SUnCNN: Sparse Unmixing Using Unsupervised Convolutional Neural Network

Behnood Rasti, *Senior Member, IEEE,* and Bikram Koirala, *Member, IEEE,*

***Abstract*—In this paper, we propose a sparse unmixing technique using a convolutional neural network (SUnCNN) for hyperspectral images. SUnCNN is the first deep learning-based technique proposed for sparse unmixing. It uses a deep convolutional encoder-decoder to generate the abundances relying on a spectral library. We reformulate the sparse unmixing into an optimization over the deep network's parameters. Therefore, the deep network learns in an unsupervised manner to map a fixed input into the sparse optimum abundances. Additionally, SUnCNN holds the sum-to-one constraint using a softmax activation layer. The proposed method is compared with the state-of-the-art using two synthetic datasets and one real hyperspectral dataset. The overall results confirm that the proposed method outperforms the other ones in terms of signal to reconstruction error (SRE). Additionally, SUnCNN shows visual superiority for both real and synthetic datasets compared with the competing techniques. The proposed method was implemented in Python (3.8) using PyTorch as the platform for the deep network and is available online: https://github.com/BehnoodRasti/SUnCNN.**

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

I. Introduction

SPECTRAL unmixing estimates the fractional abundances of different pure materials, so-called endmembers, within a hyperspectral pixel. This is done by minimizing the error between the true reflectance spectrum and the spectrum generated by a particular mixing model. In remote sensing applications, the linear mixing model (LMM) [1] is the most popular one. The main assumption of the LMM is that the incoming rays of light interact only once with a specific pure material in the pixel before reaching the sensor. The fractional abundance is the areal percentage, and therefore, no endmember can have a negative area yielding the abundance nonnegativity constraint (ANC). Additionally, the observed spectrum is entirely decomposed by endmembers which leads to the abundance sum-to-one constraint (ASC). Taking into account both ANC and ASC, the reconstruction error between the true spectrum and the linearly mixed spectrum can be minimized using the fully constrained least squares unmixing (FCLSU) [2].

In general, spectral unmixing has three steps. First, the number of endmembers available in the hyperspectral image (HSI) is estimated. Then, endmembers are extracted from the HSI. Finally, fractional abundances are estimated using the extracted endmembers [3], [4]. On the other hand, blind unmixing techniques simultaneously estimate endmembers and abundances. However, due to the non-convexity of the problem, they rely on a good initialization of endmembers using endmember extraction techniques. Therefore, both groups of approaches perform very well if the HSI contains pure pixels. They might also cope with the scenario in which there exist sufficient spectra on the facets of the data simplex allowing to geometrically locate the vertices of the data simplex. When neither the pure pixels nor the sufficient spectra on the facets of the data simplex are available in the HSI, the endmembers can not be successfully extracted/estimated that results in poor abundance estimations. An alternative approach to solve this problem is to use spectral libraries. As the number of pure materials available in the scene is fewer than the number of endmembers in the spectral libraries, only a few endmembers can reconstruct the mixed hyperspectral pixel. That leads to the sparse abundance matrix. Therefore, sparse regression techniques are exploited to estimate the abundances without extracting/estimating the endmembers. This group of methods is called sparse unmixing [5].

Sparse unmixing by variable splitting and augmented Lagrangian (SUnSAL), constrained SUnSAL (C-SUnSAL) [5] and collaborative sparse unmixing [6] are examples of sparse unmixing methods. Both SUnSAL and C-SUnSAL apply an ℓ_1 penalty on the fractional abundances. SUnSAL utilizes ℓ_2 for the fidelity term while C-SUnSAL assumes a constraint to enforce the data fidelity. Collaborative sparse unmixing is similar to SUnSAL but applies $\ell_{2,1}$ (i.e., the sum of ℓ_2 on the abundances) to promote the sparsity on the abundances. In [7], a spectral prior was added to the sparse regression problem that assumes some materials are known in the scene.

SUnSAL was improved in [8] by incorporating spatial information through applying a total variation penalty on the abundances (SUnSAL-TV). Some drawbacks of SUnSAL-TV are that it oversmoothed boundaries, blurred abundance maps, and is computationally expensive. This was somewhat addressed by developing the technique called local collaborative sparse unmixing (LCSU) [9] and a new spectral–spatial weighted sparse unmixing (S^2WSU) framework [10]. In [11], an efficient two-phase multiobjective sparse unmixing approach was presented to exploit the spatial-contextual information for improving the abundance estimation.

The problem of SUnSAL-TV was further tackled by introducing a fast Multiscale Sparse Unmixing Algorithm (MUA) [12] that promotes piecewise homogeneous abundances without compromising sharp discontinuities among neighboring pixels. This method consists of two steps. In the first step,

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, behnood.rasti@gmail.com

Bikram Koirala is with Imec-Visionlab, University of Antwerp (CDE) Universiteitsplein 1, B-2610 Antwerp, Belgium; Bikram.Koirala@uantwerpen.be

Manuscript received .....

coarse fractional abundances are estimated by grouping pixels into perceptually meaningful regions and performing sparse regression. For the pixels grouping, a binary partition tree (BPT), the simple linear iterative clustering (SLIC), and the K-means algorithm can be utilized. In the second step, the fractional abundances were estimated using a sparse regression problem in which the coarse fractional abundance matrix (estimated in the first step) is the regularizer. Similarly, in [13], SLIC was utilized for the pixels grouping, and pixel-based sparse unmixing was performed using superpixel-based graph Laplacian regularization.

Although numerous sparse unmixing techniques were developed, a limited number of algorithms consider ASC [14]. When fractional abundances are estimated without considering ASC, the estimated fractional abundances do not necessarily describe the aerial fraction of each pure material on the ground. To tackle this challenge, in this paper, we propose a sparse unmixing technique using a convolutional neural network (SUnCNN) for hyperspectral images. SUnCNN uses a deep convolutional encoder-decoder to generate the abundances relying on a spectral library. We show that sparse unmixing can be reformulated into an optimization over the deep network's parameters. Additionally, SUnCNN holds the sum-to-one constraint using a softmax activation layer. The major contributions of this paper are as follows: 1) This paper reformulates the sparse unmixing problem into an optimization over a deep network's parameters and therefore the proposed method (i.e., SUnCNN) is the first deep learning-based sparse unmixing technique; 2) SUnCNN implicitly induces an image prior while holding ASC; and 3) SUnCNN incorporates spatial information using the convolutional filters.

II. Methodology

In semi-supervised spectral unmixing, the observed spectra can be represented using a linear mixing model given by

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{N}, \tag{1}$$

where $\mathbf{Y}$ and $\mathbf{N} \in \mathbb{R}^{p \times n}$ denote the observed HSI and the model error including noise, respectively. $\mathbf{D} \in \mathbb{R}^{p \times m}$ and $\mathbf{X} \in \mathbb{R}^{m \times n}$, $p < m$. $\mathbf{D}$ is the spectral library containing m endmembers and $\mathbf{X}$ is the unknown fractional abundances. In sparse unmixing, the goal is to estimate the fractional abundances $\mathbf{X}$ by applying sparsity-enforcing penalties or constraints.

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^{n} \|\mathbf{x}_{(i)}\|_q \quad \text{s.t.} \; \mathbf{X} \geq 0, \mathbf{1}_m^T \mathbf{X} = \mathbf{1}_n^T, \tag{2}$$

where $0 < q < 1$. Therefore the problem to solve is non-convex. To have a convex problem i.e., $q = 1$ the sum to one constraint should be omitted due to the conflict with the ℓ_1 sparsity (see [14] for more detail) and therefore the problem to solve is

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda \sum_{i=1}^{n} \|\mathbf{x}_{(i)}\|_1 \; \text{s.t.} \; \mathbf{X} \geq 0. \tag{3}$$

Defining the problem as (3) without enforcing the ASC might provide meaningless abundance values. Additionally, the question that which penalty and what value for q is more suitable for sparse unmixing is still an open question and the selection of the penalty might be dependent on the the observed data. Therefore, we address this issue using an unsupervised deep network. In [15], [16], it has been shown that a general regularizer such as $R(\mathbf{X})$ can be implicitly applied using a deep network called deep image prior (DIP). In other words, the problem

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda R(\mathbf{X}) \tag{4}$$

can be reformulated as

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{2}\|\mathbf{Y} - \mathbf{D} f_\theta(\mathbf{Z})\|_F^2 \quad \text{s.t.} \quad \hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z}), \tag{5}$$

where $\mathbf{Z}$ is the network input and it is constant during the training, f_θ is the deep network with parameters θ (θ is randomly initialized) and updated through the learning process. The deep network f_θ is learned to map the input $\mathbf{Z}$ to $\hat{\mathbf{X}}$ such that $\hat{\mathbf{X}} = f_{\hat{\theta}}(\mathbf{Z})$. To perform this mapping the network parameters i.e., $\hat{\theta}$ should be estimated. Hence, they are optimized iteratively by computing the gradient of the loss function (5). Both ASC and ANC are enforced using a softmax function in the final layer of the network given by

$$softmax(\mathbf{X}) = \frac{e^{\mathbf{X}_{ij}}}{\sum_{i=1}^{m} e^{\mathbf{X}_{ij}}} \quad \forall i, j \tag{6}$$

As a result, the proposed network finds a solution to the following optimization problem

$$\hat{\mathbf{X}} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 + \lambda R(\mathbf{X}) \quad \text{s.t.} \; \mathbf{X} \geq 0, \mathbf{1}_m^T \mathbf{X} = \mathbf{1}_n^T, \tag{7}$$

In experiments, we show that if a suitable spectral library (i.e., $\mathbf{D}$) is available (which is necessary for sparse unmixing techniques why it is called semi-supervised unmixing), then a well-designed deep network can implicitly induce a regularizer ($R(\mathbf{X})$) to enforce a satisfactory sparsity on $\mathbf{X}$ while it holds the ASC.

A. Convolutional Neural Network in SUnCNN

We use a convolutional encoder-decoder for $f_{\hat{\theta}}$ having a skip connection as shown in Fig. 1. SUnCNN uses 5 convolutional layers (Conv) in the main forward path and one in the skip connection, a downsampling block, and an upsampling block. We use batch norm (BN) to speed up the learning process and obtain robustness for selecting the hyperparameters. Leaky ReLU was selected as the activation function for the first synthetic datacube (DC1) and the real dataset (Cuprite) and ReLU was selected for the second synthetic datacube (DC2). The final layer is the softmax layer. All the hyperparameters selected for the network are given in Table I. SUnCNN is an unsupervised deep network inspired by DIP and the network is trained over the observed data iteratively and, therefore, the number of iterations is a hyperparameter. In experiments, the number of iterations is set to 4000, 8000, 16000, for 20 dB, 30

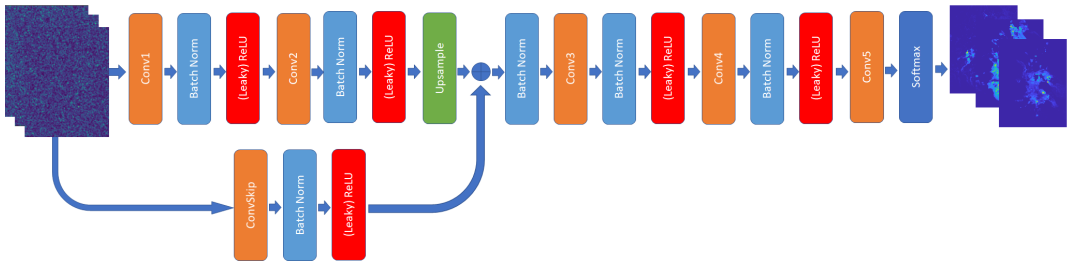

Fig. 1. The architecture of the deep network, f , used in SUnCNN. The network uses a skip connection and different layers are shown using specific colors.

dB, and 40 dB, respectively, in the case of simulated datasets. In the case of the real dataset, the number of iterations is set to 20000. We should note that the input of SUnCNN is set to the observed data, i.e., $\mathsf{Z} = \mathsf{Y}$, for DC1 and to noise, i.e., $\mathsf{Z} = \mathsf{N}$ for DC2 and the real dataset.

TABLE I
HYPERPARAMETERS USED IN THE EXPERIMENTS FOR UNDIP.

Hyperparameters				
	Input Ch.	Ouput Ch.	Filter Size	Stride
Conv1	p	256	3x3	2
Conv2	256	256	3x3	1
Conv3	260	256	3x3	1
Conv4	256	256	1x1	1
Conv5	256	p	1x1	1
ConvSkip	p	4	1x1	1
	Negative Slope			
Leaky ReLU	0.1			
	Scale Factor		Mode	
Upsample	2		Bilinear	
	Type		Learning Rate	
Optimizer	Adam		0.001	

.

III. EXPERIMENTAL RESULTS

Here, the results of SUnCNN are compared with five different sparse unmixing techniques (i.e., SUnSAL [17], SUnSAL-TV [8], S^{2}WS [10], MUA$_{\text{BPT}}$ and MUA$_{\text{SLIC}}$ [12]) applied to two simulated datasets (i.e., DC1 and DC2) and one real dataset (i.e., Cuprite). All the tuning parameters are set as default for the competing methods. We should note that the results are mean values over 10 experiments.

A. Simulated Experiments

Two popular datasets in sparse unmixing (i.e., DC1 and DC2) were used in simulated experiments. The synthetic library used for the simulated experiments is composed of 240 spectral signatures from the USGS library with the minimum pair-spectra angle of 4.44°. DC1 with 75 75 pixels was simulated using a linear mixing model with 5 endmembers. The endmembers were selected from the library and the abundance maps are composed of five rows of square regions uniformly distributed over the spatial dimension. DC2 with 100 100 pixels was simulated using a linear mixing model with 9 endmembers. The endmembers were selected from the library and the abundance maps were sampled from a Dirichlet distribution centered at a Gaussian random field to have piecewise smooth maps with steep transitions. The results are in terms of signal to reconstruction error (SRE) in dB given by

$$SRE(\mathsf{X}, \hat{\mathsf{X}}) = 10 \log_{10} \frac{k\mathsf{X}k_F}{k\mathsf{X} \quad \hat{\mathsf{X}}k_F}, \tag{8}$$

for the three levels of additive noise i.e., 20, 30, and 40 dB.

Table II shows the results obtained by applying different sparse unmixing techniques to DC1. The results confirm that SUnCNN outperforms the other techniques for 30 dB and 40 dB. In the case of 20 dB, SUnCNN provides the second-best results in terms of SRE. SUnSAL performs the poorest in terms of SRE. S^{2}WSU performs better for higher SNR and MUA performs better for the low SNRs. This could be attributed to the segmentation-based framework used in MUA which copes with the low SNRs by oversmoothing the abundances. The visual comparison in Fig. 2 confirms the argument of oversmoothed abundance maps estimated by using both segmentation-based techniques. SUnSAL-TV also gives oversmoothed abundances. Additionally, Fig. 2 shows that SUnCNN successively preserves the structures existing in the abundance maps even for low SNR i.e., 20 dB. On the other hand, all the competing methods with varying degrees fail to preserve the structures, particularly for lower SNRs. This advantage of SUnCNN can be attributed to the incorporation of spatial information via convolutional operators. We should note that SUnSAL-TV and S^{2}WSU also utilize the spatial information, however, the results confirm that SUnCNN considerably outperforms them. Moreover, S^{2}WSU, MUA$_{\text{BPT}}$, and MUA$_{\text{SLIC}}$ induce artifacts into the abundance maps which is considered a drawback associated to those techniques.

TABLE II
THE RESULTS OF DIFFERENT SPARSE UNMIXING TECHNIQUE APPLIED TO DC1 IN TERMS OF SRE. THE BEST PERFORMANCES ARE SHOWN IN BOLD.

SNR	SUnSAL	SUnSAL-TV	S^{2}WSU	MUA$_{\text{BPT}}$	MUA$_{\text{SLIC}}$	SUnCNN
20 dB	2.27	4.71	3.85	**6.70**	5.67	5.71
30 dB	4.46	7.22	7.74	9.13	7.87	**10.25**
40 dB	6.89	11.05	14.12	10.72	11.17	**15.20**

The results of the sparse unmixing techniques applied to DC2 are compared in Table III in terms of SRE and in Fig. 3 visually for the three different levels of noise. The results follow similar trends to DC1. SUnCNN outperforms the other techniques overall. It provides considerable improvements for 40 dB. In the case of 30 dB, S^{2}WSU gives the best results and

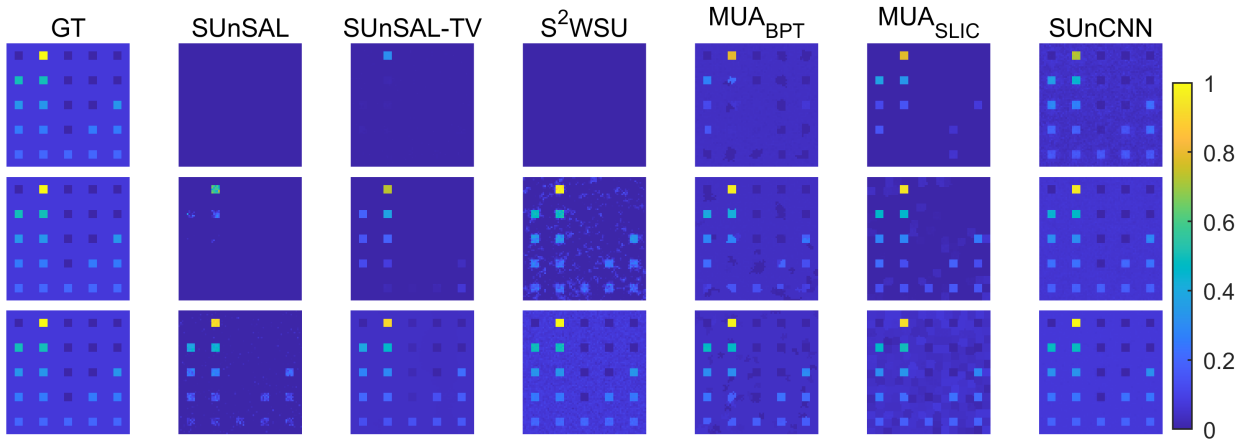

Fig. 2. DC1: The fractional abundance of endmember 2. From top to bottom SNR of 20, 30, and 40 dB.

SUnCNN is the second best with 0.14 dB difference. MUA_{SLIC} and MUA_{BPT} provide the highest SRE (unlike DC1 for 20 dB, in the case of DC2 MUA_{SLIC} outperforms the other techniques) and SUnCNN outperforms the rest of the techniques. This experiment further confirms our argument that the segmentation-based techniques provide higher SRE in the case of low SNR due to the segmenting module used in those techniques, however, the visual comparisons (in Fig. 3) reveals that they induce artifacts and oversmooth the abundances. The results confirm that SUnSAL is not robust to noise and SUnSAL-TV oversmooth the abundances, particularly for low SNRs. S^2WSU also performs poorly in the case of SNR=20 dB. The experiments in Table II and III confirm the consistency of SUnCNN to the noise power compared to the other techniques. Segmentation based techniques (i.e., MUA_{BPT} and MUA_{SLIC} [12]) perform well only for low SNRs. S^2WS [10] outperforms the other techniques only for DC2 in the case of SNR=30dB with a small advantage compared to the SUnCNN.

TABLE III
THE RESULTS OF DIFFERENT SPARSE UNMIXING TECHNIQUE APPLIED TO DC2 IN TERMS OF SRE. THE BEST PERFORMANCES ARE SHOWN IN BOLD.

SNR	SUnSAL	SUnSAL-TV	S^2WSU	MUA_{BPT}	MUA_{SLIC}	SUnCNN
20 dB	2.14	5.78	4.67	6.96	**7.38**	6.50
30 dB	5.22	8.99	**10.83**	8.48	9.16	10.69
40 dB	8.94	10.55	12.07	9.09	10.46	**14.35**

B. Real Experiment

In the real experiment, we use a spatial subset of the Cuprite dataset (250 191 pixels). The minerals in that region are well-studied and therefore are suitable to compare the abundance maps visually. The geological ground reference for the dominant minerals existing in the scene is shown in Fig. 4 (a). The library $\mathbf{D}\ 2\ \mathsf{R}^{188\ \ 498}$ is composed of 498 spectral pixels from the USGS library. The water absorption and noisy bands were removed and the final pixels are of dimension $p = 188$.

Fig. 4 (b) visually compares the abundance maps obtained by different techniques applied on the Cuprite for four dominant minerals (i.e., Alunite, Buddingtonite, Chalcedony, and Kaolinite) in the scene. Visual comparison reveals that SUnSAL-TV, MUA_{BPT} and MUA_{SLIC} performed similarly. It can be observed that they all oversmooth the mineral abundances which could be attributed to the total variation penalty in SUnSAL-TV and the segmentation-based framework in both MUA_{BPT} and MUA_{SLIC} which cannot preserve the textures. On the other hand, SUnCNN, SUnSAL, and S^2WSU provide sharper maps. Comparing to the ground truth of USGS (Fig. 4 (a)) both SUnCNN and S^2WSU perform slightly better for estimating Buddingtonite. However, the abundance maps estimated using both SUnSAL and S^2WSU are noisy compared to the ones estimated using SUnCNN. That could be attributed to the convolutional operator that exploits the spatial information while preserving the spatial textures.

IV. CONCLUSION

In this paper, we proposed a sparse unmixing technique using a deep convolutional neural network called SUnCNN. SUnCNN uses an unsupervised convolutional encoder-decoder which iteratively generates the abundance maps by implicitly enforcing a prior on the loss function. The experiments were carried out on two synthetic datasets and the Cuprite dataset. The results confirmed the advantage of the SUnCNN compared to the state-of-the-art both in terms of SRE and visually for the synthetic datasets. Additionally, the visual comparison confirmed that, unlike the competing techniques, SUnCNN provides sharp maps without artifacts while decreases the noise effect on the abundance maps even for low SNR.

ACKNOWLEDGMENT

The work of Behnood Rasti was funded by the Alexander-von-Humboldt-Stiftung/foundation.

REFERENCES

[1] J. W. Boardman, "Geometric mixture analysis of imaging spectrometry data," in *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, 1994, vol. 4, pp. 2369–2371.