A DEEP LEARNING ARCHITECTURE TO SIMUOUTALY DETECT METASTASES AND CLASSIFY INTO PN-STAGE OF LYMPH NODE USING TENSORFLOW

Bram Geelen[a], *Xian Mao*[a], *Siamak Mehrkanoon*[a]

a ESAT - STADIUS, Stadius Centre for Dynamical Systems, Signal Processing and Data Analytics
Kasteelpark Arenberg 10 - box 2446, 3001 Leuven

ABSTRACT

This paper proposes a deep learning architecture using TensorFlow that can effectively detect and classify histopathological whole-slide images(WSI) of human lymph node tissues for improved diagnosis and prognosis of breast cancer metastases. Appearance of metastasis in lymph node is of utmost importance in prognosis of breast cancer. However, the accuracy of manual detection of the metastasis is suboptimal and very costly. It is easy but non-trivial to miss out metastatic cells that are interspersed among other cells, leading to inaccurate prognosis and staging. In addition, many effort exits in assisting pathologists at a lesion-level such as cell nuclei detection and image segmentation but few of them can classify a series of images in patient-level into a cancer pN-stage. This stage is crucial and is usually based on the extent of metastases which is relatively sensitive to errors in detecting metastases. Nevertheless, our method shown improve performance than the state-of-the-art.

***Index Terms*—** deep learning, whole-slide images, lymph node, U-net, TensorFlow

1. INTRODUCTION

Crucial diagnoses and reports that directly affect patient care are made by the pathologists, mostly using microscopes and hematoxylin-eosin (H&E)–stained tissue with the golden-standard techniques established more than 100 years ago. However, there are several inherent variabilities that can affect the results of these tests, including inter-pathologists difference and various methods of image acquisition. Whole slide imaging (WSI) with algorithmic analysis is (one potential technology to improve the assessment of histopathological markers. Studies of this kind can be found from basic H&E slides to IHC to multiparametric quantum dot staining. Although WSI is gaining more and more attention, they merely serve as a digital replacement of the microscopy in most situations and settings while their great advantages of being digitalized are left behind. As every cell turned into bits and bytes, many manually intensive tasks of pathologist can be converted to a computational problem and be solved and automated. On the other hand, with the collective advancement in computer science, computational methods such as neural network and deep learning [1] have been strongly developed and utilized in many both knowledge and labor -intensive domains such as financing, astronomy, etc. There is also a growing interest in its application in radiological images as well as histopathological images since deep learning approaches are relatively good in extracting high level features of objects in images compared to other methods in the field of Artificial Intelligence and machine learning.

For instances, back in 2002, Zhou et al.[2] used feature extraction methods to grab attributes about every nucleus, which were passed into ensemble neural networks. The authors had different networks to predict whether a cell is cancerous, and later what type of (cancer) cell it is. Cireşan et al.[3] used the Sliding window approach in combination with convolutional neural networks to achieve segmentation. This means that a small window is slided over the whole image, and the window is classified. In essence, this method tries to classify the middle pixel only, which makes the method very computationally intensive (especially for large images). This method is also used in the winners of CAMELYON16. Using the same datasets, Google reached state-of-the-art results in some test statistics, but not in the statistic that is used to rank the submissions in the challenge. This team also experimented with using more complicated techniques like multi-scale inputs, but found that this did not improve the classification results. They did, however, bring the challenge more into mainstream media and demonstrate again the usefulness and difficulties of using deep learning for automatic detection of metastasis.

Recently, Ronneberger et al.[4] proposed a new structure, namely the U-net architecture. This architecture makes use of a new concept, called the fractionally strided convolution, which makes it possible to classify every pixel in a tile. This improves the classification and training speed. The first half of the network is alike to a regular convolutional neural network. In the second part, the fractionally strided convolution technique is used. Note that the key (new) feature in the U-net architecture is the copy and crop from the first half to the second.

An advantage of this method is that it is a lot faster than the sliding-window approach, because it can classify whole areas at once. A disadvantage of U-net is that it can't be made to segment images on a different detail level than their input.

This is problematic in our case, as we have very detailed input data, but the annotations are rather rough.

As for classifying the patient (classifying the group of five images), this approach will not work. The complete images are simply too large to feed into a neural network. Moreover, not all the images are of the same size, so if we use a tiling input, the number of tiles will differ.

Consequently, inspired by this U-net, we build our architecture by taking its foresaid drawbacks into account.

2. MATERIALS

2.1 WSI Data and annotation

The data set for CAMELYON17 is used. These are WSI of hematoxylin and eosin (H&E) stained lymph node sections and collected from 5 medical centres in the Netherlands. For training, 100 patients are used as training set and another set (test set) 100 patients are kept untouched until the final testing. In total, we used 1000 slides with 5 slides per patient. For the training set, detailed annotations of metastases in 50 WSI on lesion-level are known (10 training slides from each medical centre) and on a patient-level we have a pN-stage label per patient. Lesion-level annotations are in XML files and WSI are TIFF images

2.2 Library and software used

We use TensorFlow for our computational analysis and we used OpenSlide to read the WSI files. In additional, we used Automated Slide Analysis Platform (ASAP) to view the slides and annotations. We wrote our code in python 3.

3. METHODS

3.1 Preprocessing

We filtered out the non-tissue region by applying a threshold. We do not use the pixels that their peak-to-peak values is less than 10, as in Figure 4. We also build a performant rasterizer, as in Figure 3 for tiling. Moreover, we use WSI at the zoom level of 1.

3.1 Deep learning architecture

The solution we present consists of two networks. In the first network, as shown in Figure 1, we construct a latent representation for every tile of the input image. Afterwards, we feed these latent vectors into a second convolutional network, as shown in Figure 2, to predict the stage label of each node more accurately.

The first network is inspired by the U-net, this type of convolutional neural network can be used for image segmentation, and in our case, labeling the parts of an input tile that are in metastasis. Its architecture is composed of several convolutional and max-pooling layers, which decrease the representation resolution but increase the number of channels. These layers are followed by a sequence of fractionally strided convolutions, which make the network symmetric. Additionally, so-called 'skip connections' are added between the first and the second part of the network, which aim at increasing the detail of the output segmentation. However, our network will differ in some ways from the original U-net structure.

Firstly, we want to learn some abstract representation with which we can predict the stage label of the given node. For this, we will branch out the network in the middle, to two sequential fully connected layers. In principle, the neural network could learn to extract information that is relevant in predicting the label of the whole image into the middle part of the architecture. For that reason, we branch out from the middle representation of the U-net to predict this label of the whole image.

In this branch, we go from the output of the last convolutional layer into a fully connected layer of size 32. The output of this fully connected layer is the latent representation we will use in the second layer. We also use the output of this layer to predict the stage label of the image. This prediction is shown in the bottom part of Figure 1. Adding two fully connected layers will improve the prediction accuracy for the tile, but it will also make sure we catch some useful abstract features in the latent representation.

Secondly, we do not implement the 'skip connections' that are present in the U-net architecture. Because the segmentation of the image is less important than accurately predicting the stage label of the node, we want to make sure as much relevant information toward this latter goal is passed through to the middle representation, instead of being 'skipped' over it.

Lastly, we make use of batch normalization. This technique was proposed by Ioffe et al [5]. Using batch normalization layers after every convolutional layer (both normal and fractionally strided layers) has a regularizing effect on learning the network parameters, to avoid overfitting.

To train this first network, we used softmax cross entropy for both the segmentation output as the stage label output. We combined the loss of these two classifications by weighting the stage label prediction loss by 256^2 against the loss of the segmentation output, which is approximately this much larger.

Once this first network is fully optimized, we can use it to produce the abstract representations of the tiles. Using these abstract representations, we can train a second convolutional neural network, as the size of the input tensor will be much smaller than the size of the whole slide. The precise structure and sizes in this second network is illustrated in as shown in Figure 2.

The latent representations that are obtained in the first network are concatenated together, in the same order as they were in the original image. We center and crop the resulting matrix of abstract representations to construct the input tensor. This tensor is fed into a shallow convolutional neural network, with few variables. Thus, we can easily train the

network and quickly predict the output once we converted the tiles into the abstract representations.

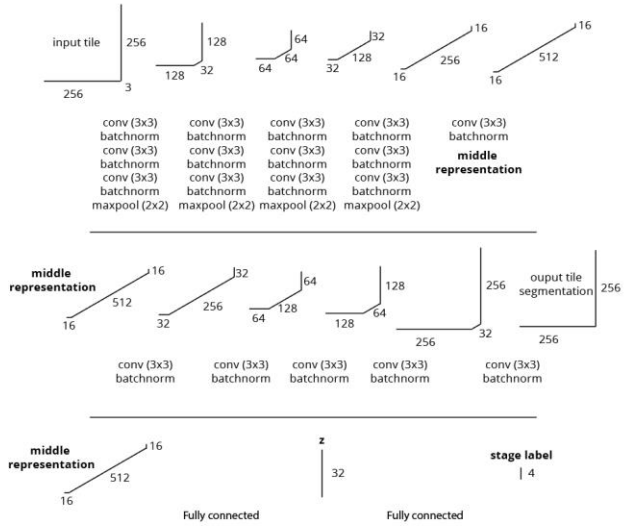

Figure 1 Deep learning architecture to obtain the latent representations of WSI.

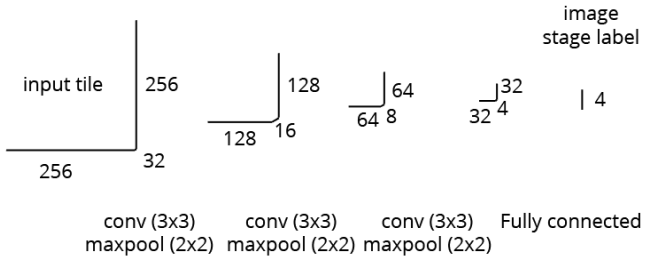

Figure 2 Deep learning architecture to predict the stage label

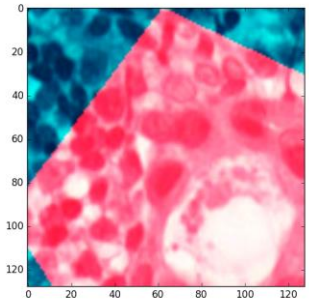

Figure 3 Effect of rasterizer showing a tile with annotated areas

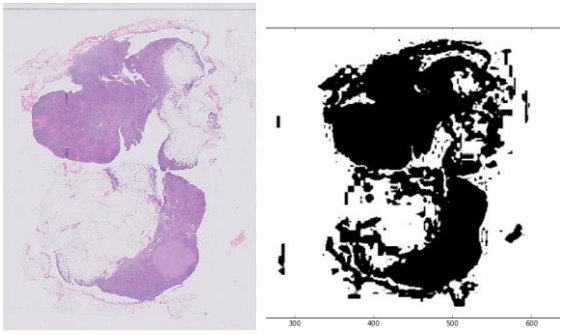

Figure 4 Effect of applying threshold for preprocessing WSI

4. RESULTS AND DISCUSSION

Please note our submitted result contains the valid prediction of the first 36 patients due to an unforeseen technical issue. On average, it took 1 minutes per WSI for obtaining its latent representation.

One of features of our architecture is that it simultaneously detects metastasis and classify the patient into pN-stage. Unlike other methods for stage classification that are rely merely and sensitive to the predicted metastases, our method take into more tumorous factors into account. For instance, more and more clinical studies have shown that the local and global environments of the tumour are as important as the tumour cells themselves such as the stroma and fibroblasts and our architecture consider these environments during training. This biomedical phenomenon might account for the fact that current segmentation efforts still are not widely used in the clinic routinely whereas more and more sophisticated and complicated models are being made

Note that we cannot be certain that there is any useful information in this part of the network; the network might train itself to segment the image purely by using the 'copy and crop' connections.

The reason that we choose RNN for the task of classifying pN-stage is that as far as neural networks are considered, this is the only method to classify sets of data that are irregular in size.

We used batch normalization frequently so a drop-out step can be omitted.

The use of max pooling avoids the risk of being over-engineering.

We experimented with random pooling with the aim to speed up the extraction of latent representations. Yet it didn't work well as expected. We think this is because the time needed for reading one tile from a WSI was not changed.

We also made a performant augmentation tool that can get the most information out of the annotations made by the doctors for future improvement. This tool can create a batch of images from a single tile by randomly flipping the image in both directions, and applying some small color variation. Applying this should make the neural network generalize more.

Furthermore, the fact that we build our algorithm via TensorFlow makes our work as generic, reliable and versatile as TensorFlow and compatible for back ends of digital pathology platform such as Pathomation.

In conclusion, using the powerful TensorFlow, we developed a promising deep learning architecture to simultaneously detect metastases for each WSI and classify series of WSI into pN-stages for each patient.

5. REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097-1105.

[2] Z.-H. Zhou, Y. Jiang, Y.-B. Yang, and S.-F. Chen, "Lung cancer cell identification based on artificial neural network ensembles," Artificial Intelligence in Medicine, vol. 24, pp. 25-36, 2002.

[3] D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in International Conference on Medical Image Computing and Computer-assisted Intervention, 2013, pp. 411-418.

[4] 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, pp. 234-241.

[5] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.