Spectral Unmixing And Sub-pixel Classification: Analysis of learning strategies

P.V. Arun K.M. Buddhiraju

Centre of Studies in Resources Engineering
Indian Institute of Technology Bombay
Mumbai, India

***Abstract*— Sub-pixel details in the hyperspectral images are generally ignored by the conventional classifiers. However, some recent approaches use this information to generate fine resolution land cover maps from images having coarse spatial resolution. Two main aspects in this regard are: 1) estimation of fractional abundances of the reference signatures at each pixel (spectral un-mixing); and 2) prediction of class distributions at sub-pixel scale (sub-pixel classification). This study proposes some spectral unmixing as well as sub-pixel mapping techniques that take in to account certain constraints which are usually ignored by the conventional approaches. In the context of spectral unmixing methods, our main contribution is the analysis of auto-encoders when compared with ELM, STM and SVM. In case of sub-pixel mapping methods, our study may be summarized as the modelling deep auto-encoders for predicting the spatial distributions at target scale. Also, we have compared the effectiveness of Auto-Encoders and their convolutional counterparts in learning the coarse image features. Among the proposed unmixing approaches, autoencoder approach has given better results when compared to that of SVM and STM. The deep learning based sub-pixel mapping approaches have also produced good results, even for complex scenes. The sensitivities of all these techniques towards various tunable parameters are also analyzed.**

Keywords—sub-pixel mapping, spectral unmixing, autoencoder, convolutional neural network

I. INTRODUCTION

Hyperspectral imaging sensors capture electromagnetic reflectance in hundreds of wavelengths, resulting in fine spectral but coarse spatial resolution. The standard fusion techniques, used for multispectral and panchromatic images, cannot be directly extended for hyperspectral datasets. Sub-pixel classification approaches, which project class-labels to finer-sized pixels, are found to be more effective in this context. These techniques utilize the pixel-wise endmember fractions to predict spatial distribution of classes. Although endmember extraction and abundance estimation approaches are widely discussed in the literature, their accuracy and reliability are dependent on availability of suitable libraries, initialization conditions, and endmember variability. These issues can be addressed by using contextual information derived from coarse images.

Most sub-pixel classification techniques [1-5] focus on optimizing class-wise spatial dependency without considering the scene specific variations [6]. Recently, some geostatistical interpolation techniques have been employed to estimate the required target distribution [6-8]. However, the difficulty in formulating an effective structural model at finer scale limits the practical implementation of these works [6]. Some supervised approaches have been proposed in this regard, but they are generally suitable for ideal conditions where representative regions can be assumed to characterize the spectral variability of the entire image [9]. In short, the prediction of sub-pixel spatial distributions based on the proximity consideration is inherently vague [7-9]. Hence, instead of only generalizing the local spatial dependencies, it is better to refine them using coarse image features.

Among the various state of the art strategies, deep learning methods have been shown to be considerable successful in modelling image features. However, these approaches have rarely been explored for spectral unmixing or sub-pixel mapping. Recently, Guo et al [10] proposed an Auto-Encoder (AE) cascade, which implicitly denoises the observed data and employs self-adaptive sparsity constraint to solve the unmixing problem. Arun et al [11] suggested the use of auto-encoders for sub-pixel mapping. Similarly, convolutional neural networks have been used for super resolution [12-13]. In the present contribution, we investigate deep learning models for improving spectral unmixing process. We also analyse the effectiveness of these architectures in learning the spatial structure models at coarser scale for refining the sub-pixel classification.

II. PROPOSED APPROACH

In this section, we reformulate some of the recently suggested deep learning architectures for spectral unmixing and sub-pixel classification.

A. Autoencoder based Unmixing

The The basic auto-encoders [10-11] are employed to learn parameters that can reconstruct an input vector with minimal error. In other words, the encoder output $y=f_\theta(x)$ is decoded back to a reconstructed vector $z = g_{\theta'}(f_\theta(x^{(i)})$ such that the parameters θ and θ' are computed by minimizing L as:

$$\underset{\theta,\theta'}{\operatorname{argmin}} \frac{1}{n}\sum_{i=1}^{n} L(x^{(i)}, g_{\theta'}(f_\theta(x^{(i)})))$$

where L is a loss function such as the traditional squared error $L(x,z) = \|x - z\|^2$ or reconstruction cross entropy

$$L(x,z) = -\sum_{k=1}^{d} x_k log z_k + (1 - x_k)\log(1 - z_k)$$

However, de-noising Auto Encoders, a variation of the basic AEs, learn to reconstruct the clean repaired versions of inputs from their corrupted versions. This is achieved by corrupting the initial inputs by means of a stochastic mapping. Geometrically, DAE can be seen as a way to define and learn a manifold. AEs or DAEs are stacked together to form deep networks (Stacked Auto-encoders and Stacked De-noising Auto-encoders respectively) which are trained using variations of gradient descent approaches.

As discussed in [10], auto-encoders can be modelled for spectral unmixing by taking the input as the observation spectra and setting the number of hidden neurons as the number of endmembers. However, in our study, the de-noising cascade is modified to capture the variation of given input from randomly selected samples of each end members. Finally, the decoding weight matrix corresponds to the abundance fractions, while the encoding weights serve as the basis for endmembers. The activation function is modified to enforce the sparsity constraint as:

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

where $g_i=W_{ij}{}^T x_i$. Through gradient learning on parameters (a_i and b_i), the lifetime sparseness of the neurons is accomplished.

Although, the model adopted above can solve the unmixing problem to a certain extent, the accuracy may be improved by using the labelled data as well as the spatial proximity information. Moreover, as discussed in our earlier works [14], a classification perspective to the unmixing problem can resolve the issues of end member spectral variability, noise effects etc. Hence, we propose the use of adversarial auto encoders [15-18] to refine the conventional models. Unlike the previous AE based approaches, here a probabilistic auto-encoder performs variational inference by matching the aggregated posterior of the hidden code vector with an arbitrary prior distribution. Such models cluster the data in accordance to the distributions learned from the samples. The inverse distance of each sample from the cluster center is considered as the corresponding fractional abundance.

B. Autoencoder based Sub-pixel Mapping

The fractional abundances estimated using unmixing approaches are filtered based on the coarse image gradients. For each class, ranks are assigned to all sub-pixel positions based on their proximity to the corresponding classes in the neighborhood. Details of the rank image computation can be found in [11]. However, it may be noted that here we consider both spectral as well as spatial affinities in the process. Thus the initial rank image ($\mathbf{R_k}$) for the $\mathbf{k^{th}}$ class will have weights ($\mathbf{W_{ik}}$) at each sub-pixel position (**i**) as:

$$\boldsymbol{W_{ik}} = \sum_{j=1}^{M} \boldsymbol{F_k(j)} e^{-(D_{ij} + G_{ij})}$$

where **M** is the number of pixels in the neighborhood, $\mathbf{G_{ij}}$ is the feature space distance between the value of the $\mathbf{j^{th}}$ coarse pixel and the coarse pixel corresponding to the $\mathbf{i^{th}}$ sub-pixel, $\mathbf{D_{ij}}$ is the distance between the centroid of the $\mathbf{j^{th}}$ coarse pixel and the

subpixel position **(i),** and $\mathbf{F_K(j)}$ is the fractional abundance at $\mathbf{j^{th}}$ coarse pixel. The subpixel positons of each coarse pixel are sorted according to their weights and only the top ranked $\mathbf{Q_{jk}}$ positons are retained. Here, $\mathbf{Q_{jk}}$ [11] is estimated as:

$$\mathbf{Q_{jk} = (Z)^2 \times F_K(j)}$$

where **Z** is the scale to which the image need to be super-resolved and $\mathbf{F_K(j)}$ is the fractional abundance of k^{th} class at $\mathbf{j^{th}}$ coarse pixel.

Further, a stacked auto-encoder (SAE) is trained to learn the resampled version of the coarse image. Gradient descent approach discussed in [18-20] is used for fine tuning the network. The constrained rank image ($\mathbf{R_{ck}}$) is then fed to the SAE whose weights represent the coarse image features. The scaled fractional maps ($\mathbf{I_{CK}}$) thus obtained are refined by setting all the top ranked $\mathbf{Q_{jk}}$ sub-pixel positions, at each coarse pixel, to one and rest to zero. The ambiguous pixels of I_{CK} as well as those whose class compatibilities [11] are less than a given threshold are reassigned to the most frequently occurring classes in their neighborhood.

Instead of using basic auto-encoders described above, we also experimented with Deep Belief Networks (Stacked Restricted Boltzmann Machines) [19] for learning image features. It may be noted that the probabilistic semantics of Restricted Boltzmann Machines (RBMs) can be defined as:

$$p(v,h) = \frac{1}{z}\exp(-E(v,h))$$

where v and h denote the state of visible and hidden layers and E is the energy function. The Contrastive Divergence algorithm [20] is applied for layer wise pre-training of the network, which is further fine-tuned using the Back-Propagation algorithm to learn the resampled coarse image.

Although, the SAEs and DBNs capture coarse image features, networks that can learn the two-dimensional spatial distributions are more effective for image processing tasks. In a recent study [11], we explored the utility of convolutional auto-encoders for the purpose. However, defining the deconvolution and un-pooling layers in this context is not trivial, and performance of the approach depends on the same. Hence, we investigate the use of supervised approaches such as Convolutional Neural Networks (CNNs) [12] and Convolutional Deep Belief Networks (CDBNs) [20] for the purpose. Here, the coarse images are downscaled at multiple coarser scales to yield low and high resolution pairs for training. Limited set of training samples have also been used to refine the approach.

III. RESULTS

The proposed unmixing algorithms are compared with some AE based as well as other state of the art approaches [13-14] using real and simulated datasets. The spectral angle distance (SAD), and abundance estimation error (AE) are adopted as the comparison metrics. Lower values of AE as well as SAD indicate better unmixing.

A. Analysis of Proposed Unmixing Algorithms

Experiments with simulated datasets

Four synthetic hyperspectral scenes are generated using the linear mixing of spectral signatures with 188 bands from the USGS digital spectral library. Sensor noise is simulated using zero-mean Gaussian random noise at different levels of signal-to-noise ratio (SNR =30, 40, and 50 dB). The accuracy analysis of the unmixing techniques over these four datasets is summarized in Table 1(values in bracket denote AE and those outside SAD).

As evident from these results, the adversarial AE is better suited for unmixing when compared to the conventional AE. Also, the proposed refinements improve the accuracy of the state of the art AE based approaches. However, as the availability of training data increases SVM based approaches outperform the proposed methods. Even using 10-15% training samples, the performance of supervised approaches (SVM and STM) are comparable to the proposed ones. Also, at lower noise levels SVM as well as STM give superior performances. Although the proposed adversarial AE based approach is more resilient to noise as compared to SVM, STM performs still better.

The results of the proposed methods (in terms of AE error) in unsupervised frameworks are summarized in Table 1. It may be noted that the method discussed in [10] gives results comparable to FCM based approach; whereas the proposed approach performs even better. The performance of FCM is sensitive to the noisy data, whereas auto-encoder based approaches seem to be resilient.

The proposed methods give stable results irrespective of the size of the test images (image1: 200 X 200, image2: 600 X 600, image3: 800 X 800, image4: 1200 X 1200) as well as the number of endmembers (image1: 4, image2: 10, image3: 15, image 4: 23). In order to analyze the sensitiveness of these methods with respect to the complexity of mixing, the size and number of uniform patches in the synthetic images are changed before applying smoothing. From the results (shown in figure 1), it can be observed that the proposed AE based approaches are less sensitive to the complexity of mixing when compared to [10].

Experiments with real-world datasets

It is important to analyze the sensitivity of the proposed approaches towards real-time atmospheric effects as well as other noises. Hence, the performances of these methods are evaluated over real world airborne/satellite datasets. Here also, SAD and AE measures are used for accuracy analyses. Since the ground truth abundance fraction is not available for validation, some representative regions are selected to identify reliable endmembers. Manual unmixing using constrained least square is then adopted to obtain the test data. Comparative analysis of different methods over standard datasets is presented in Table 2.

Table 1. Analysis of supervised unmixing techniques

S.No	**SNR**	**SVM**	**STM**	**[10]**	**AE.**	**Adv. AE**
1	30	2.24 (3.01)	1.93 (2.19)	6.26 (4.63)	4.91 (3.48)	2.11 (2.34)
	40	1.79 (1.68)	1.06 (2. 01)	5.42 (3.87)	3.54 (3.24)	1.25 (1.93)
	50	0.47 (1.42)	0.37 (0.91)	3.60 (3.16)	1.30 (2.89)	0.85 (1.60)
2	30	4.20 (3.88)	3.14 (2.97)	9.46 (4.72)	5.73 (3.92)	4.10 (3.65)
	40	3.49 (3.36)	2.81 (2.52)	7.58 (4.16)	4.77 (3.79)	2.94 (3.19)
	50	1.08 (2.14)	0.74 (1.97)	4.53 (4.02)	2.24 (3.51)	1.27 (2.21)
3	30	5.39 (3.64)	5.28 (3.05)	8.65 (5.37)	8.12 (4.28)	5.43 (3.29)
	40	4.08 (3.15)	3.32 (2.73)	5.44 (4.50)	4.35 (4.06)	3.81 (3.09)
	50	3.40 (2.76)	3.05 (2.14)	5.17 (4.23)	3.82 (3.94)	3.58 (2.94)
4	30	5.07 (4.33)	4.59 (3.70)	7.61 (5.85)	5.60 (5.18)	5.85 (4.06)
	40	4.91 (3.27)	4.52 (2.74)	6.06 (4.27)	5.43 (4.81)	5.13 (3.94)
	50	4.14 (2.89)	3.57 (2.42)	5.50 (3.89)	4.85 (4.66)	4.29 (3.21)

Table 2. Comparative analysis of proposed methods

Dataset	Method	AE error	SAD
Indian pines	SVM	2.17	3.39
	STM	1.34	2.47
	[10]	3.08	5.66
	AE	2.43	3.58
	Adversarial AE	1.96	2.64
Pavia	SVM	2.54	3.90
	STM	1.82	2.21
	[10]	4.18	5.17
	AE	3.25	4.12
	Adversarial AE	1.98	2.83
KSC	SVM	3.30	4.08
	STM	2.61	3.79
	[10]	4.28	6.55
	AE	3.57	4.30
	Adversarial AE	2.23	3.96
Salinas	SVM	1.89	2.87
	STM	0.94	1.21
	[10]	2.72	3.29
	AE	1.78	2.91
	Adversarial AE	1.44	1.40
Botswana	SVM	4.15	5.61
	STM	3.67	3.52
	[10]	5.28	8.03
	AE	4.26	5.79
	Adversarial AE	3.40	4.56

From the table, it is clear that the proposed methods give stable results that are in agreement with their performances over synthetic datasets. Visual illustration of the results on the Salinas dataset is presented in Figure 1 and the corresponding end members are presented in Figure 2. The execution times (in secs) of various methods have been measured and are summarized in Table 3. As evident, proposed Auto-encoder based approaches are computationally comparable with STM but are better than SVM as well as [10].

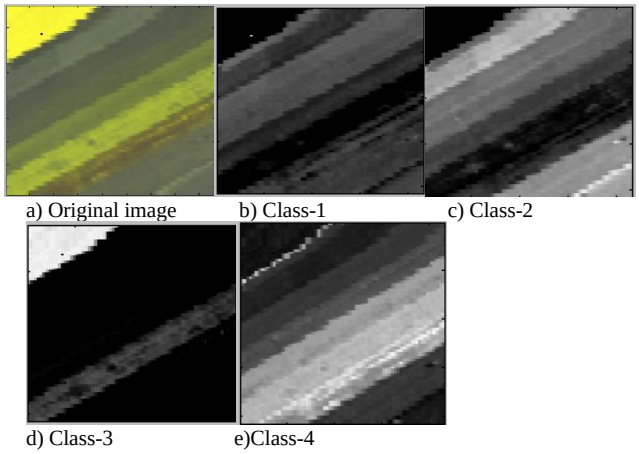


Figure 1. Results of the proposed AE method on Salinas dataset

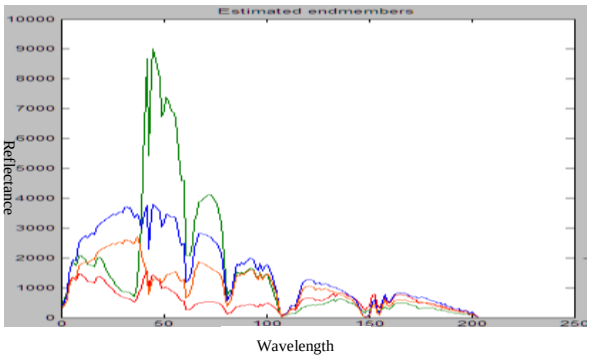


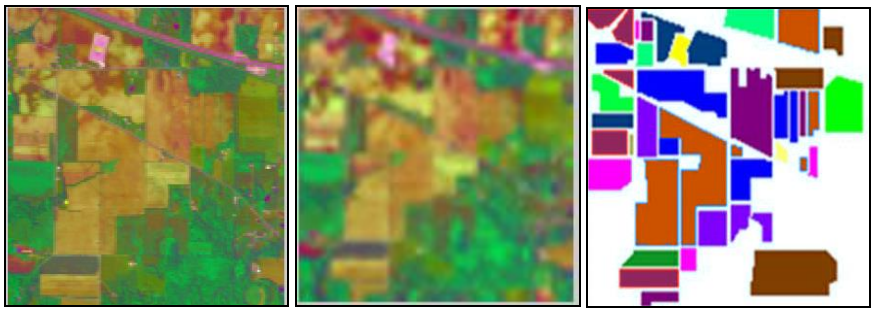
Figure 2. Endmembers derived from Salina dataset

Table 3. Computational analysis of proposed methods

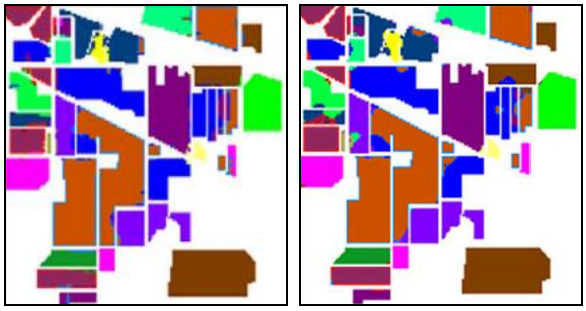
Dataset	Method	Computation time (secs)
Cuprite	SVM	94
	STM	88
	[10]	125
	AE	104
	Adversarial AE	85
Indian Pines	SVM	75
	STM	67
	[10]	108
	AE	95
	Adversarial AE	62
Salinas	SVM	46
	STM	40
	[10]	63
	AE	54
	Adversarial AE	49

B. Analysis of Proposed Sub-pixel Mapping Algorithms

To qualitatively analyze the effectiveness of the proposed deep learning strategies on real datasets, the outputs for down sampled input images are compared with the corresponding ground truth. The close similarity between the two indicates better performance. The Kappa statistics [2] and overall accuracy [3-4] measures are used to quantitatively evaluate these approaches, where higher values of both indicate better classification. The results of proposed approaches are illustrated in Figure 3 and Table 4.



a) Indian Pine image b) Reference image c) Down sampled image



d) SAE based e) CAE based

Figure 3. Results of prop. methods on Indian Pines dataset

As evident from Table 4, the use of Convolutional Auto Encoders yields the best results, followed by DBNs and SAEs. This is due to the fact that two-dimensional learning networks can better represent coarse image features than their one-dimensional counterparts. It may be noted that the proposed improvements on SAEs and DBNs lead to better performances as compared to the conventional AE based approaches; in most cases DBNs perform better than SAEs. On the contrary, CNNs and CDBNs give poor performance in most cases which may be attributed to the limited use of training data. The class wise performance of both methods on Pavia dataset, summarized in Table 5, reveals that these approaches are well suited for smooth varying classes. For instance, these methods give better results for certain classes such as brick, bare soil, and asphalt compared to heterogeneous classes such as Trees, and Tiles. This is probably because the training samples used for the latter are inadequate for capturing the spectral variability of heterogeneous classes.

It can be concluded that CNNs and CDBNs are not optimal choices for sub-pixel classification if the training samples are limited. The execution times (in secs) of various methods have been measured and are summarized in Table 6, which indicate that convolutional approaches are computationally simpler when compared to their conventional counterparts. Further, the proposed methods except convolutional auto-encoder are more complex than the STM based approach [14].

Table 4. Comparative analysis over standard datasets

Method	Dataset	Accuracy	Kappa
SAE	Indian Pines	86.19	0.84
	Pavia	83.21	0.83
	Salinas	88.92	0.87
	KSC	83.94	0.83
	Botswana	78.18	0.76
DBN	Indian Pines	88.27	0.85
	Pavia	84.19	0.82
	Salinas	89.56	0.90
	KSC	85.87	0.82
	Botswana	77.62	0.71
Convolutional AE	Indian Pines	94.95	0.92
	Pavia	84.31	0.86
	Salinas	95.59	0.93
	KSC	88.32	0.86
	Botswana	84.49	0.82
Convolutional Neural Network	Indian Pines	80.76	0.79
	Pavia	78.32	0.76
	Salinas	82.21	0.81
	KSC	75.38	0.75
	Botswana	70.29	0.68
Convolutional DBN	Indian Pines	84.92	0.85
	Pavia	81.70	0.80
	Salinas	86.15	0.84
	KSC	82.44	0.81
	Botswana	71.63	0.72

Table 5. Class-wise accuracy analysis

S. No	Class	CNN	DBN	CAE
1	Water	84.13	87.64	90.59
2	Trees	76.48	77.92	84.05
3	Asphalt	78.91	84.27	87.38
4	Bricks	71.13	76.65	85.49
5	Bitumen	69.47	73.71	80.67
6	Tiles	74.39	78.40	83.76
7	Shadows	68.54	75.27	77.04
8	Meadows	78.09	78.59	82.93
9	Bare Soil	83.73	85.63	88.39

Table 6. Summary of computational expenses

Dataset	Method	Computation time (in secs)
Cuprite	SAE	132
	DBN	124
	CAE	87
	CNN	109
	CDBN	95
Indian Pines	SAE	84
	DBN	76
	CAE	58
	CNN	71
	CDBN	64
Salinas	SAE	53
	DBN	59
	CAE	43
	CNN	61
	CDBN	53

The main factor that affects the performance of any deep learning implementation is the number of layers. In this paper, we also investigate the factors that determine the optimal number of layers required for the deep network. The result of these analyses, as presented in Figure 4, reveals that the required optimal depth of the deep network and will vary from the dataset to dataset. It may also be noted that the shallow layers generally give less accuracy, and the accuracy deteriorate as the number of layers cross beyond the optimal value. To investigate further, the accuracy of the deep network with respect to different band combinations has also been tested. The results of the same for the Indian pines dataset is summarized in Figure 5. It is worth noting that the peak performance corresponds to the optimal subset of the bands, but further addition/ reduction of bands deteriorates the performance. This trend continues to be independent of the number of layers. Further analyses also revealed that the number of layers required to get the maximum performance also increases with the complexity of the spectral mixing and the number of end members

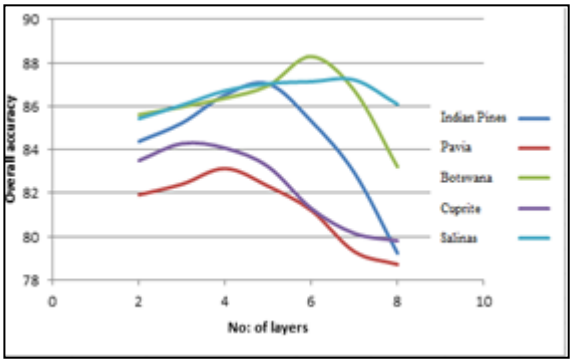


Figure 4. Analyses of deep learning framework

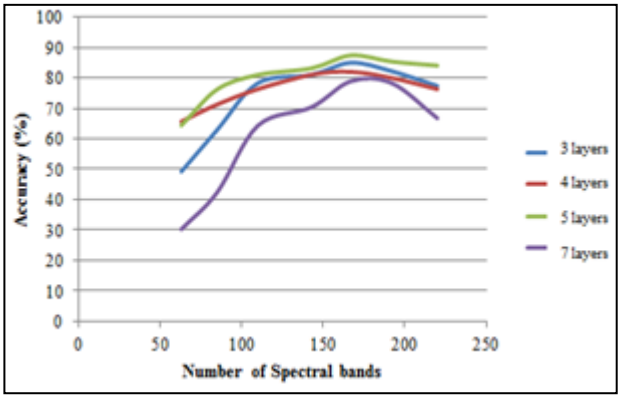


Figure 5. Analyses of learning over band combination

IV. CONCLUSION

This study investigates the use of deep learning architectures in predicting the distribution of classes at subpixel level. Our results show that adversarial AE auto-encoders are effective in estimating the end member abundances when compared to the basic stacked auto encoders. Proposed approaches give optimal performance in par with the supervised methods and even outperform the latter in cases of limited training samples. Among the various deep learning models, the convolutional autoencoders have been found to be better suitable for sub-pixel classification. It has been found that the supervised approaches such as CNN and CDBN are sensitive towards the availability of training samples as well as the homogeneity of classes. Also, the optimal depth of the learning network is found to be dependent on the number of end members, scene complexity and band combinations.

REFERENCES

[1] Q. Nguyen, P. Atkinson and H. Lewis, "Super-resolution mapping using Hopfield Neural Network with panchromatic imagery", International Journal of Remote Sensing, vol. 32, no. 21, pp. 6149-6176, 2011.
[2] X. Tong, X. Xu, A. Plaza, H. Xie, H. Pan, W. Cao and D. Lv, "A New Genetic Method for Sub-Pixel Mapping Using Hyperspectral Images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, accepted for publication, 2015.
[3] F. Ratle, G. Camps-Valls and J. Weston, "Semisupervised Neural Networks for Efficient Hyperspectral Image Classification", IEEE Trans. Geosci. Remote Sensing, vol. 48, no. 5, pp. 2271-2282, 2010.
[4] J. Ardila, V. Tolpekin, W. Bijker and A. Stein, "Markov-random-field-based super-resolution mapping for identification of urban trees in VHR images", ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 6, pp. 762-775, 2011.
[5] X. Xu, Y. Zhong and L. Zhang, "A sub-pixel mapping method based on an attraction model for multiple shifted remotely sensed images", Neurocomputing, vol. 134, pp. 79-91, 2014.
[6] Q. Wang, P. Atkinson and Wenzhong Shi, "Indicator Cokriging-Based Subpixel Mapping Without Prior Spatial Structure Information", IEEE Trans. Geosci. Remote Sensing, vol. 53, no. 1, pp. 309-323, 2015.
[7] H. Jin, G. Mountrakis and P. Li, "A super-resolution mapping method using local indicator variograms", International Journal of Remote Sensing, vol. 33, no. 24, pp. 7747-7773, 2012.
[8] A. Boucher and P. Kyriakidis, "Integrating Fine Scale Information in Super-resolution Land-cover Mapping", Photogrammetric Engineering & Remote Sensing, vol. 73, no. 8, pp. 913-921, 2007.
[9] Q. Wang, Wenzhong Shi and Liguo Wang, "Indicator Co-kriging based Subpixel Land Cover Mapping with Shifted Images", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 1, pp. 327-339, 2014.
[10] R. Guo and W. Wang and H. Qi, "Hyperspectral Image Unmixing Using Autoencoder Cascade", 7th Workshop onHyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp.1-4, 2015.
[11] P.V. Arun and B. K. Mohan, "A Deep learning based spatial dependency modelling framework for super resolution", IGARSSS 2016, accepted
[12] H. Goh, N. Thome, M. Cord and J. Lim, "Learning Deep Hierarchical Visual Feature Coding", IEEE Trans. Neural Netw. Learning Syst., vol. 25, no. 12, pp. 2212-2225, 2014.
[13] C. Dong, C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1-1, 2015.
[14] P.V. Arun and B. K. Mohan, "Classification and clustering perspective towards spectral unmixing", IGARSSS 2016, accepted.
[15] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, Pierre-Antoine Manzagol, "Extracting and composing robust features with denoising autoencoders", Proceedings of the 25th international conference on Machine learning, p.1096-1103, July 05-09, 2008, Helsinki, Finland.
[16] Kingma, Diederik P, Mohamed, Shakir, Rezende, Danilo Jimenez, and Welling, "Max. Semi-supervised learning with deep generative models", Advances in Neural Information Processing Systems, pp. 3581–3589, 2014.
[17] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio and P.-A. Manzagol, "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion", J. Mach. Learn. Res., vol. 11, no. 11, pp. 3371-3408, 2010.
[18] G. Dahl, Dong Yu, Li Deng and A. Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE Transactions on Audio, Speech, and Language Processing, vol. 20, no. 1, pp. 30-42, 2012.
[19] Y. Chen, X. Zhao and X. Jia, "Spectral–Spatial Classification of Hyperspectral Data Based on Deep Belief Network," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2381-2392, June 2015.
[20] H. Lee, R. Grosse, R. Ranganath and A. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks", Communications of the ACM, vol. 54, no. 10, p. 95, 2011.