# AUTOMATIC FUSION AND CLASSIFICATION USING RANDOM FORESTS AND FEATURES EXTRACTED WITH DEEP LEARNING

*Andreas Merentitis, Christian Debes*

AGT International
64295 Darmstadt, Germany
E-mail: {amerentitis, cdebes}@agtinternational.com

## 1. INTRODUCTION

Recently, the fusion of different sensor modalities (e.g., hyperspectral and LiDAR, thermal infrared and high resolution visual, or multispectral and infrared data) has shown to bring significant improvement in the classification of airborne imagery [1], [2]. The key idea is that two sources can be fused in a complementary way (e.g., to distinguish between streets and flat commercial buildings that can have a similar hyperspectral profile but differ in height).

However, in order to extract the most benefit from the two different sources of information that are fused, a carefully designed feature extraction scheme is required. For example, state of the art performance is often achieved using a multitude of features like spectral abundace maps, MNF transformed spectral bands, synthetic spectral (water absorption, various vegetation indices, etc.), synthetic spatial features (e.g., segmentation maps), as well as domain specific features of the sensor modalities (e.g., gradients and flatness profile for LiDAR). Extracting these features is difficult and often requires fine-tunning of parameters (e.g., regularization parameter for MNF components or the parameters of the segmentation algorithm). Furthermore, some of the features are requiring the availability of specific spectral bands (e.g., the vegetation indices) or an approximation using similar alternative indices if the required bands are not available.

As an alternative to the previous feature extraction schemes, in recent years deep learning algorithms have been proposed in order to automate feature extraction. Various definitions have been proposed for deep learning in the literature, describing variations of the deep learning family of algorithms. In general, deep learning is typically based on the (possibly unsupervised) learning of multiple levels of features or representations of the data, where higher level features are derived from lower level features to form a hierarchical representation. Deep learning has proved a remarkably successful paradigm in a wide range of application domains, from image and video recognition to sound classification.

The contribution of this paper is the use of features extracted using a deep learning approach in a random forest algorithm in [2], [3] which includes automatic feature selection. The feature extraction is unsupervised and hierarchical. Furthermore, computational efficiency (often a challenge for deep learning methods) is particularly important in order to make certain that the method can be applied in large remote sensing datasets. Finally, the deep learning feature extraction is not tightly coupled to the classification algorithm, thus different classifiers can be easily used as well.

In the remainder of the paper, first we briefly introduce deep learning before focusing on the specific variant that is used in this paper. A large number of deep learning features are generated, but using all of them as inputs for the classifier is often not the optimal choice since they can be redundant. In order to mitigate this challenge, we use relative feature relevance in random forests to perform automatic feature selection. Next, we demonstrate the efficacy of the deep learning features added to the raw spectral bands, achieving significant performance gain in challenging datasets. Finally, we show as the additional benefit from the deep learning features when added to a set of features optimized using domain level expertise.

## 2. DEEP LEARNING FEATURE EXTRACTION

As discussed previously, the goal of deep learning in general is to learn multiple levels of features or representations of the data, where higher level features are derived from lower level features to form a hierarchical representation. This can be achieved through a multitude of approaches, including Restricted Boltzmann Machines (RBMs), Deep Belief Networks (DBNs), Deep Autoencoders (DAs), Deep Convolutional Networks (DCNs), and Deep Sparse Filters (DSFs). From these methods, here we focus on the last one, introduced in [4] and [5], because it combines good performance with relatively low computational cost, as well as independence from the learning algorithm.

The main idea of Deep Sparse Filtering is to optimize a simple cost function, the l2-normalized features, as opposed to explicitly attempting to construct a model of the data distribution. This approach has been demonstrated [4] to maintain the capacity for learning meaningful high level features by greedy layer-wise stacking that is a key property of deep learning algorithms, while also scaling gracefully to large and multi-dimensional datasets. Some other desirable properties of feature extraction are summarized below:

- Sparse features per sample (Population Sparsity). Each sample should be represented by only a small number of active (non-zero) features.

- Sparse features across samples (Lifetime Sparsity). Features should have discriminative power and allow us to distinguish samples; thus, each feature should only be active for a few samples.

- Uniform activity distribution (High Dispersal). For each row, the distribution should have similar statistics to every other row; no one row should have significantly more activity than the other rows.

These properties have been studied extensively in the feature extraction literature (often drawing biological inspiration from neuroscience). Nevertheless, Deep Sparse Filtering is one of the few methods that offers all the previous properties while still being computationally efficient. In the final paper we will describe in more detail what is the practical implications of these properties for a remote sensing dataset, especially when fusion of different sensor modalities is considered. Following this we will elaborate on how Deep Sparse Filters should be applied in such a dataset to derive the best result.

## 3. RANDOM FORESTS WITH DEEP-LEARNED FEATURES

Using Deep Sparse Filters, a multitude of different hierarchical representations can be derived since the number of features is the basic parameter of the DSF algorithm. Furthermore, these features can be combined with hand-crafted features extracted using domain expertise (for example, synthetic hyperspectral features like vegetation indices, synthetic LiDAR features like gradients, etc.). In this paper we study both the efficacy of the deep features when added to the raw spectral bands, as well as their complementarity to a traditional set of state of the art features.

Feeding a large number of features to the classification algorithm may not result in optimal performance due to the curse of dimensionality effect. In our previous work [6] we have proposed to use the inherent feature ranking capability of Random Forests to learn the important features and use only them in the subsequent final classification step. Here we use again the same approach, in order to select the relevant deep features as well as to evaluate the ranking of deep features compared to the ones derived by domain expertise.

The two datasets we use in this paper are characterized by completely different properties, so that the deep learning features can be evaluated in a large spectrum of cases. The IEEE GRSS 2013 Data Fusion Contest dataset captures a challenging and varied semi-urban environment with 15 classes. It is well suited for evaluating the capacity of the algorithm to extract features from complementary sources of information (hyperspectral and LiDAR) in a data fusion setup, as well as to deal with noise on part of the input space (due to the existence of a shadow). For the University of Houston we used the same training set as in the contest.

The second dataset, Indian Pines - despite being quite old - is also a challenging dataset of an agricultural area with 16 different classes of crops. Since most of the classes are different types of crops, this dataset is well suited to assess the capability of the algorithm to extract discriminative information in a setup where many classes are of similar nature. For Indian Pines the training set consists of 50 samples for each class that have been randomly chosen from the reference data, except for classes alfalfa, grass/pasture-mowed, and oats that have very few members. Thus, only 15 samples for each of these classes were randomly picked to be used as training samples and all other samples composed the test set.

| **Indian Pines** | **Classification** | **+Post-proc.** |
|---|---|---|
| Raw | 81.9 (%) | 87.0 (%) |
| Raw + Deep | 83.2 (%) | 89.3 (%) |
| Best [6] | 94.4 (%) | 96.3 (%) |
| Best [6] + Deep | 95.0 (%) | 96.8 (%) |

**Table 1**. *Results for Indian Pines*

| **Houston University** | **Classification** | **+Post-proc.** |
|---|---|---|
| Raw | 78.7 (%) | 81.5 (%) |
| Raw + Deep | 80.2 (%) | 83.7 (%) |
| Best [2] | 88.1 (%) | 94.6 (%) |
| Best [2] + Deep | 90.3 (%) | 96.1 (%) |

**Table 2**. *Results for Houston University*

The results for the two datasets are shown in Tables 1 and 2. Each row shows the overall accuracy for a given set of features after classification, as well as after post-processing of the classification result. The latter consists of Markov Random Field (MRF) segmentation for the case of Indian Pines, while for the Houston University case an additional man-made structure correction step is applied before MRF segmentation [2]. We can see that adding the deep-learned features is improving significantly the classification result compared to using only the raw spectral bands. Particularly for the Houston University dataset, adding the deep-learned features to the raw spectral bands covers half of the performance difference in the classification step between the raw bands and the carefully optimized feature set of [2]. However, even more interesting is that the deep-learned features are able to provide some performance improvement when added to the best sets of features (MNF components, various synthetic spectral and LiDAR features, and segmentation maps) from [2] and [6]. In the final paper we will present a more detailed evaluation of the performance gain, including the relevant ranking of all features and we will discuss the complementarity of the deep learned features to the raw spectral bands, as well as to the set of features optimized using domain level expertise.

## 4. ACKNOWLEDGEMENT

The authors would like to thank the Hyperspectral Image Analysis group and the NSF Funded Center for Airborne Laser Mapping (NCALM) at the University of Houston for providing the data sets used in this study, and the IEEE GRSS Data Fusion Technical Committee for organizing the 2013 Data Fusion Contest.

## 5. REFERENCES


[1] Pall Oskar Gislason, Jon Atli Benediktsson, and Johannes R. Sveinsson, "Random forest classification of multisource remote sensing and geographic data," in *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, Alaska, USA, 20-24 September 2004*, 2004, pp. 1049–1052.

[2] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. van Kasteren, Wenzhi Liao, R. Bellens, A. Pizurica, S. Gautama, W. Philips, S. Prasad, Qian Du, and F. Pacifici, "Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 7, no. 6, pp. 2405–2418, June 2014.

[3] A. Merentitis, C. Debes, and R. Heremans, "Application of ensemble learning in hyperspectral image classification: Towards selecting favorable spots in the bias-variance plane," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, pp. 1089 – 1102, April 2014.

[4] Jiquan Ngiam, Zhenghao Chen, Sonia A. Bhaskar, Pang W. Koh, and Andrew Y. Ng, "Sparse filtering," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, Eds., pp. 1125–1133. Curran Associates, Inc., 2011.

[5] Quoc V. Le, Alexandre Karpenko, Jiquan Ngiam, and Andrew Y. Ng, "Ica with reconstruction cost for efficient overcomplete feature learning," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R.S. Zemel, P.L. Bartlett, F. Pereira, and K.Q. Weinberger, Eds., pp. 1017–1025. Curran Associates, Inc., 2011.

[6] A. Merentitis, C. Debes, R. Heremans, and N. Frangiadakis, "Automatic fusion and classification of hyperspectral and lidar data using random forests," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International*, July 2014, pp. 1245–1248.