U.P.B. Sci. Bull., Series C, Vol. 83, Iss. 2, 2021 ISSN 2286-3540

AUTOENCODER BASED METHOD FOR MULTISPECTRAL IMAGES VISUALIZATION

Iulia COCA (NEAGOE)[1], Mihai COCA[2], Mihai DATCU[3]

Current visualization techniques of Earth Observation (EO) multispectral data rely on limited settings depending on the combination of three spectral bands or the usage of the true-color representation. The visual information is incomplete, many spectral similarities or dissimilarities remain hidden to the human visual perception. We propose a deep learning-based approach translating the multispectral signatures into R – G – B images. A stacked autoencoder (SAE) is used to embed the information from all spectral bands into three bands. The resulted bands are further used for the false-color representation of the image. The experimental evidences using both Sentinel 2 and Landsat 8 datasets show that the proposed method improves the visualization performance compared with classical methods like R – G – B, minimum redundancy maximum relevance criterion (mRMR) and principal component analysis (PCA). Mutual information quantization is used for results evaluation.

Keywords: SAE, visualization, multispectral images, remote sensing

1. Introduction

The rapid technological progress in the field of remote sensing (RS) determined the exponential growth of earth observation multispectral (MS) image collections. Furthermore, due to the increased spectral diversity and resolution of the operable sensors, the observed details about the land cover scene became more numerous and more accurate. Data analysis and further exploitation in a wide category of applications (e.g. urban monitoring, oil spills or assessment of forest deforestation) became challenging tasks due to high data complexity. The visual inspection of MS images is an important step in screening, browsing, selecting regions of interest, and in all Human Machine Communication Interfaces, with a rapid expansion of active learning paradigms [1]–[3]. Remote sensing multispectral images represent products with a multitude of spectral bands of which only three from the visual part of the spectrum. Current visualization techniques of Earth Observation (EO) multispectral data rely on limited settings depending on the combination of three bands or the usage of the true-color

[1] Eng., Faculty of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania, e-mail: neagoe.iulia@yahoo.ro
[2] Eng., Faculty of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania, e-mail: coca.mihai5@gmail.com
[3] Prof., Faculty of Electronics, Telecommunications and Information Technology, University POLITEHNICA of Bucharest, Romania, German Aerospace Center DLR

representation. The visual information is incomplete, many spectral similarities or dissimilarities remain hidden to the human visual perception. Hence appears the necessity to find optimal methods capable to support the human operator to extract the most important information. Visualization techniques can highlight the utmost relevant information, enhancing the visual perception [4]. To this purpose, we propose a deep learning-based approach translating the multispectral signatures into R – G – B images.

Recently, the authors of [20] presented a method of colorization into the computer vision domain by "hallucinating a plausible color version of the photograph" and transforming a greyscale image to a colored image. Due to the fact that the dimensionality transformations are in opposition regarding the two domains, computer vision and remote sensing, the approaches to improve visualization are different. To visually analyze multispectral data the most common technique used in the past decades is based on feature extraction through dimensionality reduction followed by classification. The extracting features stage is performed using one of the existing methods: supervised or unsupervised. Using supervised feature extraction methods there are needed samples with known labels to increase dissimilarity among classes. Two of the most popular techniques of this kind are the linear discriminant analysis (LDA) [5] and nonparametric weighted feature extraction (NWFE) [6]. Unsupervised dimensionality reduction for feature extraction algorithms does not require any labels and automatically extracts features from spectral signatures data. Principal component analysis (PCA) is one of the well-known unsupervised methods and has been widely used for hyperspectral image processing [7]. Recently, deep neural networks (DNN) and in particular autoencoders (AE), have gained much attention into the field of dimensionality reduction [8]–[10]. All of these works follow the desire of obtaining the most accurate classification of the encoded representation.

Another way to visually analyze large data sets consists of studying the 3D representations of different classes obtained with different dimensionality reduction algorithms [11]. The main purpose of these methods is to diminish the information loss determined by the process.

On the other hand, a solution to visualize a multispectral remote sensing image consists of projecting it using the spectral bands acquired at wavelengths from the visual part of the spectrum, obtaining a natural color representation. But, displaying multispectral data in any three bands combination mapped to R – G – B channels generates incomplete information due to the fact that all the information contained by the other unused bands is lost. Fig.1 demonstrates this assumption by revealing through comparison the representations of true-color images and pseudo-color ones. Apparently similar regions in true color representation turn out being different in false color representation. Same statement is true for dissimilar regions from the true color representation that turn out to be similar in the false color one.

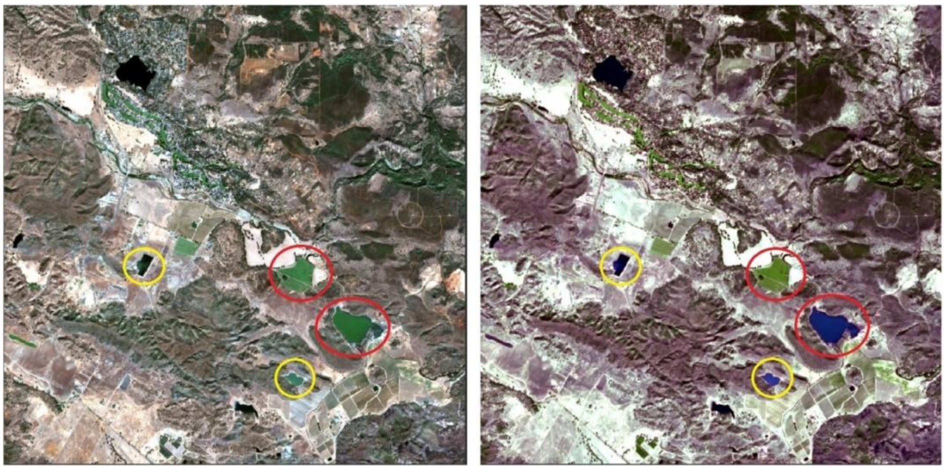

Fig. 1 Sentinel 2 scene revealing false similarities and dissimilarities in true color representation a) true-color representation with two similar regions (highlighted with red circles) and two apparently dissimilar regions (highlighted with yellow) b) the representation of our method showing true similarities and dissimilarities.

Multispectral data visualization using R – G – B channels has been the main topic of a large amount of research. Addressing the issue, the authors of [12] use minimum redundancy maximum relevance criterion (mRMR) for the selection of the optimal spectral bands. mRMR computes the mutual information between bands and determines the ones that contain the maximum amount of information relevant to a specific class. The criterion mentioned was first introduced by the authors of [13] and is based on the combination of two other distinct criteria: maximum dependency and minimum redundancy.

In [12], after determining the spectral bands that contain the largest amount of information relevant to the target class and mapping them to the R – G – B color space, the contrast between the depicted class and the surrounding classes is emphasized at both stimuli and response levels. Although at the response level, evaluation could be considered subjective due to the involvement of the human visual system, at stimuli level the quantization is performed in terms of color distance, by using the Euclidean distance in both device independent and dependent color models.

The purpose of this paper is to demonstrate that accurate visualization helps to distinguish the similarities and dissimilarities between scene objects and to better sense what the scene hides. To this goal, we propose a method which extracts the information from all spectral bands and embeds it to further render the scene content in a pseudo-color image, as close as possible to the natural color. The embedding is accomplished using a stacked autoencoder based method. A preliminary version of the method is also presented in our paper, [14]. Significant changes have been performed both in terms of neural network architecture and with respect to the training data set, the current results being more accurate. To emphasize the effective gain of the current approach, we compare the result of our method with three existing exploratory visualization methods: true-color representation, the representation resulted from the implementation of mRMR criterion and PCA representation. We evaluate our results by using both mutual information quantization and visual analysis.

The organization of the paper is as follows. Section 2 details the implementation of the proposed method while section 3 illustrates the experimental results. Some conclusions and further perspectives are exposed in section 4.

2. Stacked Autoencoder method to enhance visualization

Autoencoders are neural networks able to learn hidden representations of the input data in an unsupervised way. Due to the fact that their dimensionality is lower and still retain as much information as possible, these representations are very suitable for embedding tasks.

The autoencoders encompass two modules, an encoder and a decoder. The encoder compresses an input, *X*, and its result at the bottleneck layer is a hidden representation, *H*. The decoder computes the inverse transformation by trying to reconstruct the input *X*, using *H*. Its result, *Y*, should be as much as possible similar to *X*. AE satisfies:

$$\boldsymbol{H} = f(w_H \boldsymbol{X} + b_H) \tag{1}$$

$$\boldsymbol{Y} = f(w_Y \boldsymbol{H} + b_Y) \tag{2}$$

where H is obtained from X by the weights w_H and common bias b_H and the reconstructed output Y supposed to match X, is obtained directly from the bottleneck layer output H by w_Y and b_Y. The function *f* represents the activation function, which introduces nonlinearity to the network.

By minimizing the error between X and Y, the AE is trained, and the parameters obtained are optimized:

$$\underset{w_H, w_Y, b_H, b_Y}{\arg\min} \; [error(X, Y)] \tag{3}$$

Stacked autoencoders (SAE) represent the extended version of the basic autoencoders, the main difference being the fact that both encoding and decoding operations are computed by different sequences of layers. The symmetry with respect to the bottleneck layer is maintained. Slower in computational terms than Rectified Linear Unit (ReLU) but faster converging during training, the exponential linear unit (ELU) is an activation function that computes the output *z* by the following equation:

$$ELU_{\propto}(z) = \begin{cases} \propto (\exp(z) - 1), & z < 0 \\ z, & z \geq 0 \end{cases} \tag{4}$$

As presented in [17], a very suitable initialization method is HE (abbreviation comes from the name of the author who developed it) taking into consideration the good results obtained in combination with ELU activation function. In this paper the model of the neural network used to demonstrate the visualization gain of an embedded representation was in accordance with the theoretical aspects presented above. Likewise, due to the fact that SAE is based on an unsupervised approach, we embraced this solution. In the interest of obtaining a false-color representation as more alike as the natural color one, we augment the error function already defined in eq. (3) with the following minimization:

$$\underset{w_H, b_H}{\arg\min}[color\ difference(RGB_X, H)] \tag{5}$$

where *color difference(X,H)* is computed using the Euclidean distance and RGB_X represents the true color representation of the input. We chose this method based on the analysis performed by authors of [12] in terms of stimuli level evaluation.

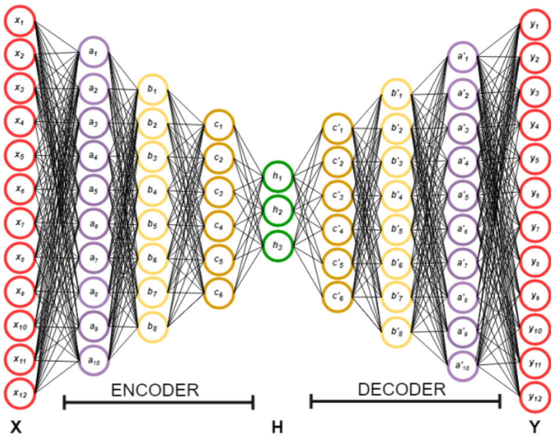

Fig. 2 SAE architecture used for the implementation of the proposed method

The proposed architecture is presented in Fig. 2 and consists of 7 hidden layers, 4 SAE marked with different colors (the bottleneck layer drawn with green is represented differently to highlight the hidden representation), the number of inputs and outputs equals 12 and the topology used is “10-8-6-3-6-8-10”. The bottleneck layer is chosen to have 3 outputs so that these values to be translated to the R – G – B channels and result a false-color representation. To obtain a model suitable for both Sentinel 2 and Landsat 8 products, the training dataset is a concatenation of two scenes, one scene for each type of sensor. As Sentinel 2 scene we used S2MSI2A product type, processed at level 2, which does not contain the cirrus band 10, thus resulting 12 bands product. The reason behind not including this band stands in the fact that it does not contain surface information.

Taking into the consideration that Landsat 8 has only 11 bands, we created a false band for this type of products and initialized it by 0 values (*i.e.* zero padding). Before concatenating the two datasets, we normalized their values separately due to values dynamic ranges. The dimension of each scene was 4600x4600x12, thus resulting in a 4600x9200x12 training set.

The tridimensional matrix *M* x *N* x *F* representing the training dataset (*M* is the number of lines, *N* is the number of columns and *F* the spectral signatures) had to be reshaped into *(M * N)* x F so that each line should represent the bands corresponding to one pixel. The differences between spectral intervals or resolutions among sensors are hidden from the network, its main purpose being to embed the whole information from the input into hidden representations; hence these details are not important. The products acquired by Sentinel 2 sensor were resampled to 10 m resolution and Landsat 8 products were resampled to 30 m resolution. The up-sampling used method is based on the value of the nearest neighbor. In order to scale out DNN training for reducing model training time, deep learning applications in RS field need to address GPU implementations to the detriment of multi-core CPU implementations. We ran the deep learning tests presented in this work leveraging high-performance cuDNN computation library on TensorFlow distributed system. Our algorithm implementation adopted data parallelism paradigm to correlate model training across a single node with 8 PCIe-connected K80 GPUs.

3. Results evaluation

In this paper we propose a method of diminishing ambiguities resulted through the visualization of remote sensing MS images. The method consists of embedding the information from all spectral bands into a hidden representation using a deep neural network. To prove the superiority of the proposed method over existing methods we chose two classical visualization methods: rendering bands from the visual part of the spectrum (4, 3, 2) and rendering the combination of the first three bands ranked using the mRMR criterion. Besides these classical visualization methods, we also used as comparison a well-known dimensionality reduction method, PCA. The bands selection according to the mRMR criterion was obtained using the DAS-Tool plugin of SNAP [18], which implements the algorithm according to [13].

PCA is one of the most popular methods for data reduction. It represents a linear transformation based on the covariance matrix and its eigenvalues. Keeping only the largest values of the number of dimensions which reduces space, *d*, the algorithm computes the covariance matrix and the eigenvalues. The kept eigenvalues associated with the eigenvectors, assemble the transformation [19].

The results evaluation involves both visual and mutual information analysis. Fig. 3 represents the visual representations of the true-color, PCA, mRMR and proposed method (SAE).

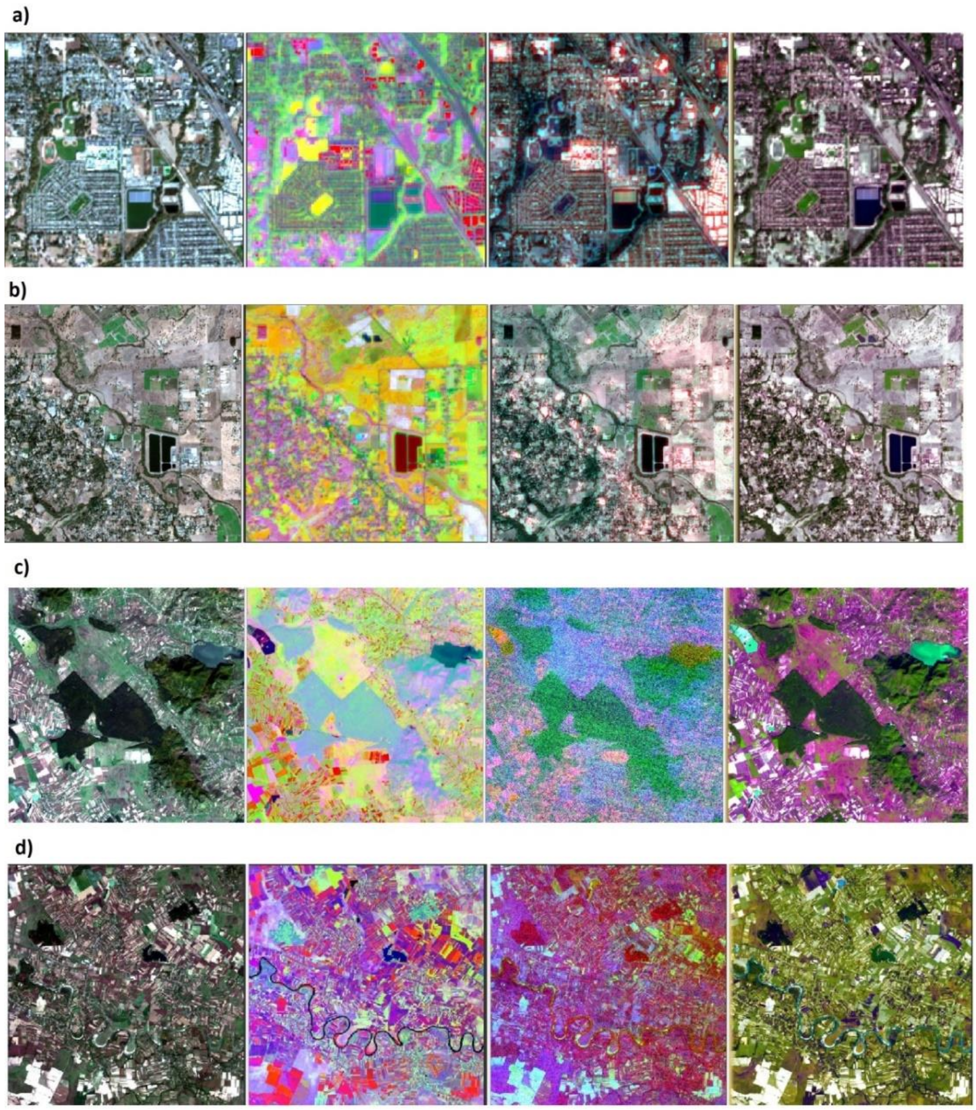

Fig.3 Experimental results of two scenes from Sentinel 2 and two scenes from Landsat 8; 1st column: true-color, 2nd column: PCA, 3rd column: mRMR and 4th column: SAE (our method); a) Sentinel 2 S1; b) Sentinel 2 S2; c)Landsat 8 S1; d) Landsat 8 S2.

The first two rows represent two Sentinel 2 sub-scenes of the image acquired on 31st of October 2020, over Sacramento, California. The last two rows represent Landsat 8 sub-scenes acquired on 16th of October 2017 over Cluj-Napoca, Romania. Visually, it may be concluded that the proposed method succeeds to embed the whole information while preserving the color code of the objects. An additional proof is represented by the colorization of the result determined by the presence of a band with a determinant spectral signature artifact, as it can be observed in Fig. 3, example c). PCA also accomplishes to show some differences that are not visible in the R – G – B representation, but the display do not resemble with a true color image. mRMR criterion could be misled by anomalies encountered in different spectral bands, like it happens in example c). One of the most relevant and less redundant bands is computed to be the 11th one, which contains a lot of noise, thus distorting the resulted representation. However, when the three bands combination selected includes at least one band from the visual spectrum, the resemblance with the true color representation is distinguishable.

To sustain the enhancement obtained with the proposed method regarding the amount of information preserved into the embedding, we chose mutual information-based quantization. Mutual information is computed band-wise between two products with the same number of spectral bands and same dimension, where dimension is represented by the product M x N, where M is the number of lines and N is the number of columns. The following equation determines the performed computation:

$$\boldsymbol{MI(A,B)} = \sum_{i=1}^{|A|}\sum_{j=1}^{|B|}\frac{|\boldsymbol{A_i} \cap \boldsymbol{B_j}|}{\boldsymbol{D}}\log\frac{\boldsymbol{D}|\boldsymbol{A_i} \cap \boldsymbol{B_i}|}{|\boldsymbol{A_i}||\boldsymbol{B_i}|} \quad (6)$$

where, A is the histogram of one band in the first product and B is the histogram of the same band number from the second product, $|A_i|$ represents the number of values in bin A_i and $|B_j|$ represents the number of values in bin B_j. D is the dimension of the products. For MI implementation we used *mutual_info_score* from the *sklearn.metrics* python library.

Table1

Mutual information quantization analysis

MI / Dataset	Input/ Decoder %	R – G – B/ Encoder %	R – G – B/ mRMR %	R – G – B/ PCA %	mRMR/ PCA %	mRMR/ Encoder %	PCA/ Encoder %
Sentinel 2 S1	**0.912**	0.786	0.718	0.795	0.692	0.679	0.803
Sentinel 2 S2	**0.936**	0.812	0.684	0.787	0.718	0.743	0.804
Landsat 8 S1	**0.847**	0.743	0.472	0.760	0.511	0.518	0.764
Landsat 8 S2	**0.832**	0.757	0.484	0.764	0.507	0.518	0.765

In Table 1 are expressed the percentages of mutual information between different obtained representations. The percentages were computed having as reference the maximum MI computed between a representation and itself.

As it can be observed, the amount of MI between the input and the decoder result is large for all testing datasets, meaning that the decoder succeeds to rebuild the input from the encoded representation. Between R – G – B (true-color) and both PCA and encoder results, the percentages are at a satisfying level because both retain information from the true color representation, although visually, only for SAE method is noticeable. mRMR and R – G – B obtain greater MI percentages only if the selected bands for mRMR include bands from the visible spectrum. As far as it goes with the amount of MI between the encoder result and PCA, it is explainable the high level because both try to embed information from all spectral bands. The small amount remained is also noticeable in the visualization analysis.

4. Conclusions

This paper aims to demonstrate a new approach to enhance the three band combination visualization of remote sensing multispectral images, by means of perceptive deep sensing of a SAE based method which embeds the information contained in all spectral bands. Based on an unsupervised training approach, a deep neural network is used to generate hidden representations, which will be further mapped into the tri-dimensional space.

The validation procedure uses as reference the natural color, mRMR and PCA representations. Visual and mutual information analysis is performed to prove the superiority of the recommended solution in terms of diminishing the number of ambiguities between similarities and dissimilarities and preserving the color code for the objects inside the scene.

The proposed method succeeds in generating accurate displays of the scenes and could represent a great help in terms of sensing and visual analysis of multispectral remote sensing images once integrated in various active learning systems.

Acknowledgement

The work has been funded by the Operational Programme Human Capital of the Ministry of European Funds through the Financial Agreement 51675/09.07.2019, SMIS code 125125.

R E F E R E N C E S

[1] *Espinoza-Molina D., Manilici V., Dumitru C., Reck C., Cui S., Rotzoll H. & Datcu M*. "The Earth Observation Image Librarian (EOLib): The Data Mining Component of the TerraSAR-X Payload Ground Segment.", Big Data from Space (BiDS'16) (2016): 228-231.

[2] *Demir B., Bruzzone L.*, A novel active learning method in relevance feedback for content-based remote sensing image retrieval. IEEE Transactions on Geoscience and Remote Sensing. 2014 Oct 13;53(5), pp.2323-34.

[3] *Xia G.S., Tong X.Y., Hu F., Zhong Y., Datcu M., Zhang L.*, "Exploiting deep features for remote sensing image retrieval: A systematic investigation", in *IEEE Transactions on Big Data*, vol. 6, no. 3, pp. 507-521, 1 Sept. 2020, doi: 10.1109/TBDATA.2019.2948924.

[4] *Wang C., Shen H.W.*, "Information theory in scientific visualization", Entropy. 2011 Jan 13(1), pp.254-73.

[5] *Chang C.I., Ren H.*, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery". IEEE Transactions on Geoscience and Remote Sensing. 2000 Mar;38(2), pp. 1044-63.

[6] *Kuo B.C., Landgrebe D.A.*, "Nonparametric weighted feature extraction for classification", IEEE Transactions on Geoscience and Remote Sensing. 2004 May 18;42(5), pp.1096-105.

[7] *Plaza A., Martínez P., Plaza J., Pérez R.*, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations", IEEE Transactions on Geoscience and remote sensing. 2005 Feb 22;43(3), pp. 466-79.

[8] *Mei S., Ji J., Geng Y., Zhang Z., Li X., Du Q.*, "Unsupervised Spatial–Spectral Feature Learning by 3D Convolutional Autoencoder for Hyperspectral Classification", IEEE Transactions on Geoscience and Remote Sensing. 2019 Apr 22;57(9), pp. 6808-20.

[9] *Koda S., Melgani F., Nishii R.*, "Unsupervised Spectral–Spatial Feature Extraction With Generalized Autoencoder for Hyperspectral Imagery", IEEE Geoscience and Remote Sensing Letters. 2019 Jul 17;17(3), pp.469-73.

[10] *Wang J., He H., Prokhorov DV.*, "A folded neural network autoencoder for dimensionality reduction", Procedia Computer Science. 2012 Jan 1; 13:120-7.

[11] *Griparis A., Faur D., Datcu M.*, "Feature space dimensionality reduction for the optimization of visualization methods", IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2015 Jul 26, pp. 1120-1123.

[12] *Bratasanu D., Nedelcu I., Datcu M., "*Interactive spectral band discovery for exploratory visual analysis of satellite images", IEEE Journal of selected topics in applied Earth Observations and Remote Sensing. 2011 Nov 8;5(1): pp. 207-24.

[13] *Peng H., Long F., and Ding C.*, "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, Aug. 2005, pp. 1667–1238.

[14] *Neagoe I., Faur D., Vaduva C., Datcu M.*, "Exploratory visual analysis of multispectral EO images based on DNN", IEEE International Geoscience and Remote Sensing Symposium (IGARSS) 2018, pp. 1-4.

[15] *Clevert D-A., Unterthiner T. Hochreiter S.*, "Fast And Accurate Deep Network Learning By Exponential Linear Units (ELUs)", International Conference on Learning Representation, 2016.

[16] *Glorot X., Bengio Y.*, "Understanding the difficulty of training deep feedforward neural networks". In Proceedings of the thirteenth international conference on artificial intelligence and statistics 2010 Mar 31, pp. 249-256.

[17] *He K., Zhang X., Ren S. and Sun J.*, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification", in Proc. IEEE international conference on computer vision 2015, pp. 1026–1034.

[18] *Grivei A.C., Neagoe I.C., Georgescu F.A., Griparis A., Vaduva C., Bartalis Z., Datcu M.*, "Multispectral Data Analysis for Semantic Assessment—A SNAP Framework for Sentinel-2 Use Case Scenarios", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2020 Jul 30;13:4429-42.

[19] *Van Der Maaten L., Postma E., Van den Herik J.*, "Dimensionality reduction: a comparative review", J Mach Learn Res. 2009 Oct 26, 10(66-71):13.

[20] *Zhang R., Isola P., Efros A. A.*, "Colorful image colorization", European conference on computer vision, Springer, Cham, October 2016, pp. 649-666.