DEEP SUPERVISED LEARNING FOR HYPERSPECTRAL DATA CLASSIFICATION THROUGH CONVOLUTIONAL NEURAL NETWORKS

Konstantinos Makantasis[1], *Konstantinos Karantzalos*[2], *Anastasios Doulamis*[2], *Nikolaos Doulamis*[2]

[1]Technical University of Crete, University campus, Kounoupidiana, 73100, Chania, Greece
[2]National Technical University of Athens, Zographou campus, 15780, Athens, Greece
kmakantasis@isc.tuc.gr; karank@central.ntua.gr; {adoulam, ndoulam}@cs.ntua.gr

ABSTRACT

Spectral observations along the spectrum in many narrow spectral bands through hyperspectral imaging provides valuable information towards material and object recognition, which can be consider as a classification task. Most of the existing studies and research efforts are following the conventional pattern recognition paradigm, which is based on the construction of complex handcrafted features. However, it is rarely known which features are important for the problem at hand. In contrast to these approaches, we propose a deep learning based classification method that hierarchically constructs high-level features in an automated way. Our method exploits a Convolutional Neural Network to encode pixels' spectral and spatial information and a Multi-Layer Perceptron to conduct the classification task. Experimental results and quantitative validation on widely used datasets showcasing the potential of the developed approach for accurate hyperspectral data classification.

Index Terms— Machine learning, Earth observation, Imaging spectroscopy, Object Recognition

1. INTRODUCTION

Recent advances in optics and photonics have allowed the development of hyperspectral imaging sensors with higher spectral and spatial resolution onboard various satellite, aerial, UAV and ground acquisition platforms. The efficient exploitation of finer spatial and spectral information can ameliorate significantly material detection and object recognition applications by revealing and modelling the subtle differences in spectral signatures of various objects.

Recognizing various materials, objects and terrain land cover classes based on their reflectance properties can be viewed as a classification task *i.e.,* classify image pixels based on their spectral characteristics. Although, hyperspectral imaging have been used in a wide variety of applications, such as agriculture, surveillance, astronomy and biomedical imaging [1], it has its own unique challenges including; i) high dimensional data, ii) limited number of labeled samples and iii) large spatial variability of spectral signatures [2].

Most of the existing work, concerning the classification of hyperspectral data, follow the conventional paradigm of pattern recognition, which consists of two separate steps; firstly, complex handcrafted features are computed from the raw data input and secondly, the obtained features are used to learn classifiers, such as Support Vector Machines (SVM) and Neural Networks (NN) [3]. In particular, for high dimensional data and when few training samples are available statistical learning methods have been employed to tackle the high dimensionality and heterogeneity of hyperspectral data [4].

However, due to the high diversity of depicted materials, it is rarely known which features are important for the classification task. In contrast to the conventional paradigm of pattern recognition, deep learning models [5–8] are a class of machines that can learn a hierarchy of features by building high-level features from low-level ones, thereby automating the process of feature construction for the problem at hand. Furthermore, for bigger datasets and quite large images with very high spatial and spectral resolution, deep learning frameworks seems to fit and address more effectively the classification problem [9]. Techniques based on deep learning have already shown promising results both for the detection of particular objects, like man-made ones [10] or vehicles [11] and for the classification of hyperspectral data [9].

More specifically, a deep learning framework was employed in [9] towards the classification of hyperspectral data with quite promising results. In particular, Autoencoders have been used as building blocks and the concept of greedy layer-wise training [8] to construct a deep architecture for hierarchically building high-level spectral features for each pixel. Spectral features were combined in a separate step with spatial-dominated information and fed as input to a logistic regression classifier.

In a similar way, we propose a deep learning framework for the classification of hyperspectral data into multiple classes. However, our approach is based on a unified framework, which combines spectral and spatial information in a

The research leading to these results has received funding from the European Union's FP7 under grant agreement n.313161, eVACUATE Project www.evacuate.eu

single step, constructing simultaneously high-level spectral-spatial features. In particular, we propose the exploitation of a modified Convolutional Neural Network (CNN), which conducts the task of high-level features construction and a Multi-Layer Perceptron (MLP), which is responsible for the classification task. Under such a formulation, the developed system constructs spectral-spatial features at once, while at the same time achieves real-time predictions due to the feed-forward nature of CNNs and MLPs.

2. APPROACH OVERVIEW

We consider the exploitation of a deep learning architecture for the classification of hyperspectral data, *i.e.* the classification of each pixel to a predefined number of classes based on their spectral and spatial properties. The spectral characteristics are associated with the reflectance properties at every pixel for every spectal band, while spatial information is derived by taking into consideration its neighbors.

Towards this direction, high-level features that encode pixels' spectral and spatial information, are hierarchically constructed using a CNN [5]. CNNs consist a type of deep models, which apply trainable filters and pooling operations on the raw input, resulting in a hierarchy of increasingly complex features. Although, it has been shown that CNNs can achieve superior performance on visual recognition tasks without relying on handcrafted features, due to their nature, they produce global image features.

A hyperspectral image is represented as a 3D tensor of dimensions $h \times w \times c$, where h and w correspond to the height and width of the image and c to its channels (spectral bands). In order to be aligned with the specific nature of CNNs, we have to decompose the captured hyperspectral image into *patches*, each one of which contains spectral and spatial information for a specific pixel.

More specifically, in order to classify a pixel $p_{x,y}$ at location (x, y) on image plane and successfully fuse spectral and spatial information, we use a square patch of size $s \times s$ centered at pixel $p_{x,y}$. Let us denote as $l_{x,y}$ the class label of the pixel at location (x, y) and as $w_{x,y}$ the patch centered at pixel $p_{x,y}$. Then, we can form a dataset $D = \{(w_{x,y}, l_{x,y})\}$ for $x = 1, 2, \cdots, w$ and $y = 1, 2, \cdots, h$. Patch $w_{x,y}$ is also a 3D tensor with dimension $s \times s \times c$, which contains spectral and spatial information for the pixel located at (x, y).

Moreover, tensor $w_{x,y}$ is divided into c matrices of dimensions $s \times s$ which are fed as input into a CNN, which hierarchically builds high-level features that encode spectral and spatial characteristics of pixel $p_{x,y}$. These features are fed to a MLP, which is responsible for the classification task.

3. SYSTEM ARCHITECTURE

In this section the developed system architecture is briefly described. Firstly, the proposed approach for the dimensionality reduction of input raw data is presented and then the structures of CNN and MLP are given.

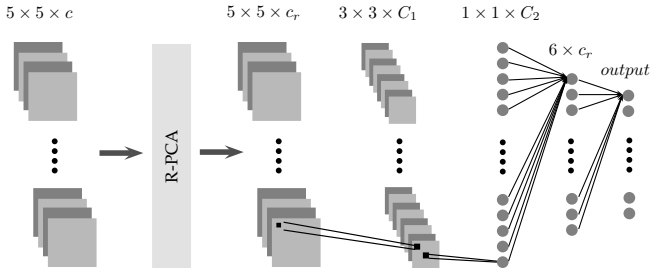

Fig. 1. Overall system architecture.

3.1. Reducing the dimensions of raw input data

Training CNNs requires the convolution of each one of the network's 2D inputs with each one of the trainable filters. The hundreds of channels along the spectral dimension (network inputs) of a hyperspectral image increase the computational cost of training and prediction processes.

However, through a statistical analysis of spectral responses of pixels that belong to the same class, we can observe that the variance of responses is very small. This suggests that pixels that belong to the same class have almost the same values at every channel. At the same time, pixels that belong to different classes present different spectral properties. Based on these characteristics a dimensionality reduction technique can be employed to reduce the dimensionality of the input data in order to speed up the training and prediction processes.

For dimensionality reduction, Randomized PCA (R-PCA) is introduced along the spectral dimension to condense the whole image. It should be noted that this step does cast away spectral information, but since R-PCA is applied along the spectral dimension, the spatial information remains intact. The number of principal components that are retained after the application of R-PCA, is appropriately set, in order to keep at least 99.9% of initial information. During the experimentation process on widely-used hyperspectral datasets, this amount of information is preserved by using the first 10 to 30 principal components, reducing this way up to 15 times the dimensionality of the raw input.

3.2. Machine learning and classification structure

After dimensionality reduction, each patch is a tensor of dimensions $s \times s \times c_r$. Parameter c_r corresponds to the number of principal components that preserve at least 99.9% of initial information, while the parameter s determines the number of neighbors of each pixel that will be taken into consideration during classification task.

During experimentation process we set the parameter s to be equal to 5, in order to take into consideration the clos-

Table 1. Quantitative Evaluation Results. The number of components and the classification accuracy for every dataset.

Dataset	No. Components	Accuracy(%)
Indian Pines	30	98.88
Salinas	10	99.53
Pavia centre	10	99.91
Pavia university	10	99.62

est 24 neighbors of each pixel. By increasing the value of s, the number of neighbors that are taken into consideration is increased and thus the computational cost of classification is increased, also. However, setting the parameter s to a value larger than 5, no further improvement on classification accuracy was reported in all experiments. On the contrary, increasing the value of s over 13, deteriorates classification accuracy.

Having estimate the values of the parameters s and c_r, we can proceed with the CNN structure design. The first layer of the proposed CNN is a convolutional layer with $C_1 = 3 \times c_r$ trainable filters of dimension 3×3. This layer delivers C_1 matrices of dimensions 3×3 (during convolution we don't take into consideration the border of the patch). In contrast to conventional CNNs, we do not use a maxpooling layer after the convolution layer, since we don't take into account any translation and scale invariance. For this reason the first convolutional layer is followed by a second convolutional layer with $C_2 = 3 \times C_1$ trainable filters. Again, the filters are 3×3 matrices.

The second convolutional layer delivers a vector with C_2 elements, which is fed as input to the MLP classifier. The number of MLP hidden units is smaller than the dimensionality of its input. In particular, we set the number of hidden units to equal $6 \times c_r$. For training the deep learning architecture the standard backpropagation algorithm was employed, in order to learn the optimal model parameters, *i.e.* minimize the negative log-likelihood of the data sets under the model parameterized by MLP weights and filters elements. The overall system architecture is presented in Fig.1.

4. EXPERIMENTAL RESULTS AND VALIDATION

In our study we experimented and validated the developed framework with AVIRIS and ROSIS hyperspectral datasets. In particular, we employed i) the Indian Pines dataset, which depicts a test site in North-western Indiana and consists of 145×145 pixels and 224 spectral reflectance bands in the wavelength range 0.4 to 2.510^{-6} meters. ii) one 224-band hyperspectral image of Salinas Valley, California, which is characterized by high spatial resolution and iii) the Pavia centre and Pavia university datasets, whose number of spectral bands are 102.

Supervised training was conducted using the ground truth images of the aforementioned datasets. In particular, we split

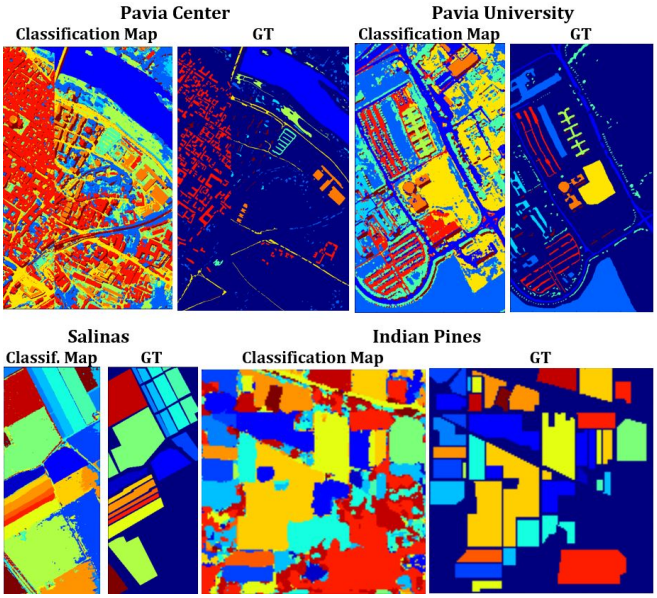

Fig. 2. Classification results after the application of the developed framework. The resulted classification map along with the ground truth (GT) are shown for the Pavia center, Pavia university, Salinas and Indian Pines datasets.

the tagged parts of hyperspectral images into three sets, *i.e.,* training, validation and testing data, with a split ratio $8:1:1$. That is, we randomly choose 80% of the tagged samples as the training set, and 10% and 10% for the validation and testing sets, respectively. It has to be mentioned that background pixels were not considered for classification purposes.

The effectiveness of our proposed deep learning model is tested in comparison to SVM-based methods. In particular, we test our model against RBF kernel SVM and Linear kernel SVM. Both models are trained twice, firstly using the raw hyperspectral data (all spectral bands) and secondly, using the reduced data (principal components that preserve at least 99.9% of raw data information). All models have been quantitatively validated in terms of classification accuracy.

Table 1 presents the classification accuracy of our system for each one of the above datasets along with the number of principal components that are used. Table 2 presents in terms

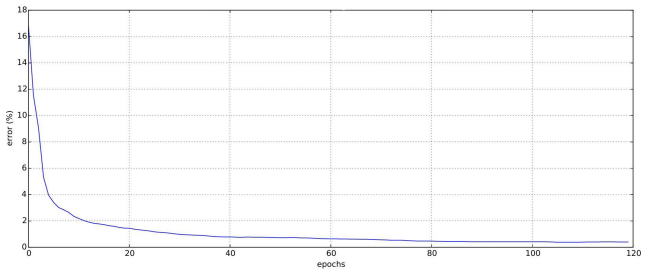

Fig. 3. Misclassification errors in regard to the training epochs for the Pavia university dataset.

Table 2. Quantitative Evaluation Results. Comparison to SVM-based methods.

	Method (classification accuracy (%))				
Dataset	Our Approach	R-PCA RBF-SVM	R-PCA Linear-SVM	RBF-SVM	Linear-SVM
Pavia centre	**99.91**	98.87	97.63	99.01	97.86
Pavia university	**99.62**	93.82	84.39	93.94	84.67
Salinas	**99.53**	93.73	90.52	93.97	90.68
Indian Pines	**98.88**	82.71	79.47	82.79	79.56

of classification accuracy the performance of our proposed method against SVM based models. Our system presents superior classification accuracy in all datasets. It should be noted that in the Pavia university dataset, our method outperforms the state-of-the-art deep learning based method of [9] for more than 1%. The classification accuracy of our method is 99.62%, while the classification accuracy of [9] is 98.52%.

Furthermore, we examine the classification accuracy from a visual perspective. Pixels, corresponding to annotated and not-annotated regions, for each one of the datasets where classified using our deep learning approach. The classification results after the application of the developed framework are presented in Fig.2. The resulted classification map along with the ground truth (GT) are shown for the Pavia center, Pavia university, Salinas and Indian Pines datasets. As we can see, by fusing spectral and spatial information for each pixel, classification process results to the formation of compact areas, eliminating noisy scatter points.

Finally, in Fig.3 the misclassification error is presented in regard to the number of training epochs for the Pavia university dataset. The training process for the proposed system converges in almost 40 epochs. Therefore early stopping criteria can be consider during the training procedure, in order to reduce computational cost, without deteriorating classification performance.

5. CONCLUSIONS

In this paper, we propose a deep learning based approach for hyperspectral data classification. Following deep learning paradigm, through the exploitation of CNNs and MLPs, our approach hierarchically constructs high-level features that encode pixels spectral and spatial information. We have compared our method to SVM-based classifiers, which are widely used and are among the state-of-the-art on four publicly available hyperspectral datasets. Experimental validation shows that our deep learning approach presents superior performance for every datasets. We present the convergence rate of the training process, which suggests that our method can be scaled to large datasets. The quite promising quantitative evaluation indicate the high potentials of the developed approach. Finally, among the future perspectives is the application of the developed framework for the detection of human behavior from hyperspectral video sequences.

6. REFERENCES

[1] Chein-I. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*, John Wiley and Sons, 2013.

[2] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. on Geos. and Rem. Sensing*, vol. 43, 2005.

[3] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*, J. Wiley and Sons, NJ, USA, 2009.

[4] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. Atli Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 45–54, Jan 2014.

[5] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[6] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, July 2006.

[7] G Hinton, S Osindero, and Y Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, July 2006.

[8] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle, "Greedy layer-wise training of deep networks," in *Advances in NIPS 19*, pp. 153–160. 2007.

[9] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu, "Deep learning-based classification of hyperspectral data," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 2094–2107, June 2014.

[10] Volodymyr Mnih and Geoffrey E. Hinton, "Learning to label aerial images from noisy data.," in *Proceedings of the 29th International Conference on Machine Learning, ICML 2012,*. 2012, icml.cc / Omnipress.

[11] Gregoire Montavon, GenevieveB. Orr, and Klaus-Robert Muller, *Neural Networks: Tricks of the Trade*, Springer Berlin Heidelberg, 2012.