

WU-NET: A WEAKLY-SUPERVISED UNMIXING NETWORK FOR REMOTELY SENSED HYPERSPECTRAL IMAGERY

Danfeng Hong^{1,2}, Jocelyn Chanussot³, Naoto Yokoya⁴, Uta Heiden⁵, Wieke Heldens⁵, Xiao Xiang Zhu^{1,2}

¹Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Wessling, Germany

²Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany

³Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France

⁴Geoinformatics Unit, RIKEN Center for Advanced Intelligence Project (AIP), RIKEN, Tokyo, Japan

⁵German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling, Germany

ABSTRACT

Recently, enormous efforts have been made to improve the performance of the linear or nonlinear mixing model for hyperspectral unmixing, yet their ability to handle spectral variability and extract physically meaningful endmembers remains limited. Based on the powerful learning ability of deep learning, we propose a weakly-supervised unmixing network, called WU-Net, to break the bottleneck. Beyond the autoencoder-like architecture, WU-Net learns an additional network from the pure or nearly-pure endmembers to correct the weights of another unmixing network towards a more accurate and interpretable unmixing solution, thus yielding a two-stream deep network. Experimental results conducted on two different datasets, one fully artificial simulation dataset and one simulated EnMap dataset generated from a real HyMap dataset, demonstrate the effectiveness and superiority of WU-Net over several state-of-the-art algorithms.

Index Terms— Deep learning, hyperspectral imagery, remote sensing, spectral unmixing, two-stream network, weakly-supervised.

1. INTRODUCTION

Spaceborne or aerial spectroscopy (or hyperspectral) imagery is characterized by rich spectral information, which enables the identification and detection of materials at a more accurate level. This has led to a growing interest in hyperspectral data processing and analysis, such as dimensionality reduction [1, 2], image fusion [3, 4], land-cover and land-use classification [5, 6], pansharpening [7, 8], spectral unmixing [9, 10], and target detection [11, 12, 13]. Due to the low spatial resolution in reality, there are a large number of mixed pixels in hyperspectral imagery (HSI), inevitably degrading spectral discrimination. Spectral unmixing is usually defined as an important pre-processing procedure that simultaneously estimates a collection of individual components (*endmembers*) and a series of corresponding fractional percentages (*abundances*).

A common but effective model, the linear mixing model (LMM), has been widely applied for hyperspectral unmixing (HU). Nevertheless, spectral variability (SV) is ubiquitous in HSI. In general, SV refers to the various deformations of a spectral signature in a certain material, due to the effects of illumination and topography, atmospheric conditions, and the intrinsic variability of the materials. SV inevitably transfers the unpredicted errors into LMM, yielding relatively poor unmixing performance. Despite numerous LMM-based methods proposed to address various spectral variabilities (e.g., ELMM [14], SULoRA [9], ALMM [10]), the modeling ability in complex conditions does not generalize well. Consequently, the linearized strategy still hinders the unmixing performance from further improving. Very recently, some nonlinear models based on deep learning (DL) techniques [15, 16] have been proposed to enhance the ability of data representation and fitting, further improving HU's performance. In practice, these methods tend to generate physically meaningless endmembers, since there is a lack of effective guidance for real endmembers in the blind HU.

To address the aforementioned limitations, we propose a weakly-supervised unmixing network, called WU-Net. WU-Net starts with a two-stream network architecture. One learns a mapping between the relatively pure endmembers extracted from the image and their corresponding abundances. The other is an autoencoder-like network similar to those previously-proposed DL-based unmixing methods. Significantly, we force the two networks in the unmixing module to share the same weights in order to transfer the potentially intrinsic properties of the endmembers into our network system, thus yielding a more physically meaningful unmixing process. In addition, the abundance non-negative constraint (ANC) and the abundance sum-to-one constraint (ASC) are also embedded into the network with the additional layers.

The paper is organized as follows. The methodology is detailed in Section 2. Section 3 presents the experimental results on two datasets. Section 4 draws a conclusion after presenting a brief summary.

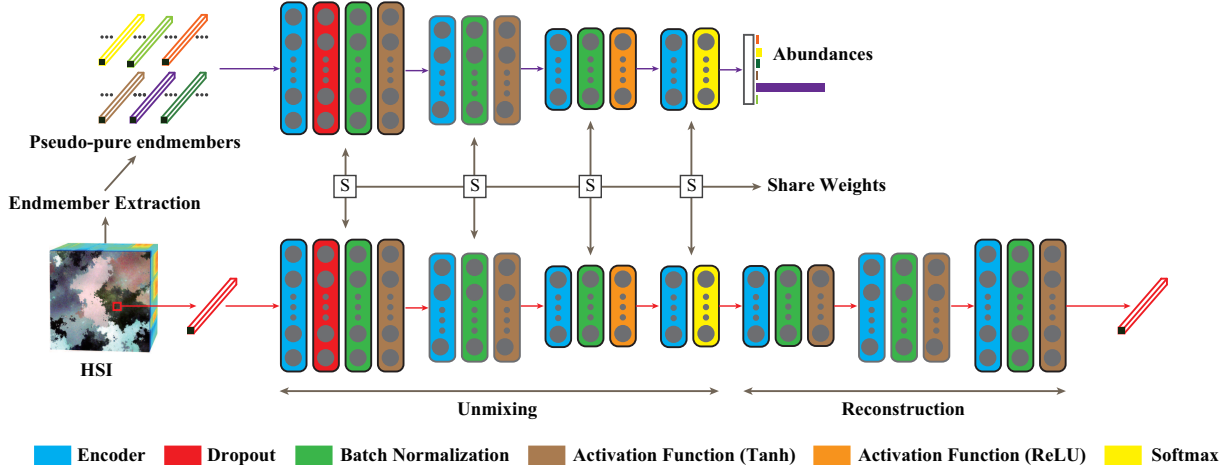


Fig. 1. An overview of the proposed network architecture (WU-Net).

2. METHODOLOGY

2.1. Network Architecture

In this section, we detail the proposed WU-Net for HU; Fig. 1 illustrates the corresponding architecture overview. WU-Net is based on a two-stream deep network, which consists of an *Endmember Network* and an *Unmixing Network*.

2.1.1. Endmember Network

We propose an *Endmember Network* (EN) for hierarchical endmember representation that maps the endmembers to their abundances, as illustrated at the top of Fig. 1. These endmembers can be extracted from the original HSI with endmember extraction algorithms (e.g., VCA [17]), while the corresponding abundances can be regressed by LMM-related methods.

2.1.2. Unmixing Network

Following the previous autoencoder-based unmixing framework, we design a similar *Unmixing Network* (see the bottom of Fig. 1) with two modules of unmixing and reconstruction. The main differences lie in two aspects. In one, a dropout layer is added in the beginning to remove the SV effectively, and then a ReLU activation function is specified for an ANC demand in the penultimate layer of the unmixing process; the softmax layer is finally used to meet the ASC with respect to the abundances. In the other, the parameters in our *Unmixing Network* (UN) are shared with those of the *Endmember Network*, making it possible to take the endmember properties fully into consideration during the process of unmixing.

2.2. Problem Formulation

Given the input $\mathbf{x}_i \in \mathbb{R}^L$, which denotes the spectral signature with L bands at i -th pixel, the l -th hidden layer (\mathbf{z}_i^l) in the

network can be formulated by

$$\mathbf{z}_i^{(l)} = \begin{cases} f(\mathbf{W}^{(l)\top} \mathbf{x}_i + \mathbf{b}^{(l)}), & l = 1, \\ f(\mathbf{W}^{(l)\top} \mathbf{z}_i^{(l-1)} + \mathbf{b}^{(l)}), & l = 2, \dots, m, \end{cases} \quad (1)$$

where f represents the activation function, and m is the number of layers. $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ stand for the weight and bias in the l -th layer, respectively. Accordingly, the overall loss function of WU-Net can be written as

$$L = L_{EN} + L_{UN} + \beta L_{reg}, \quad (2)$$

where L_{EN} is the cross-entropy loss for *Endmember Network*, while L_{UN} is the mean square loss between the input HSI and the reconstructed one for *Unmixing Network*. To avoid the overfitting of the network, a regularization term (L_{reg}) parameterized by β , is also considered in the weights.

3. EXPERIMENTS

3.1. Dataset Description

We quantitatively evaluate the unmixing performances of several state-of-the-art algorithms on two datasets: a synthetic dataset widely used in [18, 9, 10, 14] and a simulated EnMAP dataset. The first data was generated by five endmembers with 224 bands selected from the USGS spectral library. It consists of 200×200 pixels. For the second dataset, the EnMAP imagery is simulated from the HyMap data over Munich, Germany, using the EeteS – an EnMAP end-to-end simulation tool [19], with abundance maps that are computed by a high-quality ground truth [20]. We selected a region of interest (ROI) of 171×93 pixels with 221 spectral bands after noisy bands removal. In the studied scene, seven main categories are investigated, including #Roof, #Asphalt, #Polyvinyl Chloride (PVC), #Metal, #Soil, #Water, and #Vegetation. Fig. 2 shows the false color images of the two data.

Table 1. Network configuration in each layer of WU-Net.

Dataset	Pathway	Encoder	Decoder
Synthetic	EN	224-160-80-20-5	/
	UN	224-160-80-20-5	5-20-80-160-224
EnMAP	EN	221-160-80-20-7	/
	UN	221-160-80-20-7	7-20-80-160-221



(a) Synthetic data



(b) EnMAP data

Fig. 2. False color images of the two HSI used.

3.2. Implementation Details

Our network is implemented on the Tensorflow framework. Specifically, the Adam optimizer with the “poly” learning rate is adopted to update the network parameters, and the initial learning rate is 0.1 with a power of 0.99. The momentum in batch normalization is set to 0.9. The dropout technique (0.9 keeping probability) and $L2$ -norm regularization parameterized by 0.01 are used to improve the network’s generalization ability. There is a need for around 200 epochs for model convergence. Table 1 details the network architecture in each layer of WU-Net. In our case, we used around 500 endmembers extracted by VCA for training the *Endmember Network*.

3.3. Comparison with State-of-the-art Approaches

Table 2 lists the quantitative unmixing results of different methods with respect to abundance overall root mean square error (ARMSE). Fig. 3 visualizes the abundance maps of four main components (*Asphalt*, *Soil*, *Water*, and *Vegetation*) and the reconstruction error maps between real HSI and reconstructed HSI using FCLSU, ALMM and WU-Net.

Overall, FCLSU yields poor unmixing performances, as it neglects to model the SV, leading to an inaccurate abundance map estimation while strictly following the strong constraints of ANC and ASC. By considering the main SV (scaling factor), ELMM performs better than FCLSU. There is, however, a deficiency for ELMM in modeling other spectral variabilities. For that, SULoRA achieves a competitive result from the perspective of low-rank subspace unmixing, while ALMM finely models the scaling factors and other spectral variabilities by learning an SV-orientated dictionary, thereby further improving the unmixing performance. Despite this, these LMM-based approaches still fail to represent SV well in

Table 2. Quantitative performance comparisons of different algorithms in terms of ARMSE values.

Methods	Synthetic Data	EnMAP Data
FCLSU [10]	0.0630	0.2057
ELMM [14]	0.0323	0.1956
SULoRA [9]	0.0220	0.1881
ALMM [10]	0.0215	0.1804
Autoencoder [16]	0.0331	0.2040
WU-Net	0.0190	0.1653

the complex scene, due to their limited representation ability. According to the powerful learning ability of deep networks, autoencoder-like models provide new insight into the research of nonlinear unmixing. We have to admit, however, that the lack of necessary prior knowledge could generate massive bad solutions, in spite of the strong data fitting ability in autoencoder. This might explain why the autoencoder fails to effectively unmix the HSI. The proposed WU-Net outperforms the other algorithms, which indicates that the two-stream network architecture is capable of effectively addressing the SV and yields effective unmixing owing to the additional *Endmember Network* that fully embeds the semantically meaningful properties of the endmembers into the *Unmixing Network*.

4. CONCLUSION

In this paper, we propose a weakly-supervised unmixing network called WU-Net, which is a well-designed two-stream deep network for hyperspectral unmixing. Unlike the previously proposed autoencoder-like models, WU-Net additionally learns an *Endmember Network* from pure or nearly-pure endmembers and transfer its parameters into the autoencoder-based *Unmixing Network*, yielding more reasonable and superior unmixing. Notably, WU-Net is limited by endmember extraction to some extent. In our future work, we would like to develop a more general network-based framework with the aid of the multi-modal data (e.g., multispectral data [21, 22]) to address this issue more effectively.

5. REFERENCES

- [1] D. Hong, N. Yokoya, and X. Zhu, “Learning a robust local manifold representation for hyperspectral dimensionality reduction,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2960–2975, 2017.
- [2] D. Hong, N. Yokoya, J. Xu, and X. Zhu, “Joint & progressive learning from high-dimensional data for multi-label classification,” in *Proc. ECCV*, 2018, pp. 478–493.
- [3] X. Liu, C. Deng, J. Chanussot, D. Hong, and B. Zhao, “Stfnnet: A two-stream convolutional neural network for spatiotemporal image fusion,” *IEEE Trans. Geosci. Remote Sens.*, 2019.
- [4] J. Hu, D. Hong, Y. Wang, and X. Zhu, “A comparative review of manifold learning techniques for hyperspectral and polari-

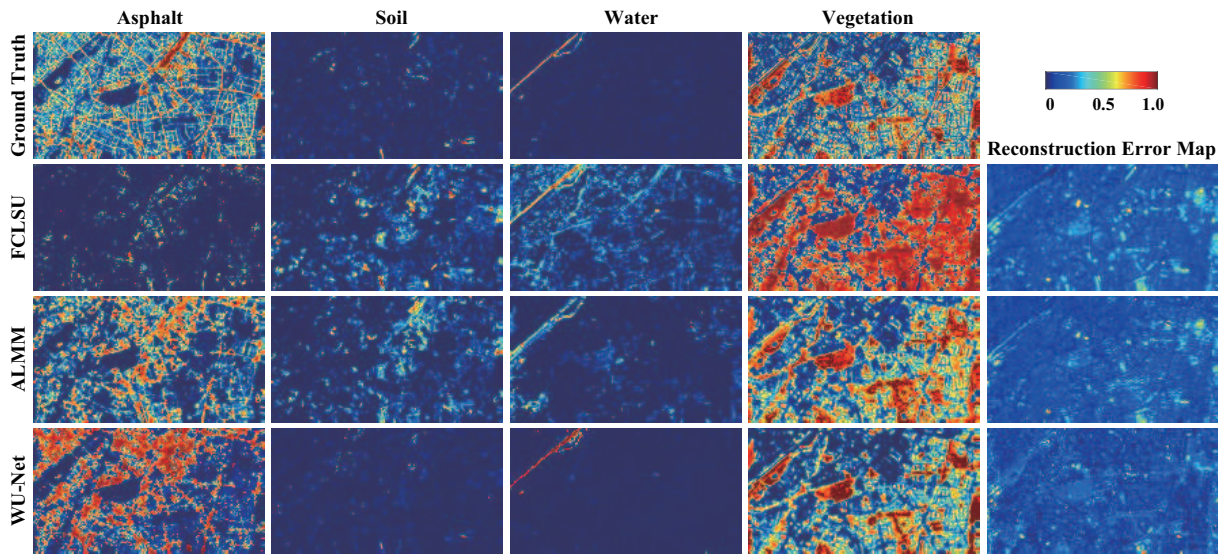


Fig. 3. Abundance maps of four main components and reconstruction error map between real HSI and reconstructed HSI.

- metric sar image fusion,” *Remote Sens.*, vol. 11, no. 6, pp. 681–2019.
- [5] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, “Cascaded recurrent neural networks for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, 2019.
- [6] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. Zhu, “Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification,” *ISPRS J. Photogramm. Remote Sens.*, vol. 147, pp. 193–205, 2019.
- [7] L. Deng, G. Vivone, W. Guo, M. Dalla Mura, and J. Chanussot, “A variational pansharpening approach based on reproducible kernel hilbert space and heaviside function,” *IEEE Trans. on Image Process.*, vol. 27, no. 9, pp. 4330–4344, 2018.
- [8] G. Vivone, R. Restaino, and J. Chanussot, “Full scale regression-based injection coefficients for panchromatic sharpening,” *IEEE Trans. on Image Process.*, vol. 27, no. 7, pp. 3418–3431, 2018.
- [9] D. Hong and X. Zhu, “SULOra: Subspace unmixing with low-rank attribute embedding for hyperspectral data analysis,” *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 6, pp. 1351–1363, 2018.
- [10] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, “An augmented linear mixing model to address spectral variability for hyperspectral unmixing,” *IEEE Trans. on Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.
- [11] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, “Joint reconstruction and anomaly detection from compressive hyperspectral images using mahalanobis distance-regularized tensor RPCA,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, 2018.
- [12] X. Wu, D. Hong, P. Ghamisi, W. Li, and R. Tao, “MsRi-CCF: Multi-scale and rotation-insensitive convolutional channel features for geospatial object detection,” *Remote Sens.*, vol. 10, no. 12, pp. 1990, 2018.
- [13] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, “ORSIm Detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features,” *IEEE Trans. Geosci. Remote Sens.*, 2019.
- [14] L. Drumetz, M. Veganzones, S. Henrot, R. Phlypo, J. Chanussot, and C. Jutten, “Blind hyperspectral unmixing using an extended linear mixing model to address spectral variability,” *IEEE Trans. on Image Process.*, vol. 25, no. 8, pp. 3890–3905, 2016.
- [15] A. Marinoni, J. Plaza, A. Plaza, and P. Gamba, “Integrating multiple nonlinear estimators into hyperspectral unmixing,” in *Proc. WHISPERS*, 2014, pp. 1–4.
- [16] F. Palsson, J. Sigurdsson, J. R. Sveinsson, and M. O. Ulfarsson, “Neural network hyperspectral unmixing with spectral information divergence objective,” in *Proc. IGARSS*, 2017, pp. 755–758.
- [17] J. MP Nascimento and J. MB Dias, “Vertex component analysis: A fast algorithm to unmix hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, 2005.
- [18] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, “Learning a low-coherence dictionary to address spectral variability for hyperspectral unmixing,” in *Proc. ICIP*, 2017, pp. 235–239.
- [19] K. Segl, L. Guanter, C. Rogass, T. Kuester, S. Roessner, H. Kaufmann, B. Sang, V. Mogulsky, and S. Hofer, “EeteS-the EnMAP end-to-end simulation tool,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 522–530, 2012.
- [20] U. Heiden, W. Heldens, S. Roessner, K. Segl, T. Esch, and A. Mueller, “Urban structure type characterization using hyperspectral remote sensing and height information,” *Landsc. Urban Plan.*, vol. 105, no. 4, pp. 361–375, 2012.
- [21] D. Hong, W. Liu, J. Su, Z. Pan, and G. Wang, “A novel hierarchical approach for multispectral palmprint recognition,” *Neurocomputing*, vol. 151, pp. 511–521, 2015.
- [22] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, “Cospace: Common subspace learning from hyperspectral-multispectral correspondences,” *IEEE Trans. Geosci. Remote Sens.*, 2019.