U-NET BASED CLASSIFICATION FOR URBAN AREAS IN ALGERIA

Semcheddine Belkis Asma [1, a], *Daamouche Abdelhamid* [1, b] *and Li Youyou* [2, c]

[1]Signals and Systems Laboratory, Institute of Electrical and Electronic Engineering, University of Boumerdes.
[2]University of Electronic Science and Technology of China
a.semcheddine@univ-boumerdes.dz
[a] Corresponding author: adaamouche@univ-boumerdes.dz
liyouyou@std.uestc.edu.cn

ABSTRACT

Nowadays, researchers in the field of remote sensing and image classification have to face the challenge of the massive amount of information contained in satellite images, especially in urban areas. These types of areas contain numerous classes, where each class is made of several groups of pixels that are not adjacent, and that are rich in texture. Convolutional Neural Networks possess the ability to handle these problems. However, CNNs require quite a very large number of annotated training samples. U-Net came as a revolutionary solution for this major drawback. This paper aims to study the ability of a pre-trained U-Net to classify a satellite image and is then compared to the performance of a Support Vector Machine classifier.

***Index Terms*—** Object-Based, SVM, U-Net, Remote Sensing.

1. INTRODUCTION

Image segmentation became reputable with the improvement of remote sensing image resolution. Recent researches have proved the weakness of pixel-based classification in manipulating high-resolution satellite images [1]. Urban areas are the main subject of our paper. Urban areas are more complex to study because they are composed of various types of objects (building-roofs, trees, grass, roads and so on) with different surfaces which are quite crowded. The segmentation process sets apart each region from its surroundings, which can be advantageous when handling images that are rich in texture [2]. Moreover, as a result of computer-vision evolution, we are now able to add meaning to the segments. Semantic segmentation relies on a mapping step where it automatically produces a matrix that covers the entire data and assigns to each group of pixels the same number. Furthermore, segments that belong to the same class can be disconnected but still have the same assignment in the semantic segmentation map. This operation became possible as a result of the Convolutional Neural Networks enhancement [3].

The evolution of Neural Networks based architectures expanded its area of application and became one of the ten breakthrough technologies of 2013 [4]. The hidden layers in a NN can even exceed the human performances in handling highly computational tasks. CNN, for example, manages to extract mid- and high-level abstract features thanks to the convolution and max-pooling operations. The convolution process aims to reduce the number of data to be treated by shrinking the feature map layer-by-layer [5], which reduces the time it takes to manage large images.

The major drawback of CNN is the number of training samples it requires in order to have a respectable performance [6] as well as producing a single class label as a means of classification. In the remote sensing field in general and urban areas in particular, we usually have to study each zone at a time and have to predict different classes for the items present in the image. The availability of the data set imposes a limitation on the CNNs. In fact, we have to extract both training and test samples from one image and deliver an algorithm that is able to recognize and locate several elements present in one image. A solution has been proposed to overcome this restriction in the form of a U-shaped Convolutional Neural Network, which is the U-net.

The U-net was first developed for the purpose of biomedical image classification. In biomedical images, the output of the network has to provide the exact coordinates of the tumor in addition to the classification of the input image [7]. The shortage of the necessary data present in the study and the need for a precise localization encouraged the establishment of an intelligent algorithm that serves such purpose. The main characteristic of U-nets is the ability to conserve the information in each epoch, valuable information that can be lost during the process of convolution and max-pooling operations, and then later, to employ such materials in the process of reconstruction [8]. The U-net is also on its own an encoder-decoder, its design consists of a narrowing path to capture context and to propagate to higher resolution layers with the use of skip connections between mirrored layers. The pattern of U-net provides a consistent and smooth output [9].

In this paper, we aim to test the functioning of U-net on a typical Algerian urban area dataset. The input is a multi-spectral satellite image of Boumerdes city. The image has diverse objects, which have been restricted to fourteen classes. Our goal is to provide a proper classification of the image and eliminate the salt and pepper effect. The rest of this paper is presented as follows: first, we will describe our approach to this specific type of image. Then, a proper description is made of the dataset. After that, we run U-Net on our image and compare the results to the ones obtained by an SVM classifier.

The ground truth is interpreted by fourteen land cover classes, namely, Water, Waves, Dry and wet Sand, Trees, Grass, Asphalt1, Asphalt2, Rocks, Red roofs, Dark roofs, White roofs, Soil and unknown. Part of the image is used for training and validation and the rest is used as a test dataset.

2. THE PROPOSED APPROACH

This section of the paper comprises a description of the methodology we adopted. The algorithm used is U-net based architecture. The two main differences are an encoder at the input that converts the four channels dataset into three channels one so that the image can be processed by the network.

The output, on the other hand, contains fourteen layers, in order to match the number of classes we want to produce. The network itself is a pre-trained model to which we apply fine-tuning.

The convolutional-based neural networks have a large number of weights to set-up. We want to benefit from a pre-trained model as an initialization of the system.

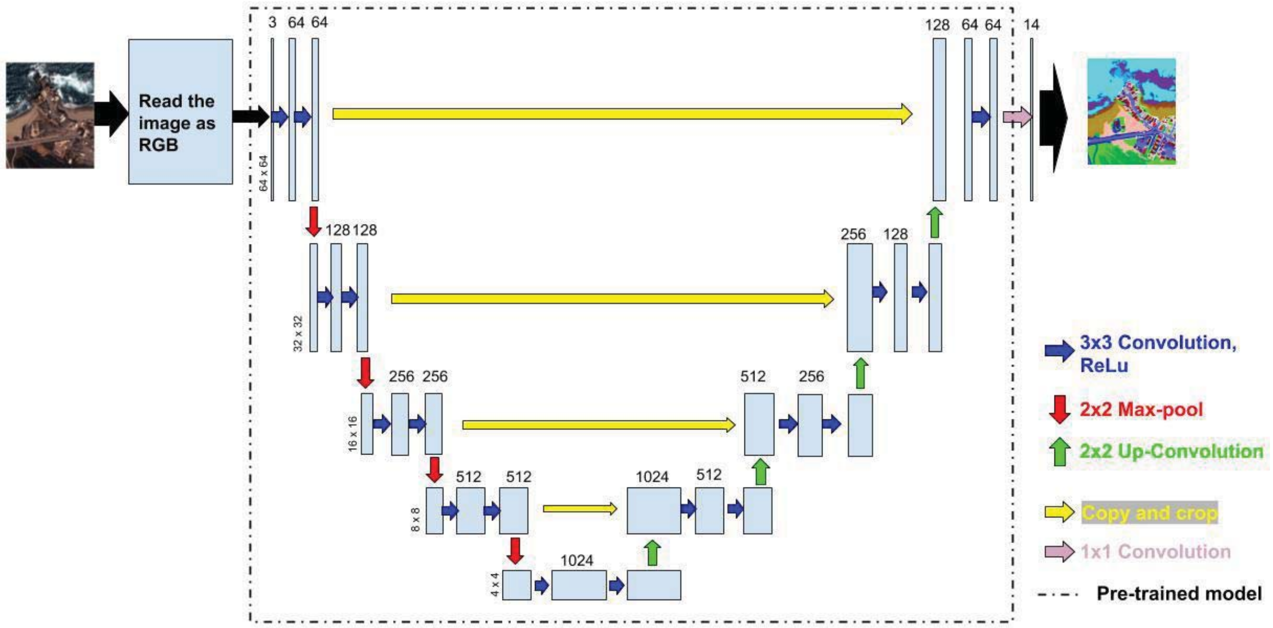

Figure 1. Architecture of the U-net based model

3. DATASET

For the validation of our model, we chose to work on a four channels satellite image obtained over the city of Boumerdes by the Quickbird sensor in 2002 (Figure 2). The resolution of the image is 1 meter. The size of the image is 600 by 500 pixels [10].

4. EXPERIMENTATION AND RESULTS

In this part of the paper, we will go through the training and the testing of our U-Net architecture. In order to use our Algerian image, we had to split it into training, validation and test square patches.

We wanted to make use of all the data available, the image was first horizontally divided into 7 strides, where each stride was 64 by 600. Then each stride was divided into 64 by 64 patches.
We decided to take the first, third, fourth and eighth stride for training. The seventh stride was used for validation, and the remaining strides were used for the test.

Figure 2. A view of Boumerdes image

We mentioned above in this article that U-Net was purposely designed to handle classification problems where there is less training data available, compared to traditional CNNs. Now in our case, the training patches that are extracted from the image are not sufficient. In order to

Table 1 Classification Results

Classes	Semantic meaning	U-Net Acc (%)	SVM Acc (%)
Class 1	Asphalt 2	60.08	83.00
Class 2	Dark Roofs	75.11	31.61
Class 3	Red Roofs	00.00	87.83
Class 4	Waves	96.17	99.59
Class 5	Wet Sand	90.00	64.26
Class 6	Sea (Water)	85.47	97.02
Class 7	Forest/Trees	90.30	97.62
Class 8	Asphalt 1	73.05	06.00
Class 9	Dry Sand	96.05	98.81
Class 10	Grass	78.59	77.95
Class 11	White Roofs	82.70	47.08
Class 12	Rocks	86.63	00.75
Class 13	Bare Soil	54.29	28.27
Class 14	Unknown Objects	41.17	67.54
	Overall Accuracy (%)	76.45	76.88
	Average Accuracy (%)	73.93	63.38

Figure 3. Ground Truth for Boumerdes image

Figure 4. U-Net mapped image

Figure 5. SVM mapped image

overcome this issue, we used a data augmenter that enlarges the training set from 40 patches to 176 patches.
The pre-trained model was a VGG19 model trained on the ImageNet data set [11]. We froze the weights of the encoder and fine-tuned the decoder part. The results of the classification, in terms of accuracy, are portrayed in Table 1. The results of mapping the whole image by our two classifiers are shown in Figures 4 and 5.

5. DISCUSSION

We can clearly see through the overall classification accuracy that U-Net and SVM classifiers gave nearly the same performance, with a slight difference of 0.43%. However, U-Net seems to outperform the SVM classifier in the per-class classification, where we can see an improvement of 10% of the average accuracy. This improvement in terms of accuracy is well justified when we take a closer look at the classification accuracy of each one of our classes. While the SVM classifier was able to determine almost all water, waves, dry sand and forest pixels, that surpassed U-Net results, U-Net, on the other hand, gave an acceptable classification of the pixels from these classes. Moreover, it was able to achieve more than 70% of accuracy for Asphalt 1 and Rocks, the ones that SVM was only able to classify 6% and less than 1%, respectively.
The only drawback of the U-Net classifier is its inability to identify the Red Roofs class. We can explain this by underlining the fact that this class has very few samples, which disappeared in the encoding part of the U-Net.

6. CONCLUSION

In this paper, we presented a comparison of two well-known classification algorithms, U-Net and Support Vector Machine. First, in the introduction part, we talked about the improvements made with Neural Networks and how they led to Convolution Neural Networks, which was followed by a U-shaped CNN called U-Net. Secondly, we explained briefly our approach and then described our dataset. After that, we showed the results obtained by U-Net and SVM classifiers. Finally, we discussed the results acquired from the proposed method.
In our future work, we would like to combine the best features of these two classifiers in order to find an even better method to classify very high resolution, remote sensing images.

7. REFERENCES

[1] Meinel, G., and M. Neubert. "A comparison of segmentation programs for high resolution remote sensing data." International Archives of Photogrammetry and Remote Sensing 35.Part B (2004): 1097-1105.

[2] Ryherd, Soren, and Curtis Woodcock. "Combining spectral and texture data in the segmentation of remotely sensed images." Photogrammetric engineering and remote sensing 62.2 (1996): 181-194.

[3] Audebert, Nicolas, Bertrand Le Saux, and Sébastien Lefèvre. "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks." Asian conference on computer vision. Springer, Cham, 2016.

[4] MIT Technology Review. (2013). 10 breakthrough technologies in 2013. [Online]. Available: https://www.technologyreview.com/lists/technologies/2013/

[5] Zhu, Xiao Xiang, et al. "Deep learning in remote sensing: A comprehensive review and list of resources." IEEE Geoscience and Remote Sensing Magazine 5.4 (2017): 8-36.

[6] Hu, Fan, et al. "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery." Remote Sensing 7.11 (2015): 14680-14707.

[7] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[8] Bai, Yanbing, Erick Mas, and Shunichi Koshimura. "Towards operational satellite-based damage-mapping using U-net convolutional network: a case study of 2011 Tohoku Earthquake-Tsunami." Remote Sensing 10.10 (2018): 1626.

[9] Shi, Qian, Xiaoping Liu, and Xia Li. "Road detection from remote sensing images by generative adversarial networks." IEEE Access 6 (2017): 25486-25494.

[10] Daamouche, Abdelhamid, *et al.* "Swarm optimization of structuring elements for VHR image classification." IEEE Geoscience and Remote Sensing Letters 10.6 (2013): 1334-1338.

[11] Pavel Yakubovskiy. "Segmentation Models". GitHub repository (2019). https://github.com/qubvel/segmentation_models