HYPERSPECTRAL IMAGE UNMIXING USING AUTOENCODER CASCADE

Rui Guo, Wei Wang and Hairong Qi

University of Tennessee, Knoxville
Department of Electrical Engineering and Computer Science
Knoxville, TN 37996, USA

ABSTRACT

Hyperspectral image unmixing is the process of estimating pure source signals (endmemebers) and their proportions (abundances) from highly mixed spectroscopic images. Due to model inaccuracies and observation noise, unmixing has been a very challenging problem. In this paper, we exploit the potential of using autoencoder to tackle the unmixing challenges. Two important facts are considered in the algorithm: first, the observation noise in the hyperspectral image generally exists and largely affects the unmixing results; second, the mixing process contains sparsity priori which should be considered to assist the endmember extraction. The proposed autoencoder cascade concatenates a marginalized denoising autoencoder and a non-negative sparse autoencoder to solve the unmixing problem which implicitly denoises the observation data and employs the self-adaptive sparsity constraint. The algorithm is tested on a set of synthetic mixtures and a real hyperspectral image. The experimental results demonstrate the proposed algorithm outperforms several advanced unmixing approaches in highly noisy environment.

Index Terms— Hyperspectral image unmixing, Autoencoder cascade, Marginalized denoising auto-encoder, Non-negative sparse auto-encoder

1. INTRODUCTION

The hyperspectral imaging technique is widely utilized in modern earth environment monitoring, mineral exploration, urban planning and more other geo-scientific applications. Through the imaging spectrometers, the hyperspectral image captures electromagnetic reflectance in hundreds of wavelength bands with high spectral resolution. However, due to the low spatial resolution at a pixel position, the measured spectrum is usually a mixture of reflectance of several materials, hence the derivation of the individual constituent materials (endmembers) and their proportions (abundances) becomes the essential task, which is often referred to as spectral unmixing, in hyperspectral image analysis [1].

The most commonly model used for the mixture formation is a linear process with the mathematical expression as

$$X = AS + E \qquad (1)$$

where $X \in R^{l \times n}$ denotes the observation matrix. Each column of X represents the spectral measurement of a pixel with l bands. $A \in R^{l \times c}$ is the source matrix containing the signatures of c endmembers. $S \in R^{c \times n}$ is called the abundance matrix, where the row vector $s_i \in R^n$ represents the fractional values of endmember i. In spectral unmixing, S is subject to two physical constraints which are termed as the abundance non-negative constraint (ANC) and the abundance sum-to-one constraint (ASC), respectively [2]. E denotes the additive uncertainty and noise in the observed data. The goal of spectral unmixing is to blindly derivate A and S given the observation set X. This is often a difficult inverse problem considering the spectral signatures are highly correlated and badly contaminated by noise in practice [11].

To tackle the unmixing challenge, a number of algorithms have been proposed. Viewing it from the signal subspace identification perspective, the unmixing process is modeled as selecting a few spectral components residing in high correlated signal space from adjacent bands [3]. Projection techniques are adopted to enable such unsupervised subspace identification, e.g., Principal Component Analysis [4], Singular Value Decomposition [5] and Independent Component Analysis [6]. As a blind source separation method, nonnegative matrix factorization (NMF) [7] was proposed and intensively studied with various additional constraints. Recently, linear sparse regression (LSR) [8] is widely investigated in spectral unmixing. The hypothesis behind is that, rather than extracting the endmembers from observation data, they are compressively sensed from a library. Then the least square regression is solved with the sparsity constraint. However, these algorithms are suffering from noise polluted data, improper initialization and heavy dependence on availability of the suitable library, making them impractical in real application scenarios [2][3].

In this paper, we introduce an autoencoder cascade to unmix the hyperspectral image. The proposed autoencoder cascade integrates the marginalized denoising autoencoder (mDA) [9] at the top layer to denoise the data and non-negative sparse autoencoder (NNSAE) [10] in the unmixing layer together to estimate the endmembers. The sparsity and a non-negative properties associated with the autoencoder cascade enable the efficient encoding and gain an excellent perfor-

mance especially in highly noisy environments.

2. AUTOENCODER CASCADE – STRUCTURE AND ALGORITHM

The autoencoder cascade is a neural network consisting of multiple layers of autoencoders in which the output of the denoising layer is fed to the input of the successive unmixing layer. The hybrid structure is designed to reduce the noise initially to prevent the mixing matrix from being ill-conditioned [3]. Given the denoised data, the subsequent unmixing layer incorporates the non-negative and sparsity constraints to accomplish the matrix factorization task which implements the endmember extraction from hyperspectral mixtures. The diagram of the cascade is illustrated in Fig. 1.
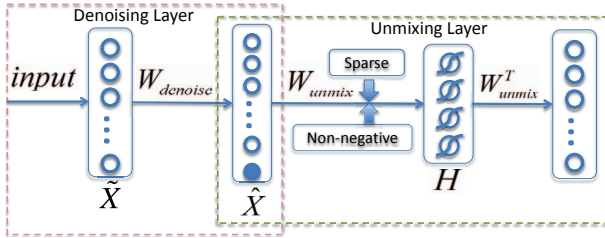

Fig. 1. The diagram of the autoencoder cascade.

2.1. Marginalized Denoising Auto-encoder

Unlike the traditional denoising autoencoder [12], mDA employs more efficient encoding process by marginalizing noise distribution and finding optimal weights for a linear transformation to reconstruct the original 'pure' signals. Suppose the columns x_i in matrix $X \in R^{l \times n}$ represent the 'clean' true spectral data. The observed noisy version of x_i is denoted as $\tilde{x}_i$. A linear transformation matrix $W : R^l \rightarrow R^l$ is introduced to recover x_is from the observation $\tilde{x}_i$s. Then, the denoising process is modeled as learning W by minimizing the squared loss,

$$\frac{1}{2n} \sum_{i=1}^{n} \|x_i - W\tilde{x}_i\|^2 \qquad (2)$$

The process is performed m-times across the data to lower the variance. Considering the entire dataset in matrix form, $X = [x_1, x_2, ...x_n] \in R^{l \times n}$, and its m-time repeated version as $\bar{X} = [X, X, ...X]$. The corrupted version of $\bar{X}$ as $\tilde{X}$. With such notation, the squared loss in Eq. (2) is converted to

$$\mathscr{L}_{sq}(W) = \frac{1}{2mn} tr[(\bar{X} - W\tilde{X})^T (\bar{X} - W\tilde{X})] \qquad (3)$$

The solution to Eq. (3) has the closed-form as

$$W = PQ^{-1} \quad with \quad Q = \tilde{X}\tilde{X}^T \quad and \quad P = \bar{X}\tilde{X}^T \quad (4)$$

which is the classic solution for traditional least square problem [13].

When $m \rightarrow \infty$, matrices P and Q converge to the expectation values following the law of large number where the mapping transformation W can be expressed as

$$W = \mathbb{E}[P] \cdot \mathbb{E}[Q]^{-1}. \qquad (5)$$

Once W is acquired, the denoising process can be finished by $\hat{X} = sigmoid(W * \tilde{X})$, where $\hat{X}$ is the denoised data. The $sigmoid$ function performed as an encoder to incorporate the non-linearity. The mathematic details of the solver to calculate the $\mathbb{E}[P]$ and $\mathbb{E}[Q]$ can be found in [9].

2.2. Non-negative Sparse Auto-encoder

The traditional autoencoder is a diabolo shaped neural network. It tries to learn the parameters that reconstruct an input vector by minimizing the cost function over the dataset,

$$\frac{1}{n} \sum_{i=1}^{n} \|g \circ f(x_i) - x_i\|^2 \qquad (6)$$

where $f(x) = \phi(W^T x)$ acts as an encoder, and $g(H) = WH$ acts as a decoder, with component-wise sigmoid activation functions deployed on the hidden layer. W is a matrix of weights shared between input-to-hidden and hidden-to-output layers. The spectral unmixing under study is modeled with autoencoder by taking the input as the observation spectra and setting the number of hidden neurons as the number of the endmembers. We refer to rows of the weight matrix W as basis for reconstruction in the decoding process. Thus, these bases form as the set of endmembers in unmixing, and the vectors in H served as abundances associated with each endmember extracted.

To consider the sparsity constraint in mixture, we modify the activation function as the augmented logistic function, which is in the form of

$$f(x) = \frac{1}{1 + e^{-(a_i g_i - b_i)}} \qquad (7)$$

where $g_i = W_{ij}^T x_i$. By adjusting the parameters a_i and b_i in the logistic function, it controls the information transformation between neurons. In the intrinsic plasticity mechanism [14], the Kullback-Leibler divergence of the neuron's output distribution with respect to a desired exponential output distribution is minimized. The sparsity measurement in term of mean activity level μ appears as a global parameter to control the learning process of a_i and b_i. The gradient rule to update a_i and b_i with learning rate η_{IP} is

$$\Delta b_i = \eta_{IP}(1 - (2 + \frac{1}{\mu})h_i + \frac{1}{\mu}h_i^2)$$

$$\Delta a_i = \eta_{IP}\frac{1}{a_i} + g_i \Delta b_i \qquad (8)$$

where h_i is the activation of the i_{th} neuron. Through the gradient learning, the lifetime sparseness of the neurons are accomplished. We use small learning rate $\eta_{IP} = 0.001$ and set $\mu = \frac{1}{c}$ for all the configurations.

To enforce the weight matrix to be non-negative, the online error correlation rule is applied

$$\Delta w_{ij} = \eta(x_i - \hat{x}_i)h_j + |\tilde{w_{ij}}| \qquad (9)$$

where $|\tilde{w_{ij}}|$ converts the negative elements in W to positive, and $\eta = 0.002$ is the learning rate.

At the end of each iteration, the activation measurement h_i will be normalized to satisfy the abundance sum-to-one constraint.

In our approach, since both the sparsity and non-negative constraints are naturally incorporated in a self-adaptive way, the encoding efficiency is largely improved.

3. EXPERIMENTS

We conduct two experiments on both the synthetic and real hyperspectral datasets to evaluate the proposed autoencoder cascade. The spectral angle distance (SAD) and abundance angle distance (AAD) are adopted as the comparison metrics.

3.1. Evaluation on Synthetic Data

The synthetic data are generated using the linear mixing [1] of spectral reflectance record with 188 bands from the US-GS digital spectral library [15]. To simulate the sensor noise and device error, the zero-mean Gaussian random noise is added to the mixture data. Defining the signal-noise-ratio as $SNR = 10\log_{10}(E[x^T x]/E[n^T n])$, we create the datasets by adding the noise varied from $30dB$ to $5dB$ as the settings used in [16].

We unmix the synthetic data by selecting four endmembers. The pure endmember signatures are absent from the data which means the unmixing processing is completely unsupervised.

In Fig. 2, we demonstrate the output of the denoising ($SNR = 20dB$). Clearly, the denoised data become more smooth and the outliers have been largely removed.

The comparison results with the popular unmixing approaches VCA-FCLS and GDME [16] are demonstrated in Table. 1. With the implicit denoising procedure and non-negative sparse autoencoder, our proposed autoencoder cascade has the best performance with average improvement by 45.38% on SAD and 42.10% on AAD over GDME. The extracted endmembers with $20dB$ data are demonstrated in Fig. 3.

3.2. Evaluation on Real Image Scene

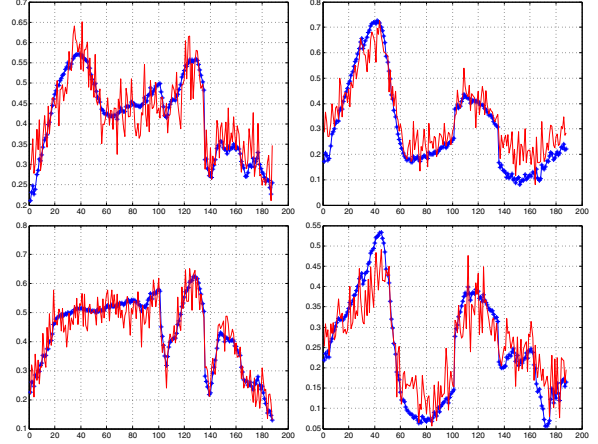The proposed autoencoder cascade is also tested on the subscene of AVIRIS data captured over Cuprite, Nevada [17]. We

Fig. 2. The denoising results. Red curves are noisy data, blue curves are denoised results. (SNR=20dB)

Table 1. SAD (upper) and AAD (lower) comparison at different noise levels using synthetic data. The auto-encoder without mDA is also compared as NNSAE [10].

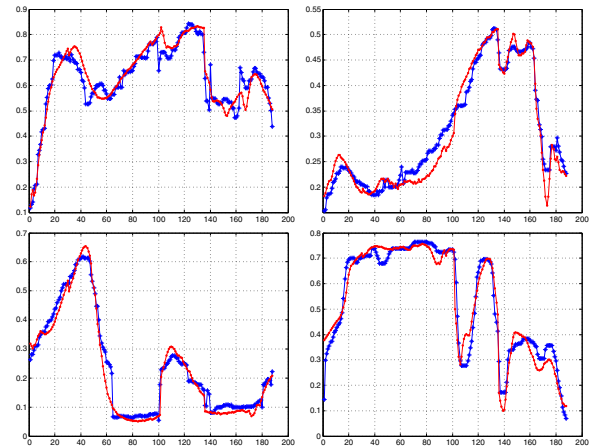| Method	5dB	10dB	15dB	20dB	30dB
VCA-FCLS	6.902	4.140	2.148	1.217	0.338
GDME	4.475	1.783	0.777	0.417	0.123
NNSAE	2.405	0.953	0.420	0.178	0.120
Proposed	**2.179**	**0.787**	**0.320**	**0.153**	**0.112**
VCA-FCLS	13.908	8.652	4.884	2.697	0.901
GDME	12.166	6.947	4.033	2.023	0.814
NNSAE	9.862	5.764	2.933	0.817	0.625
Proposed	**8.955**	**4.178**	**1.824**	**0.707**	**0.608**

Fig. 3. The demonstration of extracted endmembers. Red curves are ground-truth, and blue curves are extracted endmembers. (SNR=20dB)

assume the estimated endmembers $c = 9$ which is coincided with [1]. The abundance maps generated are illustrated in Fig. 4. Due to the lack of ground-truth, we only compare them

with the published geological papers using the same data [1]. Our estimations present more condensation appearance rather than scattered distributions for the materials of the same type.
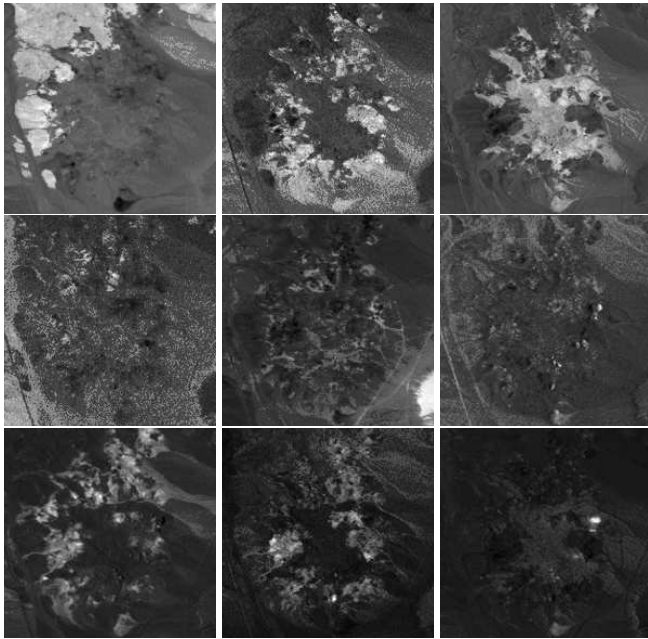

Fig. 4. Demonstration of the abundance maps extracted from Cuprite image. We use $c = 9$ endmembers for estimation. The resulting abundance maps contain condensation property for different materials.

4. CONCLUSION

This work represents the first attempt that uses deep-learning related approaches for spectral unmixing applications. We proposed an autoencoder cascade which concatenates a marginalized denoising autoencoder and a non-negative sparse autoencoder for hyperspectral image unmixing. It thus processes the advanced denoising capability and the nonnegative sparse encoding capacity optimized by the intrinsic self-adaptation rule. With such properties, the proposed autoencoder cascade outperformed some popular unmixing approaches on the synthetic hyperspectral image in highly noisy environment and received promising performance on real scene data.

5. REFERENCES

[1] Lidan Miao and Hairong Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization.," *IEEE T. Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 765–777, 2007.

[2] Lidan Miao and Hairong Qi, "A constrained non-negative matrix factorization approach to unmix highly mixed hyperspectral data.," in *ICIP (2)*. 2007, pp. 185–188, IEEE.

[3] Jos M. Bioucas-dias, Antonio Plaza, Nicolas Dobigeon, Mario Parente, Qian Du, Paul Gader, and Jocelyn Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens*, pp. 354–379, 2012.

[4] I. T. Jolliffe, *Principal Component Analysis*, Springer-Verlag, 1986.

[5] Glenn Healey and David Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions.," *IEEE T. Geoscience and Remote Sensing*, vol. 37, no. 6, pp. 2706–2717, 1999.

[6] Jessica D. Bayliss, J. Anthony Gualtieri, and Robert F. Cromp, "Analyzing hyperspectral data with independent component analysis," 1998, vol. 3240, pp. 133–143.

[7] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, Oct. 1999.

[8] Marian daniel Iordache, Jos M. Bioucas-dias, and Antonio Plaza, "Sparse unmixing of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 2014 – 2039, 2011.

[9] Minmin Chen, Zhixiang Eddie Xu, Kilian Q. Weinberger, and Fei Sha, "Marginalized denoising autoencoders for domain adaptation.," in *ICML*, 2012.

[10] Andre Lemme, Ren Felix Reinhart, and Jochen Jakob Steil, "Online learning and generalization of parts-based image representations by non-negative sparse autoencoders.," *Neural Networks*, vol. 33, pp. 194–203, 2012.

[11] W. Wang, S. Li, H. Qi, Bulent Ayhan, Chiman Kwan, and Steven Vance, "Revisiting the preprocessing procedures for elemental concentration estimation based on chemcam libs on mars rover," in *IEEE Workshop on Hyperspectral Image and Signal Processing*, 2014.

[12] Kyunghyun Cho, "Simple sparsification improves sparse denoising autoencoders in denoising highly corrupted images," in *Proceedings of ICML*, May 2013, vol. 28, pp. 432–440.

[13] Christopher M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, Springer-Verlag New York, Inc., 2006.

[14] Jochen Triesch, "Synergies between intrinsic and synaptic plasticity in individual model neurons.," in *NIPS*, 2004.

[15] R Clark, G Swayze, A Gallagher, T King, and W Calvin, "The u.s. geological survey digital spectral library: Version 1: 0.2 to 3.0 am.," *USGS Open File Report*, pp. 93–592, 1993.

[16] Lidan Miao, Hairong Qi, and H. Szu, "A maximum entropy approach to unsupervised mixed-pixel decomposition.," *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 1008–1021, 2007.

[17] W. Wang and H. Qi, "Unsupervised non-linear unmixing of hyperspectral image based on sparsity constrained probabilistic latent semantic analysis," in *IEEE Workshop on Hyperspectral Image and Signal Processing*, 2013.