Hyperspectral Image Classification with Convolutional Neural Networks

Viktor Slavkovikj, Steven Verstockt, Wesley De Neve, Sofie Van Hoecke, and Rik Van de Walle

In: Proceedings of the 23rd Annual ACM Conference on Multimedia Conference, 1159–1162, 2015.

http://doi.acm.org/10.1145/2733373.2806306

Hyperspectral Image Classification with Convolutional Neural Networks

Viktor Slavkovikj[1], Steven Verstockt[1], Wesley De Neve[1,2],
Sofie Van Hoecke[1], and Rik Van de Walle[1]

[1]Multimedia Lab, Department of Electronics and Information Systems, Ghent University-iMinds
B-9050 Ledeberg-Ghent, Belgium

[2]Image and Video Systems Lab, Korea Advanced Institute of Science and Technology (KAIST)
Yuseong-gu, Daejeon, 305-732, Republic of Korea

{viktor.slavkovikj, steven.verstockt, wesley.deneve,
sofie.vanhoecke, rik.vandewalle}@ugent.be

ABSTRACT

Hyperspectral image (HSI) classification is one of the most widely used methods for scene analysis from hyperspectral imagery. In the past, many different engineered features have been proposed for the HSI classification problem. In this paper, however, we propose a feature learning approach for hyperspectral image classification based on convolutional neural networks (CNNs). The proposed CNN model is able to learn structured features, roughly resembling different spectral band-pass filters, directly from the hyperspectral input data. Our experimental results, conducted on a commonly-used remote sensing hyperspectral dataset, show that the proposed method provides classification results that are among the state-of-the-art, without using any prior knowledge or engineered features.

Categories and Subject Descriptors

I.5.1 [**Pattern Recognition**]: Models—*neural nets*; I.5.4 [**Pattern Recognition**]: Applications—*computer vision*; I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis

Keywords

Classification, convolutional neural networks, deep learning, hyperspectral imaging

1. INTRODUCTION

Recent developments of imaging spectroscopy sensors have enabled acquisition of hyperspectral images with a high spatial resolution, a characteristic which was previously exclusive to standard electrooptical systems. However, unlike the standard color pictures, images acquired with hyperspectral sensors contain much higher spectral resolution. This is advantageous for image analysis, because each hyperspectral pixel comprises of a large number (in the order of hundreds) of measurements of the electromagnetic spectrum and carries more information as compared to color pixels, which provide data only from the visible range of the spectrum. As a result, hyperspectral image analysis has found numerous biomedical, forensic, and remote sensing applications [18, 5, 21].

One of the principal techniques in hyperspectral image analysis is image classification, where a label is assigned to each pixel based on its characteristics. Inference of class labels from hyperspectral data is challenging, however, since classification methods are affected by the curse of dimensionality (i.e., the Hughes effect[12]). That is, the classification accuracy is poor as the number of training samples required to populate the high-dimensional spectral space is limited. Therefore, many different feature extraction methods (see Section 2) have been proposed to tackle the classification problem in hyperspectral images. The goal of feature extraction is to reduce the dimensionality of the hyperspectral data while preserving as much of the discriminative information as possible, so that in a later stage a classifier can be trained on the extracted features. Since it is difficult to discern potentially relevant features from hyperspectral data, we approach hyperspectral image classification as an end-to-end learning [15, 25] task, where the assignment of class labels from hyperspectral input pixels is a single stage learning process, in which the intermediate feature representations are also learned. The contributions of our paper are twofold:

- We propose convolutional neural networks for hyperspectral image classification.
- We investigate hyperspectral data augmentation as a way of mitigating the problem of limited training samples in hyperspectral image classification.

The remainder of this paper is organized as follows. We discuss related work in hyperspectral image classification in Section 2. In Section 3, we present the architecture of the convolutional neural network that was used as basis for our

MM'15, October 26–30, 2015, Brisbane, Australia.
© 2015 ACM. ISBN 978-1-4503-3459-4/15/10 ...$15.00.
DOI: http://dx.doi.org/10.1145/2733373.2806306.

experiments. Section 4 describes hyperspectral data augmentation for alleviating the limited training samples problem. In Section 5, we report classification results obtained by the proposed method on a hyperspectral image dataset. Section 6 concludes the paper.

2. RELATED WORK

In the past, many different feature extraction and classification methods have been proposed for hyperspectral images. Some of the well-established feature extraction approaches are based on dimensionality reduction methods, such as principal component analysis (PCA) [11], or independent component analysis (ICA) [24]. These methods are aimed at projecting the hyperspectral data to a subspace, in which class separation is performed more effectively. Similarly, to be able to calculate coordinates of data in a lower-dimensional space, manifold learning methods [8, 7] try to estimate the intrinsic geometry of the manifold embedded in the high-dimensional hyperspectral data space. Discriminant analysis methods [1, 3] have been used to learn a projection matrix in order to maximize a separability criterion of the projected data. Morphological features [2], on the other hand, were introduced to take advantage of structural information present in the images. They have been successfully combined with support vector machines [6], which are known for their good generalization properties for high-dimensional data with lower effective dimensionality [19].

Recently, statistical learning models, such as neural networks, have also been investigated for the purpose of hyperspectral image classification. For instance, Li et al. [17] have proposed a deep belief network (DBN) approach for classification of hyperspectral images. The model is a stack of restricted Boltzmann machines, which are trained by greedy layer-wise unsupervised learning [9]. However, by reducing the data to the first three PCA components, the spectral characteristics of the images have not been used in a principal manner by the DBN model. Our proposed approach, by contrast, fully exploits the available spectral information in a hyperspectral image.

3. CNN ARCHITECTURE

Deep CNNs have been successfully applied in solving challenging tasks, such as image classification [13], speech recognition [16], music information retrieval [4], and text recognition [25]. However, to our knowledge, CNN models have not been studied in literature for the purpose of hyperspectral image classification.

Due to network generalization issues [10], deep CNNs for image classification tasks require a large number of images to prevent overfitting, and thus appear inadequate for the HSI classification problem, where a dataset typically consists of a single capture of a scene. Furthermore, the large number of bands in hyperspectral images pose a computational challenge for a straightforward application of a CNN model.

We propose a CNN architecture which integrates both spatial and spectral information for simultaneous spatial-spectral classification of hyperspectral images. The proposed architecture is visualized in Figure 1. The input to the network consists of the eight-connected neighborhood of a hyperspectral pixel, to account for the spatial information context. In order to exploit the original spectral information, all convolutional operations are performed across the spectral bands. The network consists of 5 layers: three convolutional layers with width 16, followed by two fully connected layers with 800 units each. Note that the size of the filters in the first convolutional layer is 9×16, where the first dimension accounts for the total number of pixels in the spatial neighborhood window of the input pixel, and the second dimension is the width of the filter. This allows for simultaneous learning from both the spatial and spectral domain.

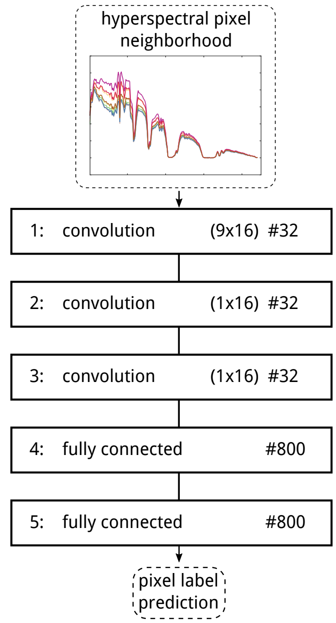

Figure 1: Diagram of the proposed convolutional neural network architecture for hyperspectral image classification. The size of the filters in the convolutional layers are indicated as $(h \times w)$, and # denotes the number of convolutional filters, as well as the number of hidden units in the fully connected layers.

In order to obtain the CNN architecture from Figure 1, we experimented with the number of layers, the number of hidden units in the fully connected layers, and the number and size of the filters in the convolutional layers. In addition, we tested several modifications of the original network. Namely, we experimented with max-pooling layers after the convolutional layers, and also with varying the stride of the convolutions. This worsened the classification results, which is indicative of non-stationarity of statistics across spectral bands. Testing the hyperbolic tangent activation function produced slightly better results than rectified linear units [20] activation. As a result, we used hyperbolic tangent activations in all layers exclusive of the last layer, where the softmax function was used. We also attempted dropout regularization [22] in the fully connected layers, however, this did not improve the classification results.

We trained the network using minibatch gradient descent and momentum [23], and we set the size of the minibatches to 50 samples. We evaluated the model on a held out validation set during training, and we report results on a separate test set for the model that achieved the best results on the validation set.

4. DATA AUGMENTATION

Identifying the classes of pixels from hyperspectral images to produce labeled training data is a manual task, which is expensive and time consuming. Therefore, available training samples for HSI classification are scarce. To try to alleviate this problem, we experimented with simple augmentation for hyperspectral data. For each class in the hyperspectral image dataset, we calculate the per-spectral-band standard deviation of the samples in the training set which belong to the class. Afterwards, we use the calculated vector of standard deviations σ as a parameter to a zero mean multivariate normal distribution $\mathcal{N}(0, \alpha\mathbf{\Sigma})$, where α is a scale factor, and $\mathbf{\Sigma}$ is a diagonal matrix containing σ along the main diagonal entries. Finally, the augmented samples for the class are generated by adding noise sampled from the distribution $\mathcal{N}$ to the original samples. We tried several values for the scaling factor in the set $\{1, 0.5, 0.33, 0.25, 0.125\}$, and fixed $\alpha = 0.25$ for the experiments. The goal of the proposed hyperspectral data augmentation is to prevent overfitting in cases where a low number of samples are used to train the network.

5. EXPERIMENTAL RESULTS

We tested our method on the commonly-used Indian Pines hyperspectral image dataset [14]. This dataset was acquired in June 1992 by NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The Indian Pines scene is a mixed forest and agricultural site in Northwest Indiana, captured at about 20 km altitude by the AVIRIS sensor. The hyperspectral image of the scene consists of 220 bands in the spectral range from 0.4 μm to 2.5 μm, with a spectral resolution of 10 nm. The whole scene consists of 145 $\times$ 145 pixels. There are in total 10,366 labeled samples. With a moderate geometrical resolution of 20 m per pixel, and 16 land cover classes, this dataset poses a challenging classification problem due to the unbalanced number of samples per class, and high inter-class similarity of samples in the dataset.

Table 1: AVIRIS Indian Pines dataset and per class training sets and corresponding test sets.

Class		Train set			Test set		
#	Name	5%	10%	20%	5%	10%	20%
1	Alfalfa	3	6	11	25	24	21
2	Corn-notil	72	144	287	681	645	573
3	Corn-min	42	84	167	396	375	333
4	Corn	12	24	47	111	105	93
5	Grass-pasture	25	50	100	236	223	198
6	Grass-trees	38	75	150	354	336	298
7	Grass-pasture-mowed	2	3	6	12	11	10
8	Hay-windrowed	25	49	98	232	220	195
9	Oats	1	2	4	9	9	8
10	Soybeans-notil	49	97	194	459	435	387
11	Soybeans-min	124	247	494	1,172	1,110	987
12	Soybeans-clean	31	62	123	291	276	245
13	Wheat	11	22	43	100	95	84
14	Woods	65	130	259	614	582	517
15	Bldg-grass-trees	19	38	76	180	171	152
16	Stone-steel-towers	5	10	19	45	42	38
Total		524	1,043	2,078	4,917	4,659	4,139

For our experiments, we evaluated the classification accuracy of the method using a balanced training set per class, with low number of training samples. We trained the network with 5%, 10%, and 20% of randomly selected labeled samples per class, and equally divided the remaining labeled samples into separate validation and test sets. In each case, we repeated the experiment with and without hyperspectral data augmentation.

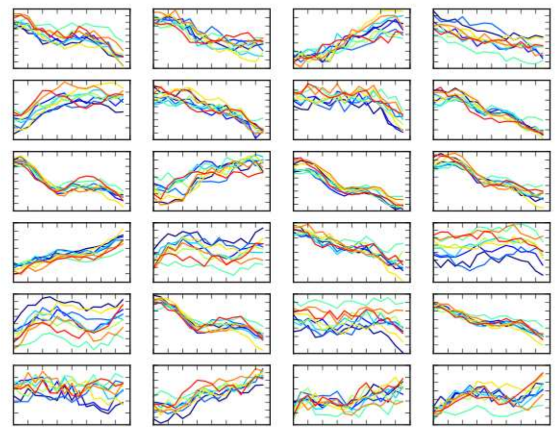

Figure 2: A subset of filters learned in the first convolutional layer of the network. Each subplot represents a (9×16) filter.

Table 2: Classification results for the Indian Pines image on the test sets.

Indian Pines		Test set		
		5%	10%	20%
Augmented	OCA(%)	86.54 $\pm$ 0.30	92.70 $\pm$ 1.00	96.58 $\pm$ 0.55
	F1 score	0.86 $\pm$ 0.00	0.93 $\pm$ 0.01	0.97 $\pm$ 0.01
Non-augment.	OCA (%)	85.46 $\pm$ 1.73	92.76 $\pm$ 0.93	96.54 $\pm$ 0.47
	F1 score	0.85 $\pm$ 0.02	0.93 $\pm$ 0.01	0.96 $\pm$ 0.00

The achieved classification results for each of the experiments are shown in Table 2. We performed 5 Monte Carlo runs, where for each run we selected a training set of 5%, 10%, and 20% of the labeled samples, as explained above, to train our model. In the cases with augmentation, we found 3 fold (per class) augmentation of the training data to give the best results. We report the average and standard error of the 5 Monte Carlo runs in terms of the overall classification accuracy (OCA), i.e., the number of correctly classified samples from the total number of samples in the test set, and the F1 score, which is weighted so that it accounts for the imbalance of the classes. From the results in Table 2, it can be seen that only when using a very low number of augmented labeled samples for training (5%), there is improvement in the classification scores over the non-augmented counterpart. However, we have observed that in all cases augmentation reduced the number of training iterations significantly, as compared with training with the corresponding non-augmented data.

We have visualized some of the learned filters from the first convolutional layer of the network in Figure 2. From the visualization, it is clear that the learned filters have a structured shape, and that some of the filters roughly resemble different spectral band-pass filters.

6. CONCLUSIONS

Due to the inherent nature of hyperspectral data, discernment of good features for hyperspectral image classification is difficult. Therefore, in this paper, we have presented a new approach towards hyperspectral image classification based on deep convolutional neural networks. To evaluate the effectiveness of the method, we performed experiments on a commonly-used hyperspectral image dataset. Our experimental results have shown that the neural network model can learn structured features resembling different spectral band-pass filters directly from the input data. These features prove useful for hyperspectral image classification, which makes end-to-end learning applicable to hyperspectral scene understanding.

7. ACKNOWLEDGMENTS

The research activities as described in this paper were funded by Ghent University and the Interdisciplinary Research Institute iMinds.

8. REFERENCES

[1] T. V. Bandos, L. Bruzzone, and G. Camps-Valls. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3):862–873, 2009.

[2] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, 2005.

[3] D. Cai, X. He, and J. Han. Semi-supervised discriminant analysis. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1–7, Oct 2007.

[4] S. Dieleman and B. Schrauwen. End-to-end learning for music audio. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 6964–6968, May 2014.

[5] G. Edelman, E. Gaston, T. van Leeuwen, P. Cullen, and M. Aalders. Hyperspectral imaging for non-contact analysis of forensic traces. *Forensic Science International*, 223(1-3):28–39, 2012.

[6] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008.

[7] X. He, D. Cai, S. Yan, and H.-J. Zhang. Neighborhood preserving embedding. In *Proceedings of the International Conference on Computer Vision (ICCV)*, volume 2, pages 1208–1213, Oct 2005.

[8] X. He and P. Niyogi. Locality preserving projections. In *NIPS*, pages 153–160. MIT Press, 2004.

[9] G. E. Hinton, S. Osindero, and Y.-W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, July 2006.

[10] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, abs/1207.0580, 2012.

[11] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24:417–441, 1933.

[12] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, Jan 1968.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105. 2012.

[14] D. A. Landgrebe. *Signal theory methods in multispectral remote sensing*. Wiley series in remote sensing. Wiley, Hoboken, N.J., Chichester, 2003.

[15] Y. Lecun, U. Muller, J. Ben, E. Cosatto, and B. Flepp. Off-road obstacle avoidance through end-to-end learning. In *Advances in Neural Information Processing Systems 18*, pages 739–746, Cambridge, MA, 2006.

[16] H. Lee, P. Pham, Y. Largman, and A. Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. In *Advances in Neural Information Processing Systems 22*, pages 1096–1104. 2009.

[17] T. Li, J. Zhang, and Y. Zhang. Classification of hyperspectral image based on deep belief networks. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 5132–5136, Oct 2014.

[18] G. Lu and B. Fei. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 19(1):010901, 2014.

[19] F. Melgani and L. Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790, Aug 2004.

[20] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel*, pages 807–814, 2010.

[21] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment*, 113, Supplement 1(0):S110–S122, 2009.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, Jan. 2014.

[23] I. Sutskever, J. Martens, G. E. Dahl, and G. E. Hinton. On the importance of initialization and momentum in deep learning. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, volume 28, pages 1139–1147, May 2013.

[24] J. Wang and C.-I. Chang. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 44(6):1586–1600, 2006.

[25] T. Wang, D. Wu, A. Coates, and A. Ng. End-to-end text recognition with convolutional neural networks. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 3304–3308, Nov 2012.