

# EXPLORATORY VISUAL ANALYSIS OF MULTISPECTRAL EO IMAGES BASED ON DNN

*Iulia Neagoie(1), Daniela Faur(1), Corina Vaduva(1) and Mihai Datcu(1)(2)*

University Politehnica of Bucharest UPB (1), German Aerospace Centre DLR (2)

## ABSTRACT

Exploratory visual analysis is often required to assist human operator to understand and interpret Earth Observation (EO) images. Optimal image representation offers cognitive support in discovering relevant facts about the scene with respect to a particular application. This is of crucial importance for training data sets selection in all Machine Learning tasks, particularly in the design of active learning tools for multispectral (MS) EO data. This paper proposes a deep neural network (DNN) based method to compress, learn and reveal the most significant information included in the spectral bands of EO data in support of relevant visualization for image content analysis. The advanced method uses a DNN to discover the most suggestive pseudo-color representation able to highlight the entire MS image content better than the particular 3 bands selection (R, G, B). We propose the use of information theory and the concept of mutual information to rank the spectral bands based on the amount of information contained, by applying the minimum-redundancy-maximum-relevance (mRMR) criterion on a the image so that we obtain the ranked bands. A DNN stacked autoencoder based paradigm is developed in order to extract and compress in three bands the overall information from the MS EO data. The developed method is demonstrated and validated for Sentinel 2 dataset.

**Index Terms**— EO images, DNN, mRMR, information theory, mutual information

## 1. INTRODUCTION

With the evolution of sensors, the complexity of remote sensing data increased, causing the processes of analyzing the information contained to become more challenging. The main reason for this is the fact that most of the tasks involved in the process are still manually performed by a human operator. Requiring accuracy, knowledge and attention, visual interpretation is the most relying part of data analysis. A more suggestive representation would improve data analysis and provide support for further human interpretation [1].

High resolution multispectral images contain information that can be useful in a wide range of applications (e.g. urban



Fig. 1 Sentinel 2 images, Musura Golf, Sulina, Romania. Represents the comparison between two image representations in: visible domain (left) and infrared domain (right) of the electromagnetic specter.

monitoring, assessment of forest deforestation). Each of the semantic classes (e.g. water, forest, urban) has a specific spectral signature, which may not include the visible spectral bands, so the visualization of the “true color” version of the image may hide important details.

The need for visualization of the most relevant features comes from the fact that the Multi Spectral Instrument (MSI) measures the Earth’s reflected radiance in 13 spectral bands (as in the case of Sentinel 2 sensor) so the usual visualization of the RGB display (4, 3, and 2) lacks the information contained in the others 10 bands.

Comparing the RGB representation (visible domain) with the IR pseudo-color representation (6, 7 and 8A bands of Sentinel-2), as in Fig. 1, we can observe that there are hidden details in the RGB display, but in the IR display these appear very clear. Also, in the IR representation we can define the water contour very accurate, in contrast with the RGB where, due to river deposits, it is hard to discriminate water from vegetation. The confusion is caused by the fact that they may seem to be color similar in the visible part of the specter, yet they have different characteristics in the IR part of the electromagnetic specter.

This paper aims to first demonstrate the performance of the mRMR criterion for the selection of spectral features that optimally describe a semantic class [2], and then to advance to the use of a deep learning approach and information theory concepts to collect the relevant, consistent information hidden in all the spectral bands imaging a scene.

The authors of [2] exploit the potential of the mRMR criterion, first introduced by [3], to discover the optimal

wavelength intervals that contain the maximum amount of information relevant to a specific class detectable in a scene.

Artificial intelligence (AI) as a part of computer science gained attention due to its success in Machine Learning (ML). AI represents the theoretical basis used by ML to develop smart software which excludes the interference of human and learn from past experience. Starting from the days when the reasoning methods were logic and understandable, it has come to a point where the question of the explain ability of AI-based systems has become complex and desirable to solve. ML is successful in a wide range of domains and demonstrates practical success, but it is seen as a “black box” due to the fact that its results cannot be explained in an understandable way. Being focused on the medical domain, the authors in [4] present and analyze understandable models.

Deep Neural Networks (DNNs) are used in a wide range of practical problems, from image recognition [5] and image classification [6] to action recognition [7]. A DNN represents in fact a collection of neurons displayed on successive layers. Every neuron receives as input the activations from the previous layer and executes a simple computation, creating a map from the input to the output, which represents the knowledge learned from this target. In [8] the authors try to open the “black box” by visualizing the layers in the Information Plane (IP) and also demonstrate that hidden layers bring a computational benefit.

The advantages of DNNs are the capability to learn and the ability to predict the same way a human does, sometimes even overrule him.

One special type of deep learning architecture is the autoencoder, which seems very alike with a communication channel and has the main advantage that is fully unsupervised. The authors of [9] try to understand, using the information theoretic methodology, the learning process and the design of autoencoders. The purpose of an autoencoder is to transform the input into output losing the minimum amount of information. It consists of two parts: an encoder and a decoder. The encoder result represents the hidden representation of the input, while the decoder result is the reconstruction of the input using the hidden representation.

Stacked autoencoders (SAE) are an extension of the basic autoencoders, but still have the same purpose. The difference is that SAE consist of multiple encoding layers followed by the same amount of decoding layers and during training the output is desired to be exactly as the input.

For a while, the standard method for reducing the dimensionality of multispectral images was Principal Components Analysis (PCA). This method consists of finding the first three principal components with  $N$  dimensions from an image, where  $N$  is the number of bands, based on the most variance of the data. These components are in fact orthonormal basis vectors. The disadvantages of this method were evaluated by authors in [10]:

- a) Mapping the amount of variance of the principal components to the R, G, B channels it would not result in an optimal transfer of information;
- b) The stretching process generated by the PCA can create highly saturated colors that may generate anomalies not linked to the input image;
- c) Spatial and contextual information is not taken into account by PCA;
- d) Spatial and spectral artifacts are introduced by PCA.

Taking into account these disadvantages, in this paper we choose not to compare PCA against the studied methods. The approaches targeted address an enhanced method of satellite images visualization using DNN and information theory concepts in order to help the users of the satellite imagery to complete the analyzing process of a scene in a short time. The proposed method aims to develop a neural network that learns how to encode the whole information contained in the spectral bands of a pixel into just three values. To achieve that, an autoencoder architecture is chosen, mostly because of the fact that it does not need labeled data and it compresses the information in such a way that the loss is minim.

For the experimental part, a Sentinel 2 dataset is used. The multispectral images must be preprocessed through a resampling due to the resolution difference between the spectral bands. Sentinel-2 images have 13 bands, out of which 4 have 10 meters, 6 on 20 and 3 on 60 m, so we compute a resampling to bring all of them to the 10 meters resolution.

## 2. VISUAL SPECTRAL RELEVANCE TRANSFORMATION BASED ON MRMR CRITERION

The mRMR, evaluating the measure of statistical dependency between the target class and the spectral bands in terms of mutual information and minimizing inter-band redundancy, ranks the spectral bands. The first three are used to represent the “false color” image (B1-B2-B3). The number of spectral bands that can be evaluated is unlimited.

Consequently, information theory concepts can be used to improve the visualization and to provide an effective solution in terms of both accuracy and robustness [1]. The notion of mutual information  $I(X,Y)$ , estimating the reduction of uncertainty of a random variable  $X$  due to the knowledge of the other variable  $Y$  will help us to quantify the achievement of the data visualization goal, that is a maximal mutual information between the input data and the resulted visualization.

The results presented in Fig. 2 demonstrate the potential of the mRMS criterion for the selection of the most relevant features to describe an image. In Fig. 2 the first column represents the RGB display and the second column represents the improved display, containing the first three more relevant bands ranked using mRMR.



Fig. 2 Visual spectral relevance transformation based on mRMR criterion. a) first row – Mănăstirea, Romania b) second row – Sulina, Romania c) third row – Tulcea, Romania.

On the first row is shown an urban example where on the RGB representation there is a confusing vegetation/water area marked with a yellow circle. The second image instead, is displayed as:

- R – B5 – NIR
- G – B4 – Visible (Red)
- B – B3 – Visible (Green),

and highlights the roads and vegetation classes very accurate.

The second row presents a case of highlighting the border between vegetation and water, proving that the spectral signatures of water and vegetation are more powerful in IR band, that being the reason for a better visibility of boundary between these two. For the new visualization are used:

- R – B7 – NIR
- G – B3 – Visible (Green)
- B – B5 – NIR.

Also, in the third row is represented a case of vegetation and water, but the difference here is the fact that in RGB visualization it cannot be really understood what the image represents because the two classes are illustrated with approximate same color. The improved visualization consists of the following bands:

- R – B7 – NIR
- G – B5 – NIR
- B – B3 – Visible (Green)

These results reveal that a better visualization of the satellite image than the one resulted by RGB can facilitate the activity of a human operator in the process of analyzing data.

#### 4. VISUAL SPECTRAL RELEVANCE TRANSFORMATION BASED ON AUTOENCODERS

Being an artificial neural network, the autoencoder learns efficient representations of the input, called *codings*, without any supervision. Usually these hidden representations have a lower dimensionality than the input data, making this type of deep learning architecture to be very useful for dimensionality reduction tasks.

The two modules of the autoencoder have different functions, the encoder compresses the input  $x$  into a hidden representation  $z$  in the bottleneck layer and the decoder uses  $z$  to reconstruct  $x$ , the result being  $y$ . The differences between  $x$  and  $y$  should be minimum.

Stacked autoencoders are in fact basic autoencoders extended so that the encoding is performed by a sequence of nonlinear layers and the decoding by the same amount of layers. The final result is also the reconstruction of input. The architecture is typically symmetrical with regards to the bottleneck layer.

SAE used in this paper contained 3, 5 and 7 hidden layers. For the 3 hidden layers network, the selected topology was “12-6-K-6-12”, where K is the same for the all three examples and equals to 3. For the 5 hidden layers network, the topology selected was “12-8-6-K-6-8-12” and for the one with 7 hidden layers the topology was “12-10-8-6-K-6-8-10-12”.

The input for the network was represented by a matrix  $X \in \mathbb{R}^{N \times m}$ , where  $N$  represents the number of input vectors and  $m$  is the number of features describing each input, in this case 12. There are not 13 features because the dataset used had only 12. The activation function used is the Exponential Linear Unit (ELU), initialization is done using HE, introduced by [11] and L2 regularizations are applied.

For this paper, the result of the bottleneck layer is the one that gives the compressed representation of the image, embedding the information from all spectral bands.

The results in Fig. 3 reveals the comparison between the true color representation, on the first column and the false color representation resulted from the stacked autoencoders.

The first row represents an urban area, where the buildings are not so discernable in the RGB representation, but in the second one their orientation is very clear. The 3 hidden layer stacked autoencoder was used to obtain the second result on the row.

The second row reveals a case where two areas that look the same in the RGB display, represent different concepts in reality. The area in the yellow square looks very similar to a vegetation region, but in reality it is water, as it can be observed from the false color result. A stacked autoencoder



Fig. 3 Visual spectral relevance transformation based on autoencoders. a) first row – Ilfov, Romania; b) second row – Ciocanesti, Romania; c) third row – Sulina, Romania

with 5 hidden layers was used to obtain the second result on this row.

The third row highlights a case in which two areas that look different, contain in fact the same concept in the image. In the RGB representation, the two yellow squares mark two surfaces that seem totally different, but in the false color image it can be observed that both of them contain water. The pseudo-color result was obtained using a 7 hidden layers stacked autoencoder.

### 3. CONCLUSION

A more explicit visualization of the EO data can help the human analyst to faster understand which are the main characteristics of the image.

Both methods presented in this paper exhibit a better result than the RGB representation, highlighting aspects that are hidden in the true color display. The first method ranks the bands using the mRMR criterion and selects the first three to represent the image. The second one uses all the information in the spectral bands, compresses it and displays the codification resulted as a pseudo-color representation.

At a quick glance, comparing the results presented in Fig. 4, it can be observed that the DNN method reveals some details about the characteristics of the water that mRMR and RGB representations hide.

As a final conclusion, the paper presented two methods to represent an EO image that are different from the true



Fig. 4 Final comparison. a) RGB b) mRMR result c) stacked autoencoder result

color one and that gradually improved the visualization for the analysis.

### 4. REFERENCES

- [1] C. Wang and H.W. Shen “*Information Theory in Scientific Visualization*” *Entropy* 2011, 13, 254-273; doi:10.3390/e13010254, 21 Jan 2011
- [2] D. Bratanu, I. Nedelcu and M. Datcu , “*Interactive Spectral Band Discovery for Exploratory Visual Analysis of Satellite Images*”, IEEE Journal of selected topics in applied Earth Observations and Remote Sensing, vol. 5, no. 1, Feb 2012
- [3] H. Peng, F. Long, and C. Ding, “*Feature selection based on mutual information. Criteria of Max-Dependency, Max-Relevance and Min-Redundancy*”, IEEE Trans. Pattern Anal. Machine Intell., vol. 24, no.8, pp. 1667–1238, Aug. 2005.
- [4] A. Holzinger, C. Biemann, C.S. Pattichis and D.B. Kell, “*What do we need to build explainable AI systems for the medical domain?*”, arXiv:1712.09923v1, 28 Dec 2017
- [5] K. Simonyan and A. Zisserman, “*Very deep convolutional networks for large-scale image recognition*”, arXiv:1409.1556, 2014.
- [6] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, “*Dermatologist-level classification of skin cancer with deep neural networks*”, Nature, 542(7639):115{118, 2017. doi: 10.1038/nature21056.
- [7] P. Wang, W. Li, J. Wan, P. Ogunbona and X. Liu, “*Cooperative Training of Deep Aggregation Networks for RGB-D Action Recognition*”, arXiv:1801.01080v1, 5 Dec 2017
- [8] R. Schwartz-Ziv and N. Tishby, “*Opening the black box of Deep Neural Networks via Information*”, arXiv:1703.00810v3, 29 Apr 2017
- [9] S. Yu and J. C. Principe “*Understanding Autoencoders with Information Theoretic Concepts*”, arXiv:1804.00057v1, 30 Mar 2018
- [10] N. Jacobson and Gupta, “*Design goals and solutions for display of hyperspectral images*”, IEEE Trans. Geosci. Remote Sens., vol. 43, pp. 2684–2692, 2005.
- [11] K. He, X. Zhang, S. Ren and J. Sun, “*Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*”, arXiv:1502.01852v1, 6 Feb 2015