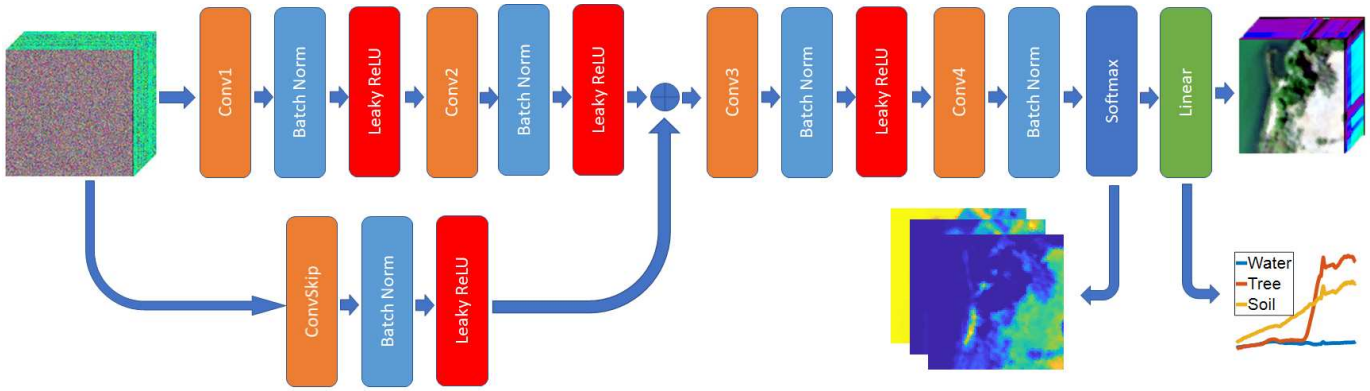
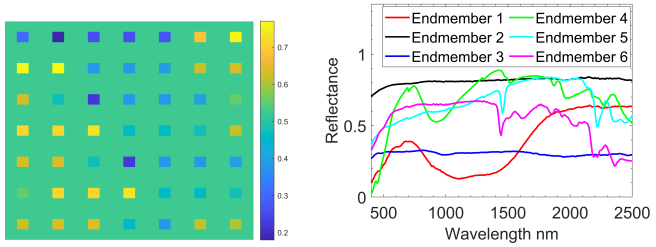
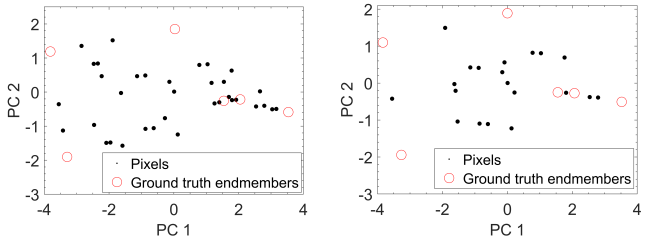


Fig. 2. The architecture of MiSiCNet. The proposed convolutional network architecture uses a skip connection and five convolutional layers. The softmax layer is used as the final activation function, followed by a linear layer that generates the abundances and endmembers, respectively.



(a) (b)

Fig. 3. Simulated Dataset 1: a) Band number 70 (1050 nm) b) Endmembers.



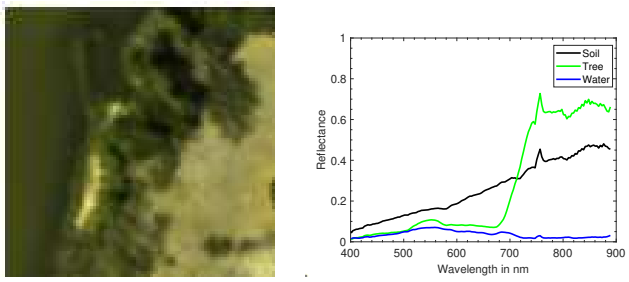
(a) (b)

Fig. 4. PCA reduced data manifolds: a) Simulated dataset 1 b) Simulated dataset 2.

2) Simulated Dataset 2: The second simulated dataset contains 105×105 pixels which are simulated by the linear combination of six endmembers (see Fig. 3(b) for endmembers). The main difference with the simulated dataset 1 is that each facet of the data manifold contains only one mixed data point (see Fig. 4(b)), making it challenging to reconstruct virtual endmembers geometrically.

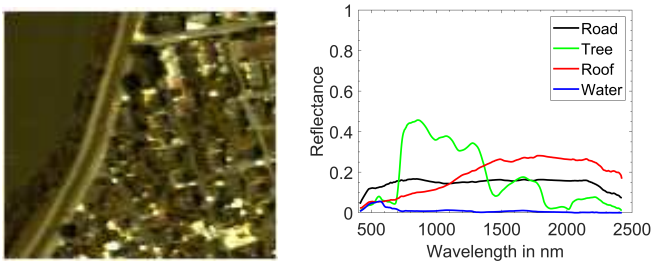
3) Samson: The Samson hyperspectral dataset ([47]) (Fig. 5(a)) contains 95×95 pixels, having 156 bands in the wavelength range from 401 to 889 nm. There are three main materials (i.e., Soil, Tree, and Water). The ground truth endmembers shown in Fig. 5(b) were manually selected from the hyperspectral image, and the ground truth fractional abundances were generated using FCLSU.

4) Apex: Fig. 6(a) shows a cropped image of the Apex dataset ([48]), as used in this paper. It contains 111×122 pixels and 285 bands covering the wavelength range from 413 to 2420 nm. Four ground truth endmembers (i.e., Water, Tree, Road, and Roof), shown in Fig. 6(b) were manually selected from the hyperspectral image, and the ground truth fractional abundances were generated using FCLSU.



(a) (b)

Fig. 5. Samson image: (a) True-color image (Red: 571.01 nm, Green: 539.53 nm, and Blue: 432.48 nm) (b) Endmembers.

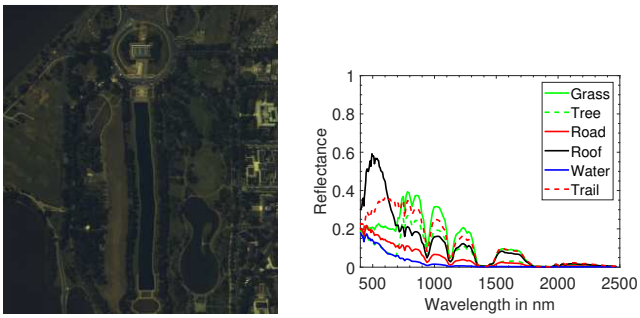


(a) (b)

Fig. 6. Apex image: (a) True-color image (Red: 572.2 nm, Green: 532.3 nm, Blue: 426.5 nm); (b) Endmembers.

5) Washington DC Mall: Washington DC Mall is an airborne hyperspectral dataset captured over the Washington DC Mall using the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor [1]. Fig. 7(a) shows the cropped data used in this paper that contains 319 × 292 pixels in 191 bands over the spectral range from 400 to 2400 nm. The ground truth endmembers for six classes, i.e., Grass, Tree, Roof, Road, Water, and Trail, were manually selected from the hyperspectral image (Fig. 7(b)), and FCLSU was used to estimate the ground truth fractional abundances.

[1]https://engineering.purdue.edu/ biehl/MultiSpec/hyperspectral.html



(a) (b)

Fig. 7. Washington DC Mall image: (a) True-color image (Red: 572.7 nm, Green: 530.1 nm, Blue: 425.0 nm); (b) Endmembers.

B. Experimental Setup

Six unmixing techniques from different unmixing categories were used as competing methods in the experiments:

- Geometrical unmixing: FCLSU [5] using VCA [8] for endmember extraction.
- Geometrical and blind unmixing: NMF-QMV [20].
- Sparse unmixing: Collaborative LASSO (Collab) [28]
- Deep unmixing methods: uDAS [33], UnDIP [41], and CyCUNet [44].

For MiSiCNet, we select all the hyperparameters as discussed in subsection III-A, and we set λ to 0.1, 0.3, and 100 for the simulated dataset 1, the simulated dataset 2, and real datasets, respectively. We choose all the parameters for the competing methods according to the reported default values. Note that for NMF-QMV, we choose the "center" as the optional penalty term to have a fair comparison with MiSiCNet.

Quantitative results are provided by the root mean squared error (RMSE) in percentage between the estimated and ground truth abundance fractions:

$$\text{RMSE}(\hat{\mathbf{A}}, \mathbf{A}) = 100 \times \sqrt{\frac{1}{rn} \sum_{i=1}^{r} \sum_{j=1}^{n} \left(\hat{\mathbf{A}}_{ij} - \mathbf{A}_{ij} \right)^2}, \quad (11)$$

and the spectral angle distance (SAD) in degree between the estimated and ground truth endmembers:

$$\text{SAD}(\mathbf{E}, \hat{\mathbf{E}}) = \frac{1}{r} \sum_{i=1}^{r} \arccos \left(\frac{\langle \mathbf{e}_{(i)}, \hat{\mathbf{e}}_{(i)} \rangle}{\|\mathbf{e}_{(i)}\|_2 \, \|\hat{\mathbf{e}}_{(i)}\|_2} \right) \frac{180}{\pi}, \quad (12)$$

where $\langle . \rangle$ denotes the inner product and $\mathbf{e}_{(i)}$ indicates the ith column of $\mathbf{E}$.

C. Unmixing Experiments

1) Experiments on Simulated Datasets: To evaluate the robustness of the methods to noise, Gaussian white noise with varying noise power is added to both simulated datasets to obtain data with SNR of 20, 30, 40 and 50 dB. Table II shows the results in terms of RMSE (in percentage) for the simulated dataset 1. CyCUNet produced the worst results. FCLSU and Collab both show poor performances. UnDIP outperforms uDAS, but its results are not competitive with NMF-QMV and MiSiCNet. Both NMF-QMV and MiSiCNet considerably outperform the other techniques in terms of RMSE. With higher amounts of of noise however, MiSiCNet gradually improves over NMF-QMV.

Table III reports the RMSE results for the Simulated Dataset 2. The trends are similar, however, all the competing methods perform poorly. MiSiCNet shows considerable improvements even compared to NMF-QMV. Additionally, the very low standard deviations reported in the tables confirm the robustness of MiSiCNet for different noise levels. That is a valuable advantage revealed from the experiments. For instance, the small (0.09%) performance improvement of NMF-QMV over MiSiCNet for the simulated dataset 1 with 20dB of noise is statistically insignificant, due to the high standard deviation of 0.2% for NMF-QMV compared to the very low standard deviation of 0.07% for MiSiCNet. Overall, NMF-QMV and MiSiCNet outperform the other techniques in terms of RMSE. However, MiSiCNet outperforms NMF-QMV for high SNRs. This could be attributed to the proposed parameter selection technique for the geometrical penalty (β) in NMF-QMV. Additionally, the search interval ($\beta \in 10^{-5}, 10^{-4}, ..., 10^{4}, 10^{5}$) might not be suitable for all datasets with varying SNRs.

Tables IV and V report SAD for simulated data sets 1 and 2, respectively. The results follow the trend of RMSE. MiSiCNet considerably outperforms the other techniques for all SNRs for Simulated dataset 2. For simulated dataset 1, NMF-QMV slightly outperforms MiSiCNet for 20 and 30 dB. As we already discussed, this could be attributed to the parameter selection technique in NMF-QMV. The search interval is too large to pick the optimum parameter.

Both simulated experiments reveal the importance of the minimum volume term in the absence of pure pixels. Both NMF-QMV and MisiCNet consider the geometry of the data simplex by minimizing the volume term while all the other techniques either ignore that or rely on the presence of pure pixels, for instance, by using VCA (in FCLSU) and SiVM (in UnDIP).

Figs. 8 and 9 depict the abundances and endmembers estimated by the different unmixing techniques on the simulated dataset 2 (40 dB). Visual comparisons confirm the advantage of MiSiCNet compared to the other methods. MiSiCNet provides excellent abundance estimations for Endmember 1 and 6 and good estimations for the other ones. NMF-QMV also outperforms the rest of the methods, but fails to estimate satisfactory abundances for this challenging mixing scenario. From Fig. 9, we can see that MiSiCNet estimates all endmembers successfully, except for a slight mismatch in Endmember 5. On the other hand, all the other techniques perform poorly for all the cases.

2) Experiments on Real Datasets: Table VI shows the abundance RMSE obtained by the different unmixing techniques to the Samson data. MiSiCNet significantly improved the abundance estimation for this dataset compared to the other methods. Additionally, MiSiCNet considerably outperformed the other techniques for all three individual abundances. MiSiCNet performed 12.3%, 3.63%, and 9.74% better on the Soil, Tree, and Water abundances, respectively, compared with the second-best results, given by Collab. Figs.10 and 11 depict the estimated abundances and endmembers, respectively, obtained by the different techniques. This again confirms that MiSiCNet

TABLE II
RMSE (SIMULATED 1). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	FCLSU	UnDIP	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
20dB	11.13±1.93	8.75±0.56	9.40±1.73	17.97±2.93	10.69±1.53	**3.81±0.20**	3.90±0.08
30dB	12.28±3.44	7.19±0.87	9.40±1.91	17.83±4.73	11.22±2.29	1.81±0.42	**1.80±0.04**
40dB	12.29±2.49	7.73±0.87	9.36±2.03	17.97±7.02	10.07±1.68	1.70±0.90	**1.23±0.05**
50dB	11.15±2.72	7.36±0.88	8.46±2.04	16.07±0.76	10.35±2.02	3.71±0.74	**1.21±0.08**

TABLE III
RMSE (SIMULATED 2). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	FCLSU	UnDIP	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
20dB	12.55±1.89	12.15±1.04	11.43±2.69	19.51±5.89	12.05±0.61	4.03±0.54	**3.96±0.04**
30dB	21.45±2.49	10.49±0.21	12.57±5.11	15.87±2.71	13.83±1.94	3.79±2.33	**2.45±0.02**
40dB	21.6±4.11	10.52±0.22	10.84±4.29	14.57±1.3	14.11±1.93	7.37±1.13	**2.15±0.03**
50dB	22.89±2.71	10.37±0.17	10.76±4.24	15.96±2.02	14.14±1.18	6.91±1.17	**2.12±0.03**

TABLE IV
SAD (SIMULATED 1). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	VCA	SiVM	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
20dB	4.64±0.85	6.55±0.6	7.72±4.43	8.37±0.89	6.37±0.81	**2.06±0.25**	2.53±0.64
30dB	4.53±0.96	6.76±0.07	7.33±5.17	8.59±1.03	6.21±1.26	**0.56±0.22**	0.95±0.1
40dB	4.26±0.57	6.84±0.06	7.75±3.74	9.32±0.63	5.67±0.82	0.83±0.51	**0.63±0.09**
50dB	4.66±0.86	6.85±0.07	6.68±3.62	9.83±0.52	5.73±1.31	1.84±0.5	**0.69±0.11**

TABLE V
SAD (SIMULATED 2). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	VCA	SiVM	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
20dB	7.83±0.88	8.03±0.08	11.77±3.62	9.41±0.63	8.36±0.63	2.47±0.48	**1.76±0.03**
30dB	7.72±1.29	7.83±0.02	14.68±4.42	9.73±0.5	7.21±0.57	8.45±7.73	**0.83±0.02**
40dB	8.24±1.02	7.85±0.01	13.95±5.97	9.96±0.6	7.82±1.2	22.95±2.75	**0.64±0.02**
50dB	7.73±1.23	7.86±0.01	14.47±5.77	10.57±0.23	7.65±0.28	24.13±1.26	**0.62±0.02**

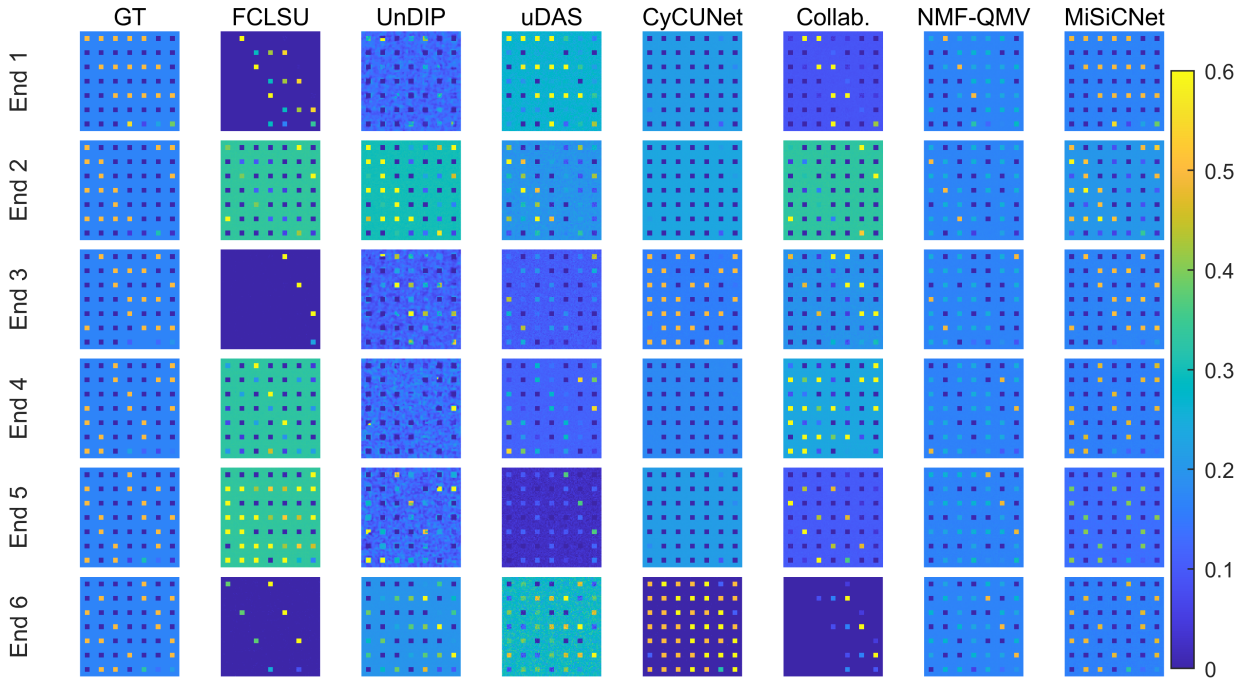


Fig. 8. Simulated dataset 2 (40dB) - The visual comparison of the abundance maps obtained by applying different unmixing techniques.

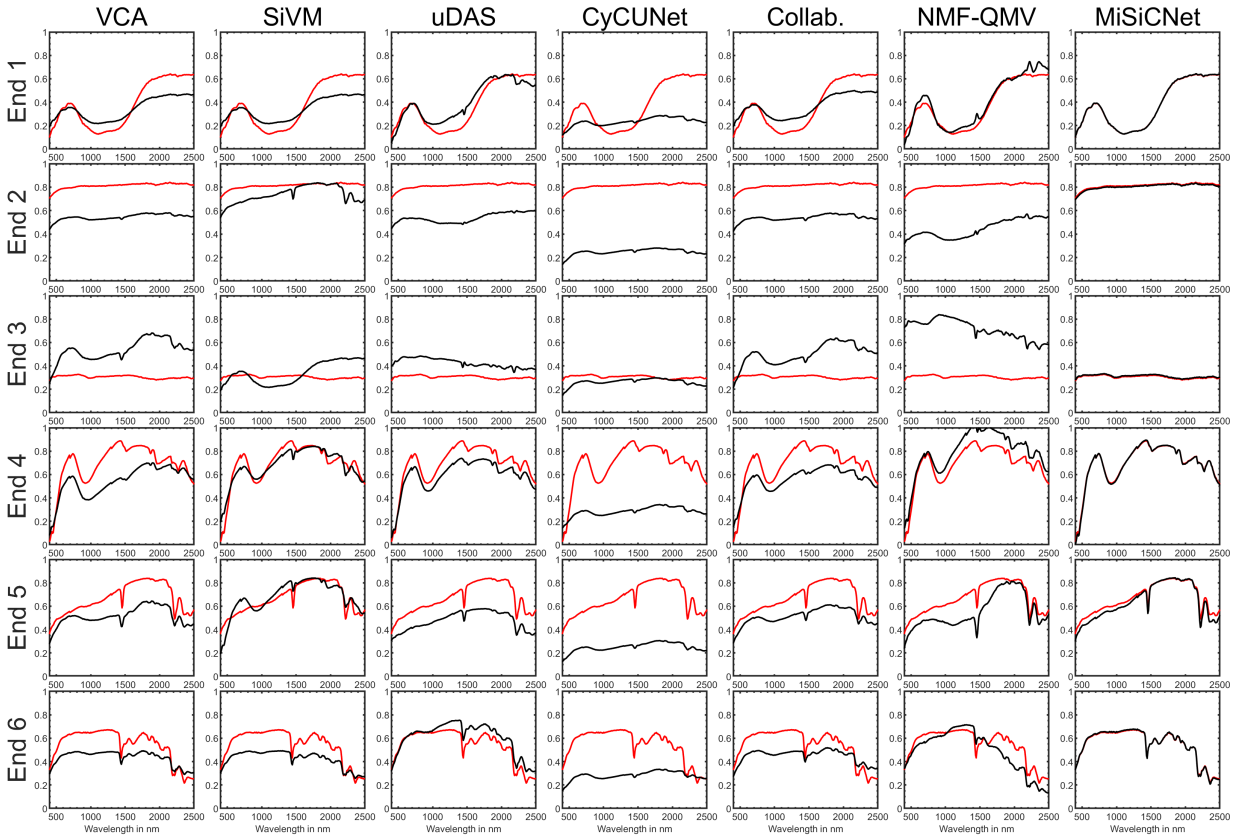


Fig. 9. Simulated dataset 2 (40dB) - The visual comparison of the endmembers obtained by applying different unmixing techniques. Red: the ground truth endmembers; Black: the estimated endmembers.

outperforms the other techniques. Table VII reports SAD values of the endmembers for Samson. MiSiCNet outperforms the other techniques in terms of SAD for the Soil and the Tree abundances, but not for the Water abundances. Overall, Collab shows 3.61 degree SAD improvement over MiSiCNet. SAD does not follow the trend of the abundance RMSE in real datasets, an effect which is caused by spectral variability. As SAD removes the norm of the endmember spectra, it ignores endmember scaling factors (caused by multiple reflections of the light and variable illumination conditions). However, such scaling factors may considerably affect the abundance estimation. Tables VIII and IX compare the abundance and endmember estimation on the Apex data of the different unmixing techniques in terms of RMSE and SAD. Figs. 12 and 13 show the estimated abundances and endmembers, respectively. From Table VIII, we can see that MiSiCNet slightly outperforms FCLSU, and they both improve the abundance estimation of Apex with 5% RMSE over the next best results. SiVM obtained the best results in terms of SAD. The effect of neglecting the endmember scaling factor by SAD is much clearer in the results of Apex. Comparing the estimated endmembers by SiVM and MiSiCNet in Fig. 13 reveals the difference between the two estimations. However, the reported results for SAD in Table IX show that SiVM obtained the best results on the Road endmember, again because SAD neglects scaling factors. On the other hand, Fig.12 confirms that MiSiCNet provided the best Road abundance estimation. Overall, the results confirm the advantage of the MiSiCNet over the other unmixing techniques.

Tables X and XI show the abundance RMSE and SAD, respectively for the Washington DC Mall dataset. On this dataset, MiSiCNet improved the overall abundance RMSE and SAD by 5.69% and 1.3 degrees over the second-best results provided by NMF-QMV and Collab, respectively. Ad-

TABLE VI
RMSE (SAMSON DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	FCLSU	UnDIP	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Soil	17.66	17.78	17.99	24.17	15.06	52.35	**2.76**
Tree	6.53	13.30	13.83	13.86	6.07	39.90	**2.44**
Water	14.92	20.96	23.03	26.54	11.81	45.98	**2.07**
Overall	13.87	17.63	18.67	22.21	11.59	46.36	**2.44**

TABLE VII
SAD (SAMSON DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	VCA	SiVM	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Soil	1.49	1.49	2.05	6.55	0.89	2.24	**0.64**
Tree	5.51	4.28	5.50	8.69	4.77	7.10	**2.60**
Water	8.91	8.91	8.75	11.92	**8.03**	87.10	21.27
Overall	5.30	4.89	5.43	9.06	**4.56**	32.14	8.17

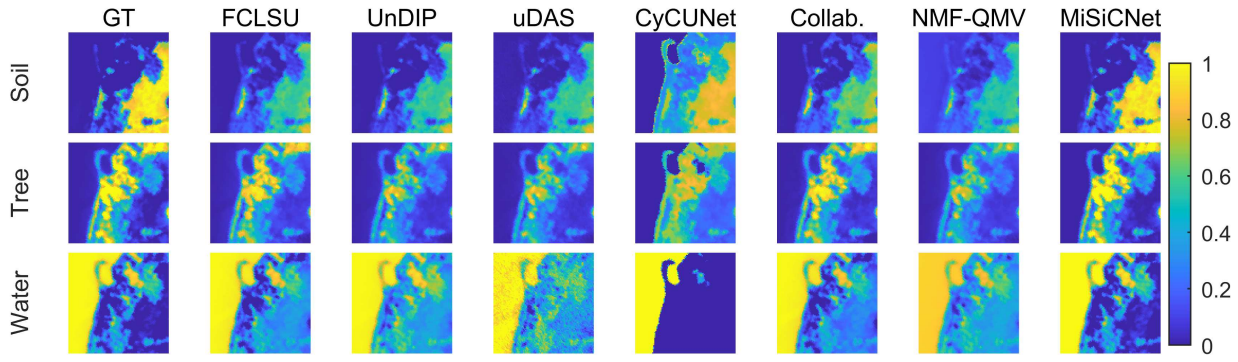


Fig. 10. Samson dataset - The visual comparison of the abundance maps obtained by the different unmixing techniques.

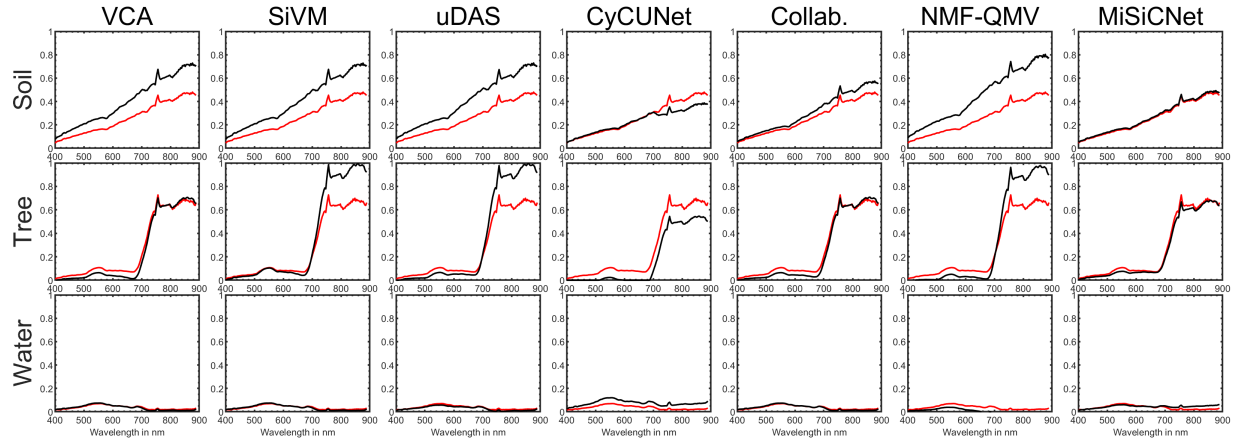


Fig. 11. Samson dataset - The visual comparison of the endmembers obtained by the different unmixing techniques. Red: the ground truth endmembers; Black: the estimated endmembers.

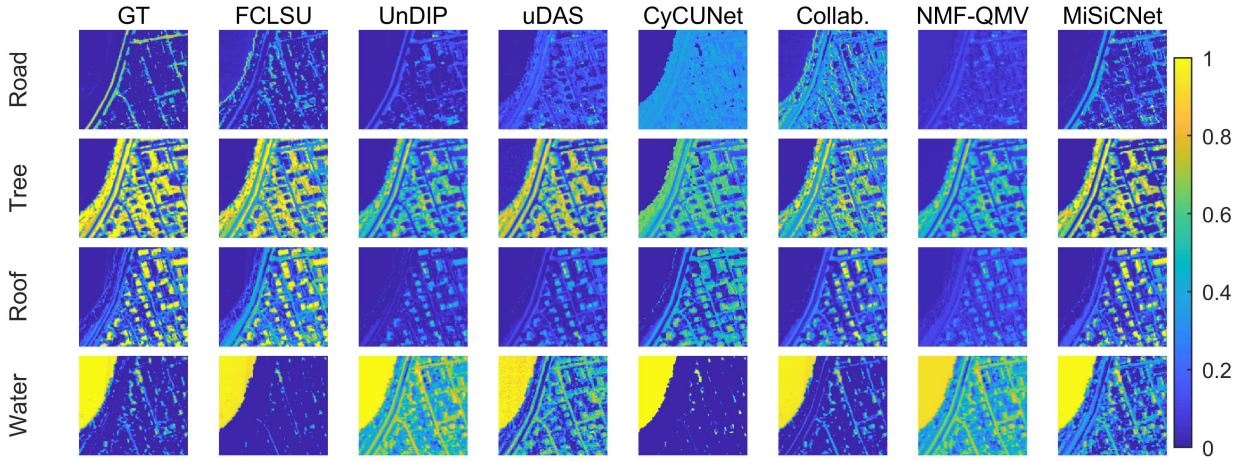


Fig. 12. Apex dataset - The visual comparison of the abundance maps obtained by the different unmixing techniques.

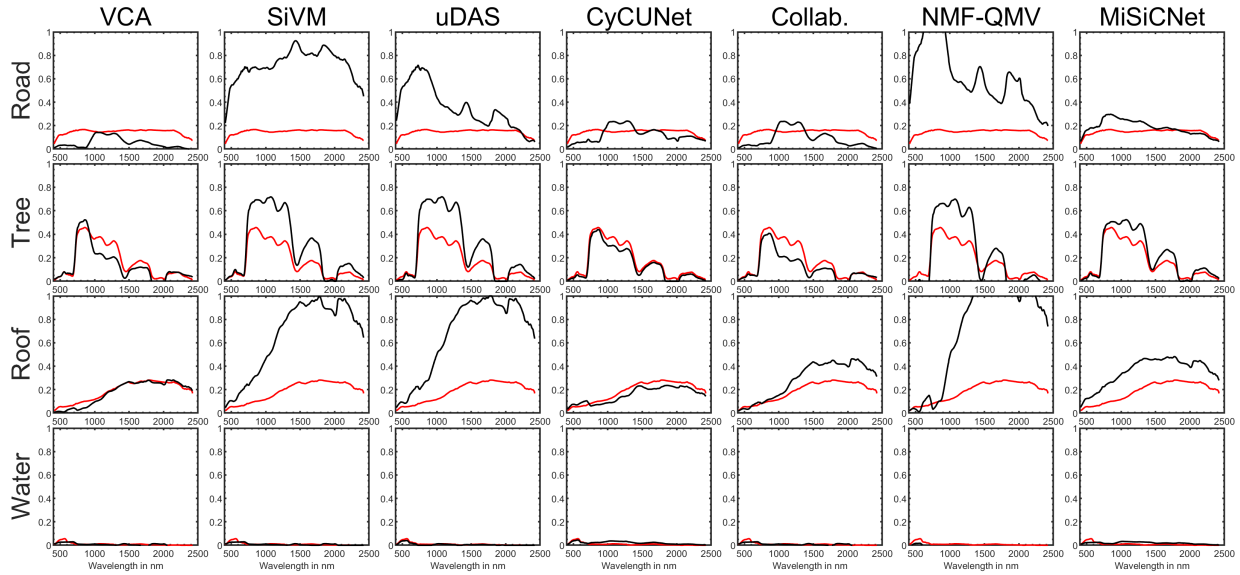


Fig. 13. Apex dataset - The visual comparison of the endmembers obtained by the different unmixing techniques. Red: the ground truth endmembers; Black: the estimated endmembers.

TABLE VIII
RMSE (APEX DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	FCLSU	UnDIP	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Road	23.28	17.42	19.81	29.38	30.83	18.14	**13.67**
Tree	**9.46**	21.54	14.28	20.15	18.92	24.69	14.92
Roof	12.06	26.78	24.27	16.78	15.54	24.87	**11.38**
Water	13.10	42.64	30.01	12.01	**7.90**	38.41	19.84
Overall	15.40	28.74	22.84	20.58	20.08	27.54	**15.27**

TABLE IX
SAD (APEX DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	VCA	SiVM	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Road	39.62	**5.20**	26.07	26.03	38.80	22.94	14.70
Tree	15.15	7.67	8.05	**4.87**	11.82	15.53	7.91
Roof	8.43	**3.94**	4.92	7.43	5.74	10.04	10.05
Water	29.66	28.88	**12.90**	35.66	29.43	74.48	43.43
Overall	23.21	**11.42**	12.99	18.50	21.45	30.75	19.02

ditionally, MiSiCNet improved the abundance estimation of Grass, Tree, Road and Water by 6%, 17.51%, 2.11% and 4.53%, respectively, over the second-best results. MiSiCNet did not improve the abundance estimations for Trail and Roof for which NMF-QMV and FCLSU, respectively obtained better abundance estimations. The estimated abundances and endmembers on the Washington DC Mall dataset are shown in Figs. 14 and 15, respectively. The visual comparison confirms the reported results for the abundance RMSE.

D. Sensitivity Analysis to Hyperparameters

In this subsection, we discuss the sensitivity of MiSiCNet to the selection of the hyperparameters. We selected the four most important hyperparameters that can affect the performance of MiSiCNet: the regularization parameter, the learning rate, the number of filters and the kernel size. Mean values of RMSE and SAD are shown, along with the standard deviations as error bars.

TABLE X
RMSE (WASHINGTON DC MALL DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	FCLSU	UnDIP	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Grass	30.18	29.09	37.34	40.20	28.38	35.81	**22.38**
Tree	39.73	34.93	33.11	28.04	41.26	27.30	**9.79**
Road	17.55	24.24	24.80	25.28	22.58	23.51	**15.44**
Roof	**3.82**	4.88	4.65	40.89	4.48	8.40	16.76
Water	29.07	38.02	51.36	39.86	30.91	20.97	**16.44**
Trail	12.49	23.72	18.26	21.32	19.34	**10.30**	20.67
Overall	25.21	27.95	31.88	33.56	26.98	23.08	**17.39**

TABLE XI
SAD (WASHINGTON DC MALL DATASET). THE BEST PERFORMANCES ARE SHOWN IN BOLD.

	VCA	SiVM	uDAS	CyCUNet	Collab.	NMF-QMV	MiSiCNet
Grass	18.16	10.60	10.87	**5.13**	18.17	11.18	16.68
Tree	16.52	41.58	24.35	15.49	19.11	25.83	**9.48**
Road	13.27	49.32	37.73	26.60	19.71	12.85	**3.37**
Roof	1.97	16.19	11.41	54.43	**1.90**	11.91	19.11
Water	44.50	54.40	13.34	24.09	**1.75**	38.60	3.80
Trail	37.08	10.05	5.39	45.30	19.74	**3.52**	20.14
Overall	21.92	30.36	17.18	28.51	13.40	17.31	**12.10**

1) Regularization Parameter: Fig. 16 (a) shows SAD and RMSE values from the estimated endmembers and abundances, respectively, when varying the tuning parameter λ from $\{10^{-3}, 10^{-2}, ..., 10^{3}\}$. The best results were obtained for $\lambda = 10^{-1}$. Note that SAD by itself (applied on the endmembers) is not sufficient to evaluate the performance. However, when combined with a quality metric such as the abundances RMSE, the performance of the proposed technique is revealed. Therefore, the low value of SAD at 10^3 along with

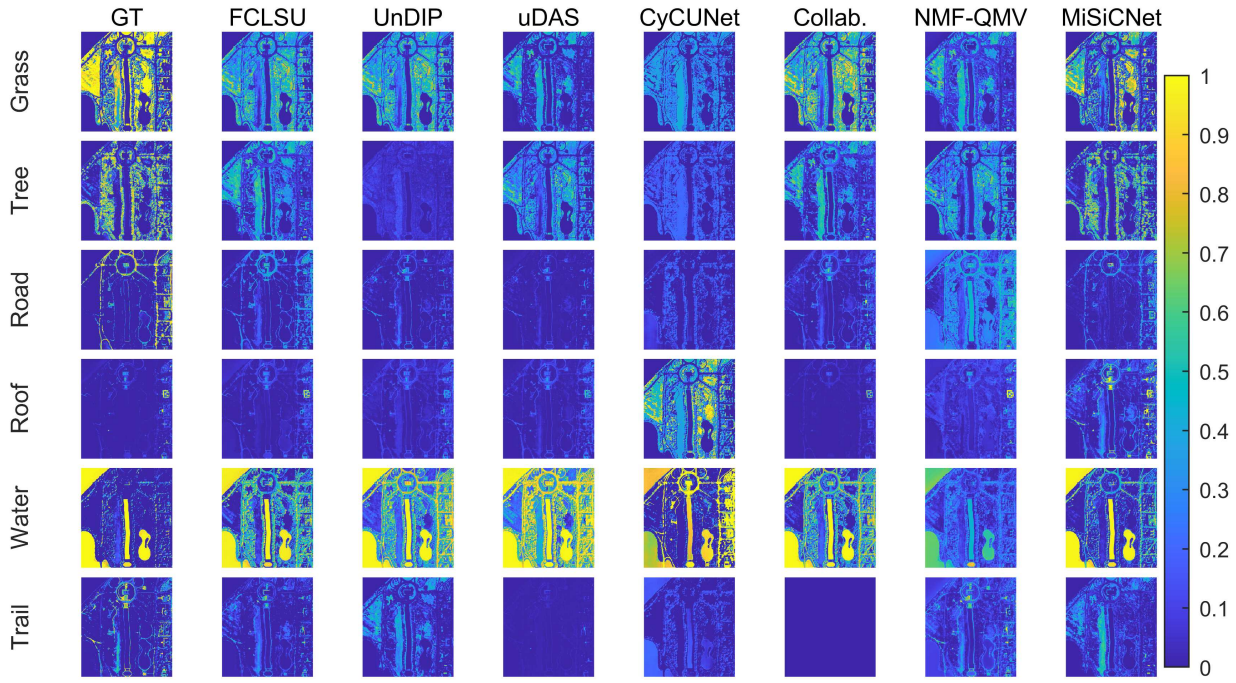


Fig. 14. Washington DC Mall dataset - The visual comparison of the abundance maps obtained by the different unmixing techniques.

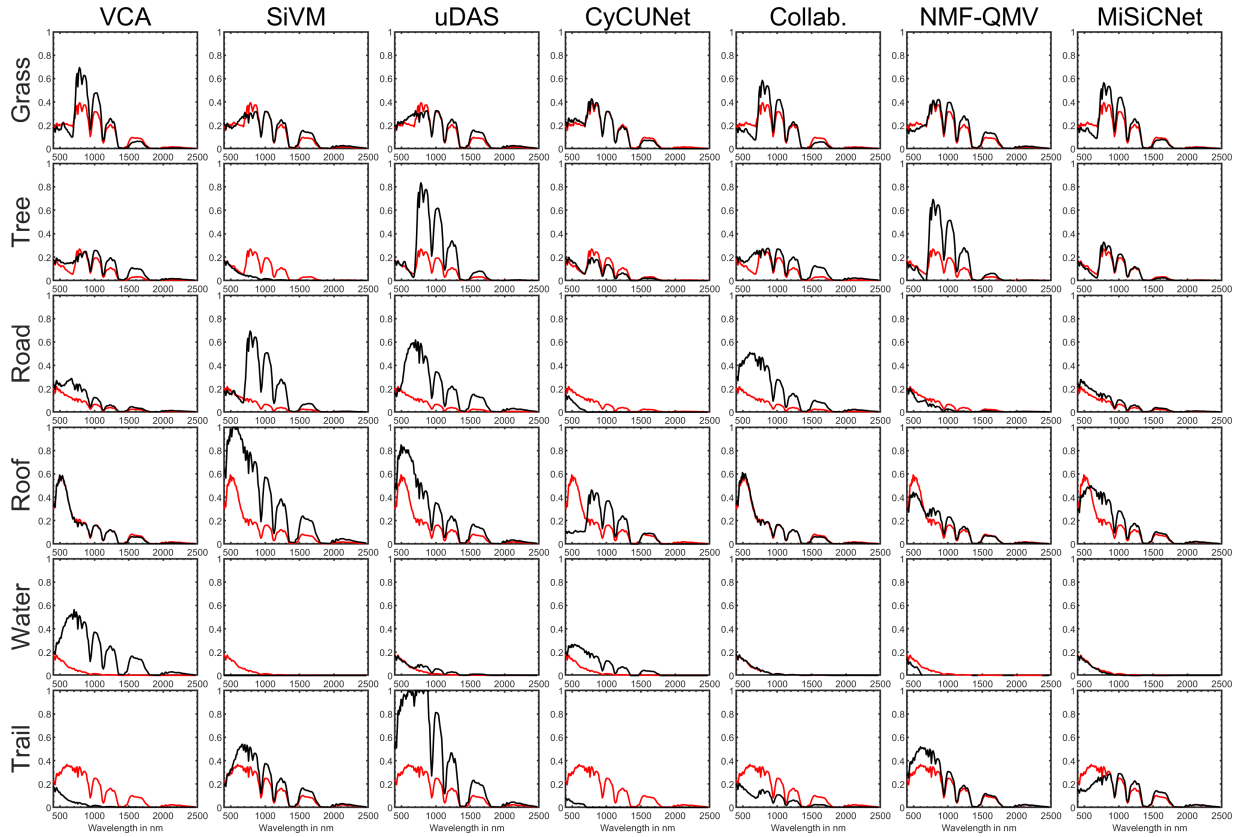


Fig. 15. Washington DC Mall - The visual comparison of the endmembers obtained by applying the different unmixing techniques. Red: the ground truth endmembers; Black: the estimated endmembers.

a high abundance RMSE does not show a good performance. We should note that λ is the only parameter that needs to be set for different datasets.

2) Learning Rate: Fig.16 (b) shows the loss function values for different learning rates (LR), i.e., 0.1, 0.01, 0.001, and 0.0001. For LR=0.1, the loss function value shows large variations. LR=0.0001 shows less variation, but the loss function converges to a higher minimum value. Both LR=0.01 and 0.001 perform similarly, but LR=0.001 shows fewer variations, and the minimum loss is slightly lower than for LR=0.01. Therefore, we suggest to use LR=0.001 for the proposed method.

3) Number of Filters: Fig.16 (c) shows the abundance RMSE and SAD with respect to the number of filters used for the convolutional layers. It can be observed that both abundance RMSE and SAD suggest the selection of 256 filters.

4) Kernel Size: Fig.16 (d) shows the abundance RMSE and SAD with respect to the size of the kernels used in the convolutional layers. Both SAD and Abundance RMSE give the lowest values for filters of size 3×3, confirming our selection.

E. Robustness w.r.t. the Endmember Initialization

A significant challenge in blind unmixing techniques is the initialization of the endmembers. Problem (2) is a non-convex optimization problem due to the non-convex fidelity term, and therefore the initialization plays a significant role in the performance of the final algorithm. In blind unmixing techniques, this is often tackled by using geometrical approaches such as VCA and SiVM. However, those approaches cannot tackle datasets without pure pixels or even more difficult scenarios like those considered in this paper. As a result, the performance of blind unmixing techniques is often not satisfactory due to poor initialization.

On the other hand, deep learning-based methods can cope with the initialization challenge since they search for optimum weights in high dimensional space. Often, the minimum becomes a settled point in that space, which can be reached by optimization algorithms such as gradient descent. We should note that the endmembers are the weights of the decoder and can also be initialized randomly, like the other weights. Fig. 17 shows the performance of MiSiCNet in terms of RMSE and SAD applied to the simulated dataset 2 (50dB) using three different initialization strategies of the weights of the decoder; 1- Random Vectors: initializing randomly in the same way as the other layers 2- Random Pixels: initializing using a random selection of spectral pixels 3- initializing using the endmembers extracted by SiVM. All the results are mean values over five experiments. The best result is given by Random Pixels with around 1% improvement in terms of abundance RMSE and 1.8-degree improvement in terms of SAD compared to the SiVM initialization. However, the standard deviation for the abundance RMSE is much higher (0.71% compared to 0.02%) because of the random selection of the pixels.

On the other hand, SiVM shows robustness to the random noise, and hence we selected it for the initialization of MiSiCNet. Additionally, the results confirm the robustness of MiSiCNet w.r.t. that initialization, and thus the performance of MiSiCNet does not rely on an endmember extraction technique and therefore is successful in the absence of pure pixels. We should note that for Random Vectors, we used LR=0.01 to have a faster convergence.

F. Processing Time

Table XII gives the processing times for the different unmixing techniques applied to the three real datasets. FCLSU, uDAS, Collab, and NMF-QMV were implemented in Matlab (2020b). UnDIP, CyCUNet, and MiSiCNet were implemented in Python (3.8). The reported processing times are mean values over five experiments and were obtained using a computer with an Intel(R) Core(TM) i9-10980 HK processor (2.4 GHz), 32GB of memory, a 64-bit Operating System, and an NVIDIA GEFORCE RTX (2080 Super) graphical processing unit. The table shows that conventional approaches are generally faster than DL-based ones, and FCLSU is the most rapid unmixing technique. However, due to the efficiency of GPU programming, the processing time of MiSiCNet is acceptable and its use can be justified by the significant performance improvements. It is worth mentioning that MiSiCNet is less affected by the increase of the spatial size from Samson to WDC than the other techniques.

TABLE XII
PROCESSING TIME (IN SECONDS) OF THE UNMIXING TECHNIQUES APPLIED TO THE REAL DATASETS.

	FCLSU	**UnDIP**	**uDAS**	**CyCUNet**	**Collab.**	**NMF-QMV**	**MiSiCNet**
Samson	**1.05**	48.19	19.57	69.21	11.45	9.65	90.11
Apex	**1.56**	87.33	262.49	128.86	25.35	16.45	135.72
WDC	**12.9**	431.95	1.05e+03	854.73	221.6	491.26	727.04

G. Discussion

FCLSU and UnDIP estimate abundances using endmember extraction techniques, i.e., VCA and SiVM. Therefore, their performances depend on the performance of the endmember extraction techniques, which are often designed to assume pure pixels. Thus, in the absence of pure pixels, such methods provide poor performances. uDAS is a deep autoencoder network which neither incorporates the spatial information nor the geometrical information, and therefore its overall performance is not satisfactory. CyCUNet includes the spatial information by using convolutional autoencoders but it ignores the geometrical information. Additionally, its poor performance can be attributed to the absence of ASC during the optimization process. CyCUNet applies the Clamp function instead of Softmax on the abundances for the optimization, and ASC needs to be enforced after estimating the abundances. Collab uses a sparse regression to estimate the abundances using a library of endmembers created from the dataset. Hence, Collab cannot perform successfully in the absence of pure pixels. NMF-QMV incorporates geometrical information, but it does not include spatial information. Additionally, the searching

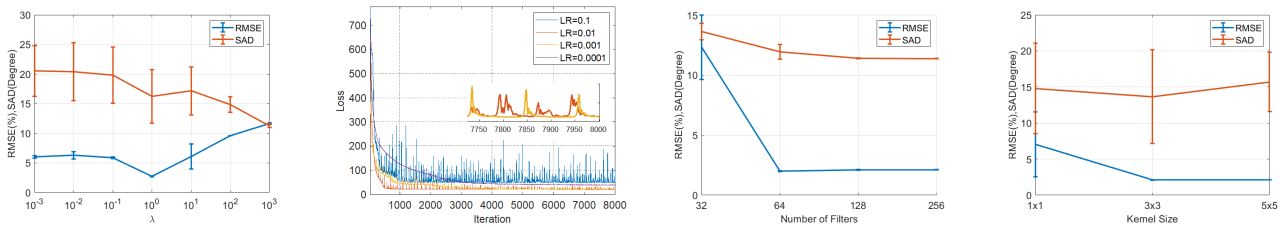


Fig. 16. Sensitivity of MiSiCNet to the hyperparameters of the network. The experiments were performed on the simulated dataset 2 (50 dB).

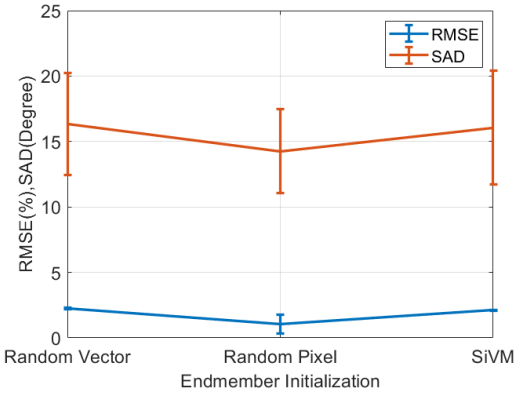


Fig. 17. Comparison of different strategies for initializing the decoder weights in MiSiCNet. The performance is given in terms of RMSE in percentage.

range for the parameter selection is too extensive such that an automatic selection technique may miss the optimal parameter values. We should note that NMF-QMV solves the unmixing problem in a subspace, i.e., it assumes that the endmembers live in an $r-1$ dimensional affine set, which best represents the data using an orthogonal projection. Such an assumption helps to remove the outliers and noise. However, the constraint on the endmembers cannot be enforced throughout the optimization, and therefore the endmembers cannot be bounded between zero and one. Accordingly, negative values for the endmembers may be estimated and should be set to zero at the end of the optimization.

On the other hand, MiSiCNet applies all the constraints in (1) throughout the optimization. More importantly, MiSiCNet incorporates spatial information by both using convolutional filters and implicitly applying a regularizer on the abundances. Additionally, geometrical information is incorporated using a penalty term. We should note that the tuning parameter λ was not optimally tuned but fixed to 100 for all the real datasets, for a fair comparison with the other techniques.

V. Conclusion

We proposed the minimum simplex convolutional network for deep hyperspectral unmixing. We have shown that a deep convolutional network can solve a blind unmixing problem defined as a constrained penalized least-squares optimization. Indeed, we show that such an optimization can be shifted towards an optimization on the deep network's parameters. We evaluated the performance of the proposed approach using two simulated and three real datasets. The simulated datasets consider realistic and challenging scenarios when there are no pure pixels and only one or two data points on the facets of the data simplex. The results on both the simulated and real datasets show considerable improvements over the state-of-the-art. Additionally, the experimental results reveal the importance of the simplex volume penalty term for unmixing when no pure pixels are available.

Acknowledgment

The work of B. Rasti was funded by the Alexander-von-Humboldt-Stiftung/foundation. The work of B. Koirala was funded by BELSPO (Belgian Science Policy Office) in the frame of the STEREO III programme – project GEOMIX (SR/06/357).

References

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[2] P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 37–78, 2017.

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "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, April 2012.

[4] N. Dobigeon, J. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 82–94, 2014.

[5] D. C. Heinz and Chein-I-Chang, "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 3, pp. 529–545, 2001.

[6] R. Heylen, D. Burazerovic, and P. Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4112–4122, Nov 2011.

[7] Tsung-Han Chan, Chong-Yung Chi, Yu-Min Huang, and Wing-Kin Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4418–4432, 2009.

[8] J. Nascimento and J. Bioucas-Dias, "Vertex component analysis: A fast algorithm to extract endmembers spectra from hyperspectral data," in *Pattern Recognition and Image Analysis*, Francisco José Perales, Aurélio J. C. Campilho, Nicolás Pérez de la Blanca, and Alberto Sanfeliu, Eds., Berlin, Heidelberg, 2003, pp. 626–635, Springer Berlin Heidelberg.

[9] J. Boardman, F. A. Kruse, and R. Green, "Mapping target signatures via partial unmixing of aviris data: in summaries," in *JPL Airborne Earth Sci. Workshop*, 1995, pp. 23–26.

[10] Michael E. Winter, "N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data," in *Imaging Spectrometry V*, Michael R. Descour and Sylvia S. Shen, Eds. International Society for Optics and Photonics, 1999, vol. 3753, pp. 266 – 275, SPIE.

[11] W. Full, R. Ehrlich, and J. Klovan, "Extended qmodel—objective definition of external end members in the analysis of mixtures," *Journal of the International Association for Mathematical Geology*, vol. 13, pp. 331–344, 08 1981.

[12] M.D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 3, pp. 542–552, 1994.

[13] Chia-Hsiang Lin, Ruiyuan Wu, Wing-Kin Ma, Chong-Yung Chi, and Yue Wang, "Maximum volume inscribed ellipsoid: A new simplex-structured matrix factorization framework via facet enumeration and convex optimization," *SIAM Journal on Imaging Sciences*, vol. 11, no. 2, pp. 1651–1679, 2018.

[14] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Collaborative nonnegative matrix factorization for remotely sensed hyperspectral unmixing," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2012, pp. 3078–3081.

[15] N. Dobigeon, S. Moussaoui, M. Coulon, J. Tourneret, and A. O. Hero, "Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery," *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4355–4368, 2009.

[16] J. Sigurdsson, M. O. Ulfarsson, and J. R. Sveinsson, "Blind hyperspectral unmixing using total variation and ℓ_q sparse regularization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 11, pp. 6371–6384, 2016.

[17] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[18] J. Li, J. M. Bioucas-Dias, A. Plaza, and L. Liu, "Robust collaborative nonnegative matrix factorization for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6076–6090, 2016.

[19] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J.F. Huntington, "Ice: a statistical approach to identifying endmembers in hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 10, pp. 2085–2095, 2004.

[20] L. Zhuang, C. Lin, M. A. T. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9858–9877, 2019.

[21] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, 2008, vol. 3, pp. III – 250–III – 253.

[22] J. M. Bioucas-Dias, "A variable splitting augmented lagrangian approach to linear spectral unmixing," in *2009 First Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, 2009, pp. 1–4.

[23] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2014–2039, 2011.

[24] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4484–4502, 2012.

[25] B. Rasti and B. Koirala, "Suncnn: Sparse unmixing using unsupervised convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2021.

[26] L. Meier, S. V. D. Geer, and P. Bühlmann, "The group lasso for logistic regression," *Journal of the Royal Statistical Society. Series B*, vol. 70, no. 1, pp. 53–71, 2008.

[27] M. Iordache, J. M. Bioucas-Dias, and A. Plaza, "Collaborative sparse regression for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 341–354, 2014.

[28] L. Drumetz, T. R. Meyer, J. Chanussot, A. L. Bertozzi, and C. Jutten, "Hyperspectral image unmixing with endmember bundles and group sparsity inducing mixed norms," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3435–3450, 2019.

[29] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. A. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep (overview and toolbox)," *IEEE Geoscience and Remote Sensing Magazine*, pp. 0–0, 2020.

[30] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427–1431, 2018.

[31] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakravortty, "Daen: Deep autoencoder networks for hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 4309–4321, 2019.

[32] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Deep generative endmember modeling: An application to unsupervised spectral unmixing," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 374–384, 2020.

[33] Y. Qu and H. Qi, "udas: An untied denoising autoencoder with sparsity for spectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, pp. 1698–1712, 2019.

[34] S. Ozkan, B. Kaya, and G. B. Akar, "Endnet: Sparse autoencoder network for endmember extraction and hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 1, pp. 482–496, 2019.

[35] 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.

[36] Ziqiang Hua, Xiaorun Li, Qunhui Qiu, and Liaoying Zhao, "Autoencoder network for hyperspectral unmixing with adaptive abundance smoothing," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 9, pp. 1640–1644, 2021.

[37] X. Zhang, Y. Sun, J. Zhang, P. Wu, and L. Jiao, "Hyperspectral unmixing via deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 11, pp. 1755–1759, 2018.

[38] B. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Spectral-spatial hyperspectral unmixing using multitask learning," *IEEE Access*, vol. 7, pp. 148861–148872, 2019.

[39] F. Khajehrayeni and H. Ghassemian, "Hyperspectral unmixing using deep convolutional autoencoders in a supervised scenario," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 567–576, 2020.

[40] B. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Convolutional autoencoder for spectral-spatial hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2020.

[41] B. Rasti, B. Koirala, P. Scheunders, and P. Ghamisi, "UnDIP: Hyperspectral unmixing using deep image prior," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–15, 2021.

[42] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[43] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," *International Journal of Computer Vision*, vol. 128, no. 7, pp. 1867–1888, Mar 2020.

[44] L. Gao, Z. Han, D. Hong, B. Zhang, and J. Chanussot, "Cycu-net: Cycle-consistency unmixing network by learning cascaded autoencoders," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14, 2021.

[45] W.-K. Lin, Ch.-H.and Ma, W.-Ch. Li, Ch.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5530–5546, 2015.

[46] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, MA, USA, 2016, http://www.deeplearningbook.org.

[47] Feiyun Zhu, Ying Wang, Bin Fan, Shiming Xiang, Geofeng Meng, and Chunhong Pan, "Spectral unmixing via data-guided sparsity," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5412–5427, 2014.

[48] Michael E. Schaepman, Michael Jehle, Andreas Hueni, Petra D'Odorico, Alexander Damm, Jürg Weyermann, Fabian D. Schneider, Valérie Laurent, Christoph Popp, Felix C. Seidel, Karim Lenhard, Peter Gege, Christoph Küchler, Jason Brazile, Peter Kohler, Lieve De Vos, Koen Meuleman, Roland Meynart, Daniel Schläpfer, Mathias Kneubühler, and Klaus I. Itten, "Advanced radiometry measurements and earth science applications with the airborne prism experiment (apex)," *Remote Sensing of Environment*, vol. 158, pp. 207–219, 2015.

Behnood Rasti (M'12–SM'19) received the B.Sc. and M.Sc. degrees both in electronics- electrical engineering from the Electrical Engineer- ing Department, University of Guilan, Rasht, Iran, in 2006 and 2009, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Iceland, Reykjavik, Iceland, in 2014. In 2015 and 2016, he worked as a Post-Doctoral Researcher with Electrical and Computer Engineering Department, University of Iceland. From 2016 to 2019, he has been a Lecturer with the Center of Engineering Technology and Applied Sciences, Department of Electrical and Computer Engineering, University of Iceland. Dr. Rasti was a Humboldt research fellow in 2020 and 2021. He is currently a Principal Research Associate with Helmholtz-Zentrum Dresden-Rossendorf (HZDR). His research interests include signal and image processing, machine/deep learning, remote sensing, and artificial intelligence.

Dr. Rasti was the Valedictorian as an M.Sc. Student in 2009. He won the Doctoral Grant of The University of Iceland Research Fund "The Eimskip University fund," and the "Alexander von Humboldt Research Fellowship Grant" in 2013 and 2019, respectively. He serves as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (GRSL).

Bikram Koirala received the M.S. degree in Geomatic Engineering from the University of Stuttgart, Germany in 2016. In 2017, he joined Vision Lab, Department of Physics, the University of Antwerp as a Ph.D. researcher. He received his Ph.D. degree in "Development of advanced hyperspectral unmixing methods" in 2021. He is currently a postdoctoral researcher at Vision Lab, Department of Physics, University of Antwerp. He is a IEEE member. His research interest includes machine learning and hyperspectral image processing.

Paul Scheunders (M'98) received the B.S. degree and the Ph.D. degree in physics, with work in the field of statistical mechanics, from the University of Antwerp, Antwerp, Belgium, in 1983 and 1990, respectively. In 1991, he became a research associate with the Vision Lab, Department of Physics, University of Antwerp, where he is currently a full professor. His current research interest includes remote sensing and hyperspectral image processing. He has published over 200 papers in international journals and proceedings in the field of image processing, pattern recognition, and remote sensing. Paul Scheunders is Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing and has served as a program committee member in numerous international conferences. He is a senior member of the IEEE Geoscience and Remote Sensing Society.

Jocelyn Chanussot (M'04–SM'04–F'12) 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 currently a Professor of signal and image processing. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning and artificial intelligence. He has been a visiting scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he is an Adjunct Professor of the University of Iceland. In 2015-2017, he was a visiting professor at the University of California, Los Angeles (UCLA). He holds the AXA chair in remote sensing and is an Adjunct professor at the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing. Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) 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 (2017-2019). 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 (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for 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 (2011-2015). In 2014 he served as a Guest Editor for the IEEE Signal Processing Magazine. He is a Fellow of the IEEE, a member of the Institut Universitaire de France (2012-2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters, since 2018).