DEEP BLIND UNMIXING USING MINIMUM SIMPLEX CONVOLUTIONAL NETWORK

Behnood Rasti[1], Bikram Koirala[2], Paul Scheunders[2], and Jocelyn Chanussot [3]

[1] Helmholtz-Zentrum Dresden-Rossendorf (HZDR),Chemnitzer Straße 40, 09599 Freiberg, Germany.
[2] Imec-Visionlab, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp, Belgium.
[3] The Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, F-38000 Grenoble, France.

ABSTRACT

This paper proposes a deep blind hyperspectral unmixing network for datasets without pure pixels called minimum simplex convolutional network (MiSiCNet). MiSiCNet is the first deep learning-based blind unmixing method proposed in the literature which incorporates both spatial and geometrical information of the hyperspectral data, in addition to the spectral information. The proposed convolutional encoder-decoder architecture incorporates the spatial information using convolutional filters and implicitly applying a prior on the abundances. We added a minimum simplex volume penalty term to the loss function to exploit the geometrical information. We evaluate the performance of MiSiCNet on simulated and 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

1. INTRODUCTION

In spectral unmixing, 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, and 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) [1].

This research was funded by Alexander von Humboldt foundation and the Research Foundation-Flanders - project G031921N.

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) [2]. Alternatively, the vertices can be selected as the extreme points after iteratively projecting the data onto a particular direction such as Vertex Component Analysis (VCA) [3]. When the endmembers are extracted prior to the unmixing problem or assumed to be known, then the unmixing problem is referred to as supervised unmixing. On the other hand, when both endmembers and abundances are estimated simultaneously, the problem is referred to as blind unmixing.

Sparse unmixing is relying on a rich and well-designed library of pure spectra and therefore it is referred to as semi-supervised unmixing. 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 [4]. On the other hand, a dictionary of endmember bundles can be generated from the data itself. An example of such approach is the Collaborative LASSO (Least Absolute Shrinkage and Selection Operator) [5]. Recently, a sparse unmixing approach was proposed using a convolutional neural network (SUnCNN) [6].

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. An untied Denoising Autoencoder with Sparsity (uDAS) was proposed in [7] 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. Unmixing using deep image prior (UnDIP) [8] utilizes a convolutional encoder-decoder architecture and a deep image prior [9]. The cycle-consistency unmixing network (CyCUNet) [10] utilizes two convolutional autoencoders, which are cascaded and performed cyclically. A major problem of the DL-based approaches is the absence of geometrical information. 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 for blind unmixing. 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.

2. METHODOLOGY

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

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

where $\mathbf{Y} 2$ $\mathbb{R}^{p \times n}$ is the observed HSI, with n pixels and p bands, $\mathbf{N}$ 2 $\mathbb{R}^{p \times n}$ is noise, $\mathbf{E}$ 2 $\mathbb{R}^{p \times r}$, and $\mathbf{A}$ 2 $\mathbb{R}^{r \times n}$, r 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.

We propose the following optimization problem for estimating both $\mathbf{E}$ and $\mathbf{A}$:

$$\arg\min_{\mathbf{A},\mathbf{E}} \frac{1}{2} jj\mathbf{Y} \quad \mathbf{EA} jj_F^2 + \lambda \left\| \mathbf{E} \quad \mathbf{m1_r^T} \right\|_F^2 + \beta R(\mathbf{A})$$
$$\text{s.t.} \mathbf{A} \quad 0, \mathbf{1}_r^T \mathbf{A} = \mathbf{1}_n^T, 0 \quad \mathbf{E} \quad 1 \quad (2)$$

where the first term is the fidelity term, the second term is the geometrical penalty term enforcing the endmembers towards the center of the data simplex, and R is the spatial penalty term applied on the abundances. λ and β control the trade-off between the penalty terms and the fidelity term. The problem is solved, subjected to the non-negativity and sum to one constraint.

Neglecting all the constraints and inspired by deep image prior (DIP) [9], the selection of a suitable regularizer for R can be substituted by optimizing the parameters of a deep network, and therefore we can turn the optimization problem (2) into an optimization of the network parameters:

$$(\hat{\theta}_1, \hat{\mathbf{E}}) = \arg\min_{\theta_1, \mathbf{E}} \frac{1}{2} jj\mathbf{Y} \quad \mathbf{E} f_{\theta_1}(\mathbf{Z}) jj_F^2 + \lambda \left\| \mathbf{E} \quad \mathbf{m1_r^T} \right\|_F^2$$
$$\text{s.t. } \hat{\mathbf{A}} = f_{\hat{\theta}_1}(\mathbf{Z}) \quad (3)$$

As $\mathbf{E}$ contains the weights of the final linear layer (i.e., the decoder), we rewrite (3) as:

$$(\hat{\theta}_1, \hat{\theta}_2) = \arg\min_{\theta_1, \theta_2} \frac{1}{2} jj\mathbf{Y} \quad \theta_2 f_{\theta_1}(\mathbf{Z}) jj_F^2 + \lambda \left\| \theta_2 \quad \mathbf{m1}_r^T \right\|_F^2$$
$$\text{s.t. } \hat{\mathbf{E}}\hat{\mathbf{A}} = \hat{\theta}_2 f_{\hat{\theta}_1}(\mathbf{Z}) \quad (4)$$

where $\hat{\mathbf{Y}} = \hat{\mathbf{E}}\hat{\mathbf{A}}$. Hence, the optimization problem (2) can be solved using a deep network with a loss function given by:

$$L(\mathbf{Y}, \hat{\mathbf{Y}}, \hat{\theta}_2, \mathbf{m}) = \frac{1}{2} jj\mathbf{Y} \quad \hat{\mathbf{Y}} jj_F^2 + \lambda \left\| \hat{\theta}_2 \quad \mathbf{m1}_r^T \right\|_F^2 \quad (5)$$

To enforce both the ASC and the ANC we use a softmax function in the final layer of the encoder 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 $\mathbf{E}$ 1.

The architecture of MiSiCNet is visualized in Fig. 1. The main architecture is based on the convolutional encoder-decoder with a skip connection. We use four convolutional layers (Conv), excluding the skip connection (ConvSkip). The number of filters for each convolutional layer is given in Table 1. 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) 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. We fix all the hyperparameters used in the experiments as given in Table 1, except the tuning parameter (λ).

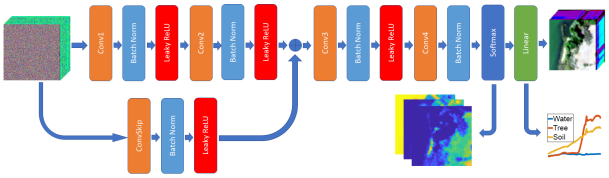

Fig. 1: The architecture of MiSiCNet. The proposed convolutional network architecture uses a skip connection and five convolutional layers.

Table 1: 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

3. EXPERIMENTAL RESULTS

The simulated dataset contains 105 105 pixels which are simulated by linear combinations of six endmembers (see Fig. 2(b) for the endmembers). Each facet of the data manifold contains only one mixed data point, making it challenging to reconstruct virtual endmembers geometrically.

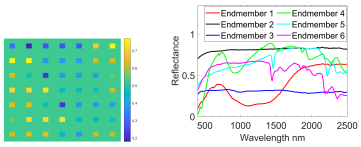

Fig. 2: Simulated Data: left: Band number 70, right: Endmembers.

The Samson hyperspectral dataset (Fig. 3(a)) contains 95 95 pixels. It contains 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. 3(b) were manually selected from the hyperspectral image, and the ground truth fractional abundances were generated using FCLSU.

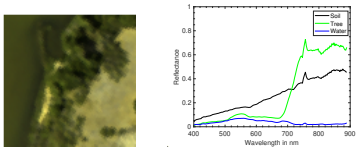

Fig. 3: Samson: left: True-color image, right: Endmembers.

Six unmixing techniques from different unmixing categories were used as competing methods in the experiments: FCLSU (supervised unmixing using VCA [3] for the endmember extraction) [1], NMF-QMV [11] (blind unmixing by incorporating geometrical information), Collaborative LASSO (Collab) [12] (sparse or semisupervised unmixing), uDAS [7] (blind deep unmixing), UnDIP [8] (supervised deep unmixing), and CyCUNet [10] (blind deep unmixing). For MiSiCNet, we selected all the hyperparameters as given in Table 1, and we set λ to 0.3 and 100 for the simulated dataset and real datasets, respectively. We choose all the parameters for the competing methods according to the reported default values. Quantitative results are provided by the root mean squared error (RMSE) in percentage and the spectral angle distance (SAD) in degree.

Table 2 reports the RMSE results for the simulated dataset. All the competing methods perform poorly. MiSiCNet shows considerable improvements even compared to NMF-QMV. Additionally, the very low standard deviations reported in the table 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 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.

Table 3 reports SAD for the simulated data set. The results follow the trend of RMSE. MiSiCNet considerably outperforms the other techniques for all SNRs for the simulated dataset. The simulated experiment reveals 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.

Table 2: Simulated Data: RMSE

	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 3: Simulated Data: SAD

	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**

Table 4 shows the abundance RMSE obtained by the different unmixing techniques on 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. 4 and 5 depict the estimated abundances and endmembers, respectively, obtained by the different techniques. This again confirms that MiSiCNet outperforms the other techniques. Table 5 reports SAD values of the endmembers on Samson. MiSiCNet outperforms the other techniques in terms of SAD for Soil and Tree, but not for Water. 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.

Table 4: 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**