# Road Segmentation using U-Net architecture

*NORELYAQINE Abderrahim*[1]
*Mohamed V University*
*Mohammedia school of engineers*
*Rabat, Morocco*
norelyaqine.abdou@gmail.com

*SAADANE Abderrahim*[2]
*Mohamed V University*
*Faculty of Sciences of Rabat*
*Rabat, Morocco*
saadaneabderrahim@gmail.com

*AZMI Rida*[3]
*Mohamed V University*
*Faculty of Sciences of Rabat*
*Rabat, Morocco*
ridaazmi@gmail.com

***Abstract* - The detection of objects has become a critical step to update ground cover information, and the availability of very high-resolution satellite images made us discover new classification methods that give us more details such as pixel classification. This study aims to explore the potential and performance of machine learning algorithms in poor urban conditions in order to show the power of the deep neural networks to detect objects and, more precisely, to detect roads. We propose a U-net architecture for road extraction from Massachusetts dataset. The results have been compared with different automatic classification learning algorithms. The results of the classification using U-net showed a high accuracy of 97.7%, more precise than all the other models, which is why it is the best method to solve classification tasks for objects detection in large-scale datasets.**

***Keywords: Remote sensing, Deep learning, Road extraction, very high resolution***

## I. INTRODUCTION

The traffic road network is one of the essential geographic elements of the urban system, he is considered as the core of the urban evolution. Road detection from satellite images is essential for researches that researchers are concerned with, it has become a crucial issue in information science and remote sensing (RS) imagery processing. Extracting object from RS data is the main goal of image analysis generally and computer vision particulary. Road are some of the most important to be axtracted from medium and very high-resolution (VHR) satellite images. Eventually, they are the primary data source for the automatic extraction of road networks [1]. Moreover, roads are seen as a fundamental and mandatory part of the transportation system, traffic safety management, land cover mapping, road monitoring, automatic road navigation, and road map updating. However, due to the fast changes that effects the road networking, specially the urban transportation system, one of the main benefit of this extraction is essential in networking for immediate road mapping. Road mapping is useful in effective removal. Consequently, proposing a new strong approach for extracting the road network from VHR imagery using RS will be an asset for intelligent transport systems and for the geographic information system in general [1, 2]. The main challenge that drives research into this line of extracting road using satellite images is the extraction problems encountered from the background features such as surface markings, shadows, vegetation and highlights. Road detection techniques and algorithms rely on edge-based techniques that consist of linear feature detection and road line verification using structures of geometry. Meanwhile, it is arduous to develop a general architecture to detect the roads because the standards change from a city to another.

Normal extraction methods generally contain a lot of human errors [3] . During the previous years, plenty of work have attempted to detect roads using remote sensing imagery. A many of detection methods have been developed using geometric, photometric and textural characteristics. To obtain a more accurate in classification Zhang et al. used a gray level co-occurrence matrix (GLCM) [4]. In urban mapping, Huang et al. have worked with an innovative multi-scale approach based on the mean shift (MS) [5]. Mean offset method has been used to detect road [6]. In classification problems the supervised methods are rather often the most precise [7]. These methods which have been learned automatically and the conditional random fields, such as SVM and random forest, have been used for classification in remote sensing [8]. And with the deep neural network which is widely used in object detection [9], visual recognition [10], in very high spatial resolution, they have shown an exceptional capacity for the classification [11] and in the detection of buildings based in VHR satellite images, a deep learning model has been proposed [12].

Recently, artificial intelligence (AI) has been important to researchers, and it's use in road detection from VHR satellite images. In recent years, with the development of deep learning, convolutional neural networks (CNN) has become a very powerful tool for classifying pixels in many areas more precisely in RS images [13] and object detection [14] which has succeeded with great accuracy. More recently, CNN is mainly used in the automatic control and classification of medical images using network architecture such as FCN [15] and U-Net [16].

To ensure excellent detection of road from the satellite images, integration of the power of multiple algorithms and available data sources need to be implemented to ameliorate the reliability and robustness of the detection results, and with the recent success in deep learning, extraction of road using satellite imagery has been very successful.

The point of this article is to build a model to extract the road from satellite images and to improve the accuracy of the classification. After testing several models, we were inspired by the U-Net architecture which it takes into consideration the spatial correlation and the geometric information of the road structure. The output of the model formed was a map of the classification of the road and the background, the whole work process will be detailed in the article sections.

## II. METHODOLOGY

### *A. Data Prepocessing*

Training our model on large images requires a lot of resources and time, and all images in our dataset have a resolution of 1500 × 1500 pixels. As a potential answer, we preprocessed the data to simplify the model training process by resizing images to smaller dimensions. The dimensions of the input images were changed to 512 × 512 pixels.

978-1-7281-5806-8/20/$31.00 ©2020 IEEE

### B. Network Architecture

In many vision tasks and image processing specifically in the segmentation of high resolution images, deep neural networks has played a very important role, so we turned to U-net [16] as it is highly used in medical image analysis, mainly in fields like cardiology or neurology. UNet has shown an excellent performance in segmenting different targets in various modalities of medical images. The architecture is made up with a total of 23 convolutional layers, which consists in a contraction path (encoder) and an expansion path (decoder). Such as SegNet [17], the encoder part is made up of repeated convolution blocks. Each block consists two convolution layers with size filters (3 × 3), each one of them is followed by activation (ReLU) and a maximum pool operation of (2 × 2). The number of feature cards is doubled after each subsampling. The expansion part of UNet is an inverted version of the contraction part. The number of characteristic cards is halved after each block. So as to connect the encoder part to the corresponding decoder part, a copy of the characteristic cards is concatenated with the corresponding cards of the decoder part. The last layer uses a convolution with size filters (1 × 1) in order to provide classification maps of the same number as the desired classes. The output size is smaller than input size as shown in Fig. 1 because of the use of convolution without adding pixels around the image. To get the same entry size, the entire image is predicted part by part using an overlap mosaic strategy.

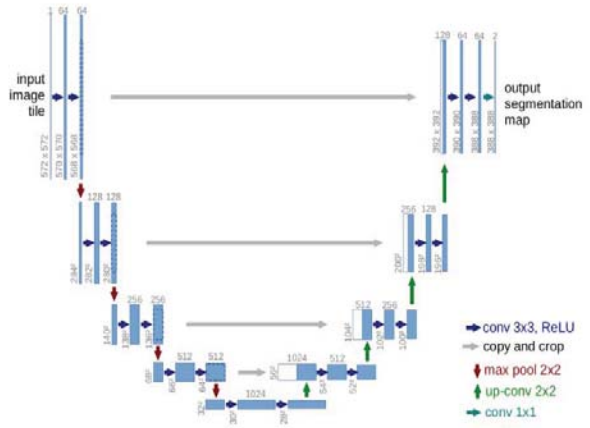


Fig. 1. Original U-net for image segmentation

### C. Loss Function

The cross entropy is the most used loss function for image segmentation. In neural networks concept, it is used as a loss function which have softmax type activations in the output layer. This loss examines each pixel individually, comparing the predicted tag to the target tag. It also calculates the distance between the output distribution predicted by the model and the actual distribution. Defined in equation (1):

$$E(P',P) = -\sum_{n=1}^{N} P'_n \log P_n \quad (1)$$

Where P is the ground truth tag of the training pixel for the n class and $P_n$ is the result of model prediction for the learning pixel for the n class. The final result is the average of the error for all the training samples in a batch. It depends of the input and of the parameters of the learning model.

### D. Data Augmentation

The number of annotated road samples is generally limited as it should be done by experts. On the other hand, learning networks in such a domain requires a very immense number of data, and it is difficult to create new images from the same distribution. To overcome this limitation, we generate artificial data allowing to improve the diversity of training data of a model in order to improve its performance and maintain a reasonable number of images using random transformations on the input images such as inversions, distortions, rotations and in particular elastic deformations. To generate more images we used albumentations library [18], it has been reported as a fast and flexible implementation for augmentation data, in our case we used basic tranformation such as flip and elasric transform as shown in the Fig. 2, which allowed us to train our model on a much larger set of images.

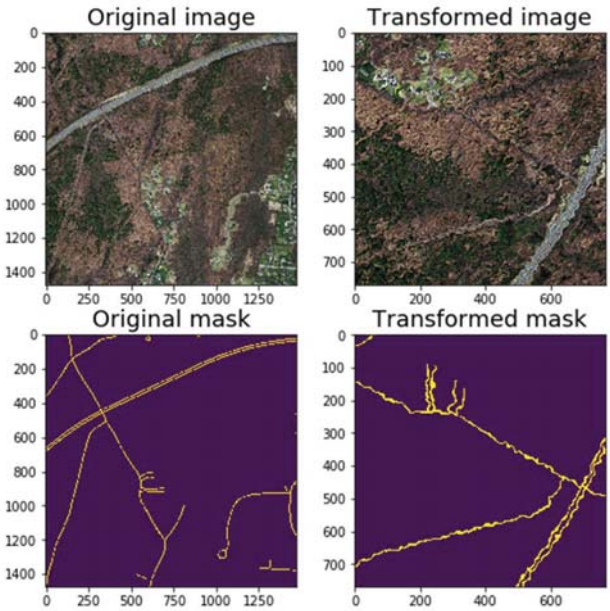


Fig. 2. Data Augmentation

## III. EXPERIMENTS AND RESULTS

### A. Dataset description

The roads dataset used for our research is the Massachusetts City dataset, he was built by Mihn and Hinton [19] and publicly available, he can be downloaded from internet. The road dataset divided into 1108 sets for training part, 49 for testing part, and 14 for validation part. All images has its binary classification, classified into road segments and non-road segment as shown in Fig. 3. The resolution of all

images is 1.2 m² and the size is about 1500×1500 pixels. The dataset covers a total area over 2600 Km² of variety urban and rural, including vegetations, buildings, roads, rivers and vehicles. The Massachusetts road dataset labels are created from OpenStreetMap.

(a) (b)

Fig. 3. (**a**) the RGB image and (**b**) the corresponding label, from Massachusetts road dataset

*B. Evaluation Metrics*

In order to evaluate the relative efficiency of the different segmentation and classification models, it is necessary to define quantitative criteria allowing comparison. In our case we will use the precision and recall method (P-R) [20].

The recall is defined as the ratio between the number of true positives and the total number of elements actually belonging to the class:

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

The precision is defined as the ratio between the number of true positives and the total number of elements assigned to the class:

$$Precision = \frac{TP}{TP+FP} \tag{3}$$

But in general, there is one measure that combines the two at the same time, it is known as the F1 score or Sorensen-Dice coefficient, is defined as the harmonic mean of precision and recall:

$$F1 = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{4}$$

Where TP is the number of true positives, FP false positives and FN false negatives.

*C. Training*

One of the main parameters of the implementation of a neural network is the number of epochs. The only way to know the optimal number is to plot the learning curve of our model during training according to the number of epochs. Loss function during training and validation data begins to saturate after 50 epochs, we considered the training to be complete when no progress after. We therefore chose, for all of the following experiments, to stop the learning process at this point.

*D. Testing and Results*

Our proposed model is compared with four road extraction methods based on deep learning [21], [22], [17] tested on the Massachusetts road dataset. Table I summarizes the results of the models compared. We can notice that our U-Net model works better than the other approaches in the table with an accuracy of 97.7% and our model produces the best F-score value with 87.5%. So, we get slightly lower accuracy if we use convolutional neural networks.

TABLE I. COMPARAISON OF PERFORMANCES IN TESTING DATA

| Model | Precision | Recall | Accuracy | F-score |
|---|---|---|---|---|
| FCN [22] | 43.5% | 68.6% | 90.4% | 53.2% |
| RSRCNN [23] | 60.6% | 72.9% | 92.4% | 66.2% |
| SegNet [24] | 77.3% | 76.5% | 95.7% | 76.8% |
| **U-Net** | **86.8%** | **88.3%** | **97.7%** | **87.5%** |

The proposed model shows clear and clean results with less noise, especially in two-land roads (Fig. 4). The model has the ability to detect each lane with high accuracy and more efficiency, and also the complex structures of roads, as we can see in the red rectangle of the second row. The model is also able to distinguish objects that have similar structure as roads, like parking and roofs of buildings, and he also can segment the roads which are covered by trees. In addition to this, in the first row of Fig. 4, it is able to segment roads that are not labeled in our source data. As we can notice in the blue rectangle, a background is labeled as a road, but our model segments with precision. Since some roads are not completely visible as it is indicated in the yellow rectangle, this is mainly because some roads are not labeled, so our model considers them as backgrounds.

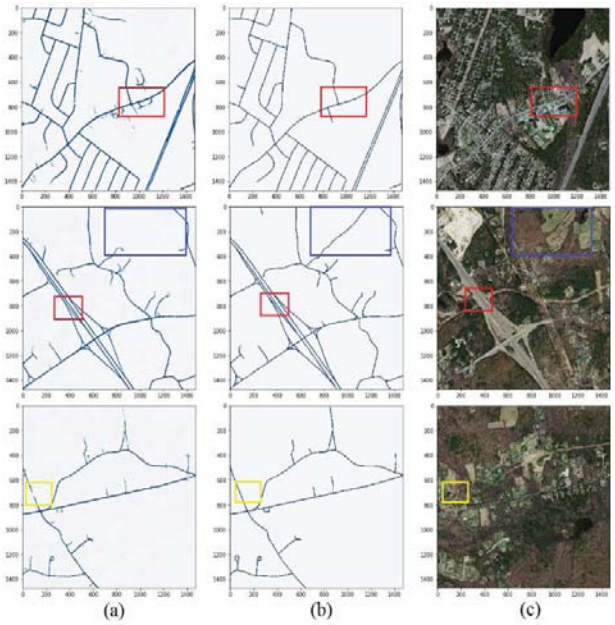


Fig. 4. Example results. (**a**) Proposed Unet. (**b**) Ground truth. (**c**) Input image

## IV. CONCLUSION

With the development of technology and science, satellite images are easier to obtain with high spatial resolution, and the detection of road information leads to better traffic management and to extract better road information. In this article we proposed deep learning framework for detection of road from Massachusetts dataset. We implemented a model based on the U-net architecture, it has been improved and trained in an efficient way so as to process dataset. Also, by a simulation experience, our model shows a good satisfying result that represents the quality of the road network. The results of the accuracy, precision, recall, and F-score show that our network can effectively rank the road network. Our model outperforms other models mentioned previously.

## REFERENCES

[1] T. Zhou, C. Sun, and H. Fu, "Road information extraction from high-resolution remote sensing images based on road reconstruction," *Remote Sensing*, vol. 11, no. 1, p. 79, 2019.

[2] C. Zhu, W. Shi, M. Pesaresi, L. Liu, X. Chen**, and B. King, "The recognition of road network from high‐resolution satellite remotely sensed data using image morphological characteristics," *International Journal of Remote Sensing*, vol. 26, no. 24, pp. 5493-5508, 2005.

[3] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *International Journal of Remote Sensing*, vol. 36, no. 12, pp. 3144-3169, 2015..

[4] Y. Zhang, "Optimisation of building detection in satellite images by combining multispectral classification and texture filtering," *ISPRS journal of photogrammetry and remote sensing*, vol. 54, no. 1, pp. 50-60, 1999.

[5] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy," *International Journal of Remote Sensing*, vol. 30, no. 12, pp. 3205-3221, 2009.

[6] Z. Miao, B. Wang, W. Shi, and H. Zhang, "A semi-automatic method for road centerline extraction from VHR images," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1856-1860, 2014.

[7] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images," *Journal of traffic and transportation engineering (english edition)*, vol. 3, no. 3, pp. 271-282, 2016.

[8] S. Tian, X. Zhang, J. Tian, and Q. Sun, "Random forest classification of wetland landcovers from multi-sensor data in the arid region of Xinjiang, China," *Remote Sensing*, vol. 8, no. 11, p. 954, 2016.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[10] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.

[11] O. A. Penatti, K. Nogueira, and J. A. Dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2015, pp. 44-51.

[12] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2015: IEEE, pp. 1873-1876.

[13] E. Maggiori, Y. Tarabalka, G. Charpiat, and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 645-657, 2016.

[14] I. Ševo and A. Avramović, "Convolutional neural network based automatic object detection on aerial images," *IEEE geoscience and remote sensing letters*, vol. 13, no. 5, pp. 740-744, 2016.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431-3440.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, 2015: Springer, pp. 234-241.

[17] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, "Road segmentation of remotely-sensed images using deep convolutional neural networks with landscape metrics and conditional random fields," *Remote Sensing,* vol. 9, no. 7, p. 680, 2017.

[18] A. Buslaev, A. Parinov, E. Khvedchenya, V. I. Iglovikov, and A. A. Kalinin, "Albumentations: fast and flexible image augmentations," *arXiv preprint arXiv:1809.06839*, 2018.

[19] V. Mnih, *Machine learning for aerial image labeling*. Citeseer, 2013.

[20] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011.

[21] Z. Zhong, J. Li, W. Cui, and H. Jiang, "Fully convolutional networks for building and road extraction: Preliminary results," in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016: IEEE, pp. 1591-1594.

[22] Y. Wei, Z. Wang, and M. Xu, "Road structure refined CNN for road extraction in aerial image," *IEEE Geoscience and Remote Sensing Letters,* vol. 14, no. 5, pp. 709-713, 2017.