Deep Learning

Project Report

Image Segmentation using Optimized UNet

Atishay Jain (2018B5A70908G)
Aman Jain (2018B5A70868G)
Shubhankar Ravindra Kate (2018B4A70786G)

Department of Computer Science and Information Systems
BITS Pilani K.K Birla Goa Campus

Abstract

We show the segmentation of images using the UNet Model. We apply this model on two different datasets. We obtain brain tumour segmentation by applying it on LGG dataset and segmentation of road maps by applying it on a dataset consisting of satellite images. The results are visualized and the accuracy is determined.

Contents

Acknowledgement

This Project becomes a reality with the kind support and the help of many individuals. We would like to extend our sincere thanks to all of them.

Foremost, we would like to thank our course instructors **Dr.Tirtharaj Dash** and **Dr.Tanmay Tulsidas Verlekar** for providing us the opportunity to explore such an interesting topic like this. Their lectures and tutorials helped us a lot in developing the strong fundamentals, and they built for us a very sound understanding of the subject.

We would also like to acknowledge the efforts of all the **Teaching Assistants (TA's)** of the course who provided regular guidance and took practical sessions which clarified a lot of concepts related to the subject.

Furthermore, we would like to extend our sincerest thanks to **BITS Pilani Computer Science Department (Goa)**, who provided us with ample opportunities to showcase our work.

Last but not the least, we would like to thank our family members, without whose motivation, this would not have been possible.

1 Introduction

Image segmentation is a method in which a digital image is broken down into various subgroups called Image segments which helps in reducing the complexity of the image to make further processing or analysis of the image simpler. Segmentation in easy words is assigning labels to pixels. All picture elements or pixels belonging to the same category have a common label assigned to them. For example: Let’s take a problem where the picture has to be provided as input for object detection. Rather than processing the whole image, the detector can be inputted with a region selected by a segmentation algorithm. This will prevent the detector from processing the whole image thereby reducing inference time.

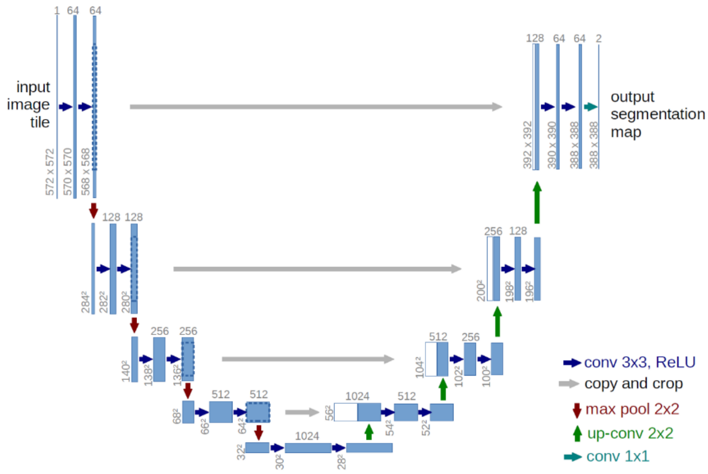
UNet is a widely used method for image segmentation, especially for multiband medical images. UNets have an ability to learn in environments of low to medium quantities of training data. The original paper[1] proposes a method to solve **Brain Tumour segmentation** problem using a modified Unet. We, in this work, showed that this optimized UNET can also be used for **Satellite Image Segmentation** problem owing to its similarity with the medical image segmentation task.

2 UNet

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 Unet 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). The encoder is about the convolution layers followed by pooling operation. It is used to extract the factors in the image. The decoder on the other hand uses transposed convolution to permit localization. It is a fully connected layers' network.

UNET — Network Architecture:

UNET is a U-shaped encoder-decoder network architecture, which consists of four encoder blocks and four decoder blocks that are connected via a bridge. The encoder network (contracting path) half the spatial dimensions and double the number of filters (feature channels) at each encoder block. Likewise, the decoder network doubles the spatial dimensions and half the number of feature channels.



Key aspects of UNet:

- **Convolution Layers:** Convolution operation are used learn information from images which then can be used as features for machine learning problems.
- **Down Sampling:** Sequence of convolution combined with max pooling results in down sampling. In down sampling size of image is reduced which means we can observe larger portion of image in a single convolution operation. Down sampling is a good approach for identifying what is present in the image. but for identifying where the object is we need to use upsampling.
- **Up Sampling:** It is just opposite of down sampling. We go from low resolution to high resolution. For up sampling UNet uses transposed covolution which is achieved by taking transpose of filter kernels and reversing the process of convolution. If you want to learn more, please check out this and this.

3 Brain Tumour Segmentation:

Brain tumour is one of the most dangerous cancers in the world. It appears when a certain type of brain cell, known as malignant, begins to grow out of control. In the past 30 years, the number of patients diagnosed with brain cancer has significantly increased, affecting many people throughout the world. This increase leads to a high risk of mortality with 241.037 cases in 2018. In 2012, eight million people died from cancer, taking all types of cancer together, while in 2013, six million people worldwide died of brain tumour. It is very important to diagnose brain cancer at an earlier stage as it allows for therapy and enhances the rate of survival. Brain tumour treatment options depend on the location, type and size of the tumour and may involve radiotherapy, surgery, chemotherapy or a combination of these options. Medical imaging is used to verify the presence and show certain characteristics of different types of brain tumours. There is a multitude of medical imaging modalities, including magnetic resonance imaging (MRI) and computerized tomography (CT), which are the most common ones used to explore brain cancer. The efficient classification and segmentation of tumours from surrounding brain tissues is a crucial task. In fact, an essential step is to exclude normal tissues by segmentation and extract more relevant characteristics of lesions for a better diagnosis. However, segmentation is a difficult task due to the wide variations in size, texture, shape and location of brain lesions.

3.1 Dataset

We present two choices of datasets that can be used for training the model:

3.1.1 Dataset Choice 1: BraTS 2021 Challenge Data

The training dataset used for the BraTS21 challenge consists of 1,251 brain mpMRI scans along with segmentation annotations of tumourous regions (annotated manually by experienced neuro-radiologist).

For each sample/patient, four modalities of mpMRI scans are given:

- Fluid Attenuated Inversion Recovery (FLAIR)
- T1-weighted pre-contrast (T1w)
- T1-weighted post-contrast (T1Gd)
- T2-weighted (T2)

Corresponding to each scan, we have one mask with four labels:

- 0 for background (voxels that are not part of the tumour)
- 1 for Necrotic tumour core (NCR)
- 2 for peritumoural edematous tissue (ED)
- 4 for enhancing tumour (ET)

PS: Label 3 is not used anywhere.

All BraTS multimodal scans are available as NIfTI files (.nii.gz)

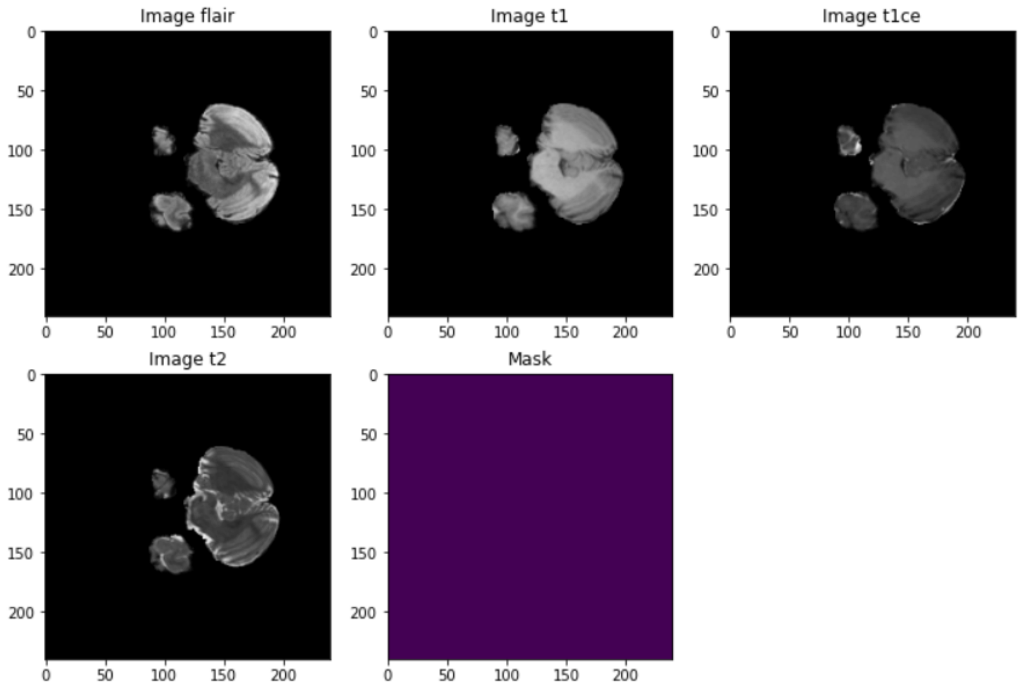


Figure 1: The images in this dataset are 3D. Note the pixel value range from 0 to 2289 (very high range of pixel intensity).

3.1.2 Dataset Choice 2: LGG Segmentation Dataset

The LGG Segmentation Dataset consists of 110 sets of mpMRI scans (each set corresponding to one patient) along with segmentation annotations of tumourous regions (annotated manually by experienced neuro-radiologist). Unlike BraTS dataset, where images were 3D and multimodal, here we have unimodal mulitple 2D RGB images (approx. 40 images per sample/patient) corresponding to a single scan, with a separate mask for each image.

MRI Modality in LGG Dataset:

- Fluid Attenuated Inversion Recovery (FLAIR)

Here, each image has a binary mask i.e., has only one label for tumour (does not categorize between the type/stage of tumour).

All LGG scans are available as TIFF files.

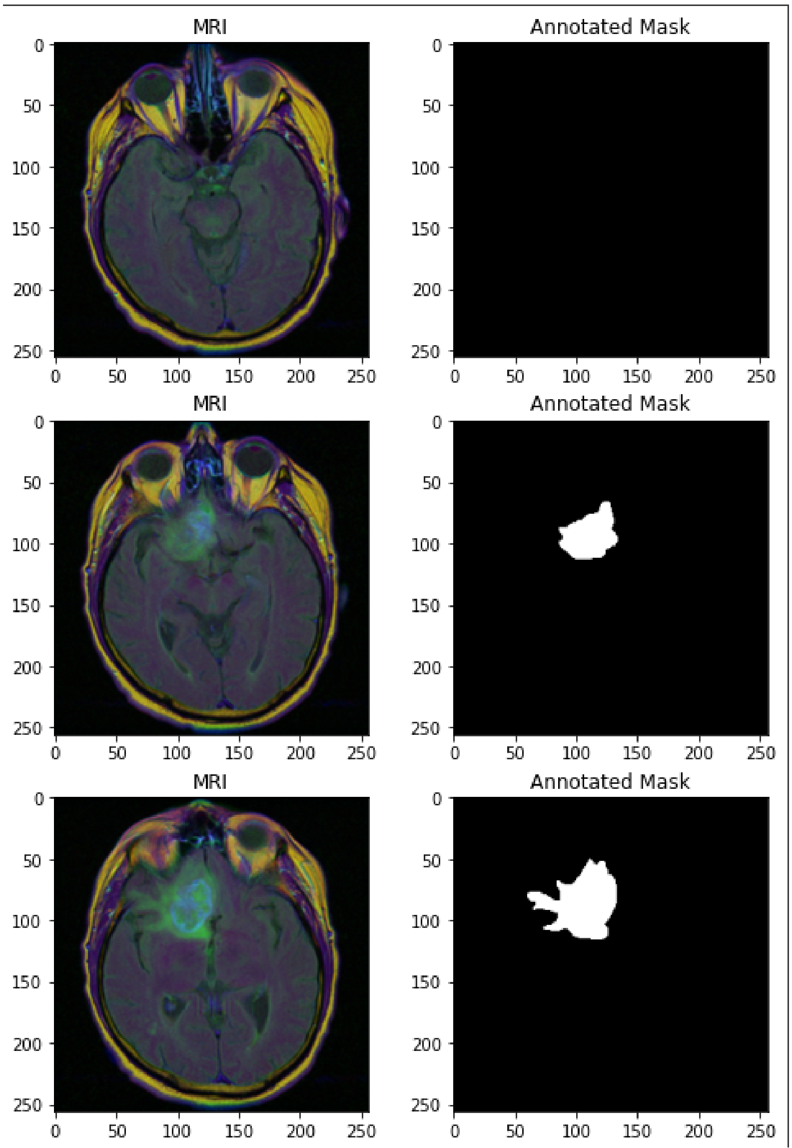


Figure 2: MRI Scan and the associated binary tumour mask.

3.1.3 Choosing the dataset

From the two datasets presented above, it is clear that the amount of data present in even one MRI image is huge which requires great computation power to process. Further, in various implementations of BraTS challenge it was found that only one high contrast modal of image is sufficient to get good results.

Since, LGG dataset is comparatively more practical to process on our machines/colab, without compromising much on the quality of the generated results, we prefer LGG dataset over BraTS in this project.

3.2 Visualizing the Results:

We have used dice loss as a loss metric for our model. The following plot shows how loss decreases for each subsequent epoch.

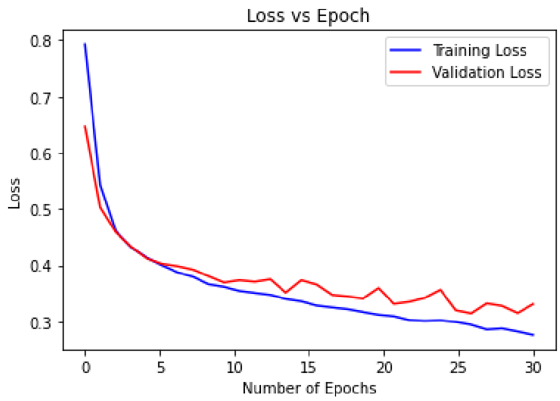


Figure 3: Plot of Loss vs Epoch

The left-hand side in the figure 4 shows the final pre-processed MRI scans that are given as input to the model. The centre column shows the binary tumour mask labels which are the target predictions we intend to achieve. The right most column shows the masks predicted by our model. It can be observed that the predicted mask has a decently accurate placement and shape of the tumour when compared to the actual mask.

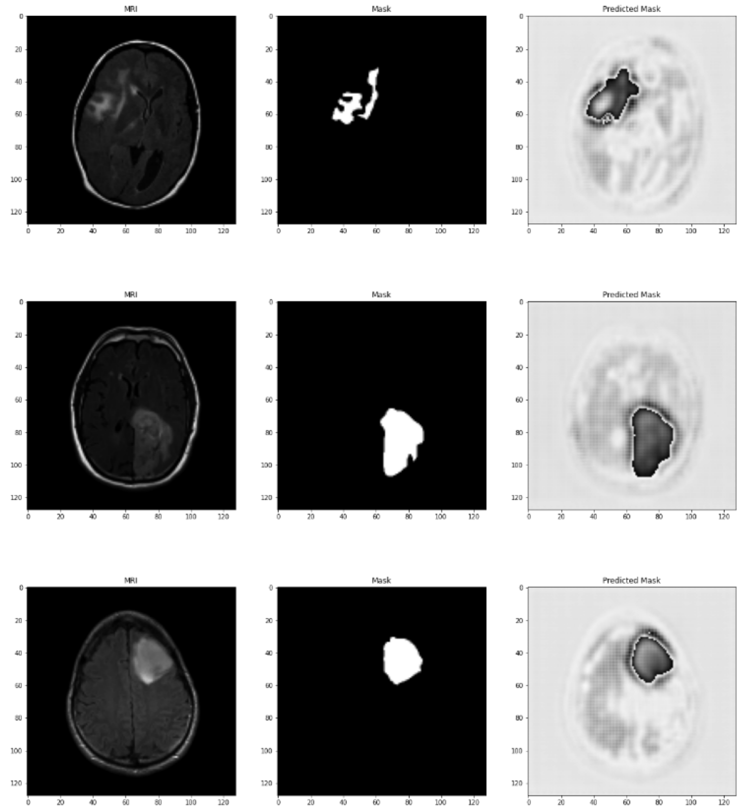


Figure 4: MRI Scan vs Actual Mask vs Predicted Mask

4.Satellite Image Segmentation

4.1 Why Satellite Imagery?

Satellite Image Segmentation, as medical image segmentation, is a challenging task because both these tasks use high resolution multimodal images. In both these tasks, annotations usually require in-depth domain knowledge which require special tools for data analysis. The deep learning model tries to analyse and extract these features without any human intervention or use of any tool. In this work, we run our optimized UNet model specifically for segmentation of road. Obstruction from nearby trees, shadows of adjacent buildings, varying texture and color of roads, road class imbalance (due to relatively few road image pixels) are among other challenges that hinder present day models in segmenting sharp road boundaries that extend from one end of the image to the other.

Further, datasets in both these tasks are often filled with different kinds of noise. Satellite Images are often unevenly oriented and cropped whereas Medical images are often filled with lots of blank space (with no label), both these issues are needed to be dealt with separately.

4.2 Dataset

The Massachusetts Roads Dataset consists of 1171 aerial images of the state of Massachusetts. Each image is 1500×1500 pixels in size, covering an area of 2.25 square kilometers. We randomly split the data into a training set of 1108 images, a validation set of 14 images and a test set of 49 images. However as discussed earlier in case of BraTS dataset, we make use of only 500 images with a training set of 400 images, and a validation set and test set of 50 images each. The dataset covers a wide variety of urban, suburban, and rural regions and covers an area of over 2600 square kilometers. The test set alone covers over 110 square kilometers. The target maps were generated by rasterizing road centerlines obtained from the OpenStreetMap project. A line thickness of 7 pixels and no smoothing was used in generating the labels. All imagery is rescaled to a resolution of 1 pixel per square meter.

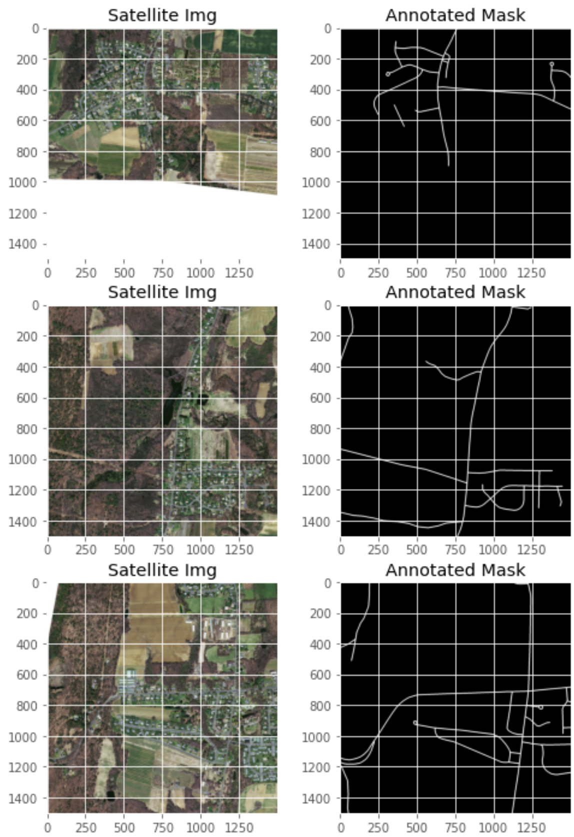


Figure 5: Satellite Image vs Annotated Mask

4.3 Visualizing the Results

We have used dice loss as a loss metric for our model. The following plot shows how loss decreases for each subsequent epoch.

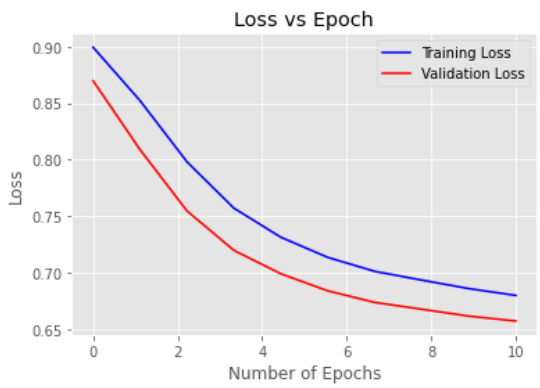


Figure 6: Plot of Loss vs Epoch

Here, we observe that the loss starts to converge to a higher value as compared to the loss in figure 3. This is mainly because here we run our model on relatively less data to use it as a proof of concept. Further, in this work, the preprocessing steps are kept same for both the datasets. Using preprocessing steps that take into account the specifics of satellite imagery would likely improve the results.

The left-hand side in the figure 7 shows the final pre-processed satellite images that are given as input to the model. The centre column shows the binary mask labels which are the target predictions we intend to achieve. The right most column shows the masks predicted by our model. It can be observed that the predicted mask has a decently accurate placement and shape of the tumour when compared to the actual mask.

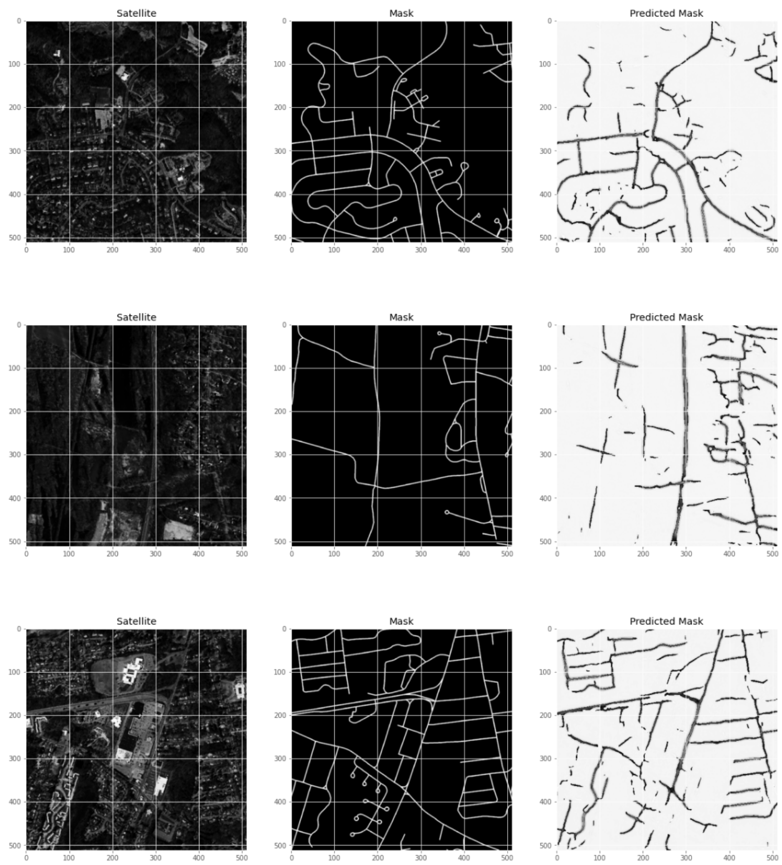


Figure 7: Satellite Mask vs Actual Mask vs Predicted Mask

5. Our Contribution

In this work, we verified that UNet image segmentation model which was essentially designed for medical imagery can also be used for Satellite image segmentation task. It draws parallel between the two important tasks and shows that optimization made for one task can be used for the other.

6. Future Scope

Now that we have tackled both these segmentation tasks we can now properly comment on how these tasks were similar yet different. We used similar techniques on the Satellite imagery dataset as we have used on the Brain tumor segmentation and got decent results which works as a proof of concept. If we use more efficient preprocessing techniques for the satellite images, something which tackles that kind of data specifically, then we can fairly assume that it will give us better if not similar results as we have obtained.

One approach for handling big images is by creating smaller “tiles” that can be easier to process, yet encapsulate the data efficiently. While resizing would have caused some information loss, by tiling we can tackle the problem more efficiently with minimal information loss.

Further we can use relatively new attention based U-et model on time series data of images like Deforestation data (satellite) and longitudinal MRI data of brain scans to observe how the model behaves when the temporal aspect is added. This only showcases one of the few possible problem domains that can be experimented with efficiently.

7. References

[1] Futrega, Michał, et al. "Optimized U-Net for Brain Tumor Segmentation." *arXiv preprint arXiv:2110.03352* (2021).

[2] U.Baid, et al., The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumour Segmentation and Radiogenomic Classification, arXiv:2107.02314, 2021.

[3] Mateusz Buda, AshirbaniSaha, Maciej A. Mazurowski "Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm." Computers in Biology and Medicine, 2019.

[4] Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention { MICCAI 2015. pp. 234} 241. Springer International Publishing, Cham (2015)

[5] Mnih, Volodymyr. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013.

[6] N. Y. Q. Abderrahim, S. Abderrahim and A. Rida, "Road Segmentation using U-Net architecture," 2020 IEEE International conference of Moroccan Geomatics (Morgeo), 2020, pp. 1-4, doi: 10.1109/Morgeo49228.2020.9121887.