Colonoscopy tissue segmentation from whole slide images

Final Project Report

El Malih David, Moalla Fatma, and Rami Hamza

CentraleSupélec, Gif-sur-Yvette 91190, FR
{david.el-malih, fatma.moalla, hamza.rami}@student.ecp.fr

Abstract. There is no doubt that very deep Convolutional Neural Networks (CNNs) have shown outstanding performance in object recognition for many applications especially for medical imaging. In this paper, we propose an application of deep semantic segmentation techniques to perform colonoscopy tissue segmentation from whole slide images (WSIs). Since these images are too big to fit into GPUs' memory, we have implemented a Patch UNet, which is a UNet trained on small patches extracted from WSIs. This model reaches quite good results but struggles with very big images, since the patches does not contain enough context to perform semantic segmentation. Hence, the next step of our implementation would be to add some context to the model as it is done in the DA-RefineNet [1]. Implementation: github.com/delmalih/WSI-Segmentation-Patch

Keywords: Semantic Segmentation · Medical Imaging · Whole-Slide Images · Colonoscopy tissue · Histopathology · UNet · CNNs.

1 Introduction

Several types of cancer like melanoma are very severe and aggressive which attack specific organs ans tissues, with a rapid decrease in survival rate if not diagnosed and treated at an early stage. That's why detecting early tumors and performing a detailed diagnosis is very important for pathologists. In fact, one of their main missions is to daily examine hundreds of tissue slices. In addition to being a time consuming and exhausting work, this method is subjective, qualitative, and seriously dependent on the professional skill of pathologists.

The development and the rise of imaging techniques, and the use of efficient and accurate automatic diagnosis algorithms will effectively supplement the skills of pathologist by giving a second opinion and consequently increase the accuracy of diagnosis greatly. Lately, digital pathology in general opened new avenues for pathologists.

In recent years, with the improvement of the performance of features extraction methods, the segmentation methods for Medical images based on Deep Convolutional Neural Network became very praised. More specifically, analysing a large number of high-resolution images has become possible and done in few minutes if we apply few pre-processing techniques.

Moreover, semantic segmentation on high-resolution medical images has proved its effectiveness in detecting early-stage tumors like the DA-RefineNet solution [1]. In addition , using a sliding window in recent papers has shown a good performance on the high-resolution of whole slide images.

2 Problem Definition

In this project, our objective is to detect tumors on colonoscopy with a performance as high as one obtained by a very-skilled doctor at least.

To accomplish that we will use semantic segmentation techniques of whole slide images of colonoscopy tissues. However some classical semantic segmentation techniques cannot solve the problem as high-resolution images can be memory-consuming for classical GPUs and therefore Unet[4] or other classical solutions cannot be applied. In addition, this task can be even more difficult to perform by human beings and class-imbalance can be a another challenge for any automated systems as we cannot have the same proportion of pixels with tumors and without tumors.

That's why we will be using many techniques that tackles these problem which are inspired from **Patch-UNet** and **DA-RefineNet**[1] methods. We will apply them on images (Figure 1) from *DigestPath2019* Dataset which contains whole slide scans of Colonoscopy tissues. Our main goal is to detect as much tumors as possible on these images.

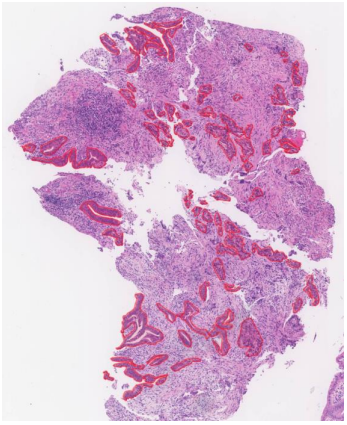

Fig. 1. Example of WSI with malignant lesion

In order to evaluate the methods that we will be using and will be explaining in the following sections, we choose the Intersection over Union metric (IoU), and the Dice Similarity Coefficient (DSC).

The IoU for a specific output of the model is defined as bellow:

$$IoU = \frac{|GT \cap P|}{|GT \cup P|}$$

With :

- GT : The ground-truth mask.
- P : The predicted mask from our model.

Then, we calculate the mean of IoU over all the images.

$$MIoU = \frac{1}{n} \sum_{i=1}^{n} IoU(image_i)$$

The second metric we choose is the Dice Similarity Coefficient (DSC) that is defined with the same notations as before as following:

$$DSC = 2\frac{|GT \cap P|}{|GT| + |P|}$$

3 Related Work

Several automated techniques have been proposed for sementic segmentation in medical imaging.

3.1 UNet-based approaches

The UNet network relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture of this network consists of a contracting path to capture context and a symmetric expanding path that enables precise localization. In Ronneberger's et Al. paper[4], its is proved that such a network can be trained end-to-end from very few images and outperforms prior best methods like a sliding-window convolutional network.

The following figure 2[4], explains the UNet architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The x-y-size is provided at the lower left edge of the box. White boxes represent copied feature maps. The arrows denote the different operations.

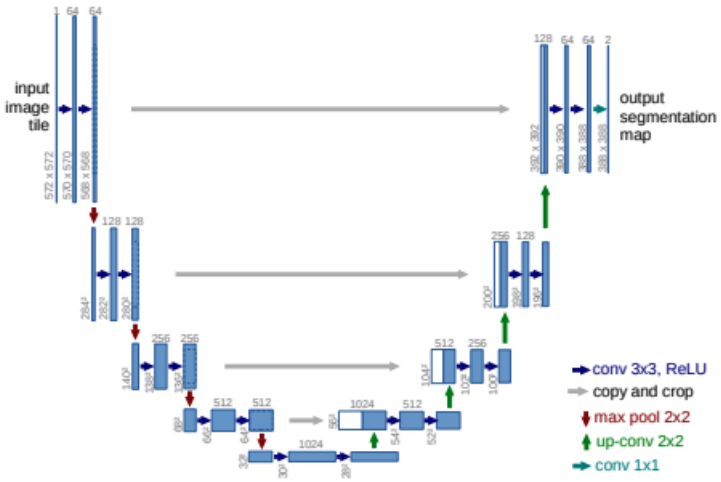

Fig. 2. UNet architecture

3.2 DA-refinenet

DA-Refinenet is a Dual Input Whole Slide Image Segmentation Algorithm published last year in July. The main contribution of this work is to propose a new feature extraction method for WSI segmentation. The idea is to extract rough global features and fine local features simultaneously over a given patch that holds that fine texture and its surrounding area (Context information).

The following figure 3 explains the architecture introduced by Li et Al. in their paper[1]. In fact, Patch/image A represents the slice image derived from the training dataset and holds the fine texture information and image B is the downsampled version of the corresponding train image and Keeps the rough contour of the image. or context info.

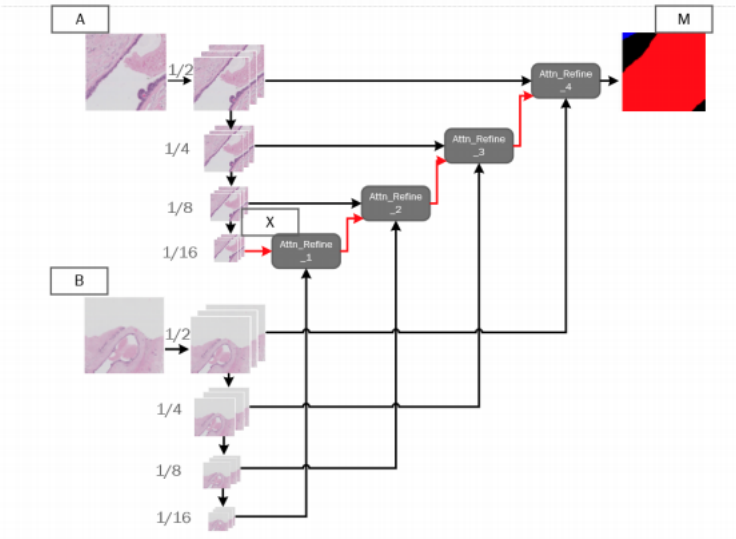

Fig. 3. DA-Refinenet architecture

3.3 Other patch-based solutions

Other papers propose a Polyp segmentation based on Fully Convolutional Network(FCN). In fact, Polyps are one of the main causes of colorectal cancer and early diagnosis of polyps by colonoscopy could result in successful treatment. That's why Akbari et Al. in their paper[5] used a wise method of patch selection for improving training phase of convolutional neural network.

FCN was used for its powerful ability in semantic segmentation. They then evaluated their method with different training sets and they achieved 81% of dice score outperformed previous methods in segmentation of colorectal polyps.

4 Methodology

Since we cannot train classical deep semantic segmentation models on the database of whole slide images, we decided to train a model that could segment small patches extracted from big images. We chose the UNet as model, hence we called the the **Patch UNet**.

The model (Figure 4) is the same as the classical UNet presented in [4] but we've added some layers that have been proved efficient for training. All blocks of the encoder (resp. decoder) part are composed of 2 residual blocks followed by a Downsampling (resp. Upsampling) layer. The Downsampling (resp. Upsampling) step is performed using Convolution (resp. TransposeConvolution) layer with a stride of 2, filter size of 3 and padding same. The residual blocks are composed of a first bottleneck block, which is a convolution layer with a filter size of 1, followed

by a Batch Normalization layer and a ReLU activation layer. This block allows the input to be in the exact same shape as the output (since we will add them). The bottleneck block is followed by these layers :

1. Convolution layer with filter size of 3 with padding
2. Batch Normalization layer with ReLU activation
3. Convolution layer and Batch Normalization layer
4. Output of previous layer summed up with the output of the bottleneck block
5. ReLU activation.

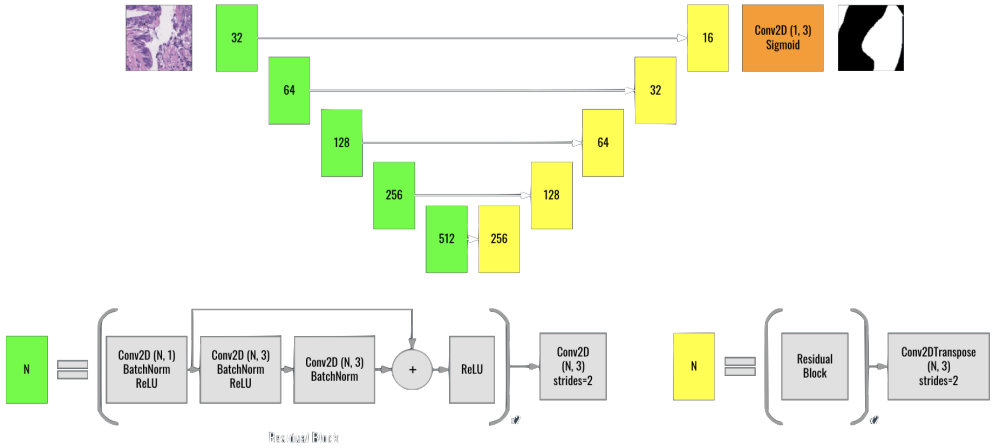

Fig. 4. Patch UNet Model

The training and inference processes are described below.

4.1 Training

To train the network, we first split our 250 whole slide images and masks into 2 subsets : Train and Val sets. The train set contains 200 images and the val set 50 images. For each image, we randomly sample 1000 patches of size 256x256 along with their masks. Hence, the train set is composed of 200.000 pairs of patches and masks and the val set has 50.000 ones.

In order to learn the weights from the patch dataset, we chose the binary crossentropy loss. Indeed, the segmentation task is, mathematically, as a pixelwise binary classification. Hence, the loss from one patch is defined by :

$$\mathcal{L} = \frac{1}{256^2} \sum_{i}^{256} \sum_{j}^{256} Y[i,j] \log P[i,j] + (1 - Y[i,j]) \log(1 - P[i,j])$$

with: Y the ground-truth mask and P the predicted one.

The model was trained with Adam optimizer (learning rate = 0.001) and a batch size of 20. To prevent overfitting, we used early stopping to stop the training. In other words, at the end of each epoch, we compute the loss on the validation set and we stop the training when the validation loss increases.

4.2 Inference

The inference process is the step of predicting a whole segmentation binary mask from a given WSI. To do so, we start by extracting patches of size 256x256 from the given WSI with a sliding window algorithm with a stride of 32. In other words, we extract the top left corner of the WSI of size 256x256, the we go right by 32 pixels extract the next patch, an so on and so forth, until we reach the bottom right corner of the image. Hence for a WSI of size 3000 x 3000, we extract 7396 patches.

We then give these extracted patches to the trained Patch UNet to predict their segmentation masks and combined these masks by putting them in the right locating on the whole slide mask. On the overlapping regions, we compute the average of all values. Finally, we get the final raw output mask of our WSI input. But this mask is not a binary one since its values are between 0 and 1. Hence, we need to apply a threshold to convert it into a binary mask (5).

Fig. 5. Examples of thresholded images. The image on the left is the raw output of the inference process. The middle (res. right) one is the raw thresholded with a threshold of 0.5 (resp. 0.9)

Since the last activation of the network is a sigmoid function, the most adapted value for the threshold should be 0.5. But we tried other values to check that assumption.

5 Evaluation and results

5.1 Experiment

Concerning the evaluation of the proposed approach, we used the public dataset *DigestPath2019* consisting of 250 WSIs of Colonoscopy tissues with an average size of 3000x3000.

5.2 Results

Here we will show results from different experiments depending on the threshold we set as we seen in Part 4.
Next, we will see how does the size of the patches affect the performance of the Patch-UNet model. We can see in the Figure below that increasing the size of the patches decreases the accuracy of the model since with large patches we start losing the context information contained within the patch.

Metric (Threshold)	IoU	DSC
Patch UNet (0.1)	63.26%	75.85%
Patch UNet (0.4)	76.39%	85.65%
Patch UNet (0.5)	77.09%	86.18%
Patch UNet (0.6)	76.56%	85.86%
Patch UNet (0.9)	59.58%	73.25%

Table 1. The performance of the Patch UNet

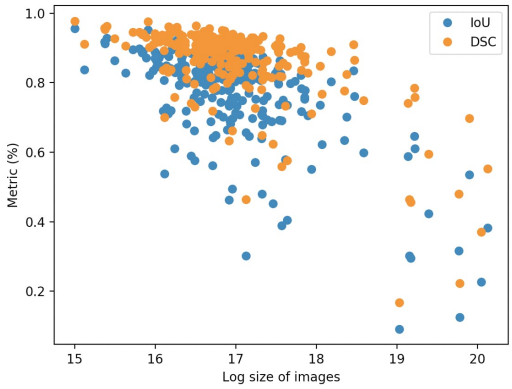

Fig. 6. Log size of patches vs the performance of the UNet

6 Discussion

In this paper, we propose **Patch UNet** which is a UNet based architecture. The training approach we are proposing consists on training the Unet model on patches extracted from the original data images in order to reconstruct the full mask of the WSI. As we have seen in the previous sections, this simple approach showed really promising results when applied to *DigestPath2019*. Yet, not good enough to answer real world problems, since the medical imaging field requires good results. Now the future work will consist on the limitations of the proposed approach; First, adding the Dice Loss or other losses in order to improve the segmentation task, then trying other sampling techniques since we are not handling the class imbalance problem in the dataset. Finally, we will think of way to get use of the WSIs that are not containing tumor cells and also capture the background context of the patches and consider rough global features.

References

1. Li Ziqiang, Tao Rentuo, Wu Qianrun, Li Bin: DA-RefineNet:A Dual Input WSI Image Segmentation Algorithm Based on Attention, (Jul 2019)
2. Guosheng Lin, Anton Milan, Chunhua Shen, Ian Reid: RefineNet: Multi-Path Refinement Networks for High-Resolution Semantic Segmentation, (Nov 2016)
3. Kay R. J. Oskal, Martin Risdal, Emilius A. M. Janssen, Erling S. Undersrud, Thor O. Gulsrud: A U-net based approach to epidermal tissue segmentation in whole slide histopathological images, (Jun 2019)
4. Olaf Ronneberger, Philipp Fischer, Thomas Brox: U-Net: Convolutional Networks for Biomedical Image Segmentation, (May 2015)
5. Akbari, M., Mohrekesh, M., Nasr-Esfahani, E., Soroushmehr, S.M., Karimi, N., Samavi, S., Najarian, K. (2018). Polyp Segmentation in Colonoscopy Images Using Fully Convolutional Network. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 69-72.

6. Digestive-system Pathological Detection and Segmentation Challenge, https://digestpath2019.grand-challenge.org/. Last accessed October 1, 2020