POLITEHNICA UNIVERSITY OF BUCHAREST

Doctoral School of Electronics, Telecommunications and Information Technology

Decision No. 907 **from** 15-09-2022

Ph.D. THESIS SUMMARY

Eng. Iulia COCA (NEAGOE)

METODE NEURONALE MULTINIVEL PENTRU TELEDETECȚIE: RECONSTRUIREA ȘI PREDICȚIA IMAGINILOR SATELITARE

SENSOR DEEP LEARNING: RECONSTRUCTION AND PREDICTION OF MULTISPECTRAL SATELLITE IMAGES

THESIS COMMITTEE

Prof.Dr.Eng. Mihai CIUC Politehnica Univ. of Bucharest	President
Prof.Dr.Eng. Mihai DATCU Politehnica Univ. of Bucharest	PhD Supervisor
Prof.Dr.Eng. Cătălin Daniel CĂLEANU Politehnica Univ. of Timisoara	Referee
Prof.Dr.Eng. Alexandru BADEA Romanian Space Agency	Referee
Conf.Dr.Eng. Daniela FAUR Politehnica Univ. of Bucharest	Referee

BUCHAREST 2022

Table of contents

Chapter 1

Introduction

Over time, the sensors used to capture information about the Earth's surface have evolved, and nowadays sensors mounted on satellites orbiting the Earth are used for this purpose.

1.1 Presentation of the field of the doctoral thesis

Remote sensing observations of land cover using optical sensors are generated in the form of multispectral images. They constitute a collection of information grouped in several bands, each band being acquired at a certain wavelength. Visual analysis is generally performed by a human operator using the mapping of the three bands of the visual spectrum onto RGB (red-green-blue) channels, thus obtaining a representation as close as possible to natural color. The purpose of this analysis is to identify the key aspects corresponding to the target application, e.g. flood monitoring, fire monitoring, detection of changes in urban planning, etc.

Nevertheless, multispectral imaging is sensitive to certain atmospheric conditions that may affect the visual appearance of the images. Clouds, smoke or fog are different types of atmospheric phenomena highlighted in the images, especially when using the classic RGB viewing technique.

Multispectral images can also be affected by artificial artefacts caused by sensor's malfunction in the form of certain types of noise or even the complete absence of a band.

1.2 Scope of the doctoral thesis

The thesis aims to address both topics: visualization improvement and band recovery, having into consideration the theoretical study and elaboration of advanced methods in RS imagery. The methods are intended for the analysis of multispectral images, having the goals to develop a framework for image visualization reconstruction containing the whole information consisted by all spectral bands, and a solution to predict a missing or corrupted band using only the information from the others spectral bands.

1.3 Content of the doctoral thesis

According to the main research directions followed during the PhD, this thesis is structured according to the following chapters:

Chapter 2 (Remote sensing basics and challenges) presents the main notions of remote sensing that are subsequently used in the thesis, the properties of multispectral satellite images and how to interpret them, distinctive features for the missions whose products were used in the experimental process, as well as the obstacles encountered in the analysis of multispectral images, which were the main premises for defining the main objectives of the thesis.

Chapter 3 (Current context and previous researches) introduces the current context and the types of solutions addressed so far to solve the two target challenges of this thesis. The last section briefly illustrates, with reference to previous state-of-the-art research, the approach used to implement the proposed new methods.

Chapter 4 (Information recovery for visualization) describes in detail the set of methods developed to value the spectral information available in all bands. The first section of this chapter proposes a suite of five methods based on a stacked autoencoder (SAE) to transfer information between bands in order to reconstruct an improved visualization. The second section proposes a multi-sensor compatible visualization enhancement method.

Chapter 5 (Information valorization for band reconstruction) describes in detail the method of reconstructing a band using prediction based on the spectral information available on the concurrent bands.

Chapter 6 (Conclusions) consists of four sections that follow the results conclusions and contributions of the thesis, as well as a list of published papers followed by future research perspectives.

Chapter 2

Remote sensing basics and challenges

In this chapter are presented the basic and general principles of remote sensing so that the context and the data being analyzed to be comprehended.

2.1 Basic understanding

A formal definition for remote sensing could be the methodology of information acquisition by a sensor about an object without being in contact with it. During the information acquisition there is an interaction of energy between the sensor and the target. In this process, the signal detected by the sensor could be represented by solar energy emitted by the Sun and reflected from the Earth's surface, energy emitted by the surface itself or even the sensor's own pulses emitted and reflected. After detecting and measuring the energy, the satellite sensor passes the information to a receiving system for a proper interpretation. The step of analyzing and interpreting the acquired data also is part of remote sensing and depends on the following knowledge: acquisition process, physical basis and the methods used to process the initial data,[2].

2.2 Remote sensing systems and signals

As presented in the previous section, remote sensing represents the activity of remotely sensing information about something without being in contact with it. The sensed information is represented by electromagnetic (EM) signals which come from different objects with particular chemical and physical properties. In remote sensing, EM energy can be sensed as reflection, emission or combined, emission-reflection. Having into consideration the three ways of sensing energy, there are two types of sensors: *passive* and *active*. The difference between these two is the fact that active sensors have the ability to emit pulses, not only to receive them. According to their wavelength, the EM radiations are split into seven categories: *gamma rays, x-rays, ultraviolet, visible, infra-red, microwave and telecommunication, TV and radio*.

2.3 Remote sensing images

Optical passive sensors measure EM signals coming from multiple spectral regions, each region being considered a spectral band. The number of bands classifies the obtained concatenated result into panchromatic, multispectral, superspectral or hyperspectral image.

In this thesis, the attention is concentrated over optical sensors and multispectral images. There are multiple satellite missions, but the products of Sentinel 2 and Landsat 8 are the ones researched into following experiments. Multispectral images are characterized by several resolution types which are defined by the sensor which acquires them: spatial, spectral, radiometric and temporal resolution.

Translating the original radiance values into brightness levels represents the process of generating a digital image. The smallest unit which contains brightness value and is the minimum spatial unit in an image, is called a "pixel". Each band of a multispectral image is defined by a matrix of pixels. As the display of an image is limited to three bands, the visualization of a multispectral EO product is usually performed by mapping the three bands of the visual part of the spectrum corresponding to the *reg - green - blue* colors to $R - G - B$ (RGB) channels.

2.4 Remote sensing multispectral image analysis obstacles

A large part of multispectral images are covered to varying degrees by various atmospheric phenomena that prevent a proper analysis. Clouds are the main obstacle encountered in most examples, but there may also be areas covered by fog, haze or smoke.

Also, in addition to the challenge of analyzing the images obtained in certain atmospheric conditions, there is also the challenge of encountering corrupted images obtained due to some physical defects of the sensor. These obstacles can also be called artificial artifacts, and can be found in the form of noises present on one or more bands. Artifacts can be represented by missing areas, lines of pixels with zero value, or even the complete absence of a band.

Chapter 3

Current context and previous researches

Two major research directions of interest in the field of remote sensing are followed in this thesis: the harness of the information contained in all spectral bands to obtain a complete visualization and the recovery of information to rebuild a damaged or absent band.

3.1 Existing information recovery for visualization purposes

After analyzing the existent methods, their data transformation approach and the purpose of their application, it may be concluded that, in order to improve visualization for visual inspection of a human operator, a method that preserves spatial information and resolution needs to be taken into account. The solution has to be capable of learning from the spectral space and not from the spatial one. The autoencoder learns to embed information in an unsupervised way and does not perform any spatial mediation, thus proving to be a suitable solution to the problem mentioned.

3.2 Lost or degraded information recovery

Over time, different methods have been discovered to approach the subject of reconstructing a degraded or missing band. They can be differentiated according to the basic used principle: the information needed to perform the reconstruction. Thus, depending on the source of this information, four types of methods can be distinguished: spatial-based, spectral-based, temporal-based (multitemporal) and mixed (hybrid) methods [13].

3.3 New methods to address remote sensing images analysis obstacles

Since there are different types of obstacles that generate diverse problems, the solution methods have also to be distinct. The obstacles types presented in the previous chapter determine the analysis overburden of multispectral images as follows:

- atmospheric phenomena obstruct mainly the visual analysis process;
- artificial artifacts influence any kind of subsequent analysis.

Thus, two different challenges can be developed: improving the visualization to facilitate the visual analysis process and recovering the missing information from a certain band to improve the subsequent digital processing. Both challenges can be addressed based on the fact that in a multispectral image, the information available from all bands must be exploited. Thus, in order to improve the visualization of images covered by clouds, haze or smoke, the information available on all spectral bands can be used, imposing on the advantages of bands with longer wavelengths. Also, to recover the information from an affected band, the existing correlation between the other spectral bands can be used to predict the missing data.

Through recent research in the field, the superiority obtained by methods based on neural networks has been demonstrated. Starting from this premise, this thesis proposes solutions based on artificial intelligence to address the challenges mentioned above. The visualization improvement is proposed to be achieved using an autoencoder that embeds the information from all the spectral bands in the form of a latent representation, mapped to the RGB channels. The prediction of the missing band is proposed to be obtained using a convolutional neural network with UNet architecture, which receives as basic information for prediction the bands complementary to the affected band.

The following two chapters explain in detail methodology, results and evaluation of the two main contributions proposed in this thesis.

Chapter 4

Information recovery for visualization

Visualization of multispectral images through band selection methods determines an information loss that in utmost cases proves to be critical for the adequate understanding of the represented scene.

4.1 Cross-bands information transfer to improve visualization

This section presents a set of five original different methods to offset the effects caused by ambiguities, fog, light clouds and smoke by transferring relevant information between bands in order to visually reconstitute those parts of the image affected by atmospheric phenomena.

4.1.1 Proposed concept

The principal objective is to improve visualization by reducing the ambiguities and obstructions generated by the lack of information in the image displayed compared to all spectral bands in the multispectral product. As such, transformations generating illuminant invariant features are applied to reduce obstruction. The aim is to accomplish these objectives without affecting the spatial resolution.

The set of proposed methods is based on a general concept: the use of an autoencoder to embed information from all bands on only three bands. These resulting bands are subsequently mapped on the RGB channels and false color visualization is thus obtained. The methods represent a variation of the actors involved. The motivation behind this diversity is given by the different utility of each of the proposed approaches.

Figure 4.1 illustrates the general concept underlying the five proposed methods. An autoencoder is the core and the common part in defining these methods because it performs the operation of embedding information from several bands to a 3-dimensional latent representation. This obtained representation is then used for mapping on RGB

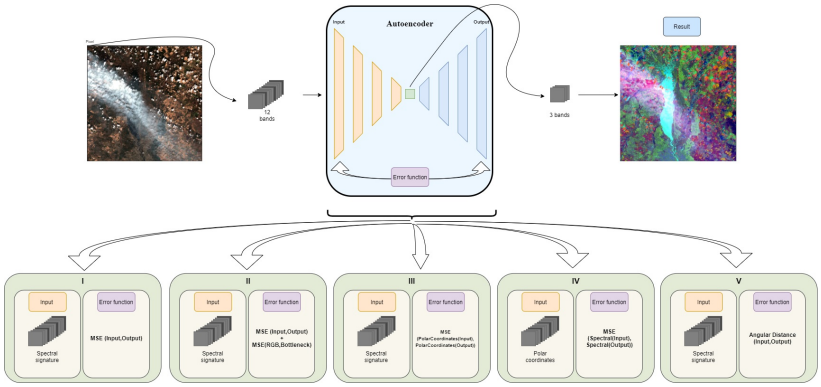


Fig. 4.1 General architecture of the five proposed methods.

channels for visualization. The actors that differentiate the five methods are the way the data enters the network, the input, and the way the error function is computed. The data can enter the network in the form of radiant values or polar coordinates transformed values. Regarding the error function, it starts from a general computation that evaluates the difference between input and output, and changes from one method to another by including additional evaluations or by transforming the compared values into polar coordinates or angles.

4.1.2 Multispectral data representation methods

Multispectral remote sensing images are conventionally represented as a plot of the image features into a multispectral vector space having the space dimension equal to the number of spectral components. The distance between two vectors, A and B, in this space, may be computed using the Euclidean or angular distances.

In order to obtain an improved analysis over multispectral remote sensing images with large cloud coverage or shadows, it was developed an illuminant invariant feature descriptor based on polar coordinate transformation of the reflectance values [10]. Polar coordinate transformation represents a computation with the help of which radiances values are transformed into angles θ and distance ρ. Having a product with N spectral bands, the obtained result consists of N-1 angles and one distance, making the dimensionality of the two objects equal.

4.1.3 Multispectral image compression into three bands

Autoencoders are neural networks that learn in an unsupervised way to reconstruct an input, obtaining a latent representation of smaller dimension inside the network, at the bottleneck layer. The autoencoders include two main structures, an encoder and a decoder. The input of the encoder, X, represents the object to be compressed, and its output is a latent embedded representation of the input, H. The output of the encoder represents the input to the decoder which has the main assignment to reconstruct X using

H. The result of the decoder, Y, is a representation that must be as similar as possible to X, having the same dimensionality.

Stacked autoencoder, SAE, represent an enlarged version of a basic autoencoder. The encoding and decoding operations are performed by sequences of layers and the symmetry relative to the bottleneck layer is preserved.

4.1.4 Enhanced visualization proposed methods

This section proposes five different visualization improvement methods for multispectral remote sensing images based on a SAE neural network, as illustrated in Figure 4.1. The proposed methods represent a diversity of combinations of the previously defined actors, namely: X and error function.

Spectral input – spectral error (SI-SE)

The purpose of this method is to accomplish the first objective of this chapter by revealing the hidden details from an apparently accurately displayed scene. The combination of actors with respect to this method consists of an input of reflectance values and a loss evaluation using MSE, which stands for mean squared error and is very often used to compute the error function in neural network models.

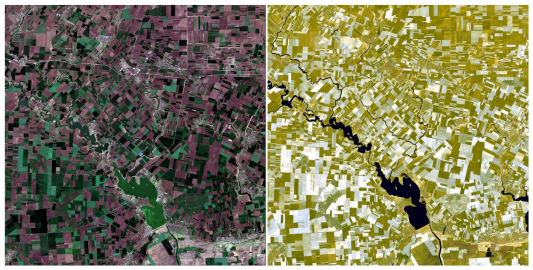


Fig. 4.2 SI-SE visualization comparison.

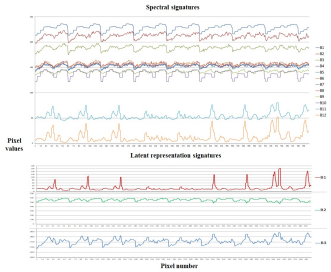


Fig. 4.3 SI-SE spectral signature comparison.

Apparently similar regions could represent different things and apparently different regions could represent the same thing, as shown in Figure 4.2. The information contained by the bands which are not involved into visualization representation may be different, this observation being also depicted from Figure 4.3, where the spectral signatures show different patterns among the spectral bands. The bands from the visible part of the spectrum share the same pattern while, as close as they get to the NIR and SWIR part of the spectrum, the pattern changes. The latent representation signatures demonstrate the embedding ability of the network, capturing all different patterns from all spectral signatures.

Spectral input – spectral error with color correction (SI-SECC)

The second method comes as an improvement to the first one, meaning that besides fulfilling the first objective of eliminating the ambiguities, this method aims to keep the visualization as close as possible to that obtained using the bands from the visual part of the spectrum mapped on the RGB channels. The actors involved are the spectral values and the loss computed over an augmented error function of the previous defined one. Augmentation consists in adding an evaluation in terms of color difference between the latent values, H, obtained by the encoder and the values of the bands from the visual part of the spectrum, RGB_X. Color difference is computed using the Euclidean distance.

Fig. 4.4 SI-SECC visualization comparison.

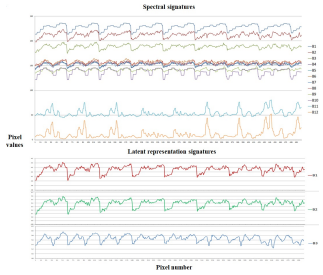


Fig. 4.5 SI-SECC spectral signature comparison.

The comparison between RGB and SI-SECC from Figure 4.4 emphasizes the capability of the autoencoder to embed information from NIR and SWIR bands because the smaller water bodies from the left of the sea are not visible in left representation, but are highlighted with a lighter blue shade in the right one. Figure 4.5 shows the latent signatures pattern preservation along with the tendency of being more similar with the visual spectral signatures pattern.

Spectral input – polar coordinates error (SI-PcE)

This method has been developed in order to satisfy both objectives, reduction of ambiguities and visual contamination caused by clouds, smoke or fog. The input is represented by the spectral features and the loss function implies an error evaluation which computes the MSE between the transformation of X to polar coordinates, $polar_X$, and the transformation of Y to polar coordinates, $polar_Y$.

Watching the graphical representation of the spectral and latent signatures from Figure 4.7, it can be observed that the signatures patterns of the input are preserved and embedded into a three bands combination, each different pattern from the input being dominant over one band in the latent.

Figure 4.6 highlights the better visualization result obtained with SI-PcE. Although the left side of the figure shows a scene covered by clouds, the right one succeeds to disclose the Earth surface.

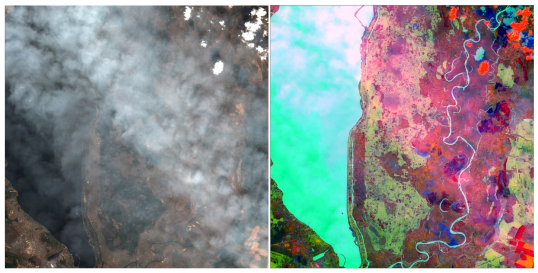


Fig. 4.6 SI-PcE visualization comparison.

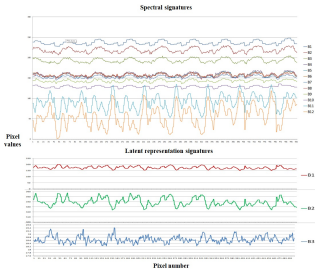


Fig. 4.7 SI-PcE spectral signature comparison.

Polar coordinates input – spectral error (PcI-SE)

This method aims to accomplish both objectives in terms of visualization enhancement of multispectral remote sensing images. Also, an additional objective of this method would be to reduce as much as possible shadows. This supplementary goal comes from the property of polar coordinates to be illuminant invariant. This method represents the inverse, in terms of actors implied, of the previous method.

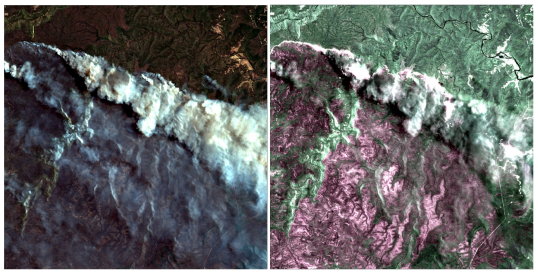


Fig. 4.8 PcI-SE visualization comparison.

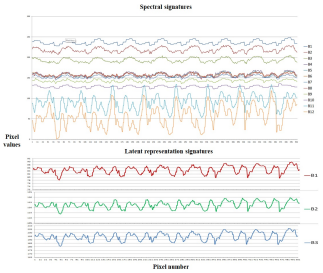


Fig. 4.9 PcI-SE spectral signature comparison.

Figure 4.9 denotes the pattern merge effect that takes place in the latent representation signature, showing that each band resulted is influenced by all input spectral features patterns. Figure 4.8 represents a scene of ongoing fire that has an emphasized visualization using the proposed method. The smoke from the RGB representation is predominantly removed, making smoked areas, remained vegetation and also the fire borderline visible.

Spectral input – angular error (SI-AE)

Visible enhancement by means of ambiguities and atmospheric phenomena reduction are the main objectives of this method. Having as auxiliary purpose to verify the angular distance property of being invariant to linearly scaled variations of spectral values, is developed a method that includes this distance in the error function of the neural network.

Latent representation signatures show a mixed preservation of the spectral ones, each band from the latent space consisting of multiple patterns from the spectral space, Figure 4.11. Ambiguities are eliminated from the visualization, as Figure 4.10 shows.

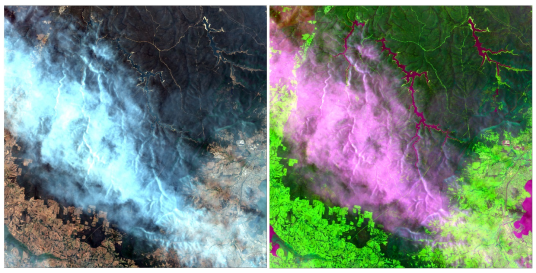


Fig. 4.10 SI-AE visualization comparison.

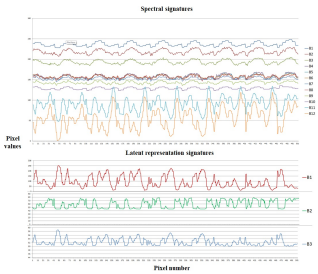


Fig. 4.11 SI-AE spectral signature comparison.

Also, smoke and shadows present in the RGB representation are reduced, demonstrating the illuminant invariance character of angular distance.

4.1.5 Experimental results and demonstration

Sentinel-2 images acquired at different moments of time are used. The footprint of the analyzed scenes covers multiple regions of the world. All the scenes were resampled before use so that all bands shall have 10m resolution. Thus, an up-sampling was performed on bands with 20 m and 60 m resolutions by setting each output pixel to the nearest input pixel value. The resulting products have the same dimension, *10980x10980x12*. The processing level 2 of Sentinel-2 sensor does not contain band 10, so the resulting product contains only 12 bands.

The SAE architecture used by all methods contains four autoencoders. The encoder is defined according to a topology that decreases from 12 inputs to 3 according to the following pattern "12-8-6-3", and the decoder is defined by an ascending topology following the same pattern. Elu activation and Adam as optimizer were used. The hidden representation consists of three values for visualization purposes. The training dataset for all methods consisted of a 4 concatenated subsets of clear, smoky, foggy and cloudy images and had a dimensionality of *10980x10980x12*.

The experimental code was implemented using Python 3.8.5 and TensorFlow 2.3.0 for GPU. To reduce model training time were used a distributed system and parallelized computation across 8 PCIe-connected K80 GPUs. All the visualization operations were performed using Sentinel Application Platform (SNAP).

Results are grouped into four scenarios: *clear images, foggy images, cloudy images and smoky images* and discussed in the next subsections.

Clear images

Even if a scene does not contain atmospheric phenomena that prevent the information about the Earth's surface to be achieved, there can be many hidden details in the spectral bands not included in the production of the visualization.

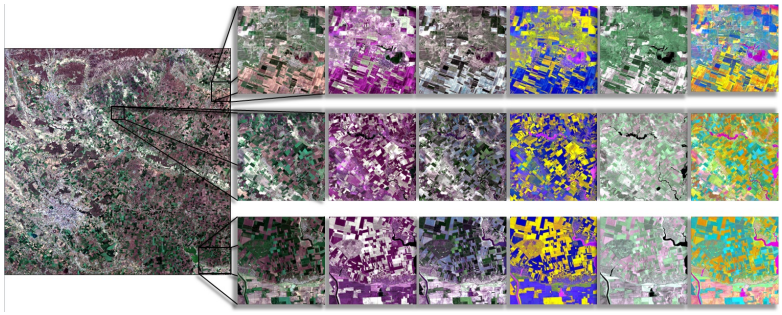


Fig. 4.12 Clear images proposed methods visualization - Bucharest, Romania.

Figure 4.12 demonstrates the benefit of visualizing a representation that contains the information from all spectral bands, making the differences between apparently similar regions to be observable. All the five methods accomplish this discrimination obtaining a contrast between dissimilar regions.

Smoky images

The bands from the visual part of spectrum are often the ones affected by disasters such as fires, and visualizing such a scene using only three bands could cause a loss of information about what is under the smoke. Therefore, enclosing the information present in the NIR and SWIR bands may improve visualization and bring additional information about the Earth's surface.

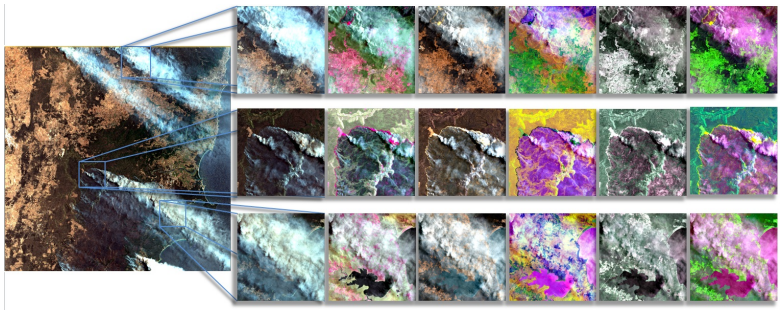


Fig. 4.13 Smoky images proposed methods visualization - Chico, California.

When information about what is under the clouds or smoke is available on at least one of the bands, a view that includes this information is very beneficial in the analysis process. Figure 4.13 shows a scene of ongoing fire and demonstrate the advantage brought by the visualization proposed methods for investigation purposes.

Foggy images

Fog is one of the atmospheric phenomena that can intervene in the visualization and analysis of the Earth's surface. The phenomenon of obstruction of visibility is similar to that encountered in smoke scenes, with the mention that due to the size of the water droplets that make up the fog, their penetration by the wavelengths available on optical

sensors, like Sentinel 2, is often impossible. Retrieving information about the terrestrial aspect becomes impossible in case of a dense fog. Figure 4.14 shows three different cases of retrieving the information contained in all spectral bands, so that the first two lines demonstrate the improvement of the contrast and regions distinction, even if the fog is visible. The third line illustrates a reduction of the surface covered by fog besides improving the contrast.

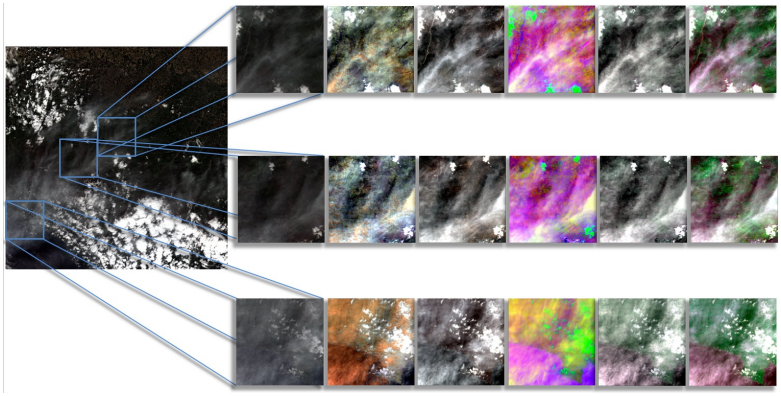


Fig. 4.14 Foggy image proposed methods visualization - Parma, Italy.

Cloudy images

Although clouds are of several types, dense or less dense, at a lower or higher altitude, there are certain situations in which longer wavelengths manage to penetrate them. Most of the time, not even the wavelengths in the SWIR range of the spectrum manage to pass through to capture information about the terrestrial aspect.

Figure 4.15 illustrates an example of a semi-transparent cloud, which allows the observation of scene details.

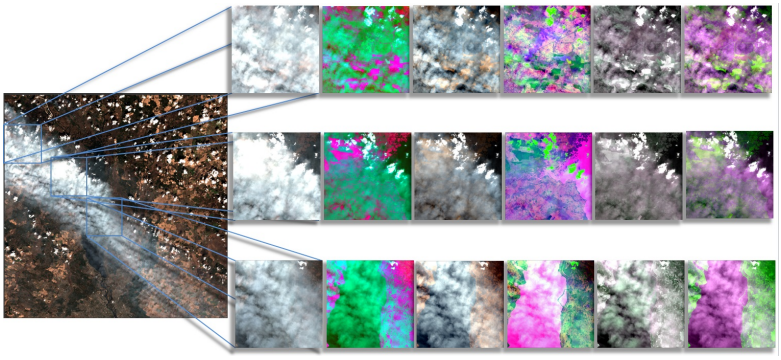


Fig. 4.15 Cloudy image proposed methods visualization - Kyiv, Ukraine.

4.2 Multi-sensor method for multispectral image visualization

The method proposed in this section, msSI-SECC, addresses the objective of eliminating false similarities and dissimilarities between objects through visualization of remote sensing MS images acquired by more than one sensor. So, the research is performed on Sentinel-2 and Landsat 8 images.

msSI-SECC is similar with the ones presented in the previous section regarding the fact that is based on SAE to embed the information from all spectral bands and it does perform a color correction following the same error function as SI-SECC. But, significant changes have been performed both in terms of neural network architecture and with respect to the training data set.

To emphasize the effective gain of the current approach, the results of this method are compared with three existing exploratory visualization methods: true-color representation, the representation resulted from the implementation of minimum redundancy maximum relevance criteria (mRMR) and principal component analysis (PCA) representation.

4.2.1 Proposed msSI-SECC SAE architecture and configuration

In this approach, the model of the neural network used to demonstrate the visualization gain of an embedded representation was in accordance with the theoretical aspects presented in Section 4.1.4, method SI-SECC.

To obtain a multi-sensor 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 is 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, a false band is created for this type of products and initialized it by 0 values (i.e. zero padding). Before concatenating the two datasets, their values are normalized separately due to dynamic ranges values. The dimension of each scene is *4600x4600x12*, thus resulting in a *4600x9200x12* training set.

The products acquired by Sentinel 2 sensor were resampled to $10m$ resolution and Landsat 8 products were resampled to $30m$ resolution. The up-sampling method used is based on the value of the nearest neighbor.

4.2.2 Classical methods in support of validation

To prove the superiority of msSI-SECC over existing methods are chosen two classical visualization methods: rendering bands from the visual part of the spectrum (4, 3, 2)

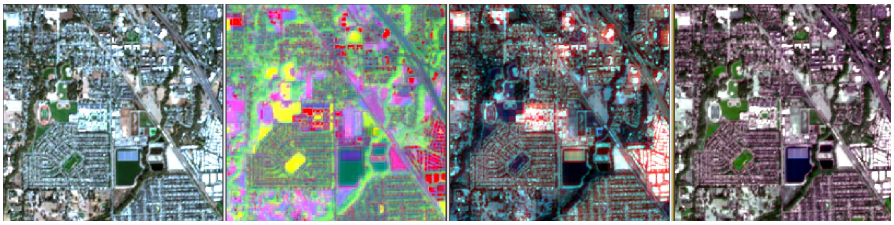
and rendering the combination of the first three bands ranked using the mRMR criteria. Besides these classical visualization methods, a well-known dimensionality reduction method, PCA, was used as comparison.

The bands selection according to the mRMR criteria was obtained using the DAS-Tool plugin of SNAP [3], which implements the algorithm according to [11].

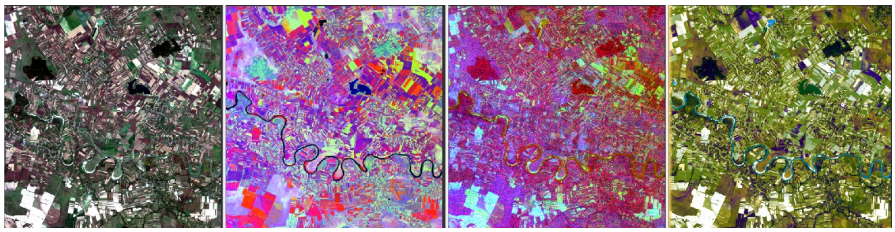
PCA is one of the most popular methods for data reduction. It represents a linear transformation based on the covariance matrix and its eigenvalues.

4.2.3 Experimental results and validation of msSI-SECC

The results evaluation involves both visual and mutual information analysis. Figure 4.16 illustrates the visual representations of the true-color(first column), PCA (second column), mRMR(third column) and proposed method(fourth column). The first row shows a Sentinel 2 scene and the second one a Landsat 8 scene.



(a)



(b)

Fig. 4.16 Experimental results for a Sentinel-2 scene and a Landsat 8 scene.

Visually it may be concluded that the proposed method succeeds to embed the whole information while preserving the color code of the objects, as it can be observed in Figure 4.16. PCA also accomplishes to show some differences that are not visible in the RGB representation, but the display do not resemble with a true color image. mRMR criteria could be misled by anomalies encountered in different spectral bands.

msSI-SE method results evaluation

msSI-SE is a multi-sensor method that respects the input - error function pair according to the SI-SE method, but follows the implementation details of the network and the training/test data set corresponding to the msSI-SECC method. msSI-SE involved the exclusion of the color correction adaptation at the error function level present in msSI-SECC.

The results of msSI-SE compared with the RGB and mRMR representation are highlighted in Figure 4.17.

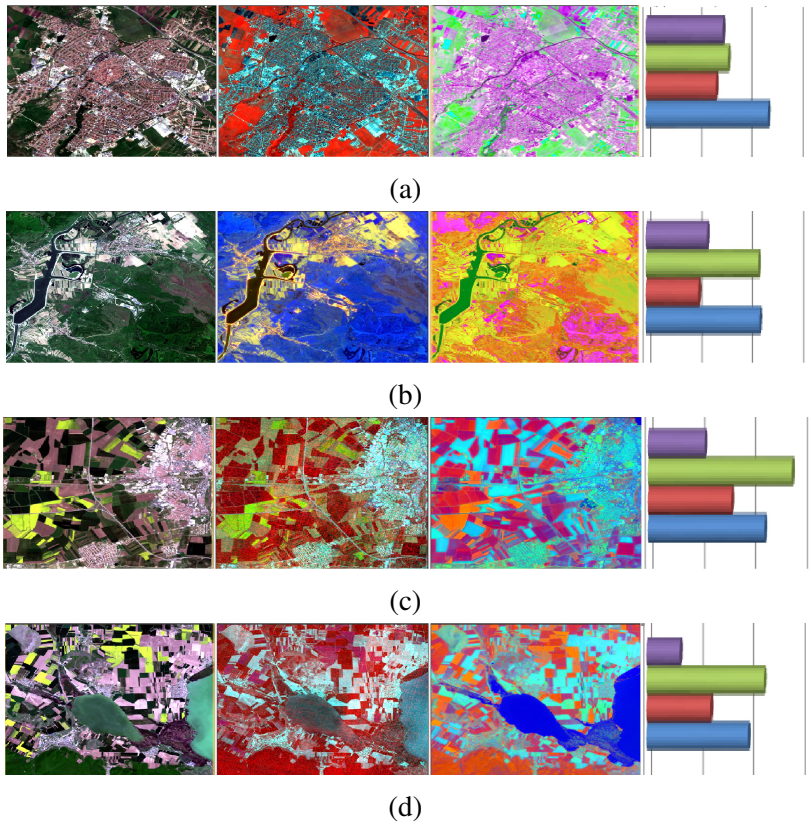


Fig. 4.17 RGB vs mRMR vs msSI-SE and mutual information quantization; a) and b) represent Sentinel-2 scenes, while c) and d) represent Landsat 8 scenes.

The most frequent top ranked bands by mRMR criteria which increase the distinction between semantic classes are 1,2,8 and 9. From the experimental results we conclude that the autoencoder method usually emphasizes details and successfully distinguishes the similarities and dissimilarities between scene objects.

Chapter 5

Information valorization for band reconstruction

Concerning the problem of a missing or degraded spectral band, this chapter presents a method of valorizing the spectral information available in the other bands in order to predict it.

5.1 Proposed concept

Having the main objective of maximizing valorization of spectral information in the interest of reconstructing through prediction a band from a multispectral image, this thesis proposes a method to extract that information from the concurrent spectral bands of the same product.

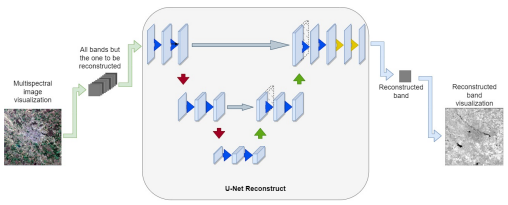


Fig. 5.1 Proposed method concept for band recovery.

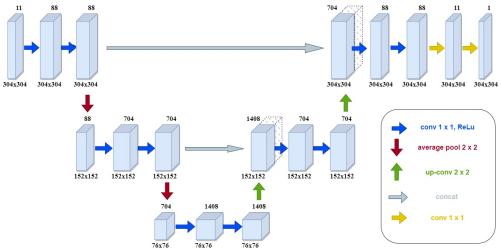


Fig. 5.2 UNetBRec architecture.

Figure 5.1 illustrates the general overview of the method. The convolutional neural network, UNetBRec, receives as input all spectral bands but the one to be predicted and returns as output a single band with the same width and height. In the training process, the network evaluates a comparison between the true and the generated bands in order to adjust its parameters and obtain a better result.

5.2 Band reconstruction for multispectral images

As a consequence of deep learning based methods emergence and usage in many fields, it was demonstrated that impossible to solve problems using a classical method, became approachable and even solvable using diverse types of neural networks.

5.2.1 Proposed deep learning architecture

U-Net [12] is a state-of-the-art CNN build upon on the "fully convolutional network" introduced by [6]. The main characteristic of this network is its u-shape architecture, containing a contracting path and a expanding path. The contracting operation is obtained through pooling operators, while the expanding is achieved through upsampling operators. The two branches, down and up, are interconnected through concatenation operations in order to pass spatial and spectral information. So, the symmetry between the two parts of the network is almost perfect. This work proposes a modified architecture of U-Net, UNetBRec, presented in Figure 5.2.

5.2.2 Physics aware multispectral image band reconstruction

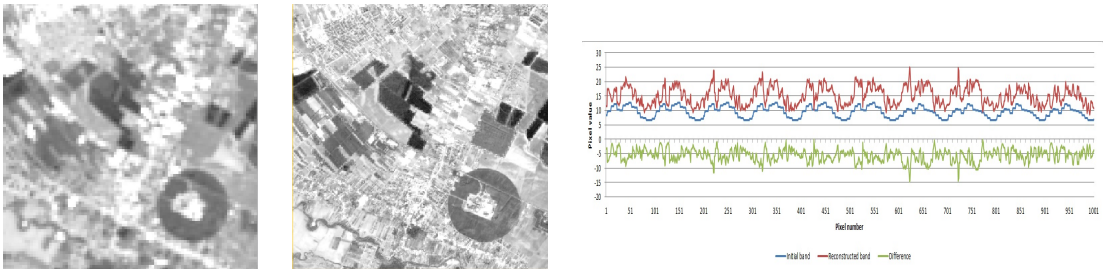
The exclusive use of information from the concurrent spectral bands can generate, in addition to band reconstruction, an improvement in terms of resolution. An important characteristic of a band is its spectral signature. In the process of a band reconstruction through prediction, the preservation of the signature demonstrates the effectiveness of the method applied.

Reconstruction of 60 m spatial resolution bands

S2 has three bands at 60 m resolution, but as we use the Level 2 products, band 10 is not included due to the fact that it does not contain surface information. The left two bands with 60 m resolution are band 1 and band 9.

Figure 5.3 presents a examples of 60 m resolution band reconstruction, emphasizing the resolution improvement as all contours are more precise. Also, next to the visualization comparison there is a graph showing the spectral signature of both initial (blue line) and reconstructed (red line) bands.

Although the spectral signatures are not identical and do not follow exactly the same pattern, the effect could be explained by the super-resolution itself. Many similar pixels from the initial band that define an area not very well delimited, in the reconstructed band may have contrasting values in order to define a better contouring of the objects from the Earth's surface.



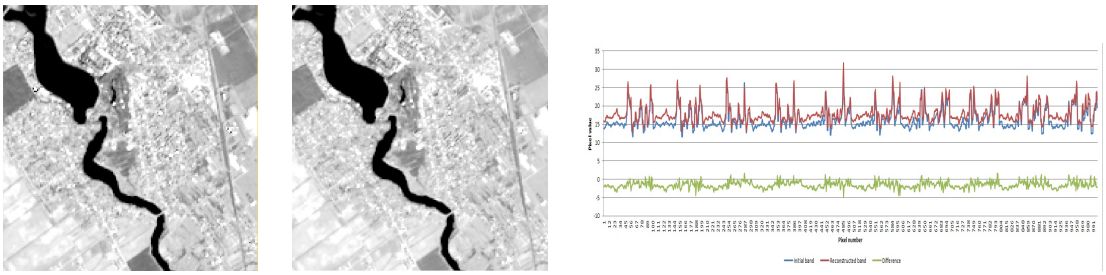
(a) Initial (b) Reconstructed (c) Spectral signature comparison

Fig. 5.3 60 m resolution band reconstruction - visual comparison to emphasize resolution improvement and spectral signature graph to highlight the pattern preservation and development.

Reconstruction of 20 m spatial resolution bands

The number of bands with 20 m resolution in S2 products is equal to six, namely the bands number $\{5, 6, 7, 8A, 11, 12\}$. Being the resolution that most of the bands are acquired at and being a very small difference compared to the highest resolution, 10 m, the reconstruction is made more accurately. The minor difference between the initial band and the prediction being visible in terms of brightness, both in spectral signature pattern and bands visualization.

Figure 5.4 illustrates an example for a 20 m resolution band prediction.



(a) Initial (b) Reconstructed (c) Spectral signature comparison

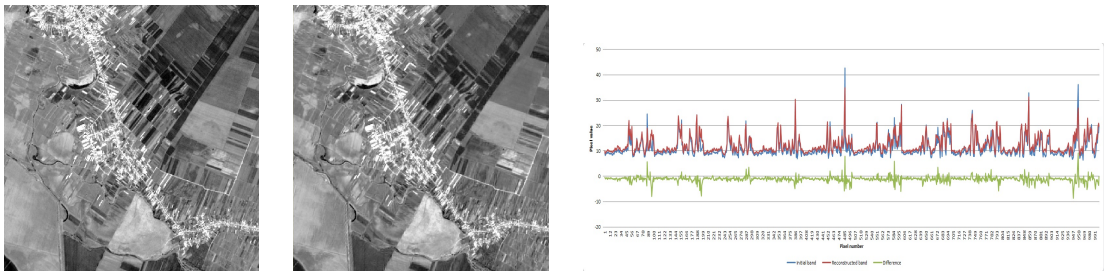
Fig. 5.4 Example of 20 m resolution band reconstruction - visual comparison to emphasize resolution improvement and spectral signature graph to highlight the pattern preservation and development.

Unlike the bands with 60 m resolution, in the case of 20 m resolution bands, from a visual point of view, pixelation differences are no longer noticed, and the spectral signatures keep exactly the same pattern, in some cases they are even overlapped, which means that they are identical from point of view of the amplitude of the pixel values.

Reconstruction of 10 m spatial resolution bands

Four bands of a S2 product have 10 m resolution, naming the bands number $\{2, 3, 4, 8\}$. First three of these bands have wavelengths from the visual part of the spectrum and are highly used for the true color representation of the image. The fourth one, band 8,

has its wavelength in the NIR part of the EM spectrum and is approximately centered positioned in the range of wavelengths available in a S2 product.



(a) Initial (b) Reconstructed (c) Spectral signature comparison

Fig. 5.5 Example of 10 m resolution band reconstruction - visual comparison to emphasize resolution improvement and spectral signature graph to highlight the pattern preservation and development.

Figure 5.5 illustrates an examples of a 10 m resolution band prediction. As regards to resolution, it was neither improved nor worsened, so that the original quality of each band was preserved. Also, the spectral signatures graphs highlight the maintenance of the pattern, registering small amplitude differences between the original and the reconstructed.

5.3 Experimental results and evaluation

In this section are presented all details regarding used datasets for training and testing, full experimental implementation and metrics used to evaluate the proposed method. Finally, experimental results are presented and analyzed.

5.3.1 Train and test datasets

Sentinel-2 products were used for both training and testing. The size of each image is 10980x10980x12.

To create the training data set, a subset of 10944x10944x12 was taken from each of the two images. Next, the subset was divided into patches of 304x304x12, resulting a number of 1269 patches from each image. Finally, the sets of patches obtained from the two images were concatenated, thus creating a training dataset of 2592 patches with a size of 304x304x12. For testing, each of the images went through the same process as for training: up-sampling, subset selection and patching process, except for the final concatenation. In that way, each image could be passed through the testing process sequentially.

5.3.2 Implementation details

The proposed method implementation was achieved using Python 3.6.13 and TensorFlow 2.3.1 for GPU. Training step was performed on a distributed system containing a Intel(R) Xeon(R) E5-2620v4@2.10GHz CPU and 8 PCIe-connected Tesla K80 GPUs, with 12GB of RAM each.

As the number of bands of a Sentinel-2 image is twelve, were trained twelve models, one for the reconstruction of each band. The networks were trained having different batch sizes and number of epochs. The duration of one model training took about 45 minutes.

The filters for the convolutional layers were set the following way: first two had 88 filters, next two 704, following another two with 1408, next three with 704, another three with 88, one with 11 and the last one with 1 filter. For numerical stability the pixel reflectance values were scaled so that the resulting interval be $[0, .., 255]$.

5.3.3 Evaluation metrics

In order to quantitatively evaluate the results obtained by the proposed method, different state-of-the-art indexes which measure the accuracy of spatial and spectral profiles preservation were used: RMSE (Root-mean-square-error), SSIM(Structural similarity index) [15], SRE(Signal to reconstruction error) [5], PSNR(Peak signal to noise ratio) [4] and SAM(Spectral angle mapper) [16]. The metrics implementation used in this thesis was the one proposed by [7]. The code, implementation details and instructions for usage are available on GitHub [14].

5.3.4 Results and discussion

The performance of UNetBRec is both quantitatively and qualitatively evaluated. For quantitative analysis, state-of-the-art image reconstruction assessment metrics are computed for UNetBRec and other state-of-the-art methods in order to critically study their achievements. The qualitative analysis is performed by visually comparing the results obtained with the ground truth.

Quantitative analysis

The main evaluation metrics of the quantitative comparison are RMSE, SSIM, SRE, PSNR and SAM. UNetBRec has twelve versions, one for each band prediction. An overall comparison, as regards to used metrics, shows that depending on the resolution of the band, the image for which it is tested and the differences between the reflectance values, the evaluation metrics lay within an acceptable range in terms of performance.

	RMSE	SRE	SAM
Superres	68.95	46.01	1.02
DSen2	25.14	51.10	0.78
VDSen2	23.31	51.73	**0.76**
UNetBRec	**12.99**	**61.89**	0.79

Table 5.1 Average computed metrics compared between SoA methods and UNetBRec

As baseline methods, are used the ones proposed by [5], DSen2 and VDSen2, and [1], Superres. UNetBRec obtains the desired band using the complementary ones, while the competition use all bands, to retrieve super-resolution ones. Average results over all test images and all bands are displayed in Table 5.1.

Qualitative analysis

The qualitative analysis of the results obtained with UNetBRec involves comparing the reconstructed band with the original one, also visualizing the difference computed pixel wise between them.

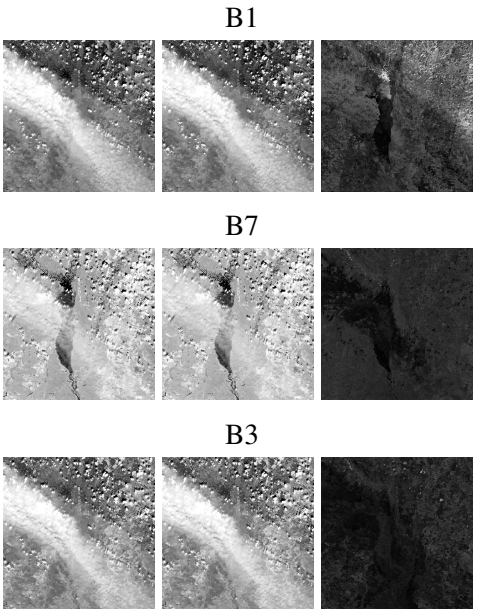


Fig. 5.6 Band wise qualitative evaluation of UNetBRec over a scene covered by fog and thin clouds, Kyiv, Ukraine; first column - initial band, second column - predicted band, last column - difference.

Figure 5.6 ilustrates one example of reconstruction which is not influenced by high contrast between the edges of the reflectance values range. The reconstruction of any of the bands is carried out with high accuracy, which can be seen from the visualization of the difference. The method manages to show only slight traces of a behavior involving large errors along high-contrast edges.

Chapter 6

Conclusions

Remote sensing data are the main sources of information for many earth observation applications. Multispectral imagery contains much information available in multiple spectral bands. But for some reasons the data may be obstructed or even missing. Depending on the nature of the obstruction, two research directions have been considered in this thesis, based on the same principle: the exploitation of the information available in all bands. The directions involved the information valorization to reconstruct and improve the visualization of a multispectral image and information valorization to predict a missing band.

6.1 Obtained results

The research directions pursued in this thesis led to the development of several methods to address the main goals.

Regarding the improvement of the visualization of a multispectral image using a SAE, with respect to all developed methods it can be concluded that:

- SAE is suitable for compressing spectral information in the form of a latent representation used for visualization;
- the illuminance invariance character of the polar coordinate transformation demonstrates its superiority in terms of removing atmospheric effects from the images;
- the color correction applied in the SI-SECC method is advantageous in the case of the elimination of similarities and dissimilarities, having an appearance closer to the natural color representation;
- the PcI-SE method also succeeds in fulfilling its additional objective of eliminating shadows;

- although the SI-SEE and SI-SECC methods were not developed with the aim of eliminating atmospheric phenomena, they achieve an improved visualization compared to RGB;
- both msSI-SECC and msSI-SE methods prove superiority regarding state-of-the-art solution;
- the methods are not mutually exclusive, either of them may have better results in certain scenarios.

Regarding the results of the band prediction method using the spectral information available in the competing bands, the following can be concluded:

- the spatial and spectral correlation between bands allows the prediction of a missing one without too much loss;
- UNetBRec achieves, in addition to band recovery, an improvement for bands with lower resolution;
- the similarity metrics (RMSE, SRE and SAM) used to compare the proposed method with baselines showed the superiority of UNetBRec;
- the interconnection through concatenation operations of the two branches of the U-shaped convolutional network is the main factor that leads to an improvement in resolution.

6.2 Original contributions

The thesis contributions consisted of:

- developing a set of SAE based methods for improving the visualization of multi-spectral spatial images by removing misleading similarities and dissimilarities or eliminating atmospheric effects [1, 2, 3, 4, 5] ;
- development of a method based on a U-shaped convolutional network for reconstructing a missing spectral band using only the complementary bands [4] ;
- understanding existing methods [3] and implementing them in operational tools to improve the analysis process of multispectral spatial images in ESA funded research projects [1, 2, 6] .

The references in this section refer to the numbers associated with the publications in the following section.

Also, a list of all thesis contributions to research projects can be found in the Appendix A.

6.3 List of original publications

Journals

1. Grivei, A. C., **Neagoe, I. C.**, Georgescu, F. A., Griparis, A., Vaduva, C., Bartalis, Z., and Datcu, M. (2020). 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, 13, 4429-4442. DOI10.1109/JSTARS.2020.3013091.

2. **Coca, I.**, Coca, M., and Datcu, M. Autoencoder based method for multispectral images visualization, U.P.B. Sci. Bull., Series C, Vol. 83, Iss. 2, 2021.

3. **Neagoe, I. C.**, Coca, M., Vaduva, C., and Datcu, M. (2021). Cross-Bands Information Transfer to Offset Ambiguities and Atmospheric Phenomena for Multispectral Data Visualization. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 11297-11310. 10.1109/JSTARS.2021.3123120.

4. **Neagoe, I. C.**, Vaduva, C., and Datcu, M. (2022) Band reconstruction using a modified UNet for Sentinel-2 images. **to be submitted to** IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

Proceedings

1. **Neagoe, I.**, Faur, D., Vaduva, C., and Datcu, M. (2018, July). Exploratory visual analysis of multispectral EO images based on DNN. In IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium (pp. 2079-2082). IEEE. DOI: 10.1109/IGARSS.2018.8518414.

2. Vaduva, C., Georgescu, F. A., Griparis, A., **Neagoe, I.**, Grivei, A. C., and Datcu, M. (2019, July). Exploratory search methodology for sentinel 2 data: a prospect of both visual and latent characteristics. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 10067-10070). IEEE. DOI: 10.1109/IGARSS.2019.8900349.

3. Coca, M., **Neagoe, I.**, and Datcu, M. (2020). Physically Meaningful Dictionaries for EO Crowdsourcing: A ML for Blockchain Architecture. In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium (pp. 3688-3691). IEEE. DOI10.1109/IGARSS39084.2020.9324361.

4. **Neagoe, I. C.**, Vaduva, C., and Datcu, M. (2021, July). Haze and Smoke Removal for Visualization of Multispectral Images: A DNN Physics Aware Architecture. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 2102-2105). IEEE. DOI: 10.1109/IGARSS47720.2021.9553735.

5. **Coca (Neagoe), I.**, Vaduva, C., Datcu, M. (2020, September). Perspective Deep Sensing: Making Visual Multispectral Earth Observation Images. ESA EO ϕ-WEEK 2020, e-Poster, section EO Applications, paper 201, ESA

6. Vaduva, C., Georgescu, F.A., Grivei, A.-C., **Coca (Neagoe), I.**, Griparis, A., Bartalis, Z., and Datcu, M. (2020, September). Sentinel-2 data analysis based on explainable features and exploratory visual analysis. ESA EO ϕ-WEEK 2020, e-Poster, section EO Applications, paper 193, ESA

6.4 Perspectives for further developments

The results obtained in the research of this thesis generate new development horizons. In terms of visualization methods, the perspectives could be the following:

- integration as alternative options for quick visualization within the Copernicus Open Access Hub platform
- integration into active learning applications

Concerning the perspectives for development of the results obtained with the band reconstruction method, the directions could be the following:

- implementation and testing of the solution for multiple sensors
- further research to develop the method into a super-resolution one

Appendix A

Contributions in funded research projects

eVADE

eVADE (Interactive Visual Analysis Tool for Earth Observation Data) is a project financed by European Space Agency, carried out in the period 2017-2019. The main objective of this project was to offer an improved and ingenious way of visualizing EO data sets content by using a visual analytics process.

The contribution of this thesis within eVADE research project consisted in: the collaboration in the implementation of 3D projection visualization within the web interface.

DAS-Tool

DAS-Tool is the abbreviation for Multispectral Data Analysis Toolbox for SNAP – ESA's SentiNel Application Platform. DAS-Tool project, financed by European Space Agency, took place during 2017 - 2019 and had the main objectives to elaborate algorithms directed for the semantic analysis and content description of Sentinel 2 images.

The contribution of this thesis in the development of this project consisted in: analysis and research over exploratory visual analysis band selection state-of-the-art methods and implementation of mRMR criteria based method for band selection and polar coordinates for pixel based processing.

xAI

Explainable Deep Learning for Earth Observation (xAI), being a national project, carries out starting from 2021 until 2023 and aims at valorizing the remote sensing data by developing new explainable neural networks based methods for EO images.

This thesis contribution within xAI research project resided in: development of unsupervised SAE based set of methods of cross band information transfer for improving multispectral images visualization, development of a DNN physics aware method for haze and smoke removal and the publication of a journal paper [8] and a conference paper [9].

References

[1] Brodu, N. (2017). Super-resolving multiresolution images with band-independent geometry of multispectral pixels. *IEEE Transactions on Geoscience and Remote Sensing*, 55(8):4610–4617.

[2] Chuvieco, E. (2016). *Fundamentals of satellite remote sensing: An environmental approach*. CRC press.

[3] Grivei, A. C., Neagoe, I. C., Georgescu, F. A., Griparis, A., Vaduva, C., Bartalis, Z., and Datcu, M. (2020). 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*, 13:4429–4442.

[4] Hore, A. and Ziou, D. (2010). Image quality metrics: Psnr vs. ssim. In *2010 20th international conference on pattern recognition*, pages 2366–2369. IEEE.

[5] Lanaras, C., Bioucas-Dias, J., Galliani, S., Baltsavias, E., and Schindler, K. (2018). Super-resolution of sentinel-2 images: Learning a globally applicable deep neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 146:305–319.

[6] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

[7] Müller, M., Ekhtiari, N., Almeida, R., and Rieke, C. (2020). Super-resolution of multispectral satellite images using convolutional neural networks. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-1-2020:33–40.

[8] Neagoe, I. C., Coca, M., Vaduva, C., and Datcu, M. (2021a). Cross-bands information transfer to offset ambiguities and atmospheric phenomena for multispectral data visualization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:11297–11310.

[9] Neagoe, I. C., Vaduva, C., and Datcu, M. (2021b). Haze and smoke removal for visualization of multispectral images: A dnn physics aware architecture. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2102–2105. IEEE.

[10] Okamura, T., Kudo, J.-i., and Tanikawa, K. (2003). Multispectral illumination invariant classification with three visible bands and a near-infrared band. *Optical Engineering*, 42(6):1665–1672.

[11] Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238.

[12] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

[13] Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., and Zhang, L. (2015). Missing information reconstruction of remote sensing data: A technical review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):61–85.

[14] UP42 (2022). Image similarity measures package. https://github.com/up42/image-similarity-measures. [Online; accessed 07-July-2022].

[15] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

[16] Yuhas, R. H., Goetz, A. F., and Boardman, J. W. (1992). Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm. In *JPL, Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop*.