UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Tissue Segmentation in Histopathological Whole-Slide images with Deep Learning

Author:
Oleksandr PRYHODA

Supervisor:
MSc. Dmytro FISHMAN

A thesis submitted in fulfillment of the requirements for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2019

Declaration of Authorship

I, Oleksandr PRYHODA, declare that this thesis titled, "Tissue Segmentation in Histopathological Whole-Slide images with Deep Learning" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Tissue Segmentation in Histopathological Whole-Slide images with Deep Learning

by Oleksandr PRYHODA

Abstract

Development of technologies led to the adoption of new digital imaging solutions in pathology field. One such innovation is whole slide imaging, the main purpose of which is digitalizing the whole glass slide with tissue into a high-resolution image. This image is then divided into sections, which are zoomed for further analysis. The main focus of examination is tissue body, but other materials such as debris, dust, and glass are also presented on the slide. In order to focus only on tissue and to make the analysis process more time- and memory-efficient, tissue location on the slide is predefined. Currently, tissue localization procedure is performed by segmentation algorithms based on classical methods of computer vision. These algorithms require manual tuning and might be inaccurate on images with a lot of debris. The issue could be solved with more adaptive methods like deep neural networks. This thesis presents tissue segmentation pipeline based on deep convolutional neural networks. Proposed pipeline showed that deep learning is capable of segmenting tissue as accurately as the currently employed approach.

Acknowledgements

I am very grateful to my supervisor Dmytro Fishman (University of Tartu, Quretec Ltd.) for spending a huge amount of time for guiding me through all thesis research, for useful advice and active discussions.

Separately, I would like to thank all organizations which have been involved in this project. This work has been funded by PerkinElmer, Inc. and supervised by members of the Institute of Computer Science of University of Tartu. Computational resources have been provided by the High Performance Computing Center of University of Tartu. Imaging data used in this work has been provided by Akoya Biosciences, Inc.

Finally, I want to thank Ukrainian Catholic University and Faculty of Applied Sciences for building awesome Bachelor's program in Computer Science and supporting my study with scholarship.

Contents

List of Figures

List of Tables

List of Abbreviations

SOTA	**S**tate-**o**f-**t**he-**a**rt
WSI	**W**hole-**s**lide **I**mage
ANN	**A**rtificial **N**eural **N**etwork
CNN	**C**onvolutional **N**eural **N**etwork
DNN	**D**eep **N**eural **N**etwork
ROI	**R**egion **o**f **I**nterest
RPN	**R**egion **P**roposal **N**etwork
IoU	**I**ntersection **o**ver **U**nion

Dedicated to my family

Chapter 1

Introduction

1.1 Motivation

Histopathology is the study of tissue structure, from the level of cells to the level of individual organs (Hani A Alturkistani, 2015). Analyzing tissue slices contributes to important areas in medicine, e.g. disease diagnostic (Farahani N, 2015). Analysis procedure typically consists of several stages. Firstly, researches take a low-resolution photo of thin tissue slice on a glass. Then, by receiving this photo, the specialized microscope focuses on all positions where the tissue is located and captures them. The focusing points microscope obtains from segmentation algorithms that separate the tissue from the rest of the debris in the picture. As a result, researchers gain a big amount of images with tissue sections at the level of individual cells which are used for further analysis. If the algorithms that determine the focusing points for the microscope do not work accurately, a lot of obtained images will be useless and should be filtered.

Currently, semi-automated classical methods of computer vision are utilized. Mostly, they are based on colors classification, edge detection, thresholding and region growing techniques. These methods have some disadvantages. They heavily rely on parametric tuning and might work inaccurately when a lot of debris is presented on the image. One of the best such methods, which is utilized by Akoya Biosciences has been compared with the proposed deep learning method in this thesis.

Relying on independence from manually choosing or tuning parameters (except training procedure) of deep learning methods, full automatization of tissue segmentation task seems promising. Moreover, deep learning showed very good performance on the segmentation problem of natural objects on COCO (Lin et al., 2014) and PASCAL (Mark Everingham, 2014) datasets (Garcia-Garcia et al., 2017). Segmentation methods based on deep learning have been successfully utilized in various medical areas, for example they were shown to perform better than the state-of-the-art traditional methods in brain MRI (Akkus Z, 2017) and cell segmentation tasks (Medeiros, 2019; Ronneberger, Fischer, and Brox, 2015).

1.2 Goals

- Test the hypothesis that deep learning method is capable of solving tissue segmentation problem in histopathological whole-slide images by developing deep learning pipeline
- Evaluate whether even with a small number of images (81 samples) deep learning could reach the performance of the current leading approach

1.3 Structure

Chapter 2: Related works
This chapter presents an overview of tissue segmentation algorithms used in histopathology field. It starts with algorithms that rely on thresholds and concludes with current deep learning methods.

Chapter 3: Background Information
This chapter introduces the main concepts and technologies used in this thesis: whole slide imaging in histopathology, artificial neural networks, convolution neural networks, object detection, object segmentation, and performance evaluation metrics for neural networks.

Chapter 4: Dataset description
In this chapter, an overview of the dataset is presented along with a description of how data was acquired and a few visual examples.

Chapter 5: Methodology
Methodology chapter describes in details the proposed pipeline. Model architectures, data preprocessing strategies, and training details are discussed.

Chapter 6: Experimental results
In this chapter, some intermediate results and the evaluation of the proposed pipeline is presented along with the comparison to SOTA approach. Results are visualized highlighting both strong and weak aspects of competing methods.

Chapter 7: Conclusions
Conclusions chapter summarizes the obtained results, discusses the importance of the work and future improvements.

Chapter 2

Background information

2.1 Whole slide imaging

Whole slide imaging is about producing digital images by scanning the glass slide. It is becoming more and more popular among pathologist for diagnostic, research and educational purposes. Whole slide imaging consists of two procedures. The first one is about utilizing specialized hardware (scanners/microscopes) for producing a high-resolution image by digitalizing glass slide. The second procedure is about analyzing produced large images by specialized software. These whole slide images are then used for medical purposes such as primary diagnostic, consultation, analysis, and researches. There are some difficulties with the adoption of whole slide imaging which relates to cost, time, digital slides storage, limiting technology and etc. (Farahani N, 2015)

One of the directions for solving faced difficulties is software algorithms development. In this thesis, the algorithm for automative tissue segmentation is presented.

2.2 Artificial Neural Networks

Artificial Neural Network (ANN) is the mathematical model created by inspiration of human brain structure. Mostly, such models are applied to pattern recognition and data classification problems. The ANN consists of linked layers of nodes. There are three types of layers: input layer - receives the initial information, output layer - produces the network result and hidden layers - information processing layers between input and output. Each hidden layer in the network receives the output of the previous layer and generates the input for the next layer. The layer itself is described by one or more nodes. The node (neuron) receives the signal (information) from the nodes of the previous layer, sums and computes this information with some nonlinear function and then transmit it forward to the nodes of the next layer. Each connection between nodes is called an edge. Simple artificial neural network architecture with fully-connected layers (every neuron of one layer is connected to every neuron in another layer) is presented in Figure 2.1. Neurons and edges have a weight that adapts for solving specific task during training. The signal strength on the connection is regulated by those weights, which in general defines the influence of the connection on the network's performance.

The network training process is about teaching the model to distinguish the patterns of data, by modifying the network's weights. This is done in order to prepare a model for solving a specific problem on specific data. Training data usually consists of pairs - input and ground truth output. During the training, the network receives

and processes the input information and return the output - prediction. Network error is evaluated by the comparison of correct ground truth values and networks prediction. Errors are then propagated back through all network to update the weights for further better performance. The procedure of forward and backward passes is repeated a lot of times in order to minimize the difference between the network output and ground truth data.

Typically ANN consists of many layers which allow learning more complex patterns. The large volumes of data are required to train model well. Usually, due to memory limits, it is not possible to pass all data to the model at once, that why the data is cropped into smaller portions called batches. The procedure of passing all batches to the model is called one iteration of training or one epoch. The required number of epochs for training depends on the model architecture and the size of the dataset.

ANNs with multiple hidden layers are called Deep Neural Networks (DNNs). Deep learning is the area of research which deals with DNNs.

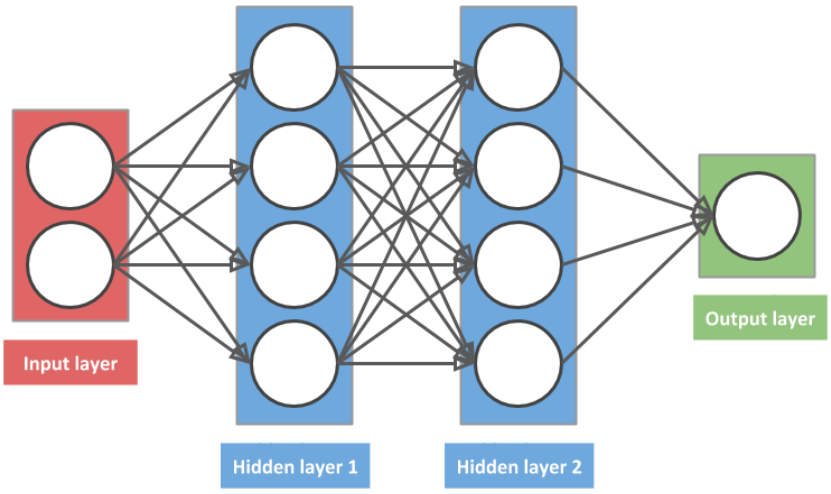


FIGURE 2.1: A schematic example of an artificial neural network with input, output, and two hidden layers

2.3 Convolutional Neural Networks

A Convolutional Neural Network (CNN) is a type of artificial neural networks primarily applied for solving computer vision tasks. The main concept of the network lies under convolution layers. Here each neuron is concentrated in own restricted region of the input image, called receptive field. Weights of the neuron are presented in the form of a matrix called the filter or kernel. Those filters are utilized for the feature extraction from the image by performing the mathematical operation. This procedure takes two inputs - two matrices: a local region of an input image and a filter, multiplies them and receives as an output the feature map - extracted features from the image. This mathematical operation is called convolution. The whole CNN consists of many convolutional layers. The first layer is responsible for capturing low-level features on the image such as color, edges, gradient orientation.

With added layers network combines the low-level features in order to learn more complex (Figure 2.2).

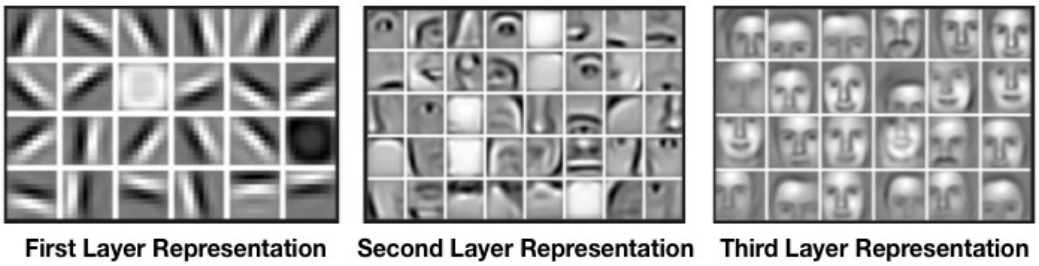


FIGURE 2.2: An example of learned filters in convolution neural network for face recognition problem. Shallow layers are responsible for capturing low-level features such as lines and edges, middle layers for more complex - eyes, nose, ears, and final layers for high-level features such as facial structures (*Image source*)

2.4 Object Detection

Object detection is a problem that deals with finding individual objects in an image. Detection is performed by predicting coordinates of the bounding box. The bounding box is the rectangular area which fully covers the object (Figure 2.3).

Recently, deep learning methods have been proven to be a powerful tool for object detection problems (Zhao et al., 2018; Ajeet Ram Pathak, 2018; Erhan, 2013). Some of the most noticeable algorithms in this area are You Look Only Once (Redmon et al., 2015), Single Shot Multibox Detector (Liu et al., 2015), Faster R-CNN (Ren et al., 2015). The latter is the state-of-the-art object detection approach used in this thesis. Faster R-CNN is the system composed of three parts: Convolution Neural Network, Region Proposal Network, and bounding box regressor and classifier. A detailed description of all parts of the architecture is presented in the Methodology chapter 5.5.1.

2.5 Semantic Segmentation

Semantic segmentation is one of the key problems in computer vision. In object detection task the goal is to define the bounding box in which the object is located, but for semantic segmentation task, each pixel on the image has to be assigned with a class label (In this thesis two classes are presented: tissue and background - Figure 2.3). Some of the most distinct deep learning approaches for semantic segmentation are U-Net (Ronneberger, Fischer, and Brox, 2015), Seg-Net (Badrinarayanan, Kendall, and Cipolla, 2015), FCNN (Long, Shelhamer, and Darrell, 2014) and Mask-RCNN (He et al., 2017). In this thesis, the U-Net model is used, which has been specifically designed for biomedical image segmentation and is capable to be trained and work well with very few images.

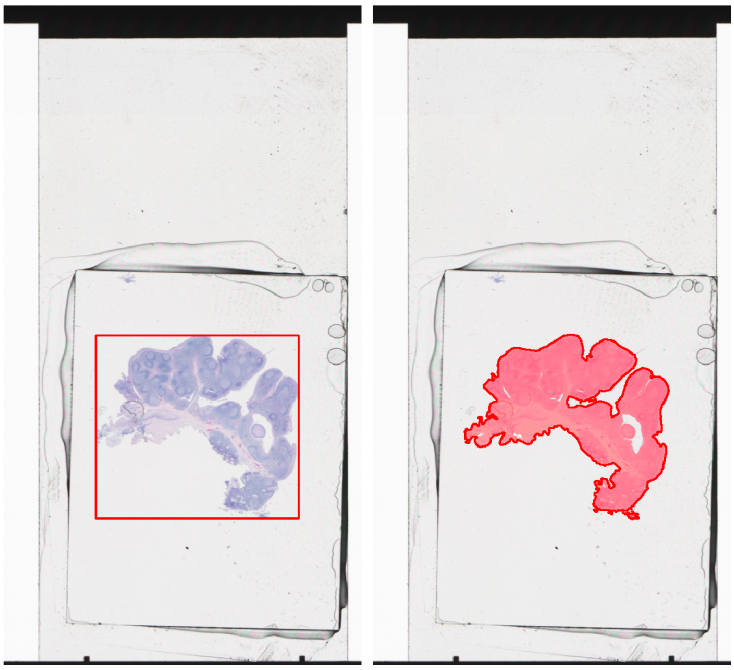


FIGURE 2.3: An example of the final result of solving two problems in computer vision. The left image represents an object detection task by defining the bounding box of the tissue. The right image represents a segmentation task by labeling each pixel which corresponds to tissue.

2.6 ANN performance evaluation

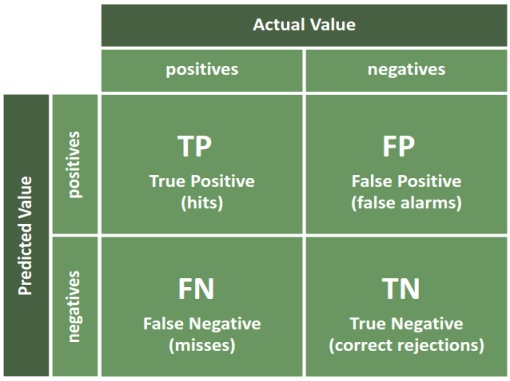
For training the artificial neural network whole data set is divided into three parts:

- **Training set** - usually, this portion of data includes the biggest number of samples from the whole data set and is used for training the model.
- **Validation set** - this set is used for intermediate evaluation of the trained model. Based on models performance on the validation set, the hyperparameters could be tuned for model retraining.
- **Test set** - the sample of data used for the final evaluation of model performance.

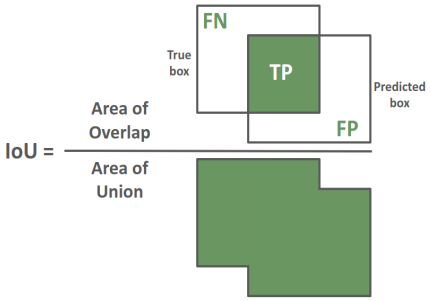
In this thesis segmentation task could be characterized as a binary classification task with two target classes of prediction for each pixel: positive - tissue, negative - background. For evaluation of binary classifier, a confusion matrix (Kohavi, 1995) is used (Figure 2.6).

From the confusion matrix the following rates are often calculated:

- **Accuracy** = (TN + TP) / (TN + TP + FN + FP)
- **Recall**, true positive rate, sensitivity = TP / (FN + TP)
- **Precision**, predicted positive value, confidence = TP / (FP + TP)
- **F1-score**, harmonic mean of precision and recall = 2 * (Precision * Recall) / (Precision + Recall)



(A) Confusion matrix



(B) Intersection over union metric

Also, for measuring the accuracy of an object detector or segmentor the intersection over union metric (Jaccard index) is used. It is described as a correlation between the area of overlap and area of the union of the predicted bounding box/mask to ground truth bounding box/mask (Figure 2.6).

- **Intersection over union** = area of overlap / area of union

Chapter 3

Related work

3.1 Tissue segmentation in medical imaging

Histopathology image analysis has been historically operated by a human expert observing the tissue body on the glass slide using a microscope. With the appearance of whole slide imaging (Pantanowitz, 2010), it became possible to analyze histopathology slides using digital algorithms. Special scanners first capture the entire slide with tissue, then save it as a digital image and then the gained image could be processed. At this point, new direction of computer-assisted diagnosis (CAD) algorithms for histopathologist started to develop (Doi, 2007). These algorithms played a huge role as additional help for histopathologist in detection, diagnosis and prediction prognosis of diseases (Gurcan, 2009).

One of the digital analysis procedures for histopathology imagery is the detection and segmentation of the tissue body. This procedure is done in order to minimize the surface that needs to be screened by the scanner. In this way, it is possible to skip glass background or detritus present on physical slides. It saves researches time and computer memory. Thus, a lot of effort has been invested into improving and automating the tissue segmentation task (Pham DL, 2000).

At first, segmentation was done using semi-manual methods like colours classification, edge detection, and region growing (Banu, 2012; G. Evelin Suji, 2013). These algorithms require manual tuning of input parameters such as threshold values of color, intensity, and texture of the pixels. But manual parameter search takes a lot of time and effort from the users. Also, these approaches work well on images that contain solid objects on contrast background, but poorly on the images with a lot of noise. These disadvantages of "thresholding" methods (Banu, 2012) motivated research and development of automated patterns recognition methods such as Markov Random field models, Artificial Neural Networks, deformable models, atlas-guided approaches (Pham DL, 2000; Chijindu, 2012; Tsechpenakis, 2011) and machine learning techniques (Komura and Ishikawa, 2017).

Raja et al., 2009, proposed the method which combines machine learning and traditional methods for feature extraction. The 21x21 pixel window was used to extract colour, pixel intensity, spatial and texture characteristics. These features were normalized, cleaned and the dimensionality was reduced using Principal Component Analysis (PCA, Jolliffe IT, 2016). Resulting feature vectors were classified into background, tissue body or detritus using simple two-layer Neural Network Architecture. Segmentation mask was generated by grouping classified feature vectors. Authors applied a median filter and border map on final mask to eliminate border connected components. Results on 2 independent sets of 294 images have shown general pixel segmentation accuracy of 96.5%.

After 4 years, Hazem Hiary with authors of the previous paper proposed an unsupervised approach for segmentation (Hazem Hiary, 2013). The new solution

was based on learning K-means cluster centers from the same set of features (colour, pixel intensity, spatial, and texture) as they extracted before (Raja et al., 2009). The whole slide image was divided into equal 5x5 pixel blocks and passed into K-means for clustering. As a clustering result, all pixels inside the block were assigned one of the three classes (background, tissue body or detritus). Than resulting mask has been smoothed with a median filter. Results on 2 independent sets of 300 images showed a general pixel segmentation accuracy of 95.5%. This method doesn't need a ground truth for training. Moreover, it has show better segmentation results on the same data compared to previously proposed supervised learning approach. But the quality of extracted features depends on data. Complex morphological structures of some tissues may result in low-quality features which in its own turn may lead to poor segmentation. Also, both supervised and unsupervised solutions algorithms may be hard to reproduce due to closed source implementation.

Bug, Feuerhake, and Merhof, 2015, proposed Foreground Extraction algorithm (FESI) for histological whole-slide images segmentation. In this paper, the foreground and background segmentation were done by a combination of basic methods like median filtering, thresholding, erosion, and dilation. The proposed algorithm was compared to the more established methods like Watershed, which is commonly applied in cell-segmentation (Veta et al., 2011) and GrabCut (Carsten Rother, 2004), applied for image regions segmentation. FESI performed better than these technically more complex methods. It was used intersection over union area ratio (IoU, Jaccard index) which measures the overlap of the calculated mask to the ground-truth to evaluate the quality of segmentation of 43 test images. FESI achieved better results with 95% IoU compared to 82.3% by Watershed and 85.9% by GrubCut.

In the last few years, deep learning has produced state-of-the-art results in the domain of natural objects detection and segmentation (Yann LeCun, 2015; Garcia-Garcia et al., 2017; Zhao et al., 2018). New complex but efficient frameworks were developed like Faster R-CNN (Ren et al., 2015) for object detection and it's extension Mask R-CNN (He et al., 2017) for instance segmentation. Therefore, researches started to apply Deep learning for medical imaging to understand its capabilities in medicine (Litjens et al., 2016; Jones et al., 2017; Rodenburg, 2016). For example, Faster R-CNN was used for detection of glomeruli - network of capillaries located in the kidney (Kawazoe et al., 2018) or U-Net (Ronneberger, Fischer, and Brox, 2015), specifically developed model for biomedical image segmentation, was used for automatic brain tumor detection (Dong et al., 2017).

Bándi et al., 2017, compared state-of-the-art deep learning models for segmentation, FCNN (Long, Shelhamer, and Darrell, 2014) and U-Net (Ronneberger, Fischer, and Brox, 2015) to FESI algorithm on tissue segmentation problem on 54 slides from three different laboratories. Deep learning approaches performed better on different tissue types and achieved a Jaccard index of 93.7% and 92.9% by FCNN and U-Net respectively compared to 87% achieved by FESI. The study showed that deep learning methods can outperform existing traditional image analysis algorithms.

3.2 Tissue finding algorithm utilized by Akoya Biosciences

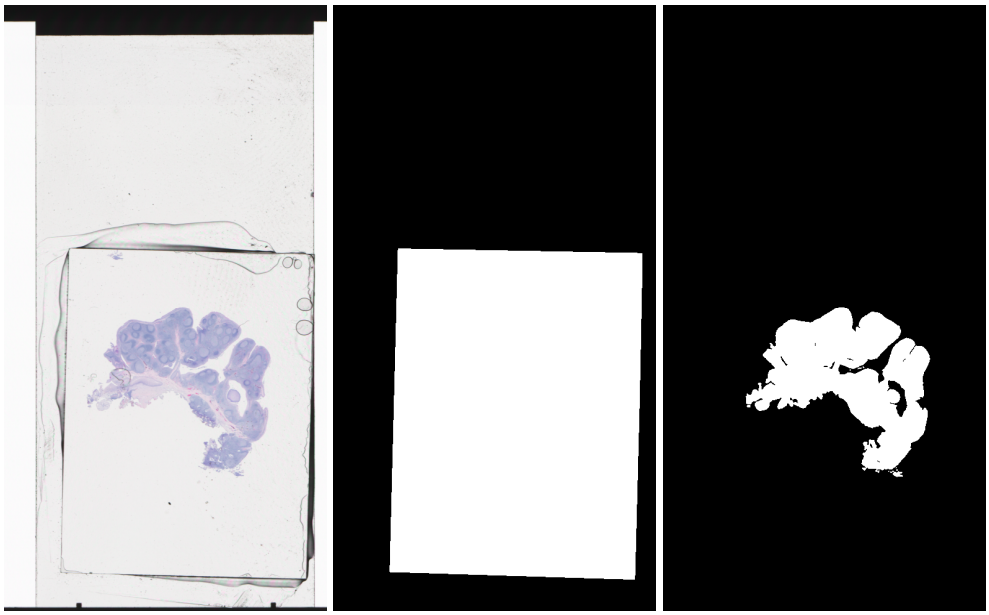
Data that has been used in the thesis was provided by Akoya Biosciences, Inc. Currently, to segment tissues on the slides, the combination of traditional methods are used. Detection procedure calls three different tissue detection algorithms. All three searches for regions which are brighter than their surroundings. The morphological transform opening, closing, and watershed are used with an aim to discard artifacts and form of the detected tissue region(s). These algorithms are not perfect and of the generated segmentations needed to be manually corrected by a human expert.

As a conclusion of the review, we could observe that the deep learning approaches are showing promising results in solving weaknesses of traditional algorithms in tissue segmentation task. It was decided to apply state-of-the-art deep learning models to our data and compare the results with performance of existing algorithm[s].

Chapter 4

Dataset description

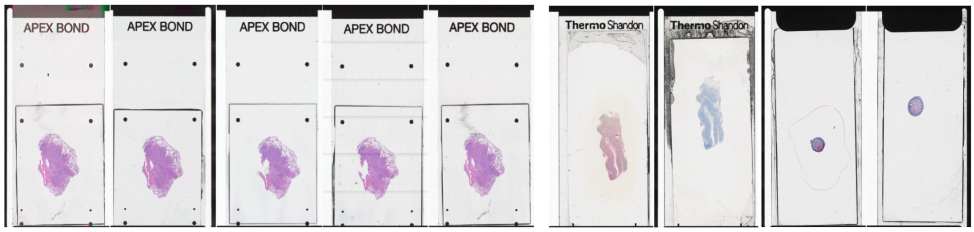
Dataset includes images of slides of formalin-fixed paraffin-embedded human and animal tissue sections with chromogenic stains applied. The images were acquired on Vectra Polaris Automated Quantitative Pathology Imaging Systems (Akoya Biosciences, Hopkinton, MA, USA). They are low-resolution brightfield scans (13.8 microns per pixel resolution) of the entire slide. Each scan is one of the following three sizes 521 x 1006, 2084 x by 4024, or 522 x 1009 pixels and represents slide marker and the coverslip, on which, one or multiple tissues are located. For each scan, two segmentation masks are provided: for coverslip area and for the tissue (Figure 4.1).



(A) Whole-slide image (B) Coverslip mask (C) Tissue mask

FIGURE 4.1: Each sample in the data is represented by the whole-slide image and two segmentation masks of the coverslip and tissue areas for this image

The whole data set consists of 81 images and represents variety of tissue types (Figure 4.3). Images have been divided into train, validation and test sets with 48, 12 and 21 images respectively. Images with almost identical tissue slices (Figure 4.2) were kept in the same data partition in order to avoid overfitting during model training.



(A) Similar images in training set (B) Similar images in validation set

FIGURE 4.2: An example of similar images in data set

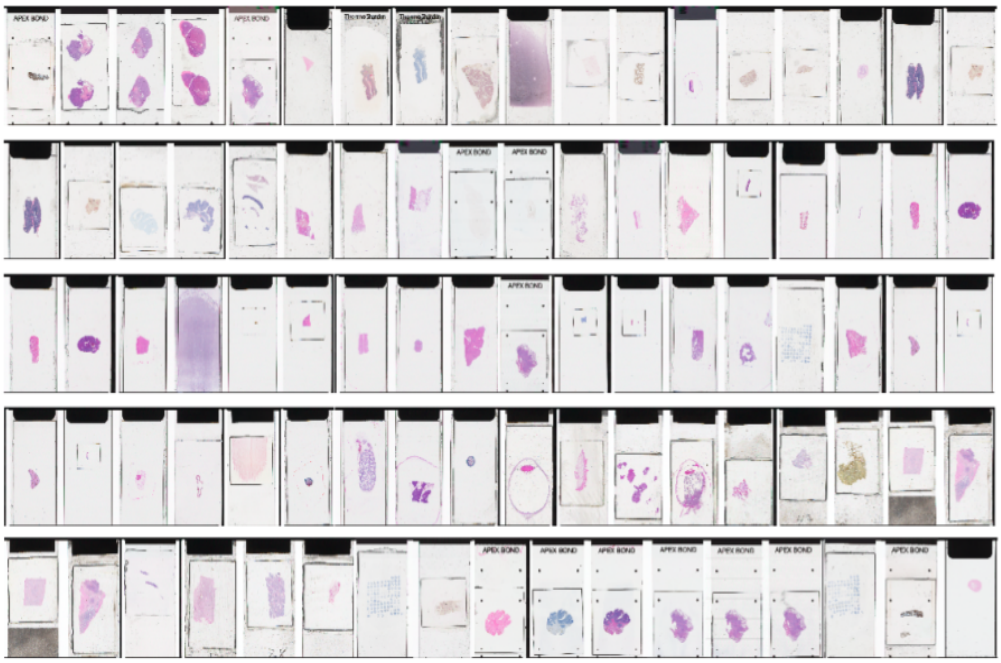


FIGURE 4.3: Presentation of a variety of tissue types in data set

Chapter 5

Methodology

5.1 Akoya Biosciences algorithm

The existing algorithm, used by Akoya Biosciences consists of two steps:

1. Extraction of the region of interest on whole-slide image defined by coverslip
2. Segmentation of the tissue inside the extracted region

Detailed description of two steps is presented in the next subsections. In this thesis, the algorithm has been used as a baseline and has been compared to the proposed method.

5.2 Rectangular coverslip extraction

Coverslip is characterized by narrow dark straight lines on the image. By these lines, two areas are defined: coverslip and bottom plate. The coverslip extraction procedure is divided into three steps. At the first step, dark lines are detected using convolution with line-kernels. At the second step based on the detected lines, a rectangular coverslip region is formed. At the final step dark areas covered by paper or holders are excluded from detected coverslip region (Figure 5.1).

5.3 Tissue segmentation on ROI defined by coverslip

Combination of three algorithms is used for segmentation. All three search for regions which are brighter than their surroundings. The morphological transform: opening, closing, and watershed are used with an aim to discard artifacts and form of the detected tissue region(s). Next step is the finding of overlap between three initially predicted regions. It has a concept of 'confident' tissue areas and 'less confident' tissue areas. In the mask, confident areas have a value of 3 (prediction overlap among all three algorithms). The less confident areas are 1 and 2. In practice, researches scan all the areas regardless of confidence (any value > 0) but only attempt to acquire focus points on the confident areas (value = 3).

5.4 Proposed pipeline overview

At first, tissue segmentation in the coverslip area has been performed. This approach led to two main disadvantages. First is the high level of false positive segmentations, since there is a lot of dirt/detritus presented on the slide. Second is the missing of some tissue parts. A model trained on the coverslip area learns not to segment the debris. As a result, tissue false negative segmentation occurs in cases where tissue

morphology is very similar to the debris.

Mostly, detritus (debris) particles are concentrated on the edges of the coverslip. Thus, to avoid above-mentioned issues, the area of segmentation should be decreased. We decided to perform it in a way of firstly defining the bounding box of the tissue and then segmenting the tissue on this smaller gained area. To accomplish this task two models have been trained: object detector for predicting bounding box area of the tissue and segmentor model for segmenting it. Also, the segmentation model has been trained only on bounding box areas, which made the model more sensitive to the tissue body and result in more accurate segmentation of tissue with similar morphology to detritus.

By using the combination of the object detection and segmentation models we present a novel approach for tissue segmentation on whole-slide images. The final proposed deep learning tissue segmentation pipeline includes the following steps:

1. Rectangular coverslip area extraction (using rectangular coverslip extraction part of the Akoya Biosciences algorithm 5.2)
2. Detection of the bounding box around the tissue
3. Generating patches for the segmentation step
4. Predicting probability maps from tissue patches
5. Merging and thresholding the probability maps

Each step is described in details below.

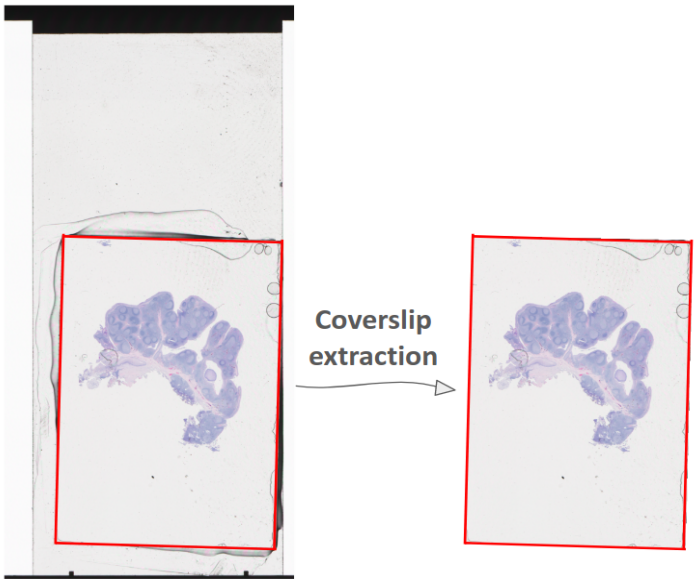


FIGURE 5.1: An example of coverslip area extraction from the whole-slide image

5.5 Tissue bounding box detection

5.5.1 Faster R-CNN

State-of-the-art object detection framework - Faster R-CNN has been used in this work for detecting tissue. This framework was introduced as an improvement of Fast R-CNN (Girshick, 2015) and consists of the following modules: feature extraction from the image, identification of candidate regions that are likely to contain objects of interest, with the subsequent classification of these objects. Figure 5.2 presents an overview of the Faster R-CNN architecture. The whole framework includes such main modules:

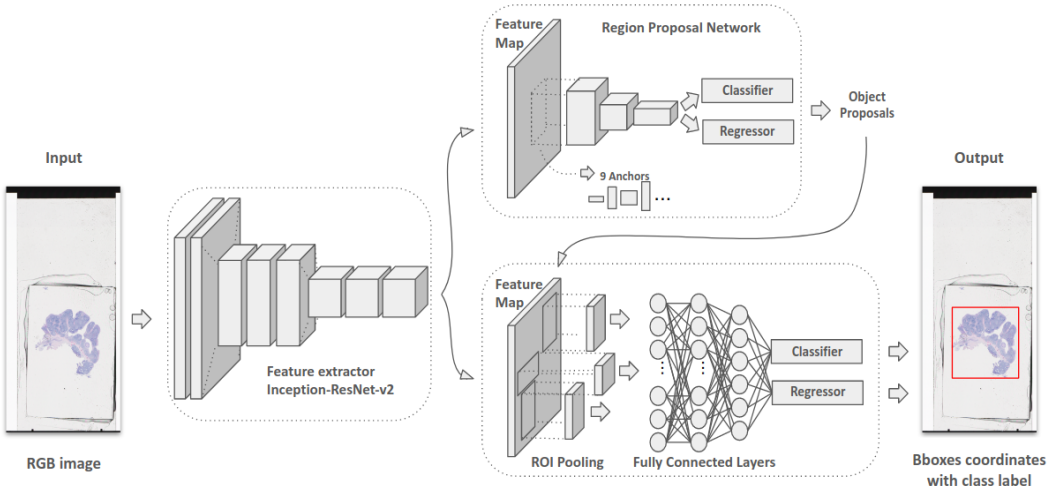


FIGURE 5.2: An overview of the Faster R-CNN architecture

- **Backbone neural network** (Inception ResNet v2 network (Szegedy et al., 2016) pretrained on COCO dataset (Lin et al., 2014)).
  The first module is the neural network which plays a role as a feature extractor from the image. In this thesis, Inception-ResNet v2 architecture backbone is used. The network generates a feature map, which is the input for the next module - Region Proposal Network.

- **Region Proposal Network (RPN)**
  RPN is a convolutional neural network which is running over a feature map with a sliding window. For each step/position, it analyzes 9 rectangular regions of different sizes and scales, called anchors. As a result, the network defines the possibility for each anchor being background or foreground and refines the shape and position of the anchor to fit the objects perfectly. Produced anchors are called Regions of Interest (ROI).

- **ROI pooling**
  As an output of RPN, regions of interest with different sizes are gained. But for the last step of classification and regression, constant size for each region is needed. ROI pooling solves this task by splitting the input feature map into a constant number of equal regions and then applying max polling on those regions.

- **ROI bounding box regressor and classifier**
  This is the las module consisted of several fully connected layers which gets as an input ROI and for each of them returns the object class and bounding box coordinates.

As a result, by getting an image as an input, Faster R-CNN returns an array of n objects. Each of these n objects consists of 5 elements: object class and 4 integer numbers representing coordinates of the bounding box around the object - minimum values of x and y, width, and height.

5.5.2 Data set preparation

Dataset for Faster R-CNN training has been created in a manner of producing one bounding box per image. It was done in order to encourage the model to predict one bounding box, which covers all tissues parts on the slide. For each image, we had only segmentation masks of coverslip and tissue. Thus, the bounding boxes coordinates were generated by getting the x-axis and y-axis minimum and maximum elements from the tissue ground truth segmentation mask. Examples of generated bounding box could be observed in Figure 5.3. For model training, bounding box coordinates has been normalized and transformed to special TFRecord format. The format stores all useful information for training - coordinates of bbox, image width, height and format, filename, object class, and other metadata. TFRecord format is required for the training Faster R-CNN model implemented by Tensorflow library (Abadi et al., 2015), used in this thesis.

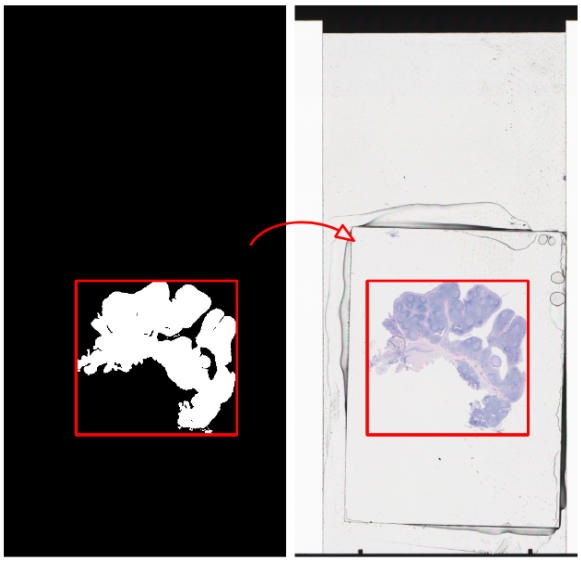


FIGURE 5.3: An example of generated bounding box around tissue mask

5.5.3 Training and evaluation

It has been used the implementation of Faster R-CNN framework based on Tensorflow library (Abadi et al., 2015) and created by Tensorflow team (Huang et al., 2016). System training has been performed on Tesla P100 hosted by High Performance Computing at the Institute of Computer Science, University of Tartu (Estonia). The

system has been trained on coverslip area images (5.5) for detecting bounding boxes for objects of one class - tissue. An evaluation has been performed on 21 test images by calculating accuracy, recall, precision, f1 score, and IoU metrics. In most cases predicted bounding bboxes miss small parts of tissue on the edges. Since, each tissue pixel is very important for this task and to avoid missing small parts of the tissue, predicted bounding boxes are increased by 25 pixels from each side.

5.6 Tissue segmentation

Segmentation procedure of proposed pipeline consists of three steps:

- Cropping of detected bounding box by Faster R-CNN into patches
- Predicting probability maps for each patch
- Merging and thresholding probability maps

5.6.1 U-Net

U-Net is a fully convolutional neural network Long, Shelhamer, and Darrell, 2014) specifically designed by Ronneberger, Fischer, and Brox, 2015 for biomedical image segmentation. The main concept lies under encoder-decoder architecture with skip connections. An encoder is used to capture the context on the image and it is basically a sequence of convolutional and max-pooling layers. The decoder is used for learning localization features and it is a combination of convolution and upsampling layers. Also, important role play skip connections between the downsampling and upsampling layers. It allows transferring information among contracting and expanding paths resulting in better information reconstruction. Figure 5.4 presents an overview of the U-Net architecture.

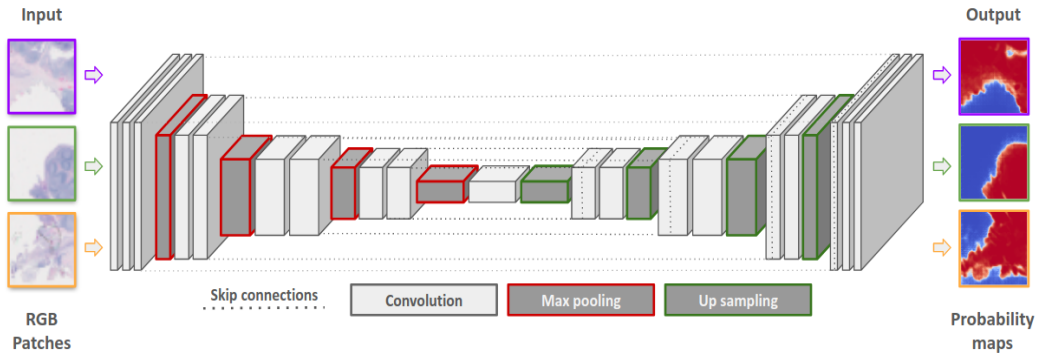


FIGURE 5.4: An overview of the U-Net architecture

5.6.2 Data set preparation

Segmentation area
For segmentation procedure two types of datasets has been tested (Figure 5.5):

1. **Coverslip data** - data set consisted of extracted coverslips from each image
2. **Bbox data** - data set consisted of extracted bounding boxes around the tissue from each image

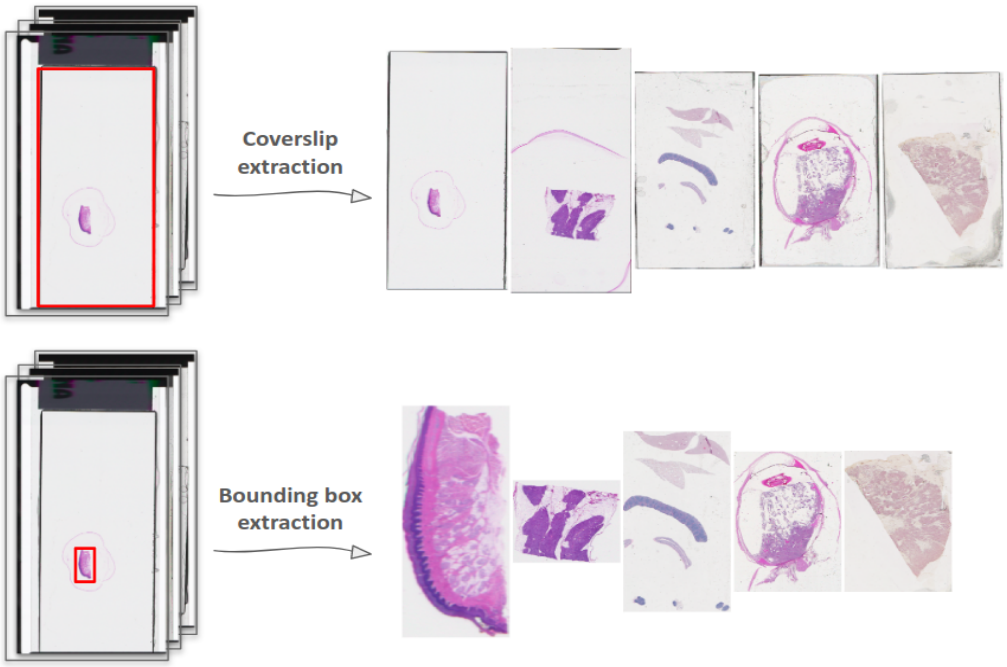


FIGURE 5.5: An example of two data sets consisted of coverslips area and tissue bounding boxes area.

Patch extraction
By reason of having only 81 high-resolution images, training data has been prepared by extracting patches from the images. Three methods for patch extraction have been tested in this work (Figure 5.6):

- **Random extraction** - extracting predefined number of patches from random locations on the image
- **Grid-like extraction** - gradually extracting patches from image in a grid manner
- **Grid-like extraction with overlap** - gradually extracting patches from image in a grid manner, but with overlap (same pixels appear on multiple patches)

Patch merging
After image is cropped into patches each patch goes as a separate input for U-Net (Figure 5.4). As a result of U-Net performance, we get the probability map for each patch. To get the probability map for whole tissue the small patch-like probability maps should be merged. This procedure is also done in three different ways depending on extraction way:

- For random cropped patches, coordinates for each patch are saved. After predicting, the probability maps are place on corresponding positions. For overlapped patches, average value for each pixel is calculated.
- For grid-like crop, patches are simply stacked in the same sequence they were cropped
- For grid-like crop with overlap, patches are stacked in the same sequence they were cropped. For overlapped patches, average value for each pixel is calculated.

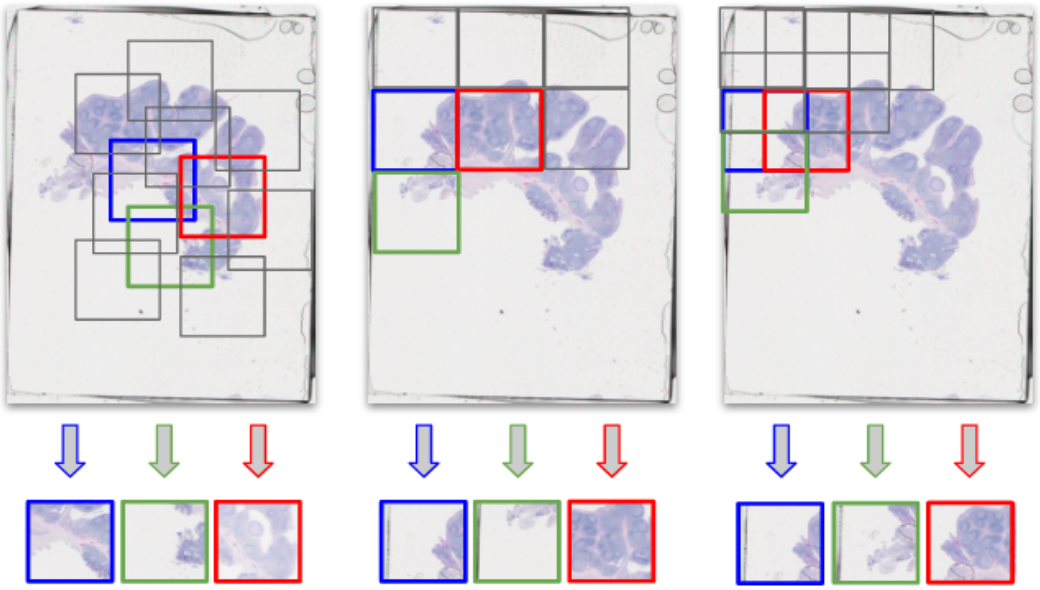
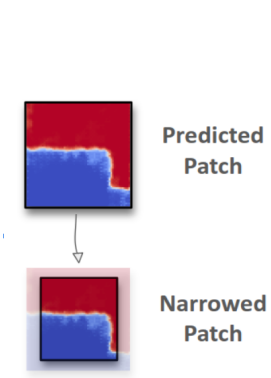


FIGURE 5.6: Patch extraction methods overview. From left to right: random extraction, grid-like extraction and grid-like extraction with overlap

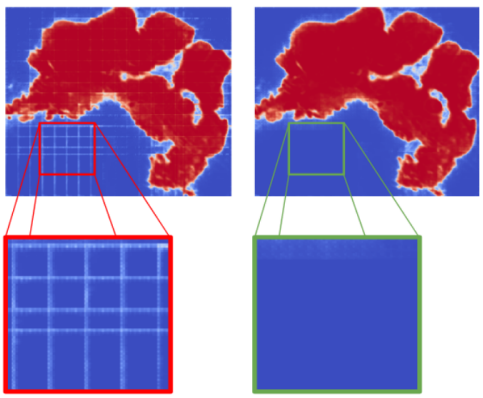
All probability patches are merged into one probability map for tissue bounding box/coverslip area. This probability map is then thresholded by 0.5 value and placed on the right position on the final mask for whole-slide image.

Patch narrowing

As a result of patch merging procedure, the so-called patch border effect appears. It is observed as small low probability lines around each patch on the final merged probability map. In some cases, it negatively affects the final mask, by reducing the probability of pixels where tissue is located. To decrease the influence of patch border effect, ignoring perimeter of some size around the patch when merging could be done. Figure 5.6.2 presents the example of patch narrowing method performance. This technique could only be applied to the merging procedure with overlapped patches.



(A) An example of narrowed patch



(B) An example of patch narrowing method performance: left - before, right - after

5.6.3 Training and evaluation

U-Net model was implemented using Python library Keras (Chollet, 2015). For training and validation, each image was cropped into 2000 patches via random extraction method. The following hyperparameters for training have been configured:

- Batch size - 512
- Maximum number of epochs - 500
- Optimizer Adam (Kingma and Ba, 2014) with starting learning rate 0.001 and binary cross-entropy loss (Janocha and Czarnecki, 2017)
- Callbacks:
  - Earlystopping - stops training if validation loss hasn't been improving for 40 epochs
  - ReduceLROnPlateau - reduce the learning rate for 0.5 factor if validation loss hasn't been changed for 10 epochs. The limit of minimum learning rate has been configured for 0.000001 value

An evaluation has been performed on 21 test images by calculating accuracy, recall, precision, f1 score, and IoU metrics

Chapter 6

Experimental results

6.1 Intermediate experiments

This section describes intermediate experiments and results of segmentation procedure considering patch sizes, segmentation areas, and patch extraction methods. All experiments evaluation has been performed on 21 images by calculating accuracy, recall, precision, f1 score and IoU metrics.

Patch size
At first, experiments for defining the best patch size has been executed. The U-Net models have been trained on three different sizes of patches (32x23, 64x64, 128x128 pixels). The best results have been achieved with patches of size 64 by 64 pixels.

Segmentation area
Two U-Net models with similar architectures have been trained on different datasets: coverslip data and bbox data (Figure 5.5). A model trained on the coverslip area performed worse. The first disadvantage appears in the prediction of a big amount of false positive segmentations since there is a lot of dirt/detritus presented on the slide. The second disadvantage relates to inaccurate segmentation of tissue borders.

Data consisted of bounding boxes area automatically avoids most of the detritus parts, since these particles are concentrated on the edges of the coverslip and the tissue is placed in the center. It leads to a lower amount of patches with detritus and bigger amount of patches with tissue body in data set, and encourage the model to better predict boundaries of tissue body. Thus the model trained on bounding boxes showed better segmentation results.

Test data for both models has been cropped into 64 by 64 pixels patches with "grid with overlap" extraction method. Predicted probability patches have been merged in the same way and then coverslips and bounding boxes areas have been placed to the right position on the final mask. Comparison of both models performance is described in Table 6.1, and some visual results are presented in Figure 6.1. (This experiment has been performed with assumption of knowing positions of bounding box areas)

Data set type	Accuracy	Recall	Precision	F1 score	IoU
Coverslip area	0.9877	0.8549	0.9029	0.8594	0.7776
Tissue bbox area	**0.9952**	**0.9667**	**0.9345**	**0.9488**	**0.9050**

TABLE 6.1: Experiments of segmentation coverslip and tissue bounding boxes areas. The best results are highlighted

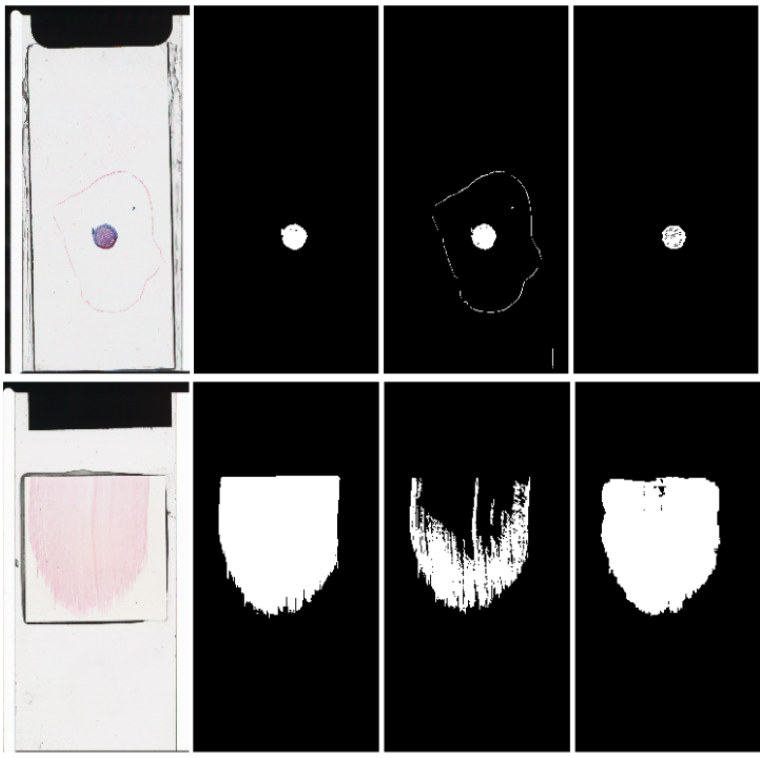


FIGURE 6.1: Visual examples of tissue segmentation. From left to right: whole-slide image, ground truth tissue mask, mask predicted by a model trained on coverslip area, mask predicted by a model trained on tissue bounding box area

Patch extraction

Five different techniques of patch extraction has been tested. Random extraction, grid-like extraction and grid-like extraction with three diferent overlap sizes - 8, 16 and 32 pixels. The best result has been acieved with grid-like extraction method with 16 pixels overlap for recall value and 32 pixels overlap for others metrcis. Ealuation metrics for this experiments is described in Table 6.2 and visual example is presented on the Figure 6.2.

Patch narrowing

For patch narrowing experiment, patches were cropped in size for 8 pixels from each side while merging. Narrowing result has been compared to the regular grid with overlap merging method. Small improvements could be observed in Table 6.3.

Patch extraction way	Accuracy	Recall	Precision	F1 score	IoU
Random	0.9071	0.8257	**0.9884**	0.8927	0.8166
Grid-like	0.9443	0.9595	0.9370	0.9466	0.9012
Grid-like with overlap (8px)	0.9424	0.9652	0.9296	0.9454	0.8990
Grid-like with overlap (16px)	0.9460	**0.9667**	0.9345	0.9488	0.9050
Grid-like with overlap (32px)	**0.9473**	0.9643	0.9389	**0.9496**	**0.9062**

TABLE 6.2: Patch extraction methods experiments. The best results are highlighted

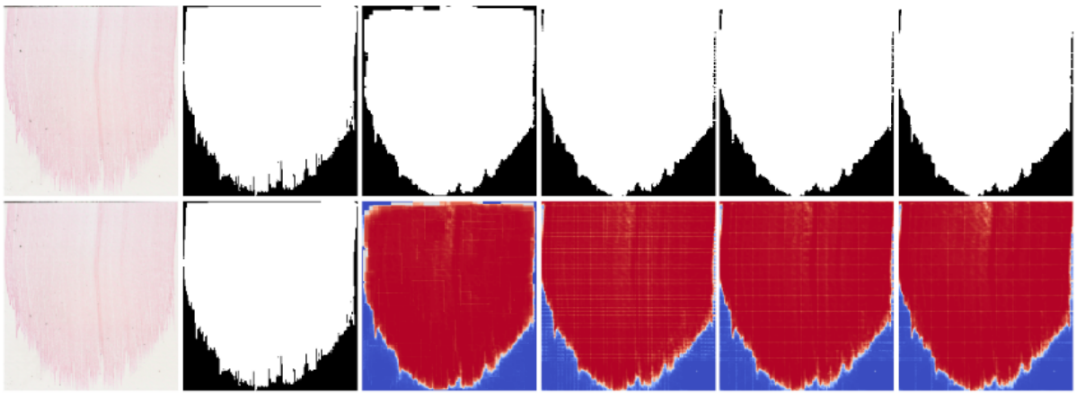


FIGURE 6.2: Visual example of model performance with different patch extraction techniques on the bounding box area. From left to right: tissue on the slide, ground truth mask, model performance with random extraction, grid-like extraction and grid-like extraction with 8, 16 and 32 pixels overlap. First row - binary masks, second - probability maps

Patch extraction and merging way	Accuracy	Recall	Precision	F1 score	IoU
Grid-like with overlap (16 px)	0.946	**0.9667**	0.9345	0.9488	0.9050
Grid-like with overlap (16 px) + Narrowing (8 px)	0.946	0.9666	**0.9351**	**0.949**	**0.9054**

TABLE 6.3: Comparsion of model performance with and without patch narrowing method. The best results are highlighted

6.2 Proposed pipeline combinations

Based on all separate experiment for each procedure (data set creation, tissue bounding box detection and segmentation) the following two pipelines variations has been created:

- Pipeline 1: Coverslip extraction from WSI + extracted coverslip segmentation
- Pipeline 2: Coverslip extraction from WSI+ tissue bounding box detection on extracted coverslip + detected tissue bounding box segmentation (If the Faster R-CNN does not predict any tissue bounding box, then the whole coverslip area is segmented)

Two proposed pipelines have been compared on 21 test images. Pipeline with tissue bounding box detection module showed more precise results. Table 6.4 represents evaluation results for both methods. All steps of the more accurate pipeline with object detection part are presented in Figure 6.3.

Pipelines	Accuracy	Recall	Precision	F1 score	IoU
Pipeline 1	**0.986**	0.8436	**0.92**	0.86	0.78
Pipeline 2	0.985	**0.949**	0.916	**0.929**	**0.874**

TABLE 6.4: Two variation of proposed pipeline performance. The best results are highlighted

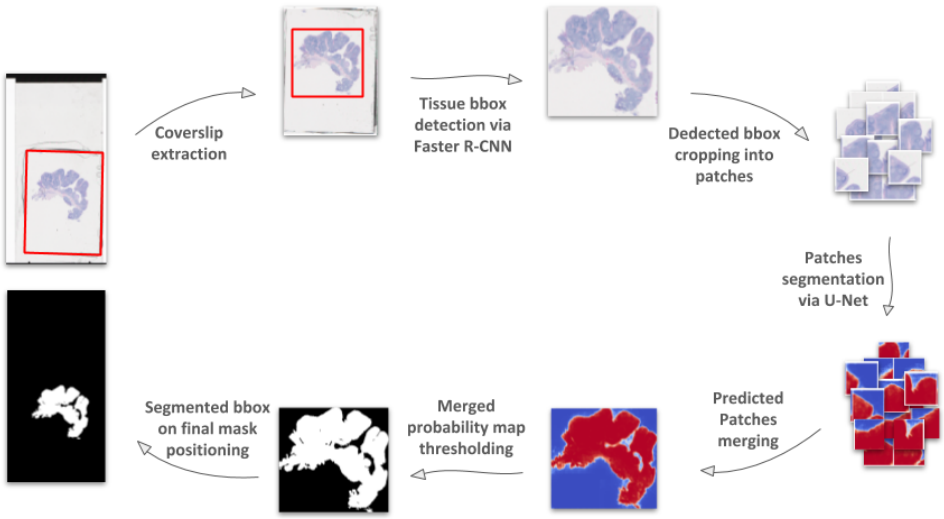


FIGURE 6.3: An overview of final proposed pipeline

6.3 Comparison with Akoya Biosciences approach

The final proposed pipeline has been compared to two variations of the current advanced algorithm (variation with overlap value = 3 and overlap value > 0 (1, 2, 3)). Table 6.5 represents the evaluation results for three methods on test images. Relying on Accuracy, F1 score and IoU metrics we beat current Akoya Biosciences algorithm on 21 test images (Figure 6.4).

Approaches	Accuracy	Recall	Precision	F1 score	IoU
Proposed pipeline	**0.985**	0.949	0.916	**0.929**	**0.874**
Current algorithm (overlap value > 0)	0.958	**0.961**	0.806	0.864	0.785
Current algorithm (overlap value = 3)	0.972	0.887	**0.927**	0.897	0.832

TABLE 6.5: Performance comparison on test set of proposed pipeline and Akoya Biosciences algorithm. The best results are highlighted

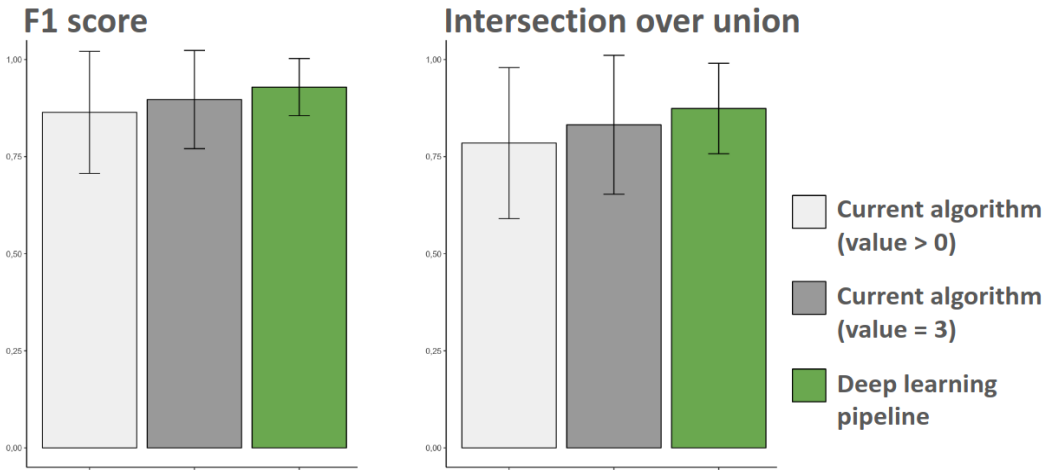
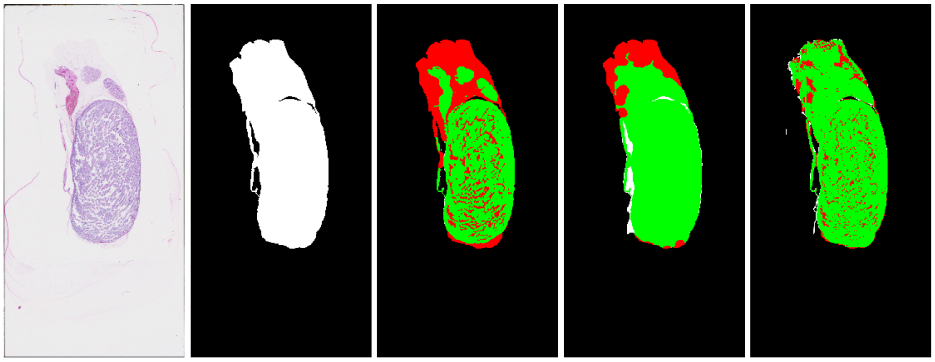
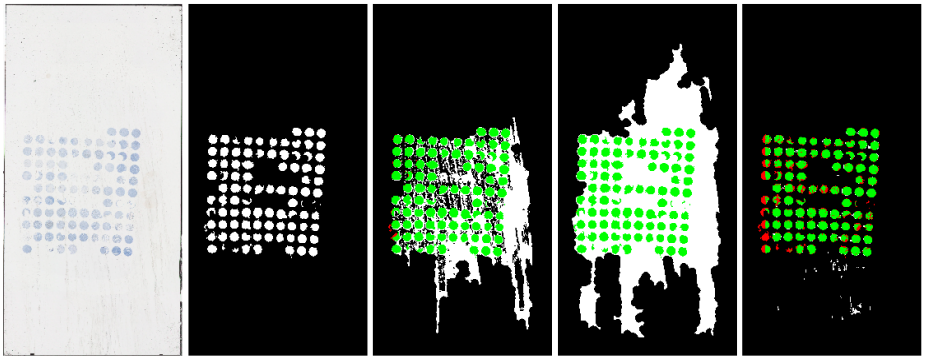


FIGURE 6.4: Bar charts with error bars based on standart deviation. F1 score and intersection over union comparison among two variation of current Akoya Biosciences algorithm and proposed deep learning pipeline

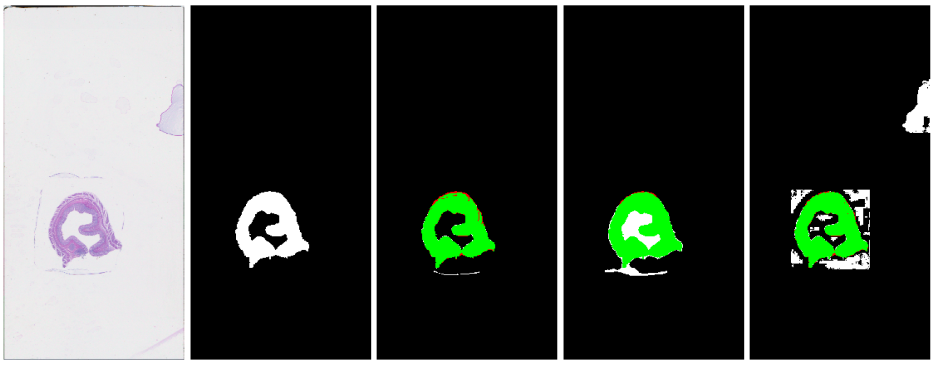
Figure 6.5 presents the visual results. Subfigures a) and b) present the cases where proposed pipeline outperforms existing algorithm. Subfigure c) presents the case, where object detector has mistakenly predicted two bounding boxes: true tissue and detritus spot. The slide, where Faster R-CNN hasn't detected any bounding box and the whole coverslip has been segmented, could be observed on subfigure d). The visual predictions of the whole test set could be found in the appendix A.



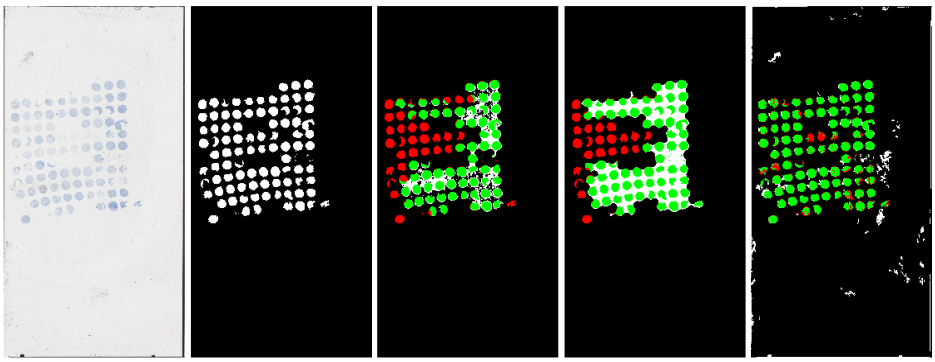
(A) Example 1



(B) Example 2



(C) Example 3



(D) Example 4

FIGURE 6.5: Visual examples of tissue segmentation. From left to right: WSI, ground truth tissue mask, mask generated by current Akoya Biosciences algorithm with overlap value = 3 , overlap value > 0, and mask predicted by proposed deep learning pipeline. In predicted masks visualization green color states for true positive, red for false negative and white for false positive predictions

Chapter 7

Conclusions

7.1 Contribution

We described the end-to-end pipeline for tissue segmentation from whole-slide images based on the combination of object detection and segmentation deep learning models. Proposed pipeline reaches the performance of Akoya Biosciences approach and seems promising due to its independence (almost complete absence of any manual tuning) and capacity to improve with more data.

7.2 Future work

The following improvements could be performed:

1. Adding more data for training the models, which might result in better performance. It would be very useful to include more complicated examples with multiple piece of tissue on a slide.

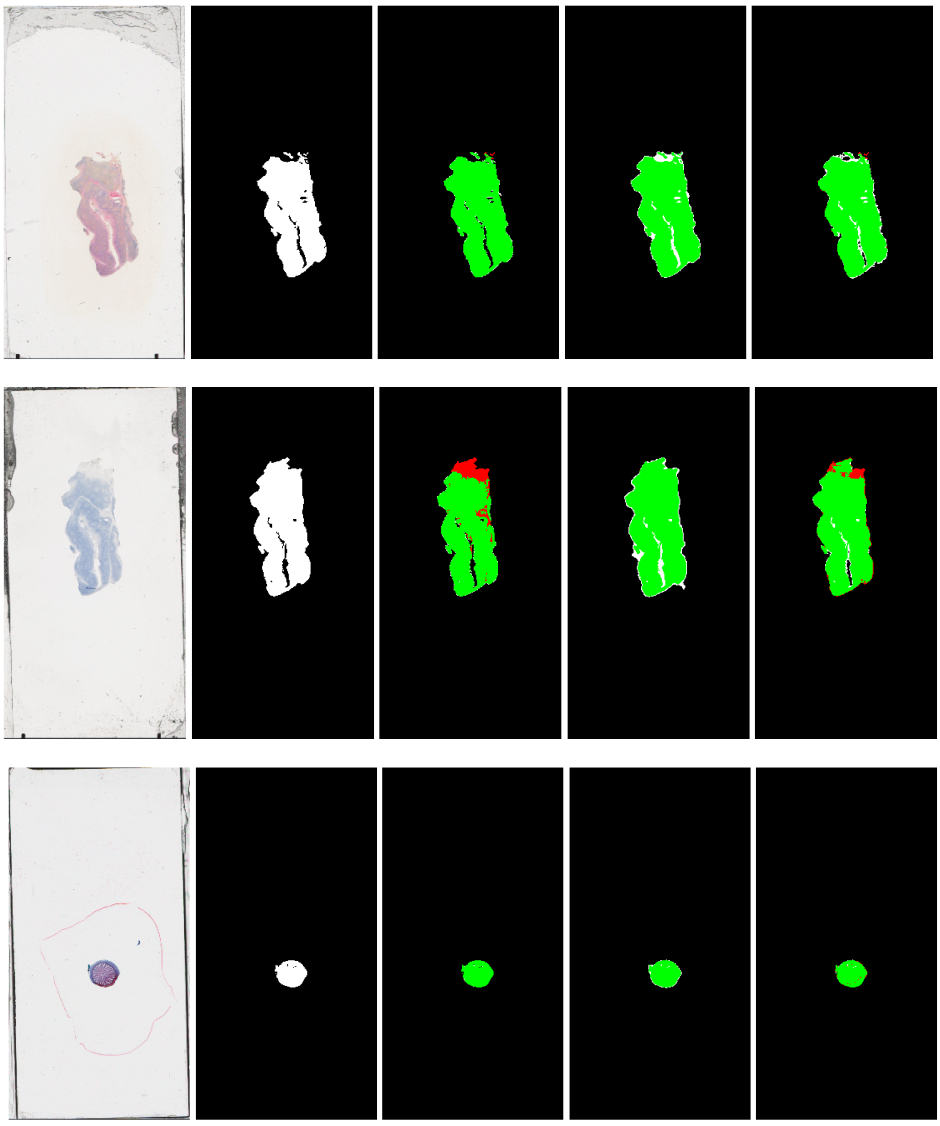
2. Searching for more appropriate model architectures. Faster R-CNN has been trained from scratch with default hyperparameters, thus tuning the model could positively affect the results. Other state-of-the-art models could be trained such as SegNet (Badrinarayanan, Kendall, and Cipolla, 2015) and Mask R-CNN (He et al., 2017)

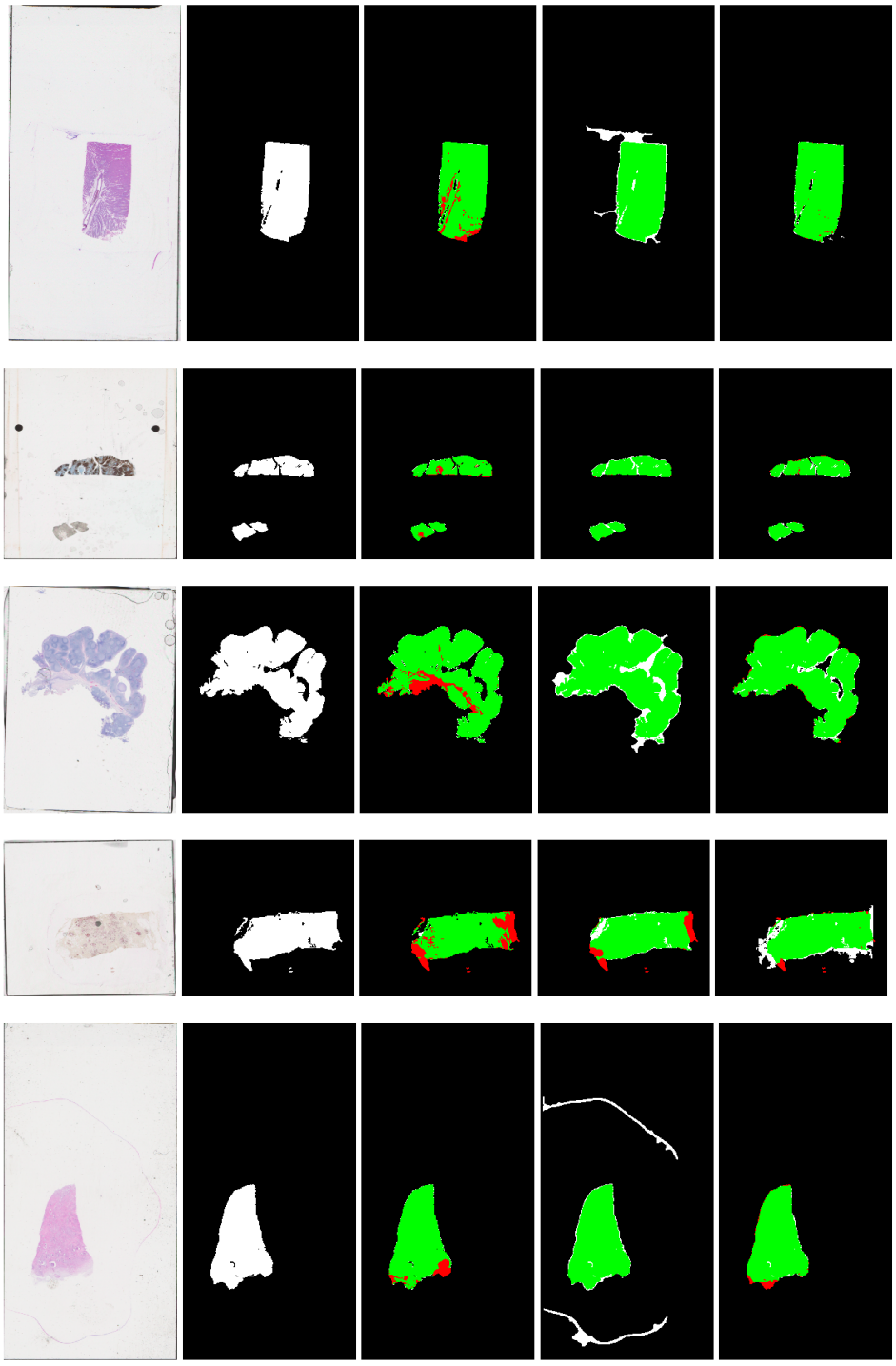
3. One more way for improvements could be done in capturing more accurately the tissue boundaries by using recently published Enhancing Segmentation Precision with Semantic Edge Aware Loss for segmentation model (Chen, Dapogny, and Cord, 2019).

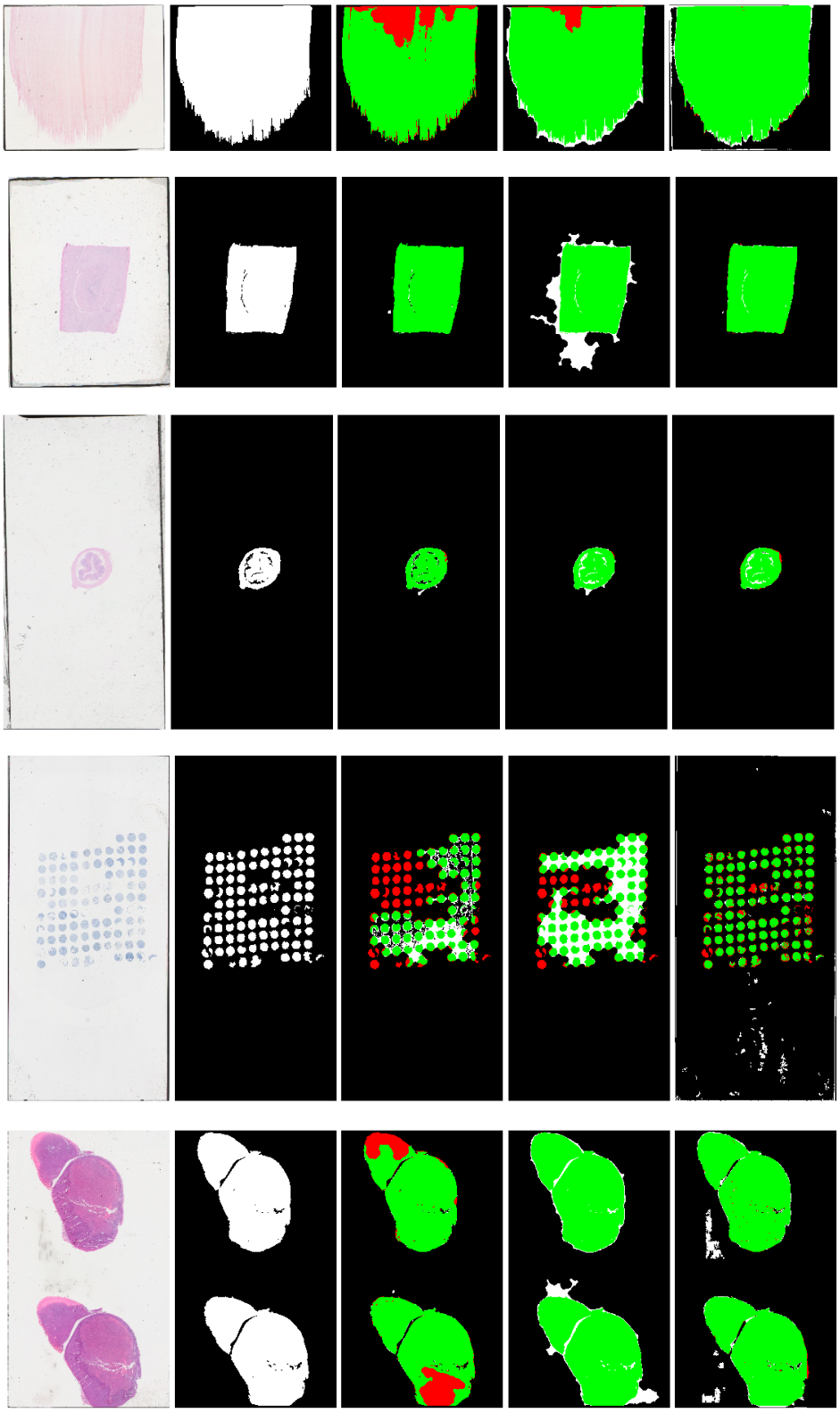
Appendix A

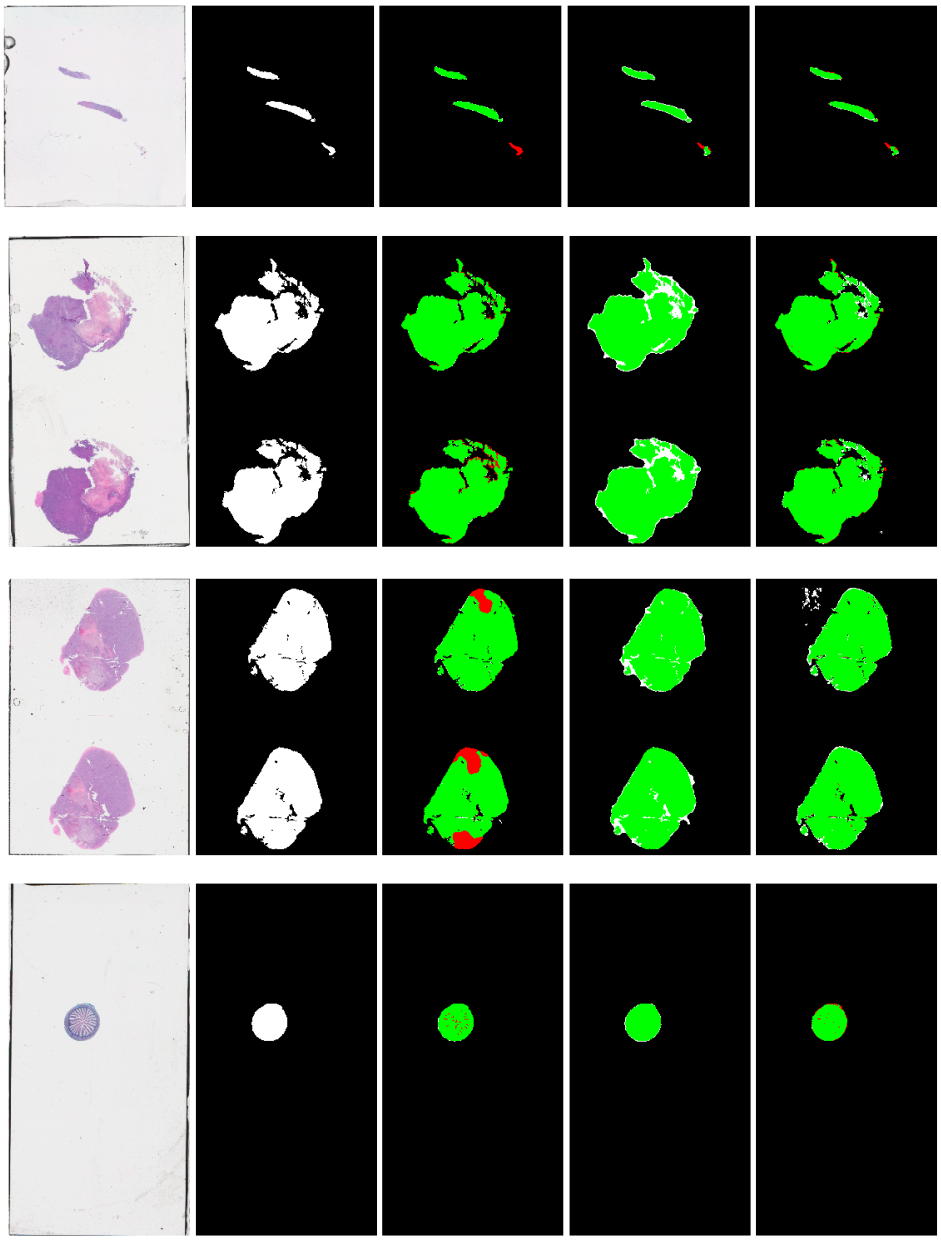
Visual results on whole test set

Visual examples of tissue segmentation in all test set images. From left to right: whole-slide image, ground truth tissue mask, mask generated by current state-of-the-art algorithm with overlap value = 3 , overlap value > 0, and mask predicted by proposed deep learning pipeline. In predicted masks visualization green color states for true positive, red for false negative and white for false positive predictions









Bibliography

Abadi, Martín et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. URL: http://tensorflow.org/.

Ajeet Ram Pathak Manjusha Pandey, Siddharth Rautaray (2018). "Application of Deep Learning for Object Detection". In: DOI: 10.1016/j.procs.2018.05.144. URL: http://dx.doi.org/10.1016/j.procs.2018.05.144.

Akkus Z Galimzianova A, Hoogi A Rubin DL Erickson BJ (2017). "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions". In: *Journal of digital imaging*. DOI: 10.1007/s10278-017-9983-4. URL: http://dx.doi.org/10.1007/s10278-017-9983-4.

Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla (2015). "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation". In: arXiv: 1511.00561 [cs.CV].

Banu, Sameena (2012). "The Comparative Study on Color Image Segmentation Algorithms". In: *International Journal of Engineering Research and Applications*.

Bug, Daniel, Friedrich Feuerhake, and Dorit Merhof (2015). "Foreground Extraction for Histopathological Whole Slide Imaging". In: pp. 419–424. DOI: 10.1007/978-3-662-46224-9_72. URL: http://dx.doi.org/10.1007/978-3-662-46224-9_72.

Bándi, Péter et al. (2017). "Comparison of Different Methods for Tissue Segmentation in Histopathological Whole-Slide Images". In:

Carsten Rother Vladimir Kolmogorov, Andrew Blake (2004). ""GrabCut": interactive foreground extraction using iterated graph cuts". In:

Chen, Yifu, Arnaud Dapogny, and Matthieu Cord (2019). "SEMEDA: Enhancing Segmentation Precision with Semantic Edge Aware Loss". In: *arXiv e-prints*. arXiv: 1905.01892 [cs.CV].

Chijindu, Engr. V. C. (2012). "Medical Image Segmentation Methodologies – A Classified Overview ". In: *African Journal of Computing and ICT*.

Chollet, François et al. (2015). *Keras*. https://github.com/fchollet/keras.

Doi, Kunio (2007). "Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential". In:

Dong, Hao et al. (2017). "Automatic Brain Tumor Detection and Segmentation Using U-Net Based Fully Convolutional Networks". In: pp. 506–517. DOI: 10.1007/978-3-319-60964-5_44. URL: http://dx.doi.org/978-3-319-60963-8.

Erhan, Christian Szegedy Alexander Toshev Dumitru (2013). "Deep Neural Networks for Object Detection". In: *NIPS*.

Farahani N Parwani A, Pantanowitz L (2015). "Whole slide imaging in pathology: advantages, limitations, and emerging perspectives". In: DOI: 10.2147/PLMI.S59826. URL: http://dx.doi.org/10.2147/PLMI.S59826.

G. Evelin Suji Y.V.S. Lakshmi, G. Wiselin Jiji (2013). "The Comparative Study on Image Segmentation Algorithms". In: *International Journal of Advanced Computer Research*.

Garcia-Garcia, Alberto et al. (2017). "A Review on Deep Learning Techniques Applied to Semantic Segmentation". In: arXiv: 1704.06857 [cs.CV].

Girshick, Ross (2015). "Fast R-CNN". In: arXiv: 1504.08083 [cs.CV].

Gurcan MN, Boucheron L Can A Madabhushi A Rajpoot N Yener B (2009). "Histopathological Image Analysis: A Review". In: DOI: 10.1109/RBME.2009.2034865. URL: http://dx.doi.org/10.1109/RBME.2009.2034865.

Hani A Alturkistani Faris M Tashkandi, and Zuhair M Mohammedsaleh (2015). "Histological Stains: A Literature Review and Case Study". In: DOI: 10.5539/gjhs.v8n3p72. URL: http://dx.doi.org/10.5539/gjhs.v8n3p72.

Hazem Hiary Raja S. Alomari, Vipin Chaudhary (2013). "Segmentation and localisation of whole slide images using unsupervised learning". In: *IET Image Processing*. DOI: 10.1049/iet-ipr.2013.0008. URL: http://dx.doi.org/10.1049/iet-ipr.2013.0008.

He, Kaiming et al. (2017). "Mask R-CNN". In: arXiv: 1703.06870 [cs.CV].

Huang, Jonathan et al. (2016). "Speed/accuracy trade-offs for modern convolutional object detectors". In: arXiv: 1611.10012 [cs.CV].

Image source. URL: https://killianlevacher.github.io/blog/posts/post-2016-03-01/img/layeredRepresentation.jpg.

Janocha, Katarzyna and Wojciech Marian Czarnecki (2017). "On Loss Functions for Deep Neural Networks in Classification". In: *arXiv e-prints*. arXiv: 1702.05659 [cs.LG].

Jolliffe IT, Cadima J (2016). "Principal component analisys: a review and recent developments". In: DOI: 10.1098/rsta.2015.0202. URL: http://dx.doi.org/10.1098/rsta.2015.0202..

Jones, William et al. (2017). "Computational biology: deep learning". In: *Emerging Topics in Life Sciences* 1, pp. 257–274. DOI: 10.1042/ETLS20160025. URL: http://dx.doi.org/10.1042/ETLS20160025.

Kawazoe, Yoshimasa et al. (2018). "Faster R-CNN-Based Glomerular Detection in Multistained Human Whole Slide Images". In: *Journal of Imaging* 4, p. 91. DOI: 10.3390/jimaging4070091. URL: http://dx.doi.org/10.3390/jimaging4070091.

Kingma, Diederik P. and Jimmy Ba (2014). "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints*. arXiv: 1412.6980 [cs.LG].

Kohavi, Ron (1995). ""A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection"". In:

Komura, Daisuke and Shumpei Ishikawa (2017). "Machine Learning Methods for Histopathological Image Analysis". In: *Computational and Structural Biotechnology Journal* 16. DOI: 10.1016/j.csbj.2018.01.001. URL: http://dx.doi.org/10.1016/j.csbj.2018.01.001.

Lin, Tsung-Yi et al. (2014). "Microsoft COCO: Common Objects in Context". In: arXiv: 1405.0312 [cs.CV].

Litjens, Geert et al. (2016). "Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis". In: *Scientific Reports* 6, p. 26286. DOI: 10.1038/srep26286. URL: http://dx.doi.org/10.1038/srep26286.

Liu, Wei et al. (2015). "SSD: Single Shot MultiBox Detector". In: arXiv: 1512.02325 [cs.CV].

Long, Jonathan, Evan Shelhamer, and Trevor Darrell (2014). "Fully Convolutional Networks for Semantic Segmentation". In: arXiv: 1411.4038 [cs.CV].

Mark Everingham S. M. Ali Eslami, Luc Van Gool Christopher K. I. Williams John Winn Andrew Zisserman (2014). "The PASCAL Visual Object Classes Challenge: A Retrospective". In:

Medeiros, Flavio H.D. Araujo Romuere R.V. Silva Daniela M. Ushizima Mariana T. Resende Claudia M. Carneiro Andrea G. Campos Bianchi Fatima N.S. (2019). "Deep learning for cell image segmentation and ranking". In: *Computerized Medical Imaging and Graphics*.

Pantanowitz, Liron (2010). "Digital images and the future of digital pathology". In: DOI: 10.4103/2153-3539.68332. URL: http://dx.doi.org/10.4103/2153-3539.68332.

Pham DL Xu C, Prince JL. (2000). "Current Methods in Medical Image Segmentation". In: DOI: 10.1146/annurev.bioeng.2.1.315. URL: http://dx.doi.org/10.1146/annurev.bioeng.2.1.315.

Raja, S et al. (2009). "Localization of tissues in high resolution digital anatomic pathology images". In: *Proceedings of SPIE - The International Society for Optical Engineering* 7260. DOI: 10.1117/12.811430. URL: http://dx.doi.org/10.1117/12.811430.

Redmon, Joseph et al. (2015). "You Only Look Once: Unified, Real-Time Object Detection". In: arXiv: 1506.02640 [cs.CV].

Ren, Shaoqing et al. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: arXiv: 1506.01497 [cs.CV].

Rodenburg, Bram (2016). "Deep Learning in Histopathology". In:

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox (2015). "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: arXiv: 1505.04597 [cs.CV].

Szegedy, Christian et al. (2016). "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning". In: arXiv: 1602.07261 [cs.CV].

Tsechpenakis, Gavriil (2011). "Deformable Model-Based Medical Image Segmentation". In:

Veta, Mitko et al. (2011). "Marker-controlled watershed segmentation of nuclei in HE stained breast cancer biopsy images". In: *Proceedings - International Symposium on Biomedical Imaging*, pp. 618 –621. DOI: 10.1109/ISBI.2011.5872483. URL: http://dx.doi.org/10.1109/ISBI.2011.5872483.

Yann LeCun Yoshua Bengio, Geoffrey Hinton (2015). "Deep learning". In: *Nature*. DOI: doi:10.1038/nature14539. URL: http://dx.doi.org/doi:10.1038/nature14539.

Zhao, Zhong-Qiu et al. (2018). "Object Detection with Deep Learning: A Review". In: arXiv: 1807.05511 [cs.CV].