

Automated Segmentation of Complex Biological Structures in Histopathology Images

by

Sina Salsabili

A thesis submitted to the Faculty of Graduate and Postdoctoral
Affairs in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Electrical and Computer Engineering

Carleton University
Ottawa, Ontario

© 2022, Sina Salsabili

Table of Contents

Chapter 1: Introduction.....	1
1.1 Motivation.....	1
1.2 Thesis aims and objectives	5
1.3 Thesis contribution	8
1.4 Thesis organization	11
Chapter 2: Literature review and background	13
2.1 Computer-aided detection/diagnosis.....	13
2.2 Histopathology imaging.....	16
2.3 Automated histopathology image analysis	20
2.4 Image segmentation in automated histopathology image analysis	26
2.5 Image segmentation methods in automated histopathology image analysis.....	29
2.6 Histopathology image segmentation applications.....	48
Chapter 3: Automated segmentation of histopathology images of human placenta	53
3.1 Introduction.....	53
3.2 Methodology.....	56
3.3 Results.....	69
3.4 Discussion.....	72
3.5 Conclusion	74
Chapter 4: Automated estimation of the mean linear intercept in histopathology images of mouse lung tissue	76
4.1 Introduction.....	76
4.2 Materials and methods.....	78
4.3 Results.....	99
4.4 Discussion.....	103

4.5	Conclusion	110
Chapter 5:	Deep learning based image segmentation in automated histopathology image analysis	112
5.1	Introduction.....	112
5.2	Deep learning-based image segmentation	114
5.3	Evaluation	125
5.4	Results.....	126
5.5	Discussion.....	134
5.6	Conclusion	136
Chapter 6:	Multi-Resolution Semantic Segmentation of Biological Structures in Digital Histopathology	138
6.1	Introduction.....	138
6.2	The multi-resolution semantic segmentation	143
6.3	Experiment.....	152
6.4	Results.....	158
6.5	Discussion.....	163
6.6	Conclusion	167
Chapter 7:	Summary and Future Work	169

Acknowledgement

The PhD is an adventurous journey, filled with countless joyful heart-warming accomplishments and bitter soring failures. My goal was to not be deceived by momentary successes and stand tall against discouraging hardships. I was privileged to get to know astonishing individuals, who shed light to the darkest moments throughout my journey. I am delighted to express my deepest gratitude to some of those, who supported me during my PhD.

First and foremost, I would like to express my sincere appreciation to Dr. Adrian Chan and Dr. Eranga Ukwatta for providing me the opportunity to join their research labs and work on fascinating projects that may impact thousands of lives. I am grateful for their invaluable guidance and astonishing mentorship, which not only empowered me to grow my technical skill set but also experience personal growth. Words fail me to describe how grateful and humble I am to have the privilege to learn from them. Thank you!

I would like to express my appreciation to our clinical collaborators, who provided us with their knowledge and guidance to complete this thesis. I would like to thank Dr. Bernard Thébaud and Marissa Lithopoulos, from the University of Ottawa, Ottawa Hospital Research Institute, and Children's Hospital of Eastern Ontario Research Institute, for providing the histopathology WSIs of mouse lung tissue, and Dr. Shannon Bainbridge-White, from the University of Ottawa, for providing the histopathology WSIs of human placenta.

I would also like to acknowledge Carleton University, Government of Ontario, and Natural Sciences and Engineering Research Council of Canada (NSERC) for the financial support, which provided me the opportunity to focus on my research.

There are no words to express my gratitude towards my father, mother, and brother, who always motivated and supported me to accomplish my goals. I am grateful for having them in my life. I appreciate and value all that you have sacrificed.

Last but not least, I would like to express my deepest gratitude to my beloved wife, Dr. Fereshteh Fakhra Firouzeh, for all the supports and sacrifices she made to pave the way for me to accomplish this outstanding achievement. I cannot imagine myself where I am today without her unconditional love and support. I love you from the bottom of my heart!

List of Figures

Figure 1.1 An example nuclei segmentation approach in histopathology images. (a) represents the original histopathology image containing multiple nuclei structures. (b) represents the corresponding segmentation of each individual nuclei structure in (a). The image courtesy of F. Mahmood <i>et al.</i> [26].....	3
Figure 1.2 Examples of complex biological structures segmentation in histopathology. (a) Two examples of glandular segmentation in histopathology images of colon [28]. (b) Segmentation of villi in placental histopathology images [30]. (c) Segmentation of tumor regions in histopathology WSI of colorectal [31].....	4
Figure 1.3 Examples of histopathology lung adenocarcinoma patches at different magnification, which is used for classification cancer types. Increasing the magnification (moving from top to bottom) enlarges the field-of-view and decreases the resolution of the patches and vice versa, which indicates the trade-off between the field-of-view and resolution with respect to the size of the patches. The image courtesy of H. Tokunaga <i>et al.</i> [164].....	7
Figure 2.1 A comparison between the resolution of MRI scans vs histopathology images. (a) A high-resolution MRI scan of size 1,024×1,024 pixels from abdomen. (b) A typical histopathology image from the mouse lung tissue, which contains approximately 20,000×18,000 pixels. (c) The zoomed in version of the (b), which contains 1,024×1,024 pixels.	15
Figure 2.2 Tissue preparation steps [62].....	16
Figure 2.3 Tissue staining using H&E, Masson’s trichrome, and Jones stain techniques. a) Unstained tissues, b) stained tissues. The image courtesy of C. Laurent <i>et al.</i> [63].....	18

Figure 2.4 A typical automated histopathology image analysis workflow.....	20
Figure 2.5 Cell nuclei detection using faster R-CNN. The image courtesy of R. Chandradevan <i>et al.</i> [71].....	22
Figure 2.6 Visualization of semantic segmentation vs instance segmentation of nuclei in Fluorescence microscopic images. The image courtesy of Y. Tokuoka <i>et al.</i> [75].....	23
Figure 2.7 Examples of imaging artifacts (marked a blue) and undesired background components (marked as red).	27
Figure 2.8 Examples of complex biological structures (placental villi and fetal blood vessels), which comprises of various histological primitives (<i>i.e.</i> , cytotrophoblast cells, syncytiotrophoblast cells, and blood cells).	28
Figure 2.9 A typical convolutional neural network [139].....	37
Figure 2.10 The original architecture of the fully convolutional neural network, proposed by [148]	38
Figure 2.11 Skip connections used in fully convolutional neural network architecture [148].	39
Figure 2.12 The R-CNN architecture [153].	41
Figure 2.13 An illustration of Faster R-CNN [155].	41
Figure 2.14 Instance segmentation using Mask R-CNN [156].	42
Figure 2.15 The U-net architecture. The Blue and white boxes represent the multi-channel feature maps and copied feature maps, respectively [159].	43
Figure 2.16 The adaptive weighting multi-field-of-view CNN architecture for semantic segmentation in histopathology whole slide images [164].	46
Figure 2.17 Histopathology image segmentation using HookNet architecture [166].	47

Figure 2.18 Deep multi-magnification CNN for semantic segmentation of cancer types in histopathology images of breast [167].	48
Figure 2.19 Different biological components of a placenta image. a) An example of a histopathology whole slide image of placenta. b) 5× magnification. c) 10× magnification.	49
Figure 2.20 a) An example of a histopathology whole slide image of mouse lung tissue. b) The manually annotated image in a). Different structures are marked with color green, blue, and red, which represent diagnostically relevant tissue compartments, bronchi, and blood vessels, respectively. c) A sample extracted patch with 5× magnification. c) A sample extracted path with 20× magnification.	51
Figure 3.1 Block diagram of our methodology.	58
Figure 3.2 Classification of color content in H&E stained images, including: syncytiotrophoblasts cells (SC) and cytotrophoblasts cells (CC) in the purple class; non-cellular mesenchymal connective tissue (MCT) in the white class; and fetal red blood cells within the villi core vasculature in the red class.	59
Figure 3.3 Example of the red and dark purple class color extraction.	61
Figure 3.4 a) Examples of artifacts and undesired objects. The blue arrows show the areas that are affected by artifacts and the red arrows show the undesired objects. b) the villi classification steps output that shows the removal of artifacts and undesired objects.	63
Figure 3.5 (a) Examples of attachment areas in villi. (b) A sample contour and an elliptical-shape structuring element to be used in calculation of top-hat transform. (c) The interpretation of the top-hat transform by fitting the SE inside the contour. The areas that	

cannot be covered by SE are shown in red. These areas are the output of top-hat transform

(d) The output of the top-hat transform, which represents all of the convexities and concavities in the contour. (e) The candidate concavity, which is the only concavity in the contour that its removal divide the original contour into two separate contours. (f) The resultant contours of removing the candidate concavity from the original contour. Both contour satisfied the minimum size threshold of a villous. 64

Figure 3.6 An example of detecting the candidate concavity by changing the characteristics of the SE. In this example, two rotation is considered for a given elliptical-shape structuring element (*i.e.*, 0° , 45°). The size of the SE is changed for each rotation to the point that the top-hat transform of the input contour becomes equal to the input contour itself. As it can be seen in this figure, the candidate concavity cannot be detected in 0° rotation. On the other hand, in 45° rotation increasing the size of the SE leads to appearance of the candidate concavity in the top-hat transform of the contour. 66

Figure 3.7 Sample segmentation of villi, a) The original image, b) The manual segmentation. The boundary of a villous structure is annotated in blue. The red annotation shows undesired objects, c) The segmentation results of our proposed method. The boundaries of villi is annotated in green. d) The generated binary mask showing the final segmentation results..... 70

Figure 4.1 Examples of extracted fields of view (FOV) images using the image analysis software. (a) This FOV image is rejected because the guideline is partially outside of the section. (b) The FOV image is rejected due to intersection of the guideline with a bronchus. (c) The FOV image is rejected because the guideline intersects with a blood vessel. (d) An

example of accepted FOV image with five intersections. The intersections of the guideline with the septa is shown using red arrows..... 80

Figure 4.2 Block diagram of the proposed methodology. The abbreviations WSI, FOV, and MLI are referred to whole slide image, field-of-view, and mean linear intercept, respectively. 81

Figure 4.3 The CNN architecture. The blue and white boxes represent the multi-channel feature maps and copied feature maps, respectively. The size of the feature maps is indicated on top of each box and the size of each input layer is denoted on the left-hand side. 82

Figure 4.4 The organization of extracted colors in a field-of-view (FOV) image. a) The original input FOV image. b) The RGB representation of each color slice. c) The visualization of each color binary mask, P_m , W_m , and R_m 84

Figure 4.5 The visualization of a bronchus (a) and a vessel (b). The corresponding lumen region (LR) and border wall region of each structure is identified by blue and red arrows, respectively. As it can be seen in (b), the LR of the vessel is densely covered by the red blood cells. 85

Figure 4.6 The block diagram of the region growing based lumen region (LR) segmentation. The notation $bwi(j)$ refers to the “ j^{th} ” selected seed region and superscript “ (j) ” indicates that bwi is expanded j times. 87

Figure 4.7 a) A sample field-of-view (FOV) image. The blue arrows indicate the border-wall region of the bronchi structures. b) Red color binary mask R_m . c) The extracted vessel’s seed region. d) The extracted vessel lumen region (LR). e) The extracted bronchi LR. f) The visualization of the ripple pattern in bronchi structures. The ripple pattern is

marked as red in this figure. g) The corresponding seed region in border-wall extraction step. h) The extracted border-wall regions of every vessel and bronchi. i) The segmented bronchi and vessel in the FOV image. 90

Figure 4.8 An example of septa segmentation in a field-of-view (FOV) image..... 93

Figure 4.9 An example of hyper-parameters grid search at each fold. At each fold, the values of *THalveoli* and *THartifact* were determined over eight training WSIs and will be tested on two remaining WSIs in the test set. This procedure continues until all WSIs used as test set. 94

Figure 4.10 An example of the complete foreground extraction approach. a) The original whole slide image. b) The manually segmented ground truth mask. c) The proposed method segmentation mask..... 99

Figure 4.11 Visualization of the automated assessment results. a) Accepted field-of-view (FOV) images. The intersections with the septa are marked as red. The blue color indicates that there are no intersections. b) Rejected FOV images..... 101

Figure 4.12 The evaluation of the proposed method in detection of the intersections against the human raters. (a), (b), and (c) represent the intersection difference between the automated method and each individual raters scores. (d) shows the comparison between the generated intersections by the proposed method and the average number of intersections reported by human raters. (e) represents the comparison in instances that all human raters agree on number of intersections (*e.g.*, all raters agree that there are five intersections with alveoli septa in the corresponding field-of-view (FOV) image). 102

Figure 4.13 The evaluation of the calculated mean linear intercept (MLI) score by the proposed method against human raters. 103

Figure 4.14 The variability in intersection counting by human raters.	108
Figure 5.1 A typical deep learning framework in automated histopathology image analysis in the training phase.	114
Figure 5.2 The architecture of SegNet [242].	119
Figure 5.3 An example semantic segmentation in single-class placental villi dataset. The ground truth segmentation and the predictions for villi class are labeled as green and blue, respectively.	127
Figure 5.4 Examples of semantic segmentation of histopathology images of mouse lung dataset. The results for Blood Vessel, Bronchi, Alveoli, Alveoli wall, and back ground classes are marked as red, blue, white, green, and black, respectively.....	129
Figure 5.5 The evaluation of the deep learning models and proposed method in Chapter 4 in detection of the intersections against the human raters. The plot for Raters Average and All Agree Cases shows intersection difference between algorithm-generated results and the average number of intersections reported by human raters and counted intersections when all raters were agree on number of intersections.....	132
Figure 5.6 The Estimated MLI scores in histopathology images of mouse lung dataset	133
Figure 6.1 The trade-off between trade-off between the field-of-view, computational efficiency, and spatial resolution. By increasing the magnification (moving from left to right), the field-of-view of the patches narrows, while the spatial resolution increases. By increasing the patch size (moving from top to bottom), the field-of-view of the patches increases, while the computation efficiency decreases.....	143

Figure 6.2 The overview of the proposed multiresolution segmentation pipeline. In contextual information extraction stage, four different expert CNNs (ECNNs) are used to generate the heat-maps associated with each class at different resolutions as shown in green, yellow, purple, blue, and red for background, alveoli lumen, alveoli border wall, bronchi, and blood vessel, respectively. In data aggregation stage, the size-based structure weighting step is applied to the heat-maps associated with each class at each magnification. The generated weighted heat-maps associated with all classes are processed using the aggregating CNN (ACNN) to generate the final heat-maps for each class. 145

Figure 6.3 Size-based structure weighting of the heat-maps of different magnification for each class. In this figure, an example WSI of lung dataset is presented, which contains five different classes. For simplicity, only the heat-map weighting process for one class (Blood Vessel) is visualized..... 148

Figure 6.4 An example of histopathology WSI of mouse lung tissue and distribution of different classes..... 156

Figure 6.5 An example semantic segmentation result in single-class placenta villi dataset. The ground truth segmentation and the predictions for villi class are labeled as green and blue, respectively..... 159

Figure 6.6 The average sensitivity of each method with respect to the structures' size in single-class placental villi dataset..... 160

Figure 6.7 An example semantic segmentation of a lung histopathology WSI using proposed method and SegNet²⁷, DilatedNet²⁸, and U-Net¹⁹ at different magnifications. The Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background classes are labeled as red, green, blue, yellow, and gray, respectively. 161

Figure 6.8 The average sensitivity of each method with respect to the structures' size for a) Alveoli class, b) Bronchus class, and c) Blood Vessel class in multi-class lung dataset. 163

List of tables

Table 2.1 The average compressed on-disk size of histopathology images in different applications [65].....	19
Table 2.2 A comparison between the performance of recently proposed color normalization approaches in different cancer datasets [66].....	21
Table 3.1 Overall F1 score and sensitivity of the previous method, the proposed method without detection of touching villi algorithm, and the proposed method with touching villi detection algorithm. The results are reported as the weighted average of five fold cross validation.....	71
Table 3.2 A comparison between our proposed method and the previous work. The results are presented for each fold for health term pregnancies and pregnancies complicated with PE. The TN row for the previously published work is not applicable as explained in section 2.4.	71
Table 4.1 The evaluation of the foreground extraction step performance using pixel-wise accuracy (AC), Dice coefficient (DSC), and Hausdorff distance.....	99
Table 4.2 Distribution of extracted field-of-view (FOV) images. The “ <i>completely inside pleural space</i> ” column represents the FOV images that contain no specific tissue compartments and therefore, are eliminated in the foreground extraction step. The remaining FOV images are used in detection of main biological structures step to identify the accepted FOV images.	100
Table 4.3 Evaluation of the proposed method performance in detection of field-of-view (FOV) images. For evaluation approach (1), the decisions are correct if the automated method agrees with at least two out of three raters. For evaluation approach (2), the	

decisions are correct if the automated method agrees with at least one rater out of three raters. The results are presented for five different folds for each group. The abbreviation STD stands for standard deviation. In this table, the abbreviations *RA* and *O2 + LPS* refers to the mice group housed in room air and mice group were exposed to a high concentration of oxygen and lipopolysaccharide, respectively. 100

Table 4.4 The evaluation of the reliability of agreement in detection of intersections in accepted field-of-view (FOV) images between the proposed method and the manual assessments using Fleiss' Kappa score. 102

Table 5.2 Overall F1 score and sensitivity metrics for detection of villous structures instances in our single-class placenta dataset. The results are presented for deep learning-based models, the proposed method in Chapter 3 with and without detection of touching villi step, and the previously published work [29]. 128

Table 5.1 Pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) accuracy metrics for semantic segmentation in single-class placental dataset. 128

Table 5.3 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) accuracy metrics for semantic segmentation in multi-class mouse lung dataset. 130

Table 5.4 Evaluation of the proposed method performance in detection of field-of-view (FOV) images. For evaluation approach (1), the decisions are correct if the automated method agrees with at least two out of three raters. For evaluation approach (2), the decisions are correct if the automated method agrees with at least one rater out of three raters. 131

Table 5.5 The evaluation of the reliability of agreement in detection of intersections in accepted field-of-view (FOV) images between the deep learning methods and the proposed method in Chapter 4 with respect to the manual assessments using Fleiss' Kappa score. 133

Table 6.1 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) metrics for our single-class placenta dataset. The metrics are reported for mean \pm standard deviation over each sample. 160

Table 6.2 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) metrics for our multi-class lung dataset. The metrics are reported for mean \pm standard deviation over each sample. 162

List of Abbreviations

AC : Pixel-Wise Accuracy

ACM : Active Contour Model

ACNN: Aggregating Convolutional Neural Network

AHIA : Automated Histopathology Image Analysis

ANN : Artificial Neural Networks

BPD : Bronchopulmonary Dysplasia

CAD : Computer-Aided Diagnosis

CC : Cytotrophoblasts Cells

CHEO : Children's Hospital of Eastern Ontario

CNN : Convolutional Neural Network

CT : Computed Tomography

DSC : Dice Similarity Coefficient

ECNN: Expert Convolutional Neural Network

FCN : Fully Convolution Neural Network

FOV : Field-of-View

H&E : Hematoxylin and Eosin

LR : Lumen Region

MCT : Mesenchymal Connective Tissue

MIA : Medical Image Analysis

MLI : Mean Linear Intercept

MPAS : Methenamine Periodic Acid Schiff

MRI : Magnetic Resonance Imaging

PA: Pixel-Wise Accuracy

PAS : Periodic Acid Schiff

PCC : Pearson Correlation Coefficient

PE : Preeclampsia

PET : Positron Emitted Tomography

PMD : Placenta-Mediated Disease

PPV: Positive Predictive Value

QSSIM : Quaternion Structure Similarity Index

R-CNN : Regional Convolutional Neural Network

RCWIH : Research Centre for Women's and Infants Health

ROI : Region of Interest

RPN : Region Proposal Network

SC : Syncytiotrophoblasts Cells

SE : Structuring Element

SMOTE : Synthetic Oversampling Of Minority Class Method

SPCN : Structure Preserving Color Normalization

SSIM : Structure Similarity Index Matrix

US : Ultrasound

VC : Villi Core

WSI : Whole Slide Image

Abstract

Analysis of high-resolution histopathology whole slide images (WSIs) is a vital step in the diagnosis and treatment of many diseases, including placenta-mediated diseases (PMDs) and respiratory illnesses. Currently, the most trusted approach is manual/semi-automated analysis of histopathology WSIs by an expert pathologist. This is problematic because high-resolution histopathology WSIs usually have a very large size (*e.g.*, 80,000×80,000 pixels) and a large number of complex biological structures. As such, applying manual/semi-automated approaches to assess histopathology WSIs can be inefficient, expensive, and subject inter- and intra-rater variability.

An alternative approach to manual and semi-automated approaches is to implement machine learning and image processing techniques to develop automated histopathology image analysis (AHIA) pipelines. A fundamental step in generating accurate AHIA approaches is the semantic segmentation of biological structures in high-resolution histopathology images. In this thesis, our main objective is to develop accurate AHIA pipelines for semantic segmentation of complex biological structures in histopathology WSIs. We specifically focus on two histopathology applications: 1) segmentation of villi in histopathology WSIs of human placenta and 2) segmentation of complex biological structures of mouse lung tissue.

Initially, we investigate the rule-based methods using conventional machine learning and image processing methods for segmentation of histopathology WSIs, developing two separate pipelines to address the challenges associated with each of our applications. We demonstrate that rule-based AHIA approaches show promising performance for the analysis of histopathology WSIs in each of applications in comparison to manual

assessment by expert pathologists and can be considered as a potential replacement for manual/semi-automated approaches. Then, we investigate deep learning methods in semantic segmentation of histopathology WSIs to further improve our developed rule-based approaches in terms of segmentation performance, generalizability, and training and testing speed. One of the bottlenecks in developing deep learning methods is the large size of histopathology WSIs, which requires implementing a patch-based approach to feed the images to deep learning models. We demonstrate that this bottleneck may limit the performance of the deep learning models due to three-fold trade-off between field-of-view, computational efficiency, and spatial resolution. As such, we propose a multi-resolution semantic segmentation pipeline to address this trade-off in AHIA using deep learning.

Chapter 1: Introduction

1.1 Motivation

The emergence of medical imaging techniques (*e.g.*, radiographic imaging, magnetic resonance imaging (MRI), computed tomography (CT), ultrasound (US), positron emitted tomography (PET), and histopathology) revolutionized the field of medicine, including diagnosis, treatment, and clinical research. Early approaches toward medical image analysis (MIA) involved manual assessment of medical images by medical specialists, which was expensive, tedious, and subject to inter- and intra- observer variability.

In recent decades, there have been major research efforts to develop automated MIA methods to address the issues associated with human analysis. Automated MIA can produce accurate and reproducible results in a short period of time, and reduce the subjectivity in manual assessment. Automated MIA is an important part of computer-aided diagnosis (CAD) systems. CAD systems have been investigated for various medical imaging modalities, including radiographic imaging [1]–[3], CT [4]–[10], MRI [11], [12], nuclear imaging [13], [14], and US [15]–[17].

Compared to other imaging modalities, there are relatively fewer works available in the literature that focus on automated MIA of histopathology images (*i.e.*, tissue images generated using microscopic imaging modality). A possible reason for the delayed appearance of automated histopathology image analysis (AHIA) systems can be related to various technological limitations involved with the digitization of histopathology slides (*e.g.*, scanner, storage, etc.). Another reason is the unique characteristics of histopathology images, which makes the developed automated MIA methods for other medical images not as easily translatable to AHIA. In contrast to other medical imaging modalities, the

histopathology images are color images, have larger image sizes, and contain greater levels of detail.

In recent years, the appearance of the whole slide image (WSI) technology has increased the availability of digital histopathology images, which has in turn resulted in a growing interest in the development of AHIA methods [18]–[21]. Moreover, advances in the field of computer vision and machine learning, have made the development of automated MIA approaches, capable of tackling the challenges involved with histopathology images, more feasible.

AHIA can be broadly categorized into four major branches: 1) image preprocessing, 2) image detection and segmentation, 3) feature extraction, and 4) classification and grading. In an AHIA system (similar to any other image processing workflow), image preprocessing prepares input images in a way that they will be compatible for further analysis. For example, color normalization is one of the basic and essential steps in most AHIA. Color normalization is performed to standardize the color intensities of each tissue type caused by variations in staining and scanning conditions. Image detection and segmentation of biological structures in histopathology images, such as cells, cells nuclei, and blood vessels, play a vital role in most AHIA applications. The detection of biological structures is to identify whether a specific type of biological component is present in a given image (*e.g.*, detection of a lesion in cancer applications). The task of segmentation in histopathology images refers to the process of determining the structures' boundaries. Feature extraction is performed in many AHIA applications to obtain discerning characteristics of the tissue (*i.e.*, texture, shape, and size), which can potentially be used for diagnostic purposes. Classification and grading of the histopathology images is often

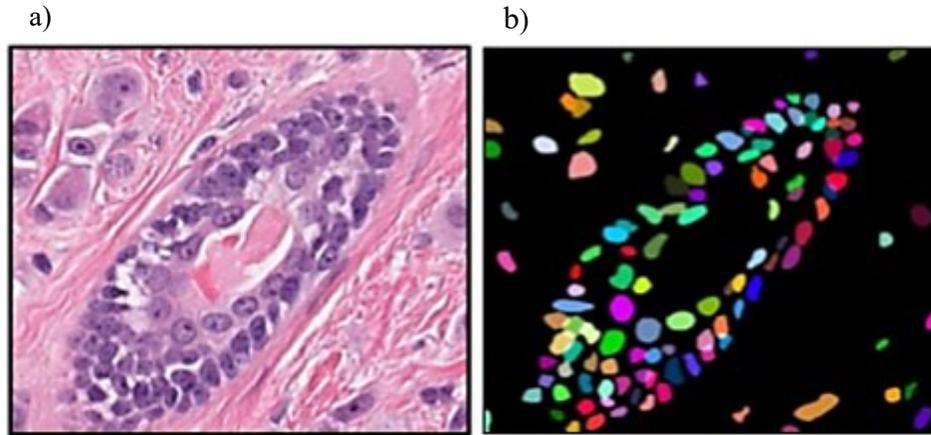


Figure 1.1 An example nuclei segmentation approach in histopathology images. (a) represents the original histopathology image containing multiple nuclei structures. (b) represents the corresponding segmentation of each individual nuclei structure in (a). The image courtesy of F. Mahmood *et al.* [26]

the ultimate goal in many AHIA. In classification and grading applications, the aim is to train a machine-learning model capable of categorizing the input images into healthy versus diseased groups, or grade the image (*e.g.*, severity of abnormality) on a continuous or discrete scale [19].

Image segmentation is a challenging task in MIA [22]. It is often a prerequisite step in many automated image analysis workflows and therefore, it directly influences the overall performance of the analysis. Previous research on segmentation in AHIA has primarily focused on the segmentation of individual cells and histologic primitives (*e.g.*, nuclei, mitotic figures, etc.) [23]–[25]. Figure 1.1 shows an example of nuclei segmentation method in histopathology images [26]. AHIA methods for the segmentation of histological primitives with high accuracies have been developed and are capable of producing results comparable to expert pathologists [27]. There are other segmentation applications in histopathology that focus on the segmentation of complex biological structures, such as gland segmentation [28], placental villi [29], [30], and cancer tissue segmentation [31],

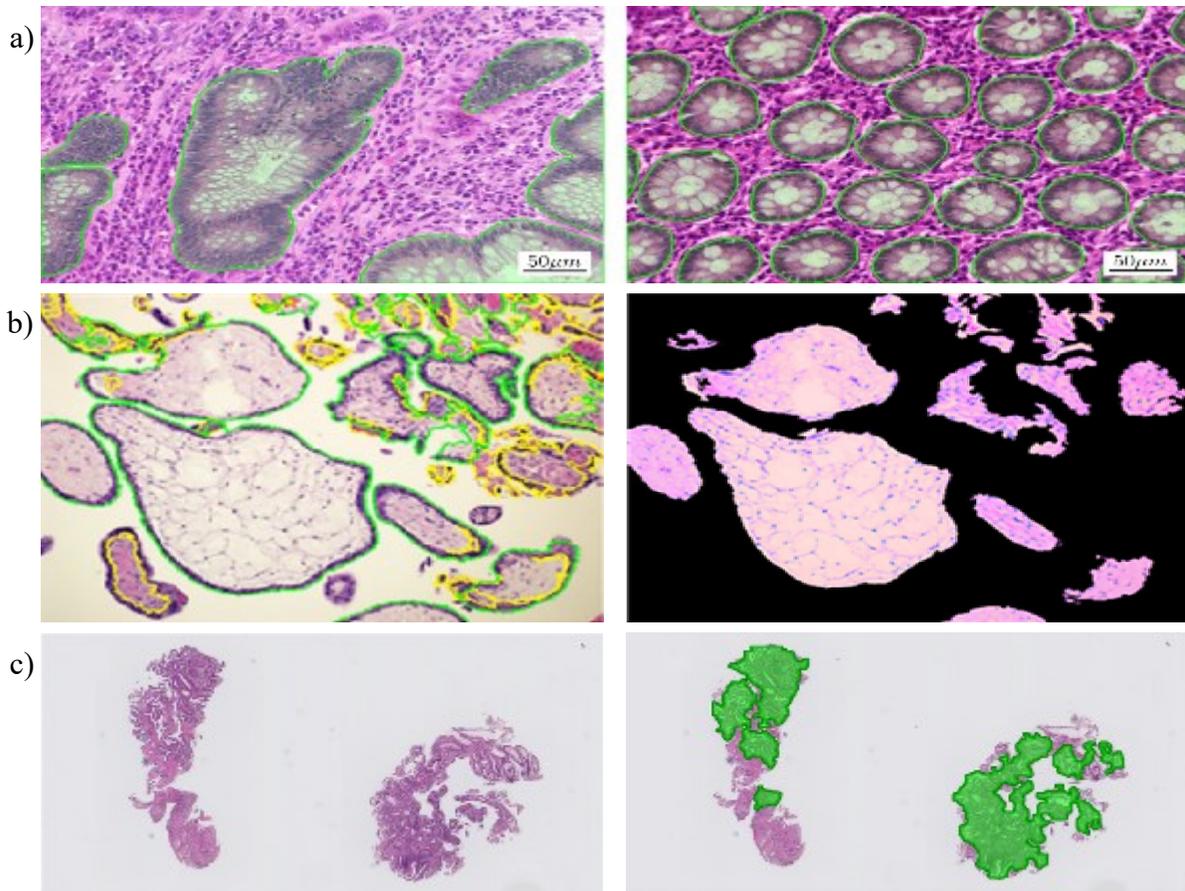


Figure 1.2 Examples of complex biological structures segmentation in histopathology. (a) Two examples of glandular segmentation in histopathology images of colon [28]. (b) Segmentation of villi in placental histopathology images [30]. (c) Segmentation of tumor regions in histopathology WSI of colorectal [31].

[32]. Compared to histological primitives, these complex biological structures vary in shape, size, and texture dramatically (Figure 1.2).

Assessment of different features associated with these complex biological structures plays a critical role in the diagnosis and/or treatment of various diseases; however, research on AHIA segmentation of complex biological structures has been limited. For instance, studies suggest that there is a strong connection between placenta-mediated diseases (PMD), such as preeclampsia, and the cardiovascular health of the baby and mother [33]. Consequently, diagnosis of PMDs through assessment of statistical measures of placenta complex biological structures (*i.e.*, placental villi structures) can potentially mitigate

cardiovascular risks by adopting early interventions. However, there are few works in the literature focusing on automated segmentation of complex biological structures of the placenta [28], [34], [35].

Another example of segmentation of complex biological structures in histopathology is the automated calculation of the mean linear intercept (MLI) score in histopathology images of the lung [36]. The MLI score is a common metric to quantify lung injury in preterm birth complications, such as bronchopulmonary dysplasia [37]. The current approaches in determining the MLI score rely on human visual assessment, which can be subjective, time-consuming, and expensive. Automating MLI scoring can be an effective strategy to resolve these issues. The prerequisite step in automating the MLI calculation is to segment the main structures of lung tissue (*e.g.*, bronchi, vessels, and alveoli). However, automated segmentation of these structures has not been comprehensively addressed in the literature.

1.2 Thesis aims and objectives

The overall objective of this thesis is to develop and evaluate generalizable and accurate AHIA methods for the segmentation of complex biological structures. In this thesis, we focus on the development of images segmentation methods in two different histopathology applications: 1) the segmentation of chorionic villi in digital histopathology images of the placenta; and 2) the segmentation of lung structures – bronchi, vessels, and alveoli- in digital histopathology images of the mice lung tissue.

As an initial step, we employ conventional image processing and machine learning techniques to develop rule-based approaches to address the segmentation task in each of our AHIA applications. One of the advantages of the rule-based AIHA approaches is the

feasibility for exact imitation of the practicing methods by expert pathologists for the diagnosis of histopathology WSIs. As such, the non-complex scheme of the developed AHIA pipeline results in a more comprehensible behavior, which can be easily justified by a human interpreter. However, there are certain shortcomings associated with rule-based approaches in terms of performance, generalizability, and processing time. Due to the complexity of the biological patterns and variability of the histopathology images, it is difficult for rule-based approaches to address the challenges associated with AHIA. Rule-based approaches usually cannot generalize across different applications, and therefore, for each new AHIA application, a new pipeline should be implemented to address the particular challenges of that application. Note that in this thesis, the term *generalizability* is mainly refers to the transferability of a developed method across different applications. In addition, compared to deep learning-based approaches, the rule-based methods requires longer run time, especially over the testing data. As a result, in the next phase of our work, we focus on developing a generic deep-learning based approach to address the challenges involved in rule-based approaches.

The implementation of deep learning methods in AHIA can be challenging because the histopathology WSIs are often too large to be directly fed into the model. As such, the input WSI is usually divided into small image patches, the extracted patches are separately processed by the segmentation model, and output patches are reconstructed to generate the segmentation results. However, smaller image patches have less context and may lack a wide range of texture patterns that might be useful for segmentation, particularly when segmenting large biological structures. In order to extend the field-of-view, while retaining a small patch size, the input image is usually down-sampled; this will result in a reduction

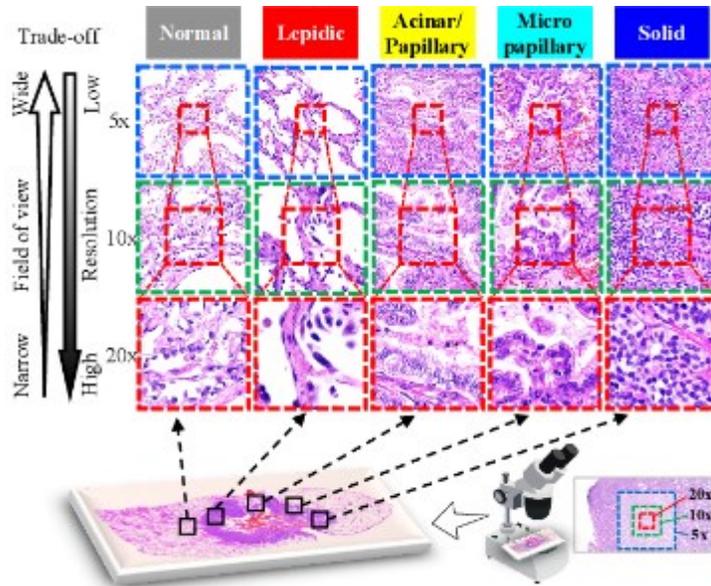


Figure 1.3 Examples of histopathology lung adenocarcinoma patches at different magnification, which is used for classification cancer types. Increasing the magnification (moving from top to bottom) enlarges the field-of-view and decreases the resolution of the patches and vice versa, which indicates the trade-off between the field-of-view and resolution with respect to the size of the patches. The image courtesy of H. Tokunaga *et al.* [164].

of the spatial resolution, which can affect the performance of the model, particularly in segmentation smaller structures (see Figure 1.3). This trade-off between the field-of-view and the resolution of the input patches makes it difficult to segment the biological structures in histopathology WSIs.

As a result, in the final phase of this thesis, we propose a multi-resolution deep learning approach capable of addressing this trade-off in the segmentation of biological structures in histopathology WSIs.

The specific objectives of this thesis are:

- Develop a rule-based automated pipeline for segmentation of chorionic villi in digital histopathology images of human placenta using image processing and conventional machine learning approaches

- Develop a rule-based automated pipeline for segmentation of complex biological structures in histopathology images of mouse lung tissue using image processing and conventional machine learning approaches, which enables automated MLI scoring
- Develop an automated pipeline for complex biological structures segmentation in histopathology images using a generic deep-learning approach. The generalizability of this pipeline will be demonstrated in the segmentation of chorionic villi in digital histopathology images of human placenta and the segmentation of mouse lung tissue
- Propose a novel multi-resolution semantic segmentation method, which enables addressing the trade-off between the field-of-view, resolution, and computational efficiency

1.3 Thesis contribution

In this thesis, we propose automated approaches for the assessment of histopathology images. First, we use conventional machine learning and image processing to develop rules-based methods for image segmentation in histopathology images of the placenta and histopathology images of mouse lung tissue. Second, we develop deep learning based methods for the segmentation of histopathology images. The conventional approaches are useful tools for: 1) providing a comparative baseline, 2) supporting the annotation of histopathology images for training deep learning methods and ground truths to evaluate the performance of these deep learning methods.

The major research contributions of this thesis are:

1. Developed a fully automated pipeline for rules-based segmentation of villi in histopathology WSIs of human placenta

2. Developed a fully automated pipeline for rules-based segmentation of biological structures in histopathology WSIs of mouse lung tissue
3. Developed a generic deep learning approach for automated segmentation of villi in placental histopathology images, which was capable of 1) acquiring the state-of-the-art performance in segmentation of histopathology images, 2) detection of touching villi, and 3) analyzing large histopathology WSIs
4. Develop a generic deep learning approach for quantification of lung injury in histopathology WSIs of mouse lung tissue, which is capable of 1) acquiring the diagnostically relevant tissue compartments, 2) segmenting the lung main biological structures, 3) automated calculation of MLI score by replicating the exact approach used in clinical approach
5. Propose a novel multi-resolution semantic segmentation method for segmentation of complex biological structures in histopathology images, which utilize the contextual information of histopathology images at different spatial resolutions and aggregate this information to generate higher overall performance. The proposed algorithm was capable of improving the state-of-the-art segmentation accuracy by addressing the dramatic size difference (*i.e.*, approximately 1,000,000 times difference)

Other research contributions of this thesis are:

1. A novel computationally efficient concavity detection algorithm based on Top-hat transform, which was used to detect candidate regions that contain touching villi in placental histopathology images
2. Generate ground truth segmentation for 96 sample images of size 2740×3964 pixels from histopathology images of human placenta and 20 WSIs of mouse lung tissue

Parts of this thesis have been disseminated in the following publications:

- Salsabili, S., Mukherjee, A., Ukwatta, E., Chan, A. D. C., Bainbridge, S., & Grynspan, D. (2019). Automated segmentation of villi in histopathology images of placenta. *Computers in Biology and Medicine*, 113, 103420. (Chapter 3)

In this research, I developed the segmentation algorithm and performed the evaluation with feedback from A.D.C. Chan and E. Ukwatta. Image data were provided by S. Bainbridge. Expertise on placental anatomy and interpretation of histopathology images were provided by A. Mukherjee, S. Bainbridge, and D. Grynspan, which helped to guide the algorithm development. I wrote the manuscript and it was edited by all co-authors.

- Salsabili, S., Lithopoulos, M., Sreeraman, S., Vadivel, A., Thébaud, B., Chan, A. D. C., & Ukwatta, E. (2021). Fully automated estimation of the mean linear intercept in histopathology images of mouse lung tissue. *Journal of Medical Imaging*, 8(2), 027501. (Chapter 4)

In this research, I developed the segmentation algorithm and performed the evaluation with feedback from A.D.C. Chan and E. Ukwatta. Image data were provided by B. Thébaud. Expertise on placental anatomy and interpretation of histopathology images were provided by M. Lithopoulos, A. Vadivel, and B. Thébaud, which helped to guide the algorithm development. I wrote the manuscript and it was edited by all co-authors.

- Salsabili, S., Ukwatta, E., & Chan, A. D. C. (2021). Deep Learning Based Histopathology Image segmentation. The manuscript is under preparation to be submitted to *Journal of Medical Imaging*. (Chapter 5)

In this research, I developed the segmentation algorithm and performed the evaluation with feedback from A.D.C. Chan and E. Ukwatta. I wrote the manuscript and it was edited by all co-authors.

- Salsabili, S., Ukwatta, E., & Chan, A. D. C. (2021). Multi-Resolution Semantic Segmentation of Biological Structures in Digital Histopathology. Submitted to *Journal of Medical Imaging*. (Chapter 6)

In this research, I developed the segmentation algorithm and performed the evaluation with feedback from A.D.C. Chan and E. Ukwatta. I wrote the manuscript and it was edited by all co-authors.

Other research related to this thesis work was disseminated in the following publications:

- Modelling Placenta Mediated Diseases with Supervised Machine Learning. A Mukherjee, S Salsabili, ADC Chan, M Leckie, D Gynspan, *Reproductive Sciences* 27 (SUPPL 1), 171A-171A 2020
- Automating Segmentation of Placenta Tissue Images for Computer Aided Diagnosis and Creating Objective Measures for Microscopic Components A Mukherjee, S Salsabili, ADC Chan, E Ukwatta, M Leckie, D Gynspan, *Reproductive Sciences* 25, 244A-245A

1.4 Thesis organization

The rest of this thesis is organized as follows: Chapter 2: presents a brief introduction to AHIA and some of the key aspects involved in automated analysis of the histopathology images. Chapter 2 also includes a broad literature review and backgrounds on image segmentation techniques used in the field of image processing and computer vision, as well

as a more focused review in AHIA. 0 presents the proposed pipeline for automated segmentation of villi in placental histopathology images, using conventional image processing and machine learning approaches. Chapter 1: presents the pipeline for automated MLI scoring in histopathology images of mouse lung tissue, using conventional image processing and machine learning approaches. In Chapter 5, a generic approach for deep learning based segmentation of histopathology images is presented. In Chapter 6, we present our novel deep learning approach for multi-resolution semantic segmentation of complex biological structures in histopathology WSIs. Conclusions and a discussion for future work are presented in Chapter 7.

Chapter 2: Literature review and background

In this chapter, an overview of the relevant methods and backgrounds in MIA are presented. In section 2.1, we briefly discuss the utilization of CAD systems for different medical image modalities and present an overview of challenges involved with the development of CAD systems in histopathology. The different steps to generate digitized histopathology images are presented in section 2.2. In section 2.3, we describe a typical workflow for a CAD system in AHIA. In section 2.4, we discuss image segmentation in AHIA applications and discuss the critical role that segmentation plays in AHIA pipelines. Common image segmentation methods are discussed in section 2.5 followed by a review of the state-of-the-art segmentation methods in the field of image processing and computer vision. In section 2.6, we provide an overview for the two AHIA applications that we focus upon in this thesis research: 1) segmentation of chorionic villi in histopathology images of human placenta; and 2) segmentation of histopathology images of mouse lung tissue.

2.1 Computer-aided detection/diagnosis

Medical imaging has proven to play a vital role in different fields of medicine since the discovery of X-rays in 1895. Medical images provide invaluable clinical information, assisting healthcare providers in the diagnosis and treatment of their patients. With technological advances in the field of medical imaging, medical images are becoming increasingly prevalent and abundant in healthcare systems. As an example, the technology used in the first generation of CT scanners involved the rotation-translation of a pencil beam, which was considerably demanding in terms of time and human effort [38]. In contrast, the latest generation (7th generation) of CT scanners are developed to be able to generate detailed high-resolution scans in a matter of seconds with minimal supervision.

As a result, such technological advances have paved the way for medical images to become a routine in the process of diagnosis and treatment of a disease. Based on a recent study in United States, the number of CT scans among older adults increased from 204 per 1000 person-year to 428 per 1000 person-year from the year 2000 to 2016 [39].

Because manual assessment is involved in many MIA applications, the considerable increase in the production of medical images may be problematic for various reasons. First, MIA by human experts is often subjective and thus susceptible to inter- and intra- expert variability in the assessment process [40]–[42]. Second, studies suggested that physicians do not detect all abnormalities on images that are visible on retrospective review, and they do not always correctly characterize abnormalities that are found [43]–[46]. As an example, the error rate in manual assessment of radiology images has been reported to be around 3%-5% (much higher in targeted studies), and a considerable portion of these errors are the result of misdiagnosis by the radiologists [47]. As a result, automated MIA is an active research area, and CAD is becoming an indispensable asset in physicians' decision-making processes [48]–[51].

The history of CAD dates back to 1955 with the work of Lee Lusted on the potential use of digital computers in large-scale data problems in medicine [52]. In 1963, Lodwick *et al.* [53] digitized chest radiographs for computer analysis. In 1964, Becker *et al.* [54], proposed the first practicable computerized approach for the diagnosis of X-ray films. In 1973, Toriwaki *et al.* [55] reported the first study on a CAD system of a focal abnormality in chest radiographs, and Roellinger *et al.* [56] reported the first study on CAD of a heart

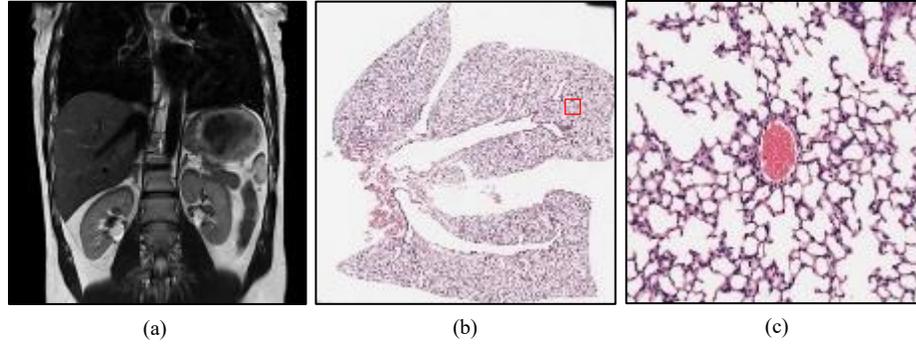


Figure 2.1 A comparison between the resolution of MRI scans vs histopathology images. (a) A high-resolution MRI scan of size $1,024 \times 1,024$ pixels from abdomen. (b) A typical histopathology image from the mouse lung tissue, which contains approximately $20,000 \times 18,000$ pixels. (c) The zoomed in version of the (b), which contains $1,024 \times 1,024$ pixels.

abnormality in chest radiographs. These ambitious research initiatives show the enthusiasm and potential for the utilization of computerized computations in the assessment of medical images. Since then, the development of CAD systems has been comprehensively addressed in different branches of medical imaging modalities, including CT imaging [57], MRI [58], nuclear imaging [59], US imaging [60].

Our focus in this thesis is on AHIA. Compared to other modalities, histopathology images pose different challenges due to their unique image characteristics. First, while most imaging modalities generate gray-scale images (Figure 2.1.a), histopathology images are color images (Figure 2.1.b). Second, the density of data that histopathology images hold is enormous in comparison to other modalities. As an example, Figure 2.1.a shows a high-resolution MRI scan (spatial resolution of 0.12mm per pixel), which comprises approximately 1024×1024 spatial elements (or ≈ 1 million pixels). Figure 2.1.b shows a typical histopathology image scanned at $40\times$ magnification (spatial resolution of $0.25\mu\text{m}$ per pixel) generates an image of approximately $120,000 \times 80,000$ spatial elements (≈ 4.8 billion pixels). Third, due to the high resolution of the histopathology images, the amount of the captured biological detail is extensive, which considerably increases the variability

and complexity of the different biological components found in these images. Figure 2.1.c shows the zoomed version of the red bounding box in Figure 2.1.b, illustrating the number of biological structures that exist in histopathology images. AHIA involves distinctive challenges that limit the extent to which automated MIA methods developed for other medical image modalities can be transferred to histopathology applications.

2.2 Histopathology imaging

In this section, we present a broad review of different preparation and processing steps for generating digitized histopathology images from tissue specimens. A more detailed description of this process can be found in [61].

2.2.1 Tissue preparation

In order to view the cellular compartments of animals or human tissue under microscope, it has to be prepared in multiple steps (see Figure 2.2 [62]):

1) **Fixation:** The tissue specimen is placed in a chemical fixative agent (*e.g.*, 10% neutral buffered formalin). This will slowly penetrate the tissue causing chemical and



Figure 2.2 Tissue preparation steps [62].

physical changes that will harden and preserve the tissue and protect it against subsequent processing steps. In general, this process takes about 6 to 24 hours for the fixative to penetrate into every part of the tissue.

2) Dehydration, clearing, and embedding: The preservation process consists of infiltrating the tissue with an embedding agent (usually paraffin wax). Because the embedding agent is immiscible with water, prior to the embedding step, the concentration of water in the tissue specimen must be reduced. As a result, the tissue is dehydrated by increasing the concentration of an alcohol solution (usually ethanol) in multiple steps, which takes around 2 to 3 hours.

After the dehydration step, an organic solvent (*e.g.*, xylene) is used to clear the alcohol solution, which enables the tissue to be subsequently infiltrated by an embedding agent. Depending on the size of the tissue, the process usually takes from one to few hours. Then, in the embedding step, the tissue is surrounded by a large block of molten paraffin wax, creating what is now referred to as the “block”. The solidification of the block provides a support matrix, which allows thin sectioning of the tissue.

3) Sectioning & mounting: Using a machine called a microtome, the block is cut into thin (usually 5 to 10 μm) sections. The thin sections obtained from the microtome are mounted upon glass slides.

4) Staining: Because most cells are transparent under microscope (Figure 2.3) histochemical stains are used to provide contrast to tissue sections, making the structures more visible and easier to evaluate. In histology, many different staining techniques have been developed for different applications. Examples of different staining techniques are shown in Figure 2.3 [63]. In clinical histology, the most widely used stain is a combination

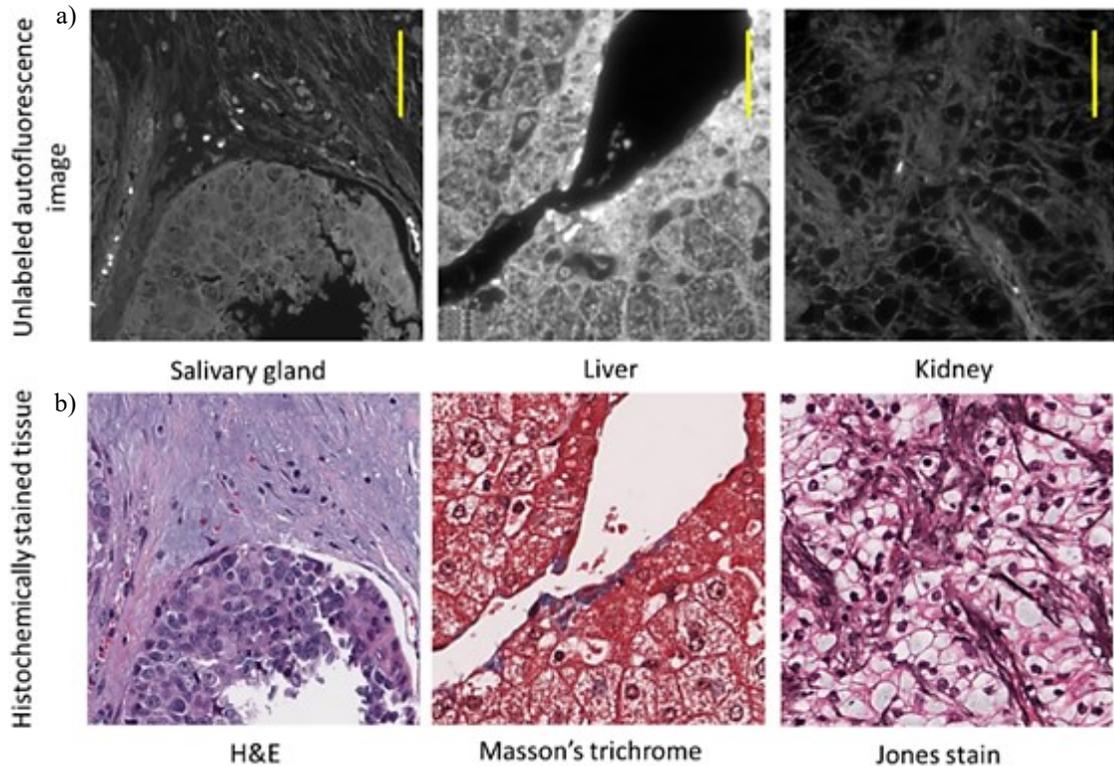


Figure 2.3 Tissue staining using H&E, Masson's trichrome, and Jones stain techniques. a) Unstained tissues, b) stained tissues. The image courtesy of C. Laurent *et al.* [63].

of hematoxylin and eosin (H&E). Hematoxylin is an alkaline stain with deep purple or blue color, and eosin is an acidic stain with a red color. With H&E staining, the chromatin rich cell nuclei are stained with a blue and/or purple shade, while the cytoplasm and other connective tissues are stained with different shades of pink. Masson's Trichrome staining (Figure 2.3.b) is used to differentiate between collagen and smooth muscle tissue in tumors. This type of staining is commonly used in skin and liver biopsies. Periodic Acid Schiff (PAS) staining is mainly used for staining structures containing a high proportion of carbohydrates such as glycogen, glycoproteins, proteoglycans typically found in connective tissues, mucus, and basement membranes. PAS staining is often used in kidney and liver biopsies. Jones' stain or methenamine PAS (MPAS) is a methenamine silve-

periodic acid schiff stain. The MPAS staining is usually used for basement membrane and is widely used in the investigation of medical kidney diseases (Figure 2.3.c).

2.2.2 Slide digitization

The early generation of histopathology scanners were simply digital cameras mounted on a standard microscope, digitizing a specific field-of-view. WSI is the process of scanning the entire glass slides to produce digital histopathological images for viewing by human or for AHIA. Since the introduction of whole slide scanners in late 1990s, WSI technology has evolved to the point where whole slide scanners are currently capable of producing high-resolution digital images within a relatively short time (a few minutes). Scanning can occur at multiple magnifications. Scanning at 20× magnification is usually acceptable for standard viewing and routine image analysis. For other applications, images should be acquired at 40× magnification to resolve information that may be separated by distances less than $0.5\mu m$ [64]. The on-disk size of the generated images ranges from a few Megabytes to several Gigabytes. Table 2.1 [65] presents an estimation of the average on-disk size of the histopathology images scanned at 20× and 40× magnification in different applications. Digitizing tissue slides at even higher resolution (*e.g.*, 60×, 63×, or 100×) is

Table 2.1 The average compressed on-disk size of histopathology images in different applications [65].

<i>Biopsy type</i>	<i>Average size (MB)</i>	
	<i>×20 resolution</i>	<i>×40 resolution</i>
Dermatopathology	1,392	-
GU	1,701	-
Head and neck	1,965	-
Hematopathology	-	40,300
Neuropathology	1,872	-
Thoracic pathology	3,240	-
Transplant pathology, needle	250 - 375	700 - 1050
Transplant pathology, wedge	600 - 900	1,500 - 2,250

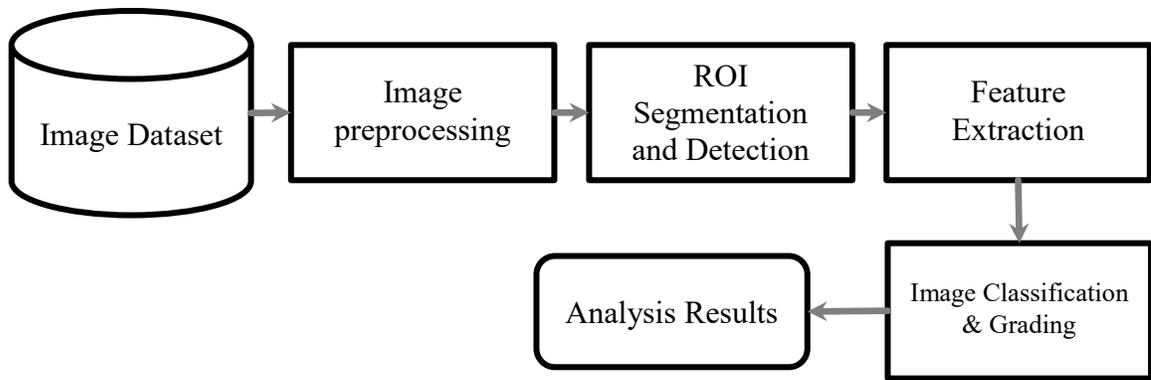


Figure 2.4 A typical automated histopathology image analysis workflow.

now available via select scanners, but is only recommended for specific use cases, such as blood smears.

2.3 Automated histopathology image analysis

A typical AHIA workflow usually comprises the following steps: 1) image preprocessing, 2) image segmentation and detection, 3) feature extraction, and 4) image classification and grading. These steps are illustrated in Figure 2.4 and discussed in the following subsections.

2.3.1 Image preprocessing

The image preprocessing step is a standard, generic procedure in AHIA, which aims to reduce image variability, artifacts, and noise, as well as transform the input image based on the requirements of the developed image analysis workflow (*e.g.*, image resizing). A common challenge in histopathology is the noticeable color variability of the digitized images, which can arise from variations in the staining and scanning conditions, such as histochemical staining time, amount of histology stain used, and use of different scanning equipment. One of the key aspects of the image preprocessing step in histopathology is color normalization, which has been comprehensively addressed in the literature over the past few decades. Various examples of color normalization techniques can be found in

comparative review studies such as Roy *et al.* [66]. Here, we present some of the most widely used histopathology color normalization techniques in recent years.

Reinhard *et al.* [67] proposed a color normalization technique, which uses a reference image to normalize the target images. In their work, the color and intensity information of the images is extracted in L*a*b color space using principle component analysis. Then, the color characteristics of the reference images are imposed on the target images, considerably reducing the color variation in different histopathology images. Macenko *et al.* [68] proposed a singular value decomposition approach to address the inconsistencies in the staining process. Khan *et al.* [69] proposed a supervised color classification method to identify the locations where each stain is present. Li and Plataniotis [70] proposed a robust color normalization approach based on an intensity matching algorithm capable of addressing illumination and stain variations in histopathology images. In Table 2.2, we present a comparison between the above-mentioned color normalization techniques, which were applied on different histopathology image datasets. In Table 2.2, the structure similarity index matrix (SSIM), quaternion structure similarity index (QSSIM), and Pearson correlation coefficient (PCC) are different quality evaluation metrics used in [66]

Table 2.2 A comparison between the performance of recently proposed color normalization approaches in different cancer datasets [66].

<i>Color Normalization Method</i>	<i>Liver Cancer</i>			<i>Breast Cancer</i>			<i>Kidney Cancer</i>			<i>Colorectal Cancer</i>		
	<i>QSSIM</i>	<i>SSIM</i>	<i>PCC</i>	<i>QSSIM</i>	<i>SSIM</i>	<i>PCC</i>	<i>QSSIM</i>	<i>SSIM</i>	<i>PCC</i>	<i>QSSIM</i>	<i>SSIM</i>	<i>PCC</i>
<i>Reinhard approach</i>	0.9254	0.9040	0.9630	0.9713	0.9480	0.9798	0.9623	0.9480	0.9798	0.9560	0.9295	0.9695
<i>Macenko approach</i>	0.9540	0.9411	0.9815	0.8540	0.9470	0.9988	0.9908	0.9470	0.9988	0.9319	0.8874	0.9879
<i>Stain color descriptor</i>	0.6534	0.6323	0.7621	0.8534	0.1567	0.8756	0.5873	0.1567	0.8756	0.8294	0.6318	0.9597
<i>Complete color normalization</i>	0.8739	0.8492	0.8957	0.9739	0.9516	0.9728	0.9668	0.9516	0.9728	0.9775	0.9678	0.9794
<i>Structure preserving color normalization</i>	0.9603	0.9417	0.9959	0.9763	0.9707	0.9834	0.9741	0.9707	0.9834	0.9779	0.9695	0.9849

for assessing the performance of color normalization techniques. One of the key observations from Table 2.2 is that the performance of color normalization techniques can dramatically be affected by the dataset being used. However, as shown in Table 2.2, the Structure preserving color normalization (SPCN) technique shows the best overall performance on all datasets.

2.3.2 Image detection and segmentation

Image detection is the acquisition of the approximate spatial location of a specific object within an image. In general, image detection consists of two main steps: 1) object localization, and 2) image classification. In the object localization step, the goal is to place a bounding box or enclosing region around a specific object to indicate the approximate spatial location of the object. Image classification is a complementary step in some image detection applications, where different objects are available in the image. Image classification assigns each bounding box to a label from a predefined set of labels. Figure 2.5 shows an example of image detection: cell nuclei detection in a histopathology WSI

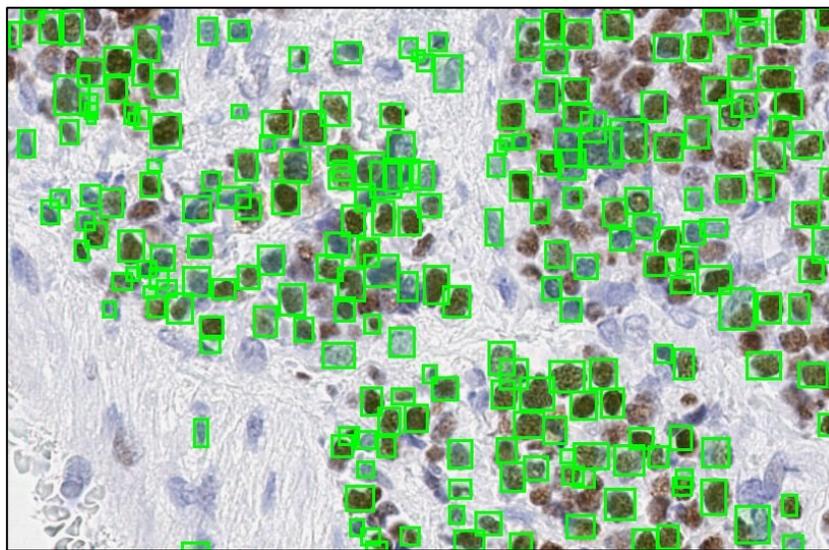


Figure 2.5 Cell nuclei detection using faster R-CNN. The image courtesy of R. Chandradevan *et al.*[71]

[71]. Other examples of histopathology image detection applications can be found in recently published review papers such as [25], [72]–[74].

Similar to image detection, the objective of image segmentation is to determine the spatial location of a specific object in a given image. However, instead of acquiring the approximate location of the object, in image segmentation, the exact boundary of the object is determined by classifying each individual pixel in the image. In general, image segmentation can be broadly categorized into two main groups: 1) semantic segmentation and 2) instance segmentation. In semantic segmentation, the main objective is to detect different objects within an image and group them based on defined categories. For instance, in semantic segmentation of biological structures of lung tissue (*i.e.*, alveoli, bronchi, and vessels), the algorithm has the ability to differentiate between objects from different groups

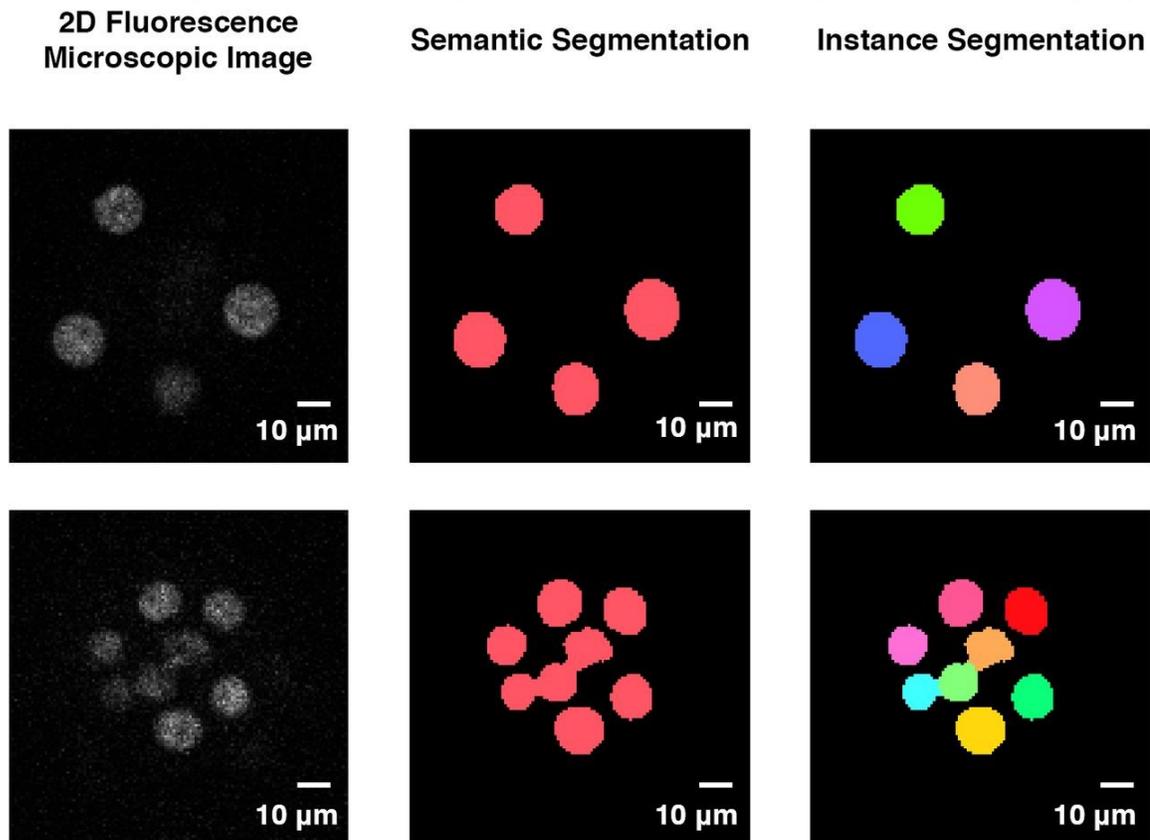


Figure 2.6 Visualization of semantic segmentation vs instance segmentation of nuclei in Fluorescence microscopic images. The image courtesy of Y. Tokuoka *et al.* [75]

(*e.g.*, vessels versus bronchi), but it does not separate different objects from the same group (*e.g.*, two or more bronchi). In instance segmentation, on the other hand, the algorithm has the ability to differentiate between objects from different groups and objects from the same group. The importance of instance segmentation is more evident in cases where the objective is to count the total number of objects from the same group and can be quite challenging when objects are touching each other. The difference between semantic and instance segmentation is visualized in Figure 2.6 for a nuclei segmentation application [75]. Instance segmentation task is important in AHIA, where the segmentation algorithm is required to identify overlapping objects [76].

As image segmentation is the focus of this thesis, a more detailed overview of histopathology image segmentation is presented in Section 2.4.

2.3.3 Feature extraction

Another important step in a typical AHIA workflow is to extract desirable histopathological features from the regions of interest (ROIs). Research on useful features has often been inspired by visual attributes defined by clinicians, which are particularly important for disease grading and diagnosis. As Belsare [77] suggested, these features can either be extracted from cellular- or tissue- level to measure morphological characteristics of the images. The cellular-level features are usually used for quantifying the properties of individual cells without considering spatial dependency between them. The tissue level features quantify the distribution of the cells across the tissue by using spatial-level and gray-level dependencies [78]. Gurcan *et al.* [73] categorized these features as object-level and spatially related features. The object-level feature (*e.g.*, size, shape, and texture) often refers to locally extracted characteristics of a group of pixels satisfying some similarity

criterion. The spatially related features, such as Veronoi features [79], represent the structural information of a histopathology tissue sample by defining a large set of topological characteristics.

Regardless of the level of view, the histopathology features can be categorized into one of four general categorizes:

- 1) Morphological (*e.g.*, size and area [80]–[82], shape[83], [84], and boundary [85])
- 2) Topological (*e.g.*, Voronoi diagram [79], minimum spanning tree [86], and skeleton[87])
- 3) Intensity based features (*e.g.*, color[88] and intensity statistics[86])
- 4) Texture (*e.g.*, density[89], wavelets[88], and entropy[90])

While handcrafted visual features allow for explicit modeling of specific characteristics that pathologists look for when identifying anomalies, there are alternative approaches in the assessment of histopathology images, inspired by convolutional neural networks (CNNs) [91], [92]. In contrast to handcrafted features, CNNs are fully data-driven, therefore may be more accurate in representing training samples and able to find feature patterns that handcrafted features may fail to describe.

2.3.4 Image classification and grading

The classification or grading of histopathology imagery is often the ultimate goal in many HIA applications. Histopathology image classification or grading refers to categorizing the presence or extent of a certain disease in a given image. As an example, in the assessment of breast histopathology images, the objective is to categorize the biopsy sample as cancerous or benign and determine the extent of the disease by grading the cancerous tissue. The conventional approach in image classification or grading in AHIA is

to classify the extracted handcrafted features derived from segmented ROIs of histopathology images using a supervised machine learning method. However, the extracted handcrafted features are usually unable to correctly represent the complex patterns and high-level information regarding the state of the disease. Therefore, in recent years, there has been a shift towards using the CNNs in image classification or grading in AHIA, which are fully data-driven.

Various review papers are available, which focus on classification or grading applications in AHIA [73], [93]–[96]. There are also organ-specific reviews that target the developed classification methods in a specific organ. As an example, Veta *et al.* [74] proposed a comprehensive review of the recent methods in automated analysis of breast cancer histopathology. Chen *et al.* [97] proposed a survey on computer vision methods in the assessment of cervical histopathology images. In their thorough review paper, the classification methods were presented under three main groups: 1) linear-based classifier, 2) weight-based classifier, and 3) other classifiers (*e.g.*, decision trees, *k*-nearest neighbours, and deep learning). Mosquera-Lopez *et al.* [98] proposed a review on automated methods for detection and grading of prostatic carcinoma in prostate histopathology images.

2.4 Image segmentation in automated histopathology image analysis

Image segmentation is a vital aspect of AHIA. The main aim here is to break down the given information into more meaningful regions, which can potentially enhance the performance of the image analysis and reduce the complexity of the assessment. This can be particularly important in AHIA applications, where the input image sizes are large, and images have substantial variability and great detail (*e.g.*, complex patterns, imaging and



Figure 2.7 Examples of imaging artifacts (marked a blue) and undesired background components (marked as red).

biological artifacts, dense appearance of biological structures, etc.). In this section, we briefly review different aspects of image segmentation AHIA.

An important segmentation task in AHIA is the segmentation of the diagnostically relevant tissue specimens. A typical histopathology image consists of undesired regions, which represent artifacts and background components (*i.e.*, areas with no specific tissue compartments). For example, in Figure 2.7, the white space (marked with blue arrow) is the background region, the areas marked as blue represent examples of imaging artifacts, and the areas marked as red represent undesired biological components of mouse lung tissue, which is not the ROI region in some AHIA applications, such as estimation of MLI score. This leaves the remaining areas of the WSI as the diagnostically relevant tissue compartments of mouse lung tissue. Consequently, it is often desirable to extract foreground regions prior to any further analysis. In many applications of AHIA, the extraction of the foreground regions is accomplished as a preprocessing step, which can be done either manually by expert pathologists or through automated approaches. Examples of automated foreground segmentation can be found in the Bandi *et al.* [99] review paper,

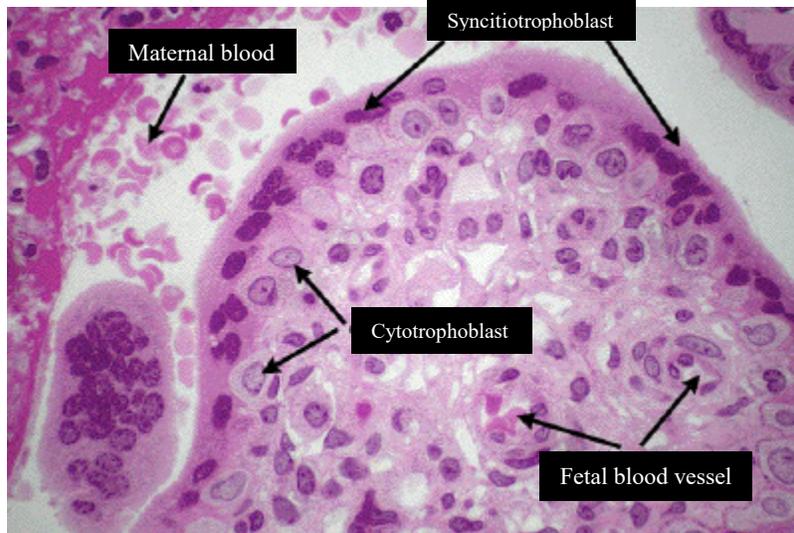


Figure 2.8 Examples of complex biological structures (placental villi and fetal blood vessels), which comprises of various histological primitives (*i.e.*, cytotrophoblast cells, syncytiotrophoblast cells, and blood cells).

where the main focus of the authors is to highlight the performance of different tissue segmentation algorithms in histopathology.

Beyond foreground segmentation in histopathology images, there is the segmentation of the main biological structures, such as some small cells (*e.g.*, syncytiotrophoblast cells, cytotrophoblast cells, and blood cells) and cell nuclei. Assessment of the morphological properties of these structures and their spatial organization are informative on many aspects of the disease (*e.g.*, cancer subtype and the grade or reaction of the patient's immune system) and determination of the next step in the patient's treatment. The segmentation of biological structures can be categorized into two main groups: 1) histological primitives, and 2) complex biological structures. Figure 2.8 shows examples of histological primitives and complex biological structures in histopathology images of placental villi. Histological primitives (*e.g.*, nuclei, lymphocytes, mitosis) are usually the smallest and simplest biological structures in a given histopathology image, which are visually less variable in size, shape, and color. In the segmentation of histological primitives, the main objective is to identify the most basic components of the tissue in a histopathology image, such as cells

nuclei, blood cells, and collagen. These applications have been subject to extensive research for over a decade [26], [85], [100], [101]; in particular, nuclei segmentation (Xing *et al.* [25] present a comprehensive review on the state-of-the-art nuclei segmentation).

Complex biological structures (e.g., glandular cells, placental villi) are usually comprised of multiple biological subcomponents that, in comparison to histological primitives are more diverse in visual characteristics (e.g., shape, size, texture, color, etc.). Consequently, automated segmentation of these structures is a more complicated task, which usually involves the identification of complex biological patterns. Although, segmentation of complex biological structures has not been addressed as comprehensively as other AHIA applications, various applications have been reported in recent years [28], [34], [35], [102]. In particular, there appears to have been a concerted effort in the automated segmentation of glandular cells. Automated gland segmentation will allow the extraction of quantitative features associated with gland morphology from digitized histopathology images, which can potentially increase the reproducibility in cancer grading applications [28].

2.5 Image segmentation methods in automated histopathology image analysis

Image segmentation methods are widely used in various applications, such as MIA, face recognition/detection, and satellite imaging [103]. Numerous segmentation methods have been proposed throughout the years. Examples of these methods can be found in surveys and reviews papers [22], [104], [105]. The image segmentation methods in AHIA can be broadly categorized into conventional and deep learning-based approaches, which are briefly reviewed in section 2.5.1 and 2.5.2, respectively. An important bottleneck in this

work was the trade-off between resolution and field-of-view in the segmentation of very large and very small biological structures. In literature, this problem is mainly addressed through multi-resolution image analysis. As such, in section 2.5.3, we present a brief overview of recently proposed multi-resolution approaches in AHIA.

2.5.1 Conventional image segmentation

Despite the growing popularity of deep neural networks, conventional image processing techniques are considered the foundation of the field of computer vision and are still utilized in various applications. In this section, we briefly introduce some conventional image segmentation methods, which have been widely used over the past few decades.

2.5.1.1 Edge-based methods

Edges in an image can be defined as the border between two regions that differ in the level of intensity [106]. These edges can often represent boundaries between objects. The edge-based segmentation methods use edge detectors to find discontinuities in gray level or color. Some of the well-known edge detectors are Sobel [107], Canny [108], Prewitt [109], and Roberts [110].

In order to detect edges in an image, two main approaches can be used [111]: 1) search-based algorithms, and 2) zero-crossing algorithms. In search-based algorithms, the gradient magnitude is calculated using first-order derivative expression. Therefore, it searches for local directional maxima of the gradient magnitude using the gradient direction. In zero-crossing based method, it first searches for zero crossings in the second derivative of the image. It can detect edges by locating the zeros in the second derivative of when the first derivative is at a maximum, the second derivative is zero. This method is also known as

Laplacian based edge detection. Examples of edge-based segmentation methods in histopathology images can be found in Nazeran *et al.* [112] review paper.

2.5.1.2 Region-based methods

In a given image, a region is defined as a subset of connected homogenous pixels of the image, which possess similar visual features such as texture, color intensity, and geometry. The region-based segmentation methods group the pixels with similar characteristics into the same regions. These methods are also required to be followed by an appropriate thresholding technique. Compared to edge detection methods, region-based segmentation algorithms are more immune to noise [113].

The region-based algorithms can be divided into two main categories: 1) region growing algorithms, and 2) region splitting and merging algorithms. In region growing algorithms [114], [115], first, a set of similarity measures, such as gray level intensity or color are calculated for a group of seed pixels in the original image. Then, the seed regions grow by appending those neighboring pixels that have similar characteristics to seed regions. Finally, the procedure is stopped when no more pixels meet the criterion for inclusion in that region (*i.e.*, size, likeness, and likeness between a candidate pixel and pixel grown so far). In region splitting and merging methods, rather than choosing seed points, users can divide an image into a set of arbitrary unconnected regions and then merge the regions in an attempt to satisfy the conditions of reasonable image segmentation. Region splitting and merging is usually implemented with theory based on quad tree data [116]. Examples of region-based histopathology image segmentation methods can be found in [77].

2.5.1.3 Threshold-based methods

Thresholding is an approach to segment a given image into two regions: foreground and background. Thresholding is the most basic, intuitive, and most commonly used image segmentation approach. Thresholding converts a grayscale image into a binary mask as follows:

$$BW(i, j) = \begin{cases} 1 & I(i, j) > T \\ 0 & otherwise \end{cases} \quad 2.1$$

where, (i, j) represents a pixel coordinate in the image I or binary mask BW , and T is a pre-defined threshold. When T is a constant, the approach is called global thresholding [117], [118]; otherwise, it is called local thresholding. Global thresholding methods can fail when the background illumination is uneven. In local thresholding, multiple thresholds are used to compensate for uneven illumination [119]. The value of the T can be automatically selected using various methods, including mean, histogram, and edge maximum technique [120]. One of the most widely used threshold methods is Otsu's [121] thresholding method, which is often used automatic segmentation of an image into foreground and background. Otsu's threshold is determined by minimizing intra-class intensity variance, or equivalently, by maximizing inter-class variance. Different applications of histopathology image segmentation using thresholding can be found in review papers, such as [76].

2.5.1.4 Clustering methods

A cluster is a collection of similar pixels that are dissimilar to the pixels in the other clusters [122]. Clustering methods operate either by partitioning or by grouping pixels. In partitioning type, the process begins by dividing the whole image into successively smaller

clusters. Alternatively, in the grouping type, it begins with each element as a unique cluster and merge these individual clusters to obtain larger clusters. The clustering segmentation methods can be categorized into two groups: 1) supervised clustering, and 2) unsupervised clustering. In supervised clustering, human interaction is required to determine the clustering criteria but in unsupervised clustering methods, the clustering criteria is determined by itself [123]. Here, we briefly introduce two of the most widely used clustering algorithms in the literature: 1) K-means clustering, and 2) fuzzy C-means clustering.

1) *K-means clustering algorithm*

The *K*-means clustering algorithm is an iterative and unsupervised clustering algorithm that partitions the dataset into *K* pre-defined, distinct, and non-overlapping subgroups (clusters), where each data point belongs to only one group. The algorithm aims to minimize the inter-class distance of data points (compactness of the data points), while maximizing the intra-class distance of the clusters' centroids (separation of clusters). Equation 2.2 describes the distance metric objective function:

$$J = \sum_{j=1}^m \sum_{i=1}^n d(x_i^{(j)}, c_j) \quad 2.2$$

where, $d(x_i^{(j)}, c_j)$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster center c_j , and J represent the distance of n data points from their respective cluster centers. Depending on the application, different distance metrics are used in *K*-means clustering algorithm, including Euclidean distance, Manhattan distance, Chebychev distance, and Minkovski distance. Some of the most common distance metrics can be found in [124]. The *K*-means algorithm has the following steps [125]:

1. Determine k as the number of clusters.
2. Select k as the number of centroids and randomly chose the centroids location.
3. Assign each data point to the nearest centroid.
4. Recalculate the cluster's centroid c_j .
5. Repeat step 3 and 4 until convergence (the clusters centroids do not change).

2) *Fuzzy C-means algorithm*

Fuzzy C-means is a clustering method, which was first introduced by Dunn [126]. In this method, a data set is grouped into N clusters, with every data point in the dataset belonging to every cluster to a certain degree. For example, a data point that lies close to the center of a cluster will have a high degree of membership in that cluster, and another data point that lies far away from the center of a cluster will have a low degree of membership to that cluster. The fuzzy C-means algorithm is based on the minimization of the objective function in Equation 2.3.

$$J = \sum_{i=1}^N \sum_{j=1}^C u_{i,j}^m \times \|x_i - c_j\| , \quad 1 \leq m < \infty \quad 2.3$$

where, m is the degree of fuzziness, which can be any real number greater or equal to 1, $u_{i,j}^m$ is the degree of membership of x_i in the cluster j , x_i is the i^{th} measured data point, c_j is the centroid of the j^{th} cluster, and $\| \cdot \|$ is any norm expressing the similarity between any measured data and the corresponding centroid.

There are various similarities between K -means and fuzzy C -means clustering algorithms. However, the fundamental differences that set these algorithms apart are as the following:

- 1) In K -means clustering, the data is divided into distinct clusters, where each data point can only belong to exactly one cluster. In fuzzy C -means clustering, data points can potentially belong to multiple clusters.
- 2) The fuzzy C -means objective function has an additional degree of membership term $u_{i,j}^m$, which control the fuzzy behavior of the algorithm.

In literature, the K -means clustering based algorithms have been comprehensively addressed in AHIA. Examples of histopathology images segmentation using K -means clustering algorithms can be found in some review papers [127].

2.5.1.5 Active contour methods

Active contour models (ACMs) are quite popular for several applications, such as object boundary detection, image segmentation, object tracking, and classification [128]. The general idea behind these methods is to evolve a curve or a surface under constraints from image forces so that it is attracted to features of interest in an intensity image. Conventional active contour models are based on energy minimization techniques. The classical snake model [129] is a spline under the influence of internal and external image forces. Image forces push the snake towards the image salient features such as edges and lines [130]:

$$E = E_{int} + E_{ext} \tag{2.4}$$

$$E_{int} = \alpha \sum_{i=0}^{n-1} |v_{i+1} - v_i|^2$$

$$E_{ext} = \sum_{i=0}^{n-1} |\nabla I(v_i)|^2$$

where, α is an adjustable constant, v_i is the i^{th} point of the evolving curve, $\nabla I(v_i)$ is the image gradient at point v_i , E_{int} represents the internal energy term, and E_{ext} represents the external energy term.

Several works have applied the ACM for ROI segmentation in microscopic images. These works are developed for cell segmentation [131], [132], nuclear and glandular structures segmentation [133], [134]. Although these techniques are reported to provide good performance, the results are very sensitive to the initialization of ACM and local intensity variations in images. In addition, ACMs are known to have high computational complexity.

2.5.2 Deep learning-based image segmentation

With the introduction of deep neural networks in recent years, image segmentation applications have experienced remarkable performance improvements, often achieving the highest accuracy rates on popular benchmarks, resulting in what many regard as a paradigm shift in the field [135]. The deep learning image segmentation models usually consist of two major components: 1) encoder and 2) decoder. The encoder is usually a CNN, which uses convolutional filters to extract features of the input image (*i.e.*, feature maps) and pooling layers to down-sample the feature maps into more comprehensible information for the network. The decoder part is responsible for generating a segmentation map, which corresponds to the target region in the input image. The decoder up-samples the feature maps generated by the encoder in multiple steps, outputting a segmentation map that is the same size as the input image.

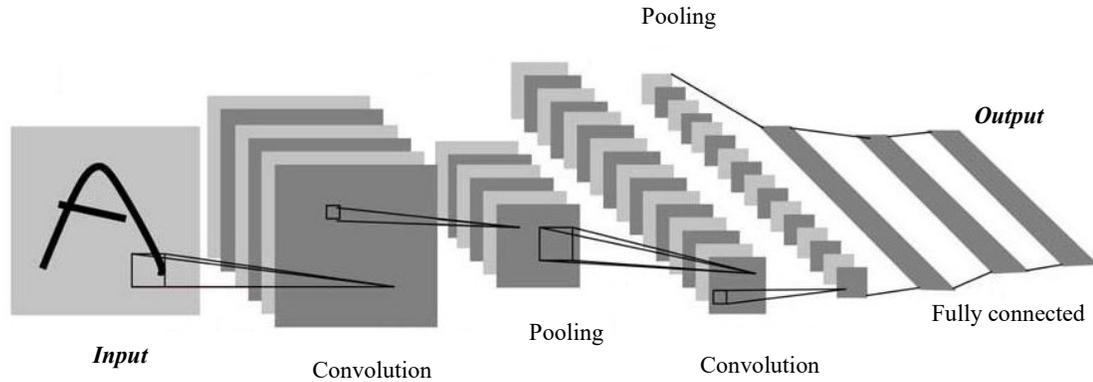


Figure 2.9 A typical convolutional neural network [139].

In this section, our aim is to 1) review some of the key concepts in deep learning based image segmentation, and 2) introduce some of the prominent segmentation algorithms, which have been widely used in AHIA.

2.5.2.1 Convolutional neural networks (CNNs) overview

CNNs are among the most successful architectures in the deep learning community, especially in image processing and computer vision applications [136]. The concept of CNN was originally proposed by Fukushima *et al.* [137], which was based on Hubel and Wiesel's work on simple and complex cells in the hierarchical receptive field model of the visual cortex. Inspired by Fukushima's work, the first work on modern CNNs was proposed in 1990s by LeCun *et al.* [138] in their work on the classification of handwritten digits. In their work, the authors demonstrated that a CNN model, which aggregates simpler features into progressively more complicated features, could be successfully used for handwritten digit recognition. Throughout the past several years, CNNs have achieved excellent performance describing natural images, performing facial recognition [139], and analyzing medical images [140].

A typical CNN architecture comprises of three main components (Figure 2.9): 1) convolutional layer, 2) nonlinearity layer, and 3) pooling layer. The convolutional layer consists of a kernel (also known as a filter) to extract features by performing convolution operations. The nonlinearity layer is responsible for applying an activation function (*e.g.*, sigmoid, ReLU, and tanh) on extracted feature maps in order to introduce nonlinear functionality to the network. The pooling component enables the spatial resolution of the feature maps to be reduced by replacing a small neighborhood of a feature map with statistical information (*e.g.*, mean, max, etc.) about the neighborhood. The units in layers are locally connected; each unit receives weighted inputs from a small neighborhood, known as the receptive field, from units in the previous layer. By stacking layers to form multi-resolution pyramids, the higher-level layers learn features from increasingly wider receptive fields. The main computational advantage of CNNs is that all the receptive fields in a layer share weights, resulting in a significantly smaller number of parameters in comparison to fully connected neural networks. Some of the most well-known CNN

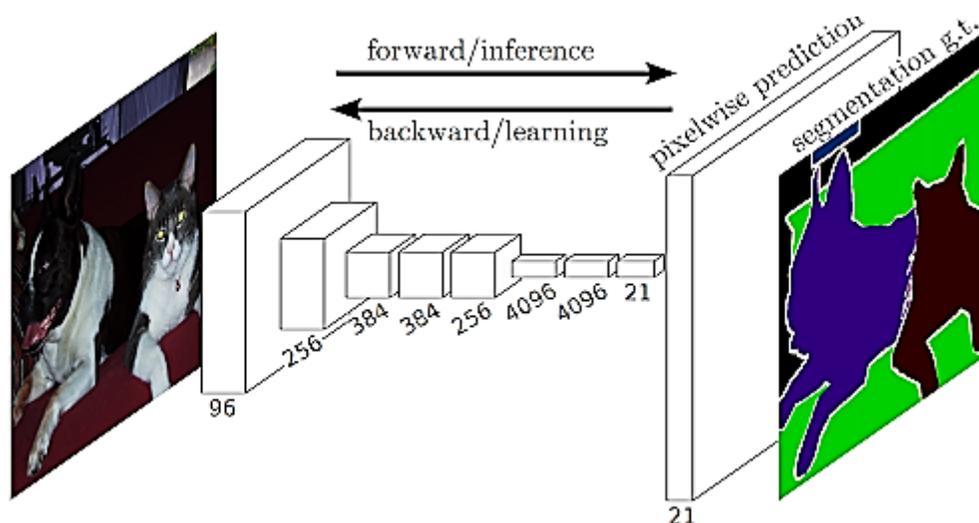


Figure 2.10 The original architecture of the fully convolutional neural network, proposed by [148]

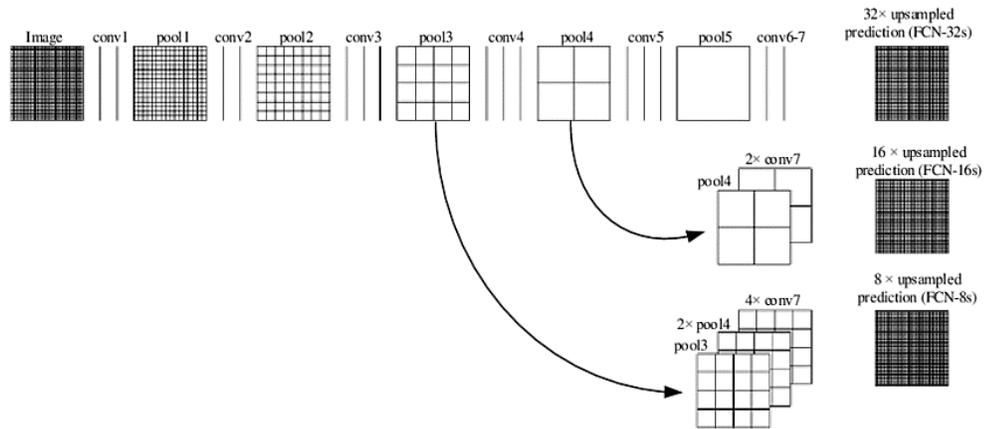


Figure 2.11 Skip connections used in fully convolutional neural network architecture [148].

architectures are AlexNet [141], VGGNet [142], ResNet [143], GoogLeNet [144], MobileNet [145], and DenseNet [146].

Here, we reviewed CNN as one of the main components of deep learning-based image segmentation models. For the rest of this section, we introduce some of the well-known deep learning-based image segmentation models.

2.5.2.2 Fully convolution neural networks

As illustrated by Long *et al.* [147], a fully convolution neural network (FCN) is a typical deep learning based semantic segmentation architecture (Figure 2.10), which is comprised of only convolution layers. The FCN is capable of receiving an arbitrary size input image and producing a segmentation map of the same size. The authors in [147] proposed modified CNN architectures based on VGG16 and GoogLeNet, to manage non-fixed sized input and output by replacing all fully-connected layers with convolutional layers. As a result, the outputs of the model are semantic segmentation maps rather than classification scores.

One of the revolutionary characteristics of the FCN is the use of skip connections [147] in the CNN architecture. In every CNN model, a fraction of the information is naturally

lost due to the pooling of the activations in each layer. Through the use of skip connections, in which feature maps from the final layers of the model are up-sampled and fused with feature maps of earlier layers (Figure 2.11), the model combines semantic information (from deep, coarse layers) and appearance information (from shallow, fine layers) in order to compensate for the loss of information in the previous layers.

FCNs have been applied to a variety of segmentation problems, including brain tumor segmentation [148], instance-aware semantic segmentation [149], skin lesion segmentation [150], and histopathology image segmentation [20].

2.5.2.3 Regional convolutional network based models

The regional convolutional network (R-CNN) and its extensions (Fast R-CNN, Faster R-CNN) are among the most successful deep learning models. Although the R-CNN family can be categorized among image detection algorithms, they gave rise to mask-RCNN, which is one of the most successful deep learning based image segmentation algorithms in the field of computer vision. As such, in this section, we briefly introduce R-CNN and its extensions.

Some extensions of R-CNN have been used in state-of-the-art real-world applications, including histopathology image analysis [151]. The R-CNN was originally proposed by Girshik *et al.* [152] for object recognition tasks (see Figure 2.12). The R-CNN architecture is comprised of two main steps: 1) selective search, and 2) feature extraction. In the first step, the algorithm extracts a manageable number of bounding-boxes (approximately ~2,000) around the candidate ROIs. Then, it independently extracts CNN features from each bounding-box. The important limitation of the R-CNN is that the R-CNN model is

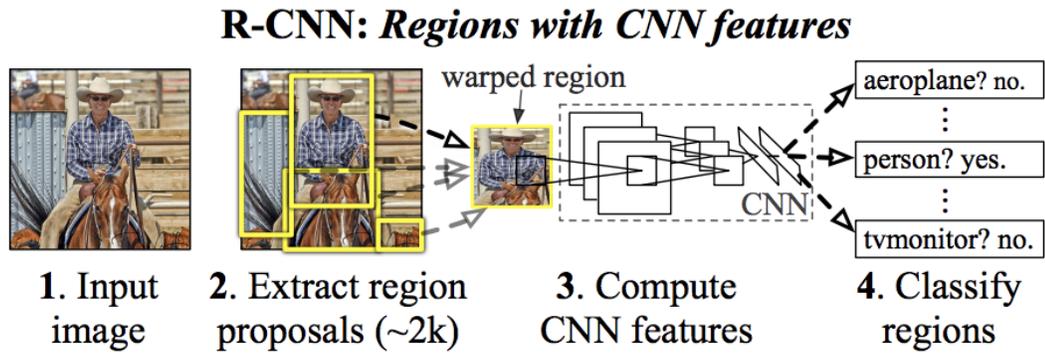


Figure 2.12 The R-CNN architecture [153].

extremely computationally expensive and slow to train. Therefore, in main extensions of the R-CNN (*i.e.*, Fast R-CNN [153] and Faster R-CNN [154]), the computational efficiency of the algorithm were the main priority. In particular, the Faster R-CNN architecture (see Figure 2.13), a systematic search unit, called region proposal network (RPN), is applied to reduce the number of analyzed bounding-boxes.

In one extension of this model, He *et al.* [155] proposed a Mask R-CNN for object instance segmentation, which outperformed all previous benchmarks on many COCO challenges (COCO is a public natural image recognition dataset provided by Microsoft with over 200k annotated images[156]). This model efficiently detects objects in an image

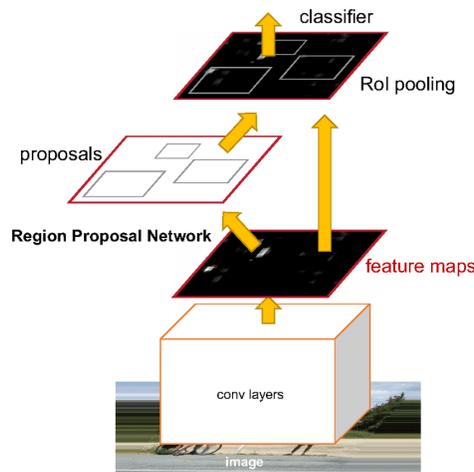


Figure 2.13 An illustration of Faster R-CNN [155].

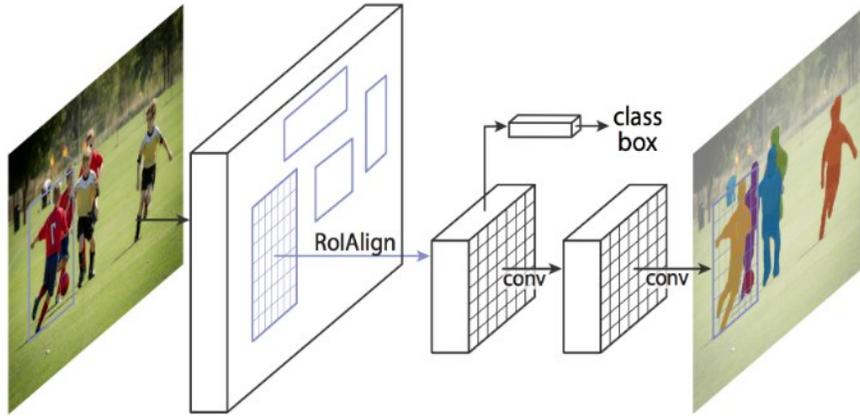


Figure 2.14 Instance segmentation using Mask R-CNN [156].

while simultaneously generating a high-quality segmentation mask for each instance. Mask R-CNN is essentially a Faster RCNN with three output branches (see Figure 2.14). The first computes the bounding box coordinates, the second computes the associated classes, and the third computes the binary mask to segment the object. The Mask R-CNN loss function combines the losses of the bounding box coordinates, the predicted class, and the segmentation mask, and trains all of them jointly.

2.5.2.4 Medical image segmentation models

There are several models initially developed for medical/biomedical image segmentation models, which are mostly inspired by FCNs. One of the very first deep learning based image segmentation models in biomedical image analysis was developed by Ciresan *et al.* [157], which used a sliding window CNN as a pixel classifier to segment neuronal membranes in patches of electron microscopy images. Although the authors were able to propose a segmentation method with human-level error rate, the proposed approach involves a trade-off, whereby smaller patches sacrifice contextual information for location

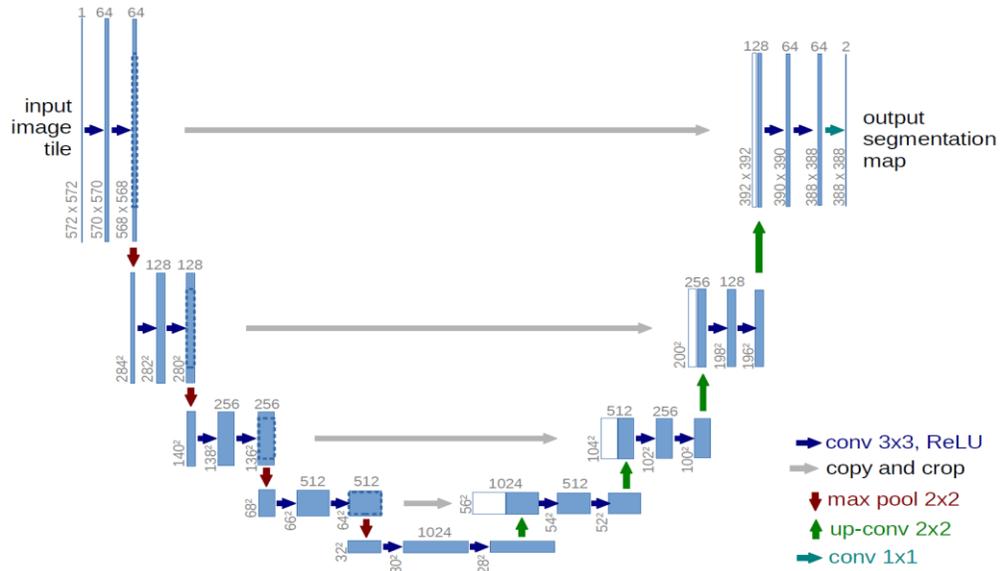


Figure 2.15 The U-net architecture. The Blue and white boxes represent the multi-channel feature maps and copied feature maps, respectively [159].

accuracy and vice versa. To resolve this, Ronneberger *et al.* [158] presented a more elegant network architecture, referred to as U-Net, which comprises of two parts, a contracting path to capture context, and a symmetric expanding path that enables precise localization (see Figure 2.15). In their proposed FCN-based architecture, 3×3 convolutions were used to extract features in the contraction path. In the expanding path, deconvolution operations were used to reduce the number of feature maps and increase their dimension. An important aspect of the original U-Net architecture was the use of skip connections (*i.e.*, concatenating connection) at each expanding path, which enabled the network to compensate for the loss of information in the down sampling (contracting) path. The final layer in the U-Net architecture were a 1×1 convolution of the feature maps to generate a segmentation map of the input image. The original U-Net won the ISBI cell tracking challenge in 2015 by a large margin ($\sim 9\%$ difference over PhC-U373 dataset [159] and $\sim 31\%$ over DIC-HeLa dataset [152]). Since then, it has become one of the popular segmentation approaches in different applications. As an example, Cicek [160] proposed a

U-net architecture for 3D images. Zhou *et al.* [161] developed a nested U-Net architecture. U-Net has also been applied to various other problems. For example, Zhang *et al.* [162] developed a road segmentation/extraction algorithm based on U-Net.

Despite the success of deep learning-based methods in computer vision applications, such as AHIA, there are certain challenges that cannot be addressed through conventional CNNs. An important bottleneck in AHIA is the trade-off between field-of-view and resolution with respect to the patch size, which dramatically affects the segmentation performance, particularly in an application where the size difference between biological structures is large. In literature, this issue has been undertaken through multi-resolution approaches, which is reviewed in detail in section 2.5.3.

2.5.3 Multi-resolution semantic segmentation

Histopathology WSIs are usually large (*e.g.*, 80,000×80,000 pixels) and have high-resolution (*e.g.*, 20× magnification, corresponding to 0.5 μm^2 per pixel), which enables the WSI to contain a detailed representation of biological structures with extreme size differences. As an example, in high-resolution histopathology WSIs of lung tissue, there may be blood vessels as small as 24 μm^2 and bronchi as large as approximately 1.2 mm^2 . The very large size of the histopathology WSIs usually prevents direct utilization of the input WSIs in CNN models due to memory limitations and long training time. As such, a standard practice in AHIA using CNNs is to use a patch-based approach by dividing the input WSIs into smaller patches (*e.g.*, 256×256 pixels). In patch-based CNNs, dividing the input WSI into patches can be potentially problematic because of the incompatibility between the size of the extracted patches and the size of the target biological structures. If the patches are too small, there may be insufficient contextual information available due to

the limited field-of-view for the CNN to make accurate predictions, particularly in the segmentation of large biological structures. On the other hand, if patch sizes are too large, the computational efficiency of the CNN model dramatically decreases, which may result in insufficient memory and training speed. To resolve this trade-off between field-of-view and computational efficiency, methods may reduce the spatial resolution of the patches to provide a sufficient field-of-view, while keeping a reasonable patch size to maintain the computational efficiency of the training process. However, reducing the spatial resolution of the patches may lower the performance of the CNN in the segmentation of small structures. To resolve this three-fold trade-off between field-of-view, computational efficiency, and spatial resolution, a multi-resolution approach is often utilized. Here, we briefly review some of the recent multi-resolution approaches.

Multi-resolution analysis of histopathology WSIs in AHIA mimics a pathologist approach towards the assessment of histopathology WSIs slides under microscope. Pathologists usually make a diagnosis by analyzing WSIs at different magnifications. This approach enables them to check a wide range of texture patterns in a low-magnification image, whereas they use high-magnification images to check details, such as the shapes of individual cells that are too small to be clearly visible in the low-resolution images. A similar strategy has been adopted in developing deep learning methods, enabling the models to learn high-level features of the input WSI at low-magnifications as well as details of the small biological structures, which can potentially boost the overall performance of the deep learning model.

Tokunaga *et al.* [163] proposed an adaptive weighting multi-field-of-view CNN for semantic segmentation of adenocarcinoma in lung histopathology WSIs. The overview of

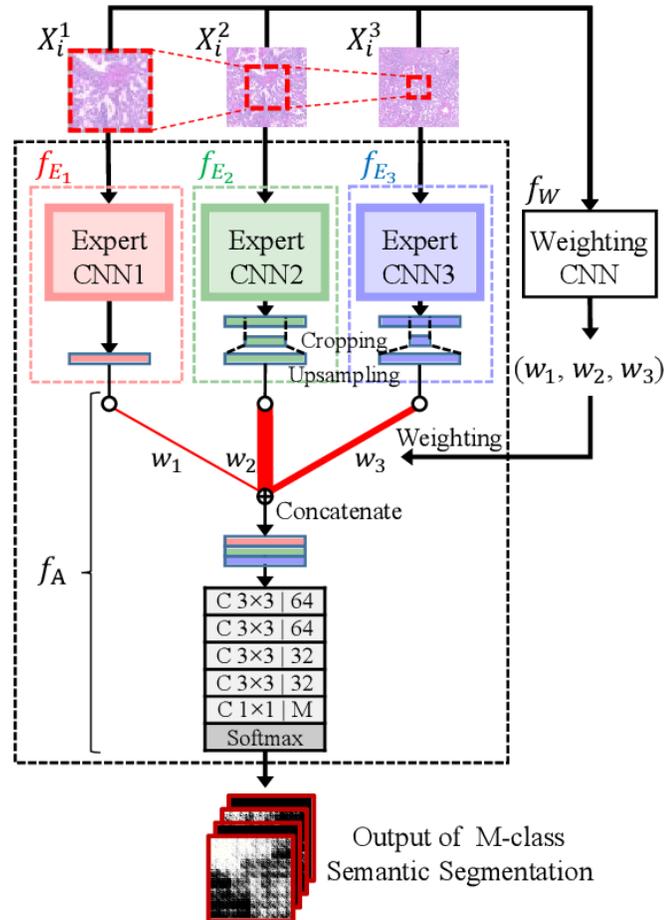


Figure 2.16 The adaptive weighting multi-field-of-view CNN architecture for semantic segmentation in histopathology whole slide images [164].

their proposed method is illustrated in Figure 2.16. In their multi-resolution approach, the authors used patches of size 256×256 at three different magnifications (*i.e.*, $20\times$, $10\times$, and $5\times$), where at each magnification, each patch has a different resolution and field-of-view. The input patches were used to train three separate CNNs, each of which is an expert in the extraction of contextual information from its respective magnification, generating a heat-map for each input patch. To improve the overall performance of the segmentation model, the authors used a weighting CNN to adaptively weight each heat-map. A simple CNN is used to aggregate the weighted contextual information to generate the final segmentation maps. Their proposed method achieved 82.1% and 53.6% mean intersection-over-union,

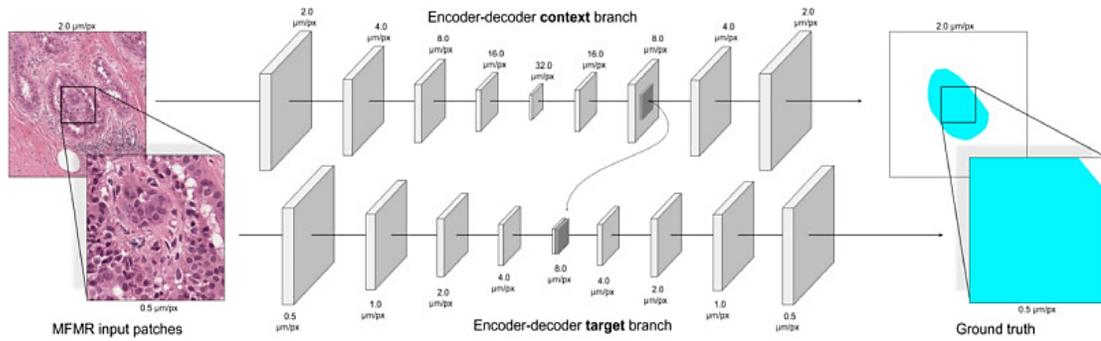


Figure 2.17 Histopathology image segmentation using HookNet architecture [166].

which showed 3% and 19% improvement against Hard-Switch-CNN [164] method for single-class and multi-class segmentation tasks, respectively.

Rijthoven *et al.* [165] proposed HookNet for semantic segmentation of histopathology WSIs via multiple branches of encoder-decoder CNN. They used two different datasets to evaluate their work; 1) histopathology WSIs of breast cancer and 2) histopathology images of tertiary lymphoid structures of the lung. The overview of their proposed method is shown in Figure 2.17. The authors trained multiple CNNs at different resolutions (0.5, 1.0, 2.0, 4.0, and 8.0 $\mu\text{m}/\text{pixel}$) using patches of size 284×284 . A hooking mechanism was implemented to effectively combine the feature maps generated by different CNNs, improving the overall performance of the model. Their proposed method achieved F1-Score of 0.91 and 0.72 for semantic segmentation in breast and lung datasets, respectively.

Joon Ho *et al.* [166] proposed a deep multi-magnification network for semantic segmentation cancer types in histopathology WSIs of breast tissue. The proposed segmentation model consists of three parallel encoder units that simultaneously extract feature maps from the input patch and a decoder unit, which generates the final segmentation heat-maps. To improve the segmentation performance, the network concatenates the extracted feature maps of the intermediate layers by the encoders, providing the model with richer contextual information regarding the resolution and field-

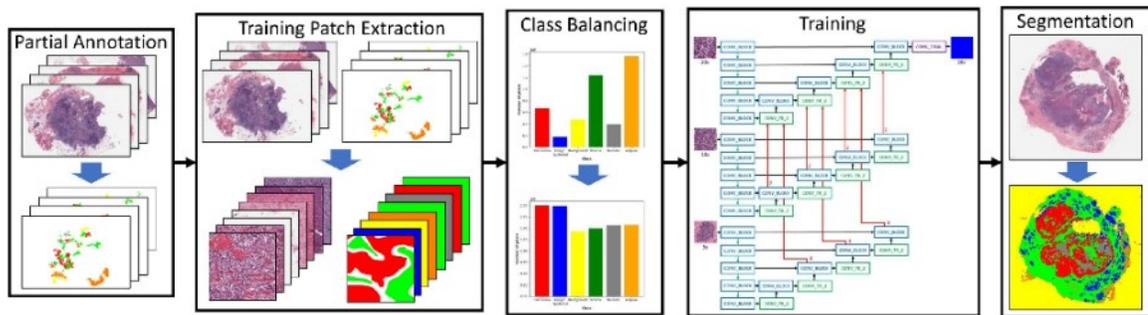


Figure 2.18 Deep multi-magnification CNN for semantic segmentation of cancer types in histopathology images of breast [167].

of-view (see Figure 2.18). To evaluate the segmentation model, the authors used two separate datasets: 1) WSIs of triple-negative breast cancer containing high-grade invasive ductal carcinoma and 2) WSIs from lumpectomy and breast margins containing invasive ductal carcinoma and ductal carcinoma in situ of various histologic grades. Their proposed model was able to obtain 0.87 and 0.71 mean intersection over union over the first dataset and second dataset, respectively. The authors also compared their model to U-Net architecture, which showed 7% improvement over the first dataset and fell short by 2% over the second dataset.

2.6 Histopathology image segmentation applications

The emergence of WSI techniques, which enabled the acquisition and storing of high-resolution digitized scans of histopathological slides, has opened a window of opportunity for histopathological studies to take advantage of computerized analysis more than ever before. As a result, there is a notable increase in research interest in different applications of AHIA, including color and illumination normalization [167]–[169], automated detection/segmentation of histopathology images [170]–[172], feature extraction [173]–[175], classification and subcellular quantification [173]–[177]. As a gateway to most AHIA applications, the development of sophisticated and comprehensive detection and

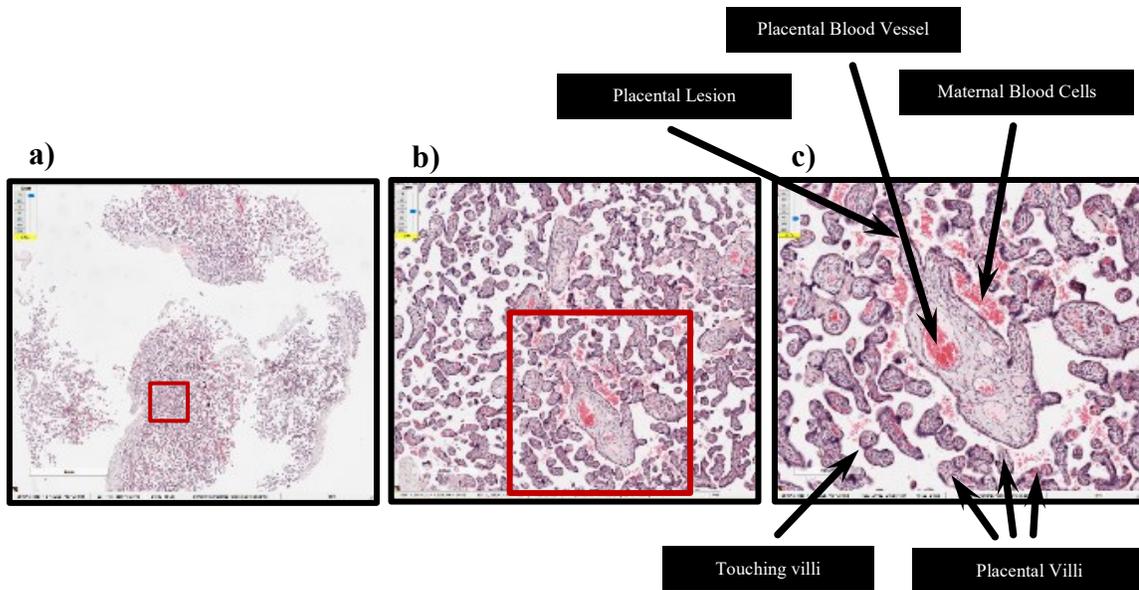


Figure 2.19 Different biological components of a placenta image. a) An example of a histopathology whole slide image of placenta. b) 5× magnification. c) 10× magnification.

segmentation algorithms have a deciding role in most histopathological studies. In this thesis, we will be focusing on two distinct segmentation applications in AHIA: 1) segmentation of villi structures in histopathology images of human placenta and 2) segmentation of biological structures in histopathology images of mouse lung tissue for quantification of MLI score.

2.6.1 Segmentation of villi structures in histopathology images of human placenta

The human placenta is a pregnancy related organ, which is responsible for exchanging gases, nutrients and waste elimination. Analysis of placental structures, such as placental villous structures and placental lesions (*i.e.*, abnormal regions in placental villi), are known to play an important role in the diagnosis and treatment of PMDs, including fetal intrauterine growth restriction, pre-term birth, stillbirth, preeclampsia, and have also been linked to long-term health outcomes, such as premature cardiovascular disease in mothers and metabolic disturbances in offspring [178]. Different biological components in a high-resolution placental histopathology image are visualized in Figure 2.19. Typical ROIs in

placental histopathology images, where pathologists analyze to assess the placental histopathology images, are usually the villous structures. As such, in developing AHIA methods for placental WSIs, a common objective is to acquire the boundary of the villous structures among other biological components in the WSI. However, this can be a challenging task due to the complexity and variability of the placental structures, specifically when the dense appearance of the biological structures can be seen in the WSIs.

In our first AHIA segmentation application, we aim to present an accurate and fully automated approach for segmentation of villi, which is able to account for main challenges in the analysis of histopathology images of human placenta, including identification of touching villi specimens, accounting for a wide range of imaging and biological artifacts, and erratic variability of placental structures.

2.6.2 Segmentation of biological structures in histopathology images of mouse lung tissue for quantification of MLI score

Analysis of injuries in histopathology images of mouse lung tissue has a central role in clinical research towards diagnosis and treatment of respiratory illnesses, such as Bronchopulmonary dysplasia (BPD), which is the most common complication of preterm birth [37]. A common metric for quantification of injury in lung histopathology images is MLI score, which represents the mean distance between alveolar septa within the lung. Currently, manual and semi-automated assessment of lung structures under microscope is the acceptable approach for the calculation of MLI score in mouse lung WSIs, which is inefficient, expensive, and subject to inter- and intra-rater variability. As such, using AHIA can potentially pace the clinical research in this area; however, similar to placental WSIs, developing AHIA methods in lung histopathology WSIs involve unique challenges due to

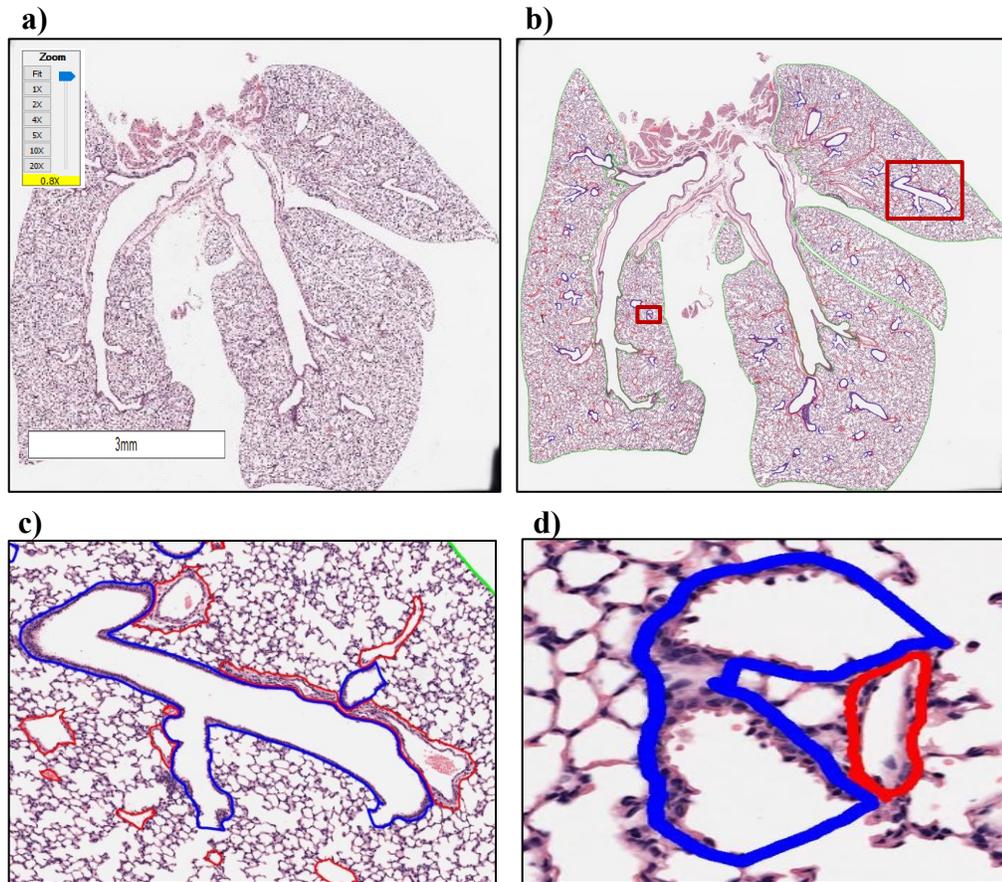


Figure 2.20 a) An example of a histopathology whole slide image of mouse lung tissue. b) The manually annotated image in a). Different structures are marked with color green, blue, and red, which represent diagnostically relevant tissue compartments, bronchi, and blood vessels, respectively. c) A sample extracted patch with 5 \times magnification. c) A sample extracted path with 20 \times magnification.

the complexity and variability of the biological lung structures. A typical high-resolution histopathology image of mouse lung tissue is shown in Figure 2.20.a. As shown in Figure 2.20.b, considerable regions of the lung tissue consists of imaging artifacts and undesired components (*i.e.*, outside of the area marked in green). An interesting observation that can be drawn from Figure 2.20.b is that there is a dramatic difference in the size of the biological structures (see Figure 2.20.c and Figure 2.20.d). This may be problematic in developing automated segmentation approaches, especially deep learning approaches, where it is required to extract multiple image patches for the segmentation model. As a

result, limited field-of-view of the image patches may affect the performance of the segmentation approach.

In this application, we are interested in automating the assessment of histopathology images of mice lungs to estimate the MLI score. The developed AHIA approach requires accurate segmentation of different biological structures of mouse lung tissue, which plays a key role in MLI estimation.

Chapter 3: Automated segmentation of histopathology images of human placenta

In this chapter, an automated pipeline for segmentation of placental villous structures in histopathology WSIs of human placenta is presented. Conventional image processing and machine learning approaches are applied to extract foreground regions (*i.e.*, villi) from background regions (*i.e.*, non-villi, including imaging artifacts and undesired biological components). The proposed pipeline is capable of identifying the villous instances, which is acquired by detecting the touching villous structures. This chapter has been published as Salsabili, S., *et al.* “Automated Segmentation of Villi in Histopathology Images of Placenta.” *Computers in Biology and Medicine*, vol. 113, 2019, p. 103420, doi: <https://doi.org/10.1016/j.combiomed.2019.103420>

3.1 Introduction

The human placenta is a vital organ of pregnancy, exchanging all gases, nutrients, and waste products between the mother and developing fetus. This complex organ is essential for fetal survival. Developmental abnormalities and/or in utero damage to this organ can have detrimental short- and long-term health effects for both the mother and her fetus. Placental damage or dysfunction is known to play a central role in the development of fetal intrauterine growth restriction, pre-term birth, stillbirth, preeclampsia, and has also been linked to long term health outcomes, such as premature cardiovascular disease in mothers and metabolic disturbances in offspring [179]–[186]. The effective application of placental histopathology following a poor pregnancy outcome can inform ongoing clinical management of mothers and neonates, and offer insight into etiology of placental disease, the risk of recurrence in future pregnancies, and may provide indications regarding which

mothers and offspring are at highest risk of cardio-metabolic disease later in life [187]–[192]. However, unlike some other fields of clinical pathology, there are a limited number of perinatal pathologists who are highly specialized in the field of placental histopathology, and the quality and reproducibility of placental pathology examinations are currently very poor[193]–[195]. Collectively, this has limited the clinical utility of this diagnostic/prognostic tool. The ability to apply an unbiased quantitative approach to the field of placental histopathology would substantially reduce the high degree of inter-observer variability and quality control issues currently plaguing this field [196]–[198], allowing for highly reproducible and reliable findings from this clinical modality. It is anticipated that the contextually rich information that could be gained from automated forms of placental histopathology would allow for a highly integrated use of placenta pathology findings in the continuum of care – allowing for effective translation of findings from the microscope to the clinical management of mothers and babies.

In recent years, there has been increased interest in and research addressing automated detection and analysis of various features within the complex architecture of the human placenta [29], [30], [199]–[202]. The chorionic villi are the basic functional units of the placenta, anatomically described as a branched “tree-like structure” covered in a multinucleated layer of syncytiotrophoblast cells, which encases the fetoplacental vasculature, embedded within a non-cellular connective tissue core. These villi structures are bathed in maternal blood, found within the intervillous space, *in vivo*. Previous work has focused on automated detection of the fetoplacental vascular space within the villi – as poor fetoplacental vascular development has been linked to several obstetrical diseases. Almoussa *et al.* [200] used a segmentation approach based on artificial neural networks

(ANNs) to automatically extract feto-placental blood vessel features from digital histological images of the placenta. The ANNs were successful in detecting these most prominent vascular spaces within the placental villi. Chang *et al.* [201] proposed an automatic filtering method, which locally detects pixels containing curvilinear structures and reduces non-vessel noise. Compared to ANN-based methods, Chang *et al.* [201] proposed a faster and more accurate approach for feto-placental vessel detection.

A fuzzy C-means clustering method was applied successfully to distinguish cellular vs. extracellular components of the chorionic villi and to identify areas normally filled with maternal blood (intervillous space) [202]. Kidron *et al.* [29] used ImageJ software (<https://imagej.net/>) to extract features, such as size and number of chorionic villi, from histological images of the placenta and tested the feasibility of automated diagnosis of delayed or accelerated villous maturation. However, the images analyzed appear to be selective (*e.g.*, there was no indication that artifacts in the histological images were present). In addition, this method does not appear to properly segment villi that are in very close proximity, touching, or overlapping (*e.g.*, Figure 3.1 in Kidron *et al.* [29] shows touching villi identified as a single villous structure). Swiderska-Chadaj *et al.* [30] described a method of automatic segmentation of placental villi structures for assessment of edema within the placenta. These authors used texture analysis, mathematical morphology, and region growing operations to extract different structures from placental images. Although they presented a comprehensive pipeline for automated analysis of placental histology, a small sample size of 50 villi from placentas with a variety of pathologies was used, with selective/optimal image selection (*e.g.*, no image artifacts).

The objective of this paper is to describe an image analysis pipeline for automated analysis and segmentation of chorionic villi structures in histopathology images of human placentas. The main contribution of this work is the development of a new automated, segmentation protocol, which operates with a high degree of accuracy over large sample size. We also address the issue of adjacent, touching, and overlapping chorionic villi in our work, a common clinical finding that has not been properly addressed in the literature. The proposed method is validated on a set of healthy control placentas, as well as those complicated with the placenta-mediated disease of preeclampsia, to ensure the developed algorithm performs effectively for analyzing placenta specimens from healthy and diseased subjects. We also compare our work with a previously published automated method to detect villi in histological images of the placenta [29]. We apply their provided source code to our dataset and compare the results to our method.

3.2 Methodology

3.2.1 Placenta Histopathology Images

Our dataset comprises high-resolution digital scans of 12 placental histopathology specimens obtained from the Research Centre for Women’s and Infants Health (RCWIH) Biobank (Mount Sinai Hospital, Toronto, ON). The ethics approval to perform sub-analyses on the Biobank samples was obtained from the Ottawa Health Science Network Research Ethics Board and the Children’s Hospital of Eastern Ontario (CHEO) Research Ethics Board.

The placental specimens were fixed in paraffin wax, stained with hematoxylin, washed in a 0.3% acid alcohol solution, and counterstained with eosin following the standard protocol for hematoxylin and eosin (H&E) staining at the Department of Pathology and

Laboratory Medicine at the CHEO. Slides were scanned (Aperio Scan Scope), and high-resolution color images (20× magnification) were obtained. In total, 12 placental scans were generated from three healthy term placental specimens and nine placental specimens collected from term pregnancies complicated with preeclampsia (PE). From each placental scan, three sample images of size 2740×3964 pixels were extracted, amounting to a total of 36 sample images (nine images from healthy term pregnancies and 27 from term pregnancies complicated with PE).

Manual segmentation of these images was performed using ImageJ software by A.M. and verified by D.G. (a clinical perinatal pathologist). For each sample image, a “chorionic villi mask” was manually generated by A.M (verified by D.G), identifying individual villi structures and artifact objects. These masks were generated by manually drawing a line around the area of each target object (*i.e.*, villi or undesired objects). The manual segmentation provided an estimation of adjacent villi boundaries and boundaries of undesired objects (*i.e.*, objects that are not considered villi). The manual segmentation served as a ground truth for training and evaluation. From the 36 sample images, there were 4946 villi in total.

3.2.2 Image Segmentation

The image processing pipeline for villi segmentation in a placental image is shown in Figure 3.1. The various steps are discussed in detail below.

3.2.2.1 Image preprocessing

A number of factors can contribute to variations in the color content of histological images (e.g., histochemical staining time, amount of histology stain used, etc.). We applied a comprehensive color normalization [203] to input images to compensate for such

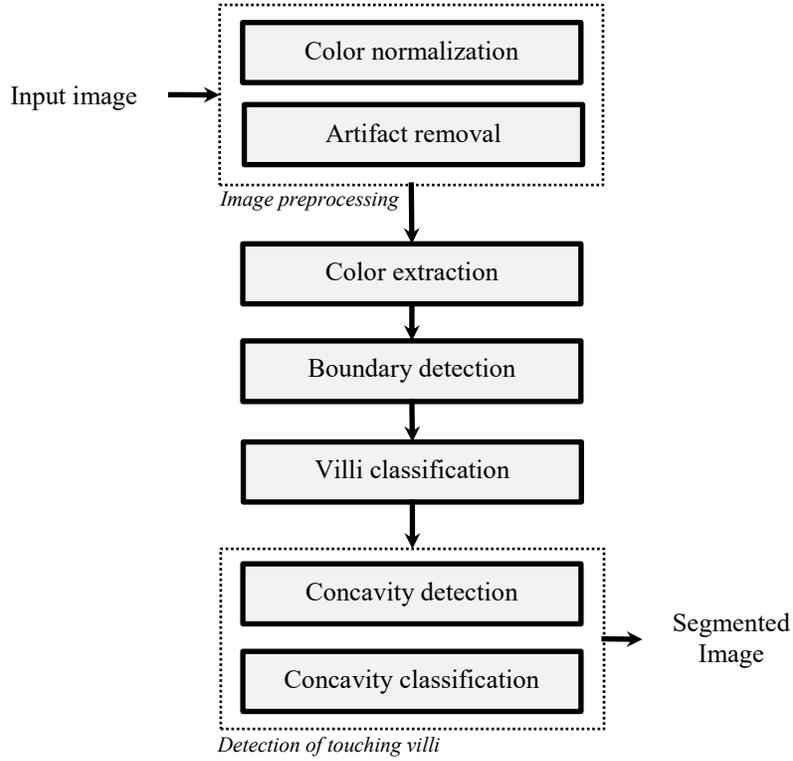


Figure 3.1 Block diagram of our methodology.

variations. Imaging artifacts might also exist in the histological images. In our work, imaging artifacts appeared as darkened areas of the image, and therefore needed to be removed prior to segmentation of the villi boundary. In Figure 3.4, examples of these imaging artifacts are identified with blue arrows. Using Equation 3.1, pixels that are part of the imaging artifacts were identified; the intensity of these pixels was then adjusted to be white.

$$\left\{ \begin{array}{l}
 bw_1(k, l) = |I_i(k, l) - I_j(k, l)| \leq C1 \\
 bw_2(k, l) = |I_i(k, l)| \leq C2 \\
 BW_{artifact} = bw_1 \cap bw_2
 \end{array} \right. \quad 3.1$$

$$\left\{ \begin{array}{l}
 where \left\{ \begin{array}{l}
 i \neq j \\
 i, j \in (R, G, B) \\
 k \in (1, m), l \in (1, n)
 \end{array} \right.
 \end{array} \right.$$

In Equations 3.1, I is an $m \times n \times 3$ intensity matrix and
$$I_i = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{bmatrix}$$

represents the i^{th} matrix in RGB color space (i.e., red, green, and blue). The constant C_1 is a threshold to ensure that the intensity differences between the R, G, and B channels are small (i.e., detected pixel is essentially grey). The constant C_2 is a threshold to ensure the detected pixels are relatively dark. The values assigned to C_1 and C_2 were 5 and 200, respectively, which were determined empirically. The $BW_{artifact}$ is a binary mask, which is the intersection of the bw_1 and bw_2 binary masks and contains all of the pixels in I that are contaminated with artifacts, which are adjusted to be white.

3.2.2.2 Color extraction

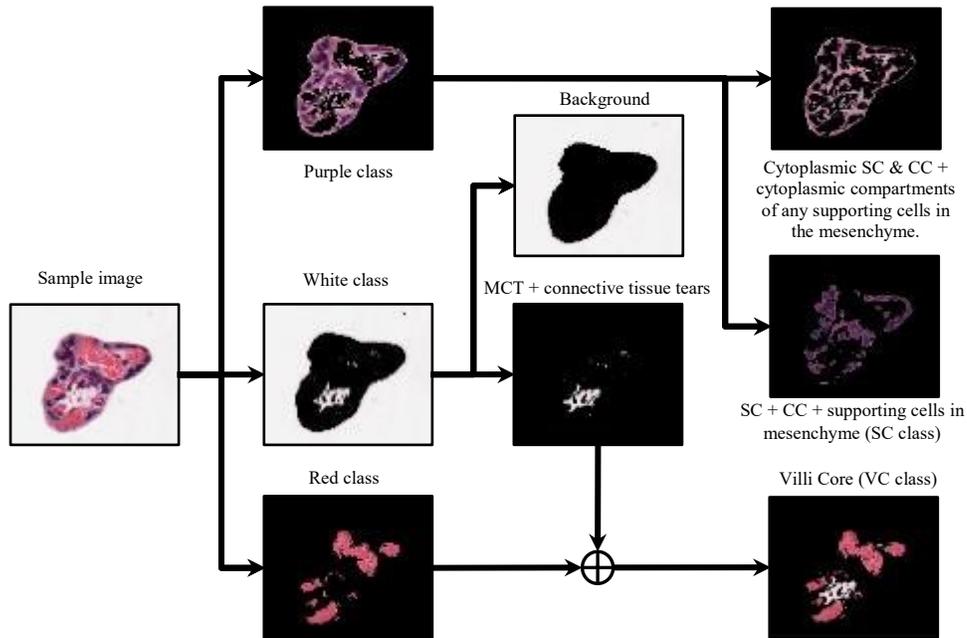


Figure 3.2 Classification of color content in H&E stained images, including: syncytiotrophoblasts cells (SC) and cytotrophoblasts cells (CC) in the purple class; non-cellular mesenchymal connective tissue (MCT) in the white class; and fetal red blood cells within the villi core vasculature in the red class.

Pixels in the histological images of the placenta were divided into three classes based on color: red, white, and purple. Examples of these color classes and their subdivisions can be seen in Figure 3.2. The first step in color classification is the extraction of the white class, found primarily outside of the villi, which is the maternal side of the placenta, called the intervillous space. This is deemed the background of the image. White spaces also exist within the core of some villi, primarily made of a non-cellular compartment of mesenchymal connective tissue (MCT) and/or tears in the connective tissue, which represent no specific tissue compartment. The white class is extracted using Equation 3.2:

$$\left\{ \begin{array}{l}
 bw_1(k, l) = |I_i(k, l) - I_j(k, l)| \leq C3 \\
 bw_2(k, l) = |I_i| \geq TH \\
 BW_{white} = bw_1 \cap bw_2 \\
 \\
 where \left\{ \begin{array}{l}
 i \neq j \\
 i, j \in (R, G, B) \\
 k \in (1, m), l \in (1, n)
 \end{array} \right.
 \end{array} \right. \quad 3.2$$

In Equation 3.2, TH is a threshold, whose value is determined using Otsu's method[204]. This threshold ensures that the extracted pixels have high intensities (close to white). The constant C3 is a threshold to ensure that the intensity difference between the R, G, and B channels is small. The value assigned to C3 was 15, which was determined empirically. The BW_{white} is a binary mask, which is generated by the intersection of bw1 and bw2 binary masks, and contains the white class pixels.

The second step was the extraction of the red class. The red regions represent areas rich in fetal red blood cells, found within the fetoplacental blood vessels encased within villi. In this work, all placental samples were washed several times during histological processing to remove blood from the tissue; however, some red blood cells often remain in each specimen. The combination of the red and white regions within the core of villi is

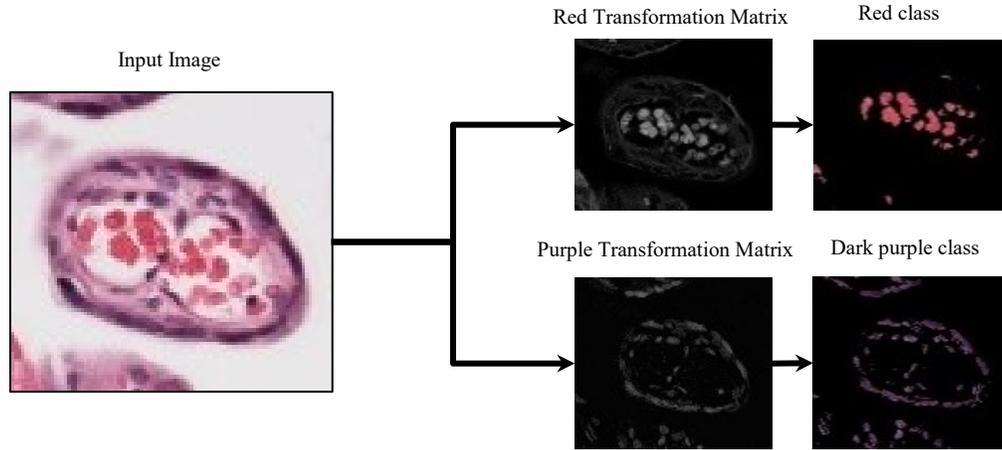


Figure 3.3 Example of the red and dark purple class color extraction.

referred to as the villi core (VC). The realization of the red class transformation is illustrated in Figure 3.3. A transformed matrix T is computed (Equation 3.3), which highlights the red class, and a threshold is applied to extract the red class. The value for this threshold “TH” is determined using Otsu’s method [204].

$$\begin{cases} T \triangleq I_R - \text{mean}(I_G + I_B) \\ BW_{red} = (T \geq TH) \end{cases} \quad 3.3$$

The remaining pixels are assigned to the purple class, which represents several other cellular compartments. The dark purple regions primarily identify the nuclei of the syncytiotrophoblasts cells (SC), cytotrophoblasts cells (CC), along with the nuclei of any supporting cells, which might be present in the mesenchyme connective tissue core (*i.e.*, fibroblasts, immune cells). SCs and CCs are densely packed along the villi borders. Nuclei of mesenchymal support cells, on the other hand, have a sparse appearance in the core regions of the villi. Therefore, the dense appearance of dark purple areas in non-border regions of a contour indicates that this is likely a border subdividing two villi in close

proximity within the contour. The light purple area identifies the cytoplasmic compartment of the SCs and CCs, along with mesenchymal support cells.

$$\begin{cases} T \triangleq I_B - \text{mean}(I_G + I_R) \\ BW_{SC} = (T \geq TH) \end{cases} \quad 3.4$$

The purple class is further classified into dark purple and light purple classes. For simplicity, we call the dark purple class the “SC class”. Equation 3.4 is used to extract the SC class, where T is the transformed matrix, and “TH” is a threshold, whose value is determined using Otsu’s method [204]. The remaining pixels are assigned to the light purple class.

3.2.2.3 Boundary detection:

After extraction of the color content, the input image was separated into the background (*i.e.*, white class) and foreground (*i.e.*, all classes excluding the white class). The foreground can be adopted as a rough estimate of the boundary of the objects in the image. The binary mask from the foreground was used as initial contours in the level set algorithm, based on the Chan-Vase formulation[205], to obtain the exact boundary of each contour. The algorithm accurately segments objects of interest (*i.e.*, villi candidates).

3.2.2.4 Villi classification:

To be considered “placental villi”, objects of interest needed to surpass a minimum threshold size of $1000 \mu\text{m}^2$. This threshold size was selected based upon well-established, healthy placenta villi dimensions previously described in the literature [206], [207]. This villi size threshold is the same as the one used in work by Kidron et al. [29], which is of relevance as we are comparing our method to their method. Objects of interest $< 1000 \mu\text{m}^2$

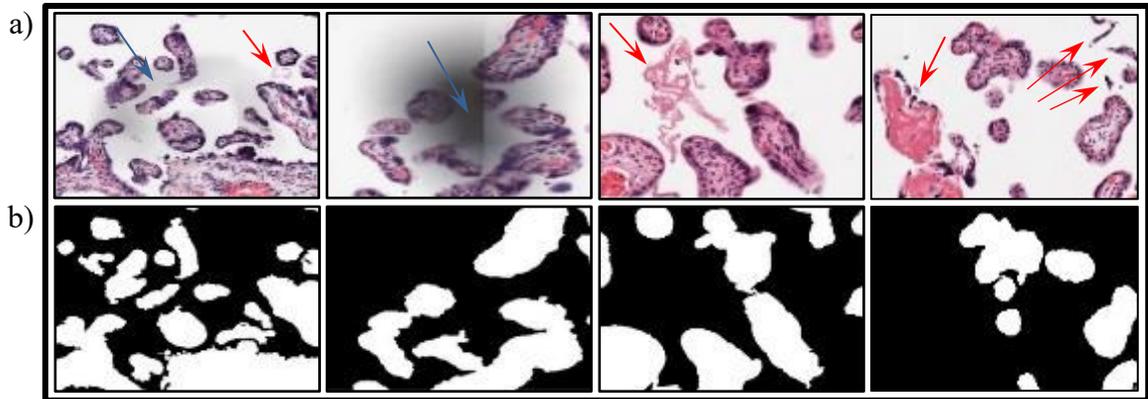


Figure 3.4 a) Examples of artifacts and undesired objects. The blue arrows show the areas that are affected by artifacts and the red arrows show the undesired objects. b) the villi classification steps output that shows the removal of artifacts and undesired objects.

might include red and white blood cells found in the intervillous space, fibrin deposition, syncytial knots (pieces of apoptotic placenta shed from the villi surface), or artifacts from tissue processing and histological preparation.

It is important to note that there are undesired objects of interest $> 1000 \mu\text{m}^2$, as demonstrated in Figure 3.4.a. Therefore, further descriptors needed to be included in order to identify true placenta villi objects versus undesired objects. Mature placenta villi contain white and/or red regions within their core, indicative of the presence of connective tissue and fetal vasculature and/or fetal red blood cells (VC class). Moreover, placental villi are surrounded by a continuous layer of SC cells and a sparse underlying layer of CC cells, all of which contain dark purple nuclei (*i.e.*, SC class). Conversely, undesired objects do not have these anatomical features and therefore demonstrate a considerably different density profile for the SC and VC classes. The SC and VC class density thresholds were tuned over the training set, using density profiles of the SC and VC class of true placenta villi objects (ground truth) as identified through manual segmentation.

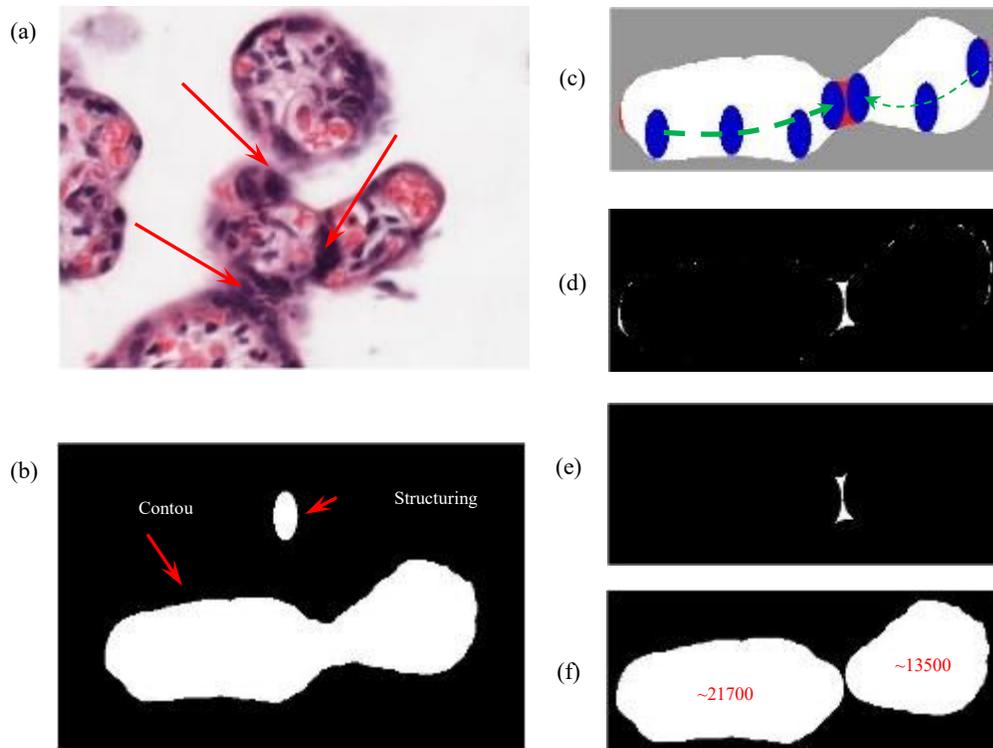


Figure 3.5 (a) Examples of attachment areas in villi. (b) A sample contour and an elliptical-shape structuring element to be used in calculation of top-hat transform. (c) The interpretation of the top-hat transform by fitting the SE inside the contour. The areas that cannot be covered by SE are shown in red. These areas are the output of top-hat transform (d) The output of the top-hat transform, which represents all of the convexities and concavities in the contour. (e) The candidate concavity, which is the only concavity in the contour that its removal divide the original contour into two separate contours. (f) The resultant contours of removing the candidate concavity from the original contour. Both contour satisfied the minimum size threshold of a villous.

3.2.2.5 Detection of touching villi:

In this work, a novel approach was used to detect pairs of touching villi. The postulated boundary separating the two touching villi was identified using the presence of concavity at the site where the two villi meet, which is visually distinguishable from other concavities in a standard villous structure. Specifically, at these concave sites, you additionally see the dense presence of dark purple nuclei within the SC class. Figure 3.5.a shows a representative example of four placenta villi that are touching each other, with three red arrows highlighting the visual features of the concave sites of separation. In this work, villi separation boundaries were identified in two steps: 1) detection of candidate concavities

through contour analysis, and 2) acceptance or rejection of each candidate concavity as a separate villous boundary, through color analysis (*i.e.*, density of different color classes in the boundary area, and the density of color classes in the two resultant villi).

3.2.2.5.1 Detection of candidate concavities

The concavity detection step employs the top-hat transform, which is defined as the subtraction of the original contour from its opening with a structuring element (SE) [208]. This definition of top-hat transform can be interpreted as an attempt to cover the areas within the contour by fitting a SE into the contour, under the condition that the entirety of the SE remains within the boundary of the contour. The areas that cannot be covered by the fitted SE are the concavities/convexities in the contour. An example is shown in Figure 3.5.c, where the regions that are marked in red show convexities/concavities in the contour. The output of the top-hat transform (Figure 3.5.d) might contain the candidate concavity (Figure 3.5.e), which must satisfy two conditions: 1) the removal of the candidate concavity from the contour must divide the contour into two new contours, and 2) the size of each of these new contours must satisfy the minimum villi size threshold (*i.e.*, $1000 \mu\text{m}^2$). Figure 3.5.e shows the target concavity, which satisfies these two conditions.

Identifying candidate concavities using the top-hat transform depends on the characteristics of the SE (*i.e.*, size, shape, and alignment). Figure 3.6 shows the effect of varying the characteristics of the SE on the detection of candidate concavities. In our work, an elliptical SE with different sizes and orientations (*i.e.*, 0° , 45° , 90° , and 135°) is used to detect concavities in the contour.

The procedure of changing the characteristics of the SE is continued until a candidate concavity is identified, or the top-hat transform of the contour becomes equal to the contour

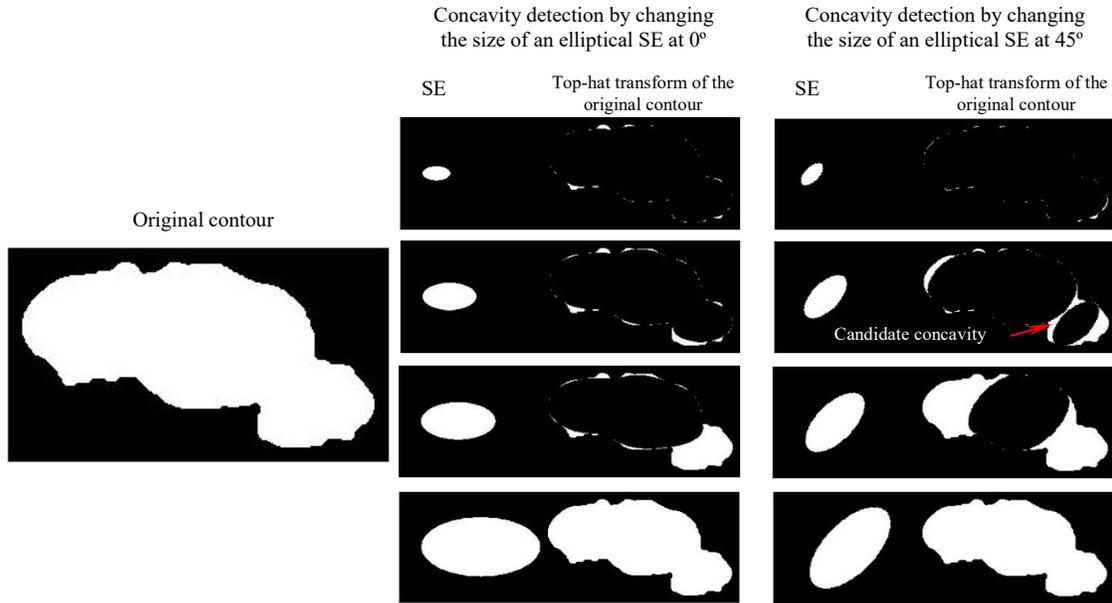


Figure 3.6 An example of detecting the candidate concavity by changing the characteristics of the SE. In this example, two rotation is considered for a given elliptical-shape structuring element (*i.e.*, 0° , 45°). The size of the SE is changed for each rotation to the point that the top-hat transform of the input contour becomes equal to the input contour itself. As it can be seen in this figure, the candidate concavity cannot be detected in 0° rotation. On the other hand, in 45° rotation increasing the size of the SE leads to appearance of the candidate concavity in the top-hat transform of the contour.

itself. This might happen if the size of the structuring element becomes large enough that the opening of the contour with the SE can become zero (*i.e.*, the contour is completely eroded by the structuring element). As a result, the top-hat transform of the contour (*i.e.*, subtraction of the contour from its opening with a SE) becomes equal to its opening. If a candidate concavity is identified, the candidate concavity is removed to produce two new contours, and the concavity detection process is iteratively applied on these contours. Algorithm 1 shows the pseudocode for concavity detection.

3.2.2.5.2 Color analysis

Not all of the identified candidate concavities are actual separation boundaries between two villi. Only candidate concavities that possess certain color features are accepted as the separation boundaries. Based these color features, the border regions of a villous structure have: 1) a high density of SC cells and CC cells (*i.e.*, SC class), and 2) sparse appearance

of the VC class. Therefore, candidate concavities that have high SC-class density and low

Algorithm 1 Concavity detection

Input: Binary mask (BW), elliptical Structuring Element (SE)

Output: A binary mask with all candidate concavities removed

Initialization:

Label the contours in BW to generate L , $L = \{L_0, L_1, \dots, L_N\}$, $N = \text{number of contours}$

For each contour in L

Set size of the $SE_{size} = I$

Loop of the *SE characteristics*

For each alignment of the $SE = \{0^\circ, 45^\circ, 90^\circ, \text{ and } 135^\circ\}$

Calculate the Top-hat transform for contour L_i using SE (*i.e.*, T_i)

If $T_i = L_i$

Exit the nested loop of SE characteristics and go to next contour

End if

If the candidate concavity found in T_i

Remove the candidate concavity from L_i to produce two new contours

Replace L_i with one of found in T_i

Add the other new contour L_{N+1} to list of labels, $L = \{L_0, L_1, \dots, L_N, L_{N+1}\}$

Increment the number of contours $N = N + 1$

Exit the nested loop of SE characteristics and start analyzing the new contour L_i

End if

End for

$SE = SE + I$

End loop

End for

density of VC-class are labeled as separation boundaries. The thresholds for SC-class density and VC-class density were tuned over the training set of images, using the manual segmentation as ground truth.

3.2.3 Evaluation

Our data set consisted of nine images obtained from healthy term pregnancies and 27 placenta images from term pregnancies complicated with PE. We performed stratified 5-

fold cross-validation, assigning the images into training and test sets with a 2:1 ratio (*i.e.*, six images from healthy patients and 18 images from PE patients were assigned to the training set, and three images from healthy patients and nine images from PE patients were assigned to the test set).

Villi detection accuracy was determined using manual segmentation as ground truth. The F1 score and sensitivity were calculated based on the following definitions:

$$Sensitivity = \frac{TP}{TP + FN} \quad 3.5$$

$$F1\ score = \frac{2 \times TP}{2 \times TP + FP + FN} \quad 3.6$$

True positive (TP): number of villous structures identified correctly

True negative (TN): number of objects in the binary mask generated in the boundary detection step that are correctly labeled as other objects (*i.e.*, non-villi)

False positive (FP): number of villous structures identified incorrectly

False negative (FN): number of villi structures identified in the manual segmentation that were not detected

For instance, if a contour contained two villi based on manual segmentation, and our algorithm correctly detected these two villi, the algorithm scored two TPs. However, if the algorithm detected only one or none of the villi, the algorithm scored one TP and one FN, or two FNs, respectively. When a contour contained only one villus based on manual segmentation, but our algorithm detected two villi, the case was scored as one TP and one FP.

3.2.4 Comparison to a previous method

The results of our method were evaluated against a recently published method [29] for placenta villi segmentation, based on the ImageJ software platform. The authors of the previous method publicly provided their source code online. In this previously published method, the authors extracted boundaries of villi to determine the number of existing villi in a given image, and standard color thresholding was used to segment the villi; we optimized this threshold over our training set. To achieve a fair comparison, we used the same values for the common thresholds in both methods (*i.e.*, villi size threshold, which is set to $1000 \mu\text{m}^2$). We then evaluated their method by comparing algorithm-generated results with manual segmentation. We also evaluate our method's segmentation results against manual segmentation and compare it with generated results from the previous method.

As stated in section II.C, TNs are defined as the number of objects in the binary mask that correctly labeled as non-villous objects. TNs were not computed for this comparison method, as it does not have a similar binary mask that would allow for a sensible TN comparison.

3.3 Results

As is typically observed in clinically relevant image datasets, our experimental image dataset contained artifacts and undesired objects, including imaging artifacts and maternal red blood cells in the intervillous space (12 out of 36 images were contaminated by imaging artifacts (Figure 3.4)). In the villi classification step, the contours were classified based on their size and density of VC class and SC class. The proposed pipeline was able to

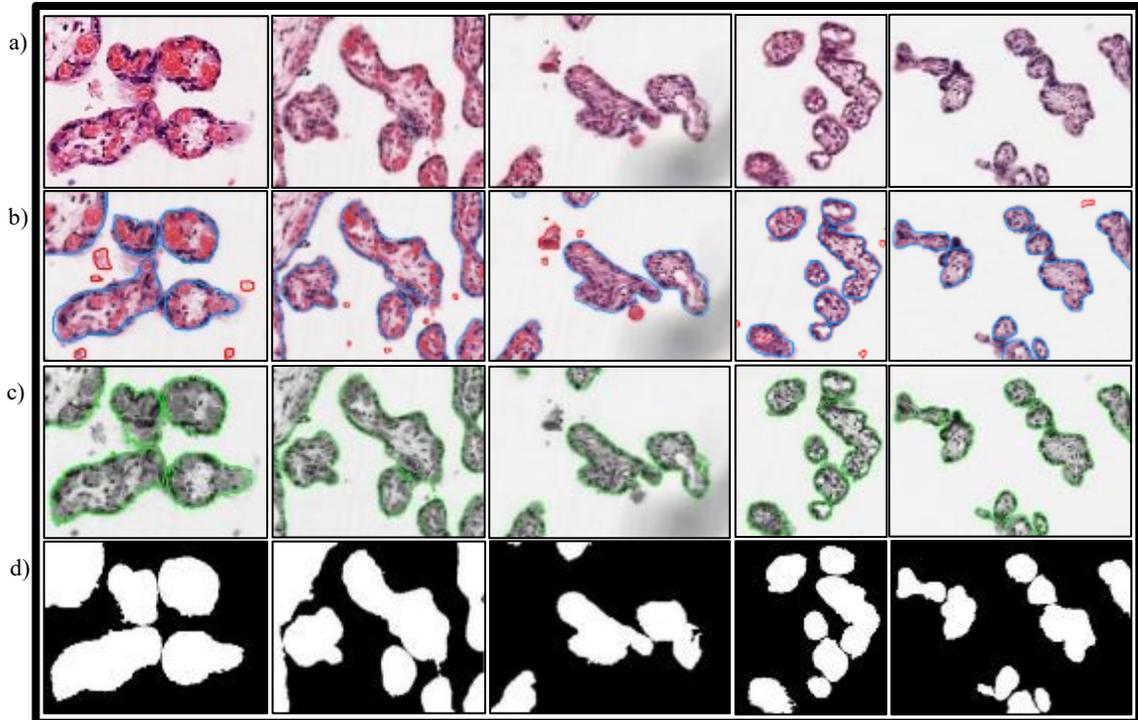


Figure 3.7 Sample segmentation of villi, a) The original image, b) The manual segmentation. The boundary of a villous structure is annotated in blue. The red annotation shows undesired objects, c) The segmentation results of our proposed method. The boundaries of villi is annotated in green. d) The generated binary mask showing the final segmentation results.

recognize desired contours with 92.86% accuracy (3625 out of 3904 contours were correctly identified).

Concavity detection was an important step in the detection of touching villi step. In comparison to manually identified concavities by experts, our algorithm detected concavities with an accuracy of 95.47% (3908 out of 4093 concavities were correctly identified).

Figure 3.7 shows multiple sample segmentation of villi, which is performed by our proposed method. Table 3.1 shows the F1 scores, sensitivity, and truth table components of a recently published method [29] and the proposed method, separating the results for healthy control patients and patients diagnosed with PE. In Table 2, the overall results across all placental images are presented. In comparison to the previous method, our

Table 3.2 A comparison between our proposed method and the previous work. The results are presented for each fold for health term pregnancies and pregnancies complicated with PE. The TN row for the previously published work is not applicable as explained in section 2.4.

		Healthy					Diagnosed with PE				
		fold #1	fold #2	fold #3	fold #4	fold #5	fold #1	fold #2	fold #3	fold #4	fold #5
<i>Proposed method</i>	<i>TP</i>	227	255	246	228	247	421	532	518	541	572
	<i>TN</i>	96	176	102	150	134	227	177	193	261	247
	<i>FP</i>	50	69	68	46	57	128	141	131	132	161
	<i>FN</i>	42	28	68	29	23	138	124	144	133	92
	<i>F1 score (%)</i>	83.15	84.02	78.34	85.88	86.06	75.99	80.06	79.02	80.33	81.89
	<i>Sensitivity (%)</i>	84.39	90.11	78.34	88.72	91.48	75.31	81.10	78.25	80.27	86.14
<i>Previously published method[29]</i>	<i>TP</i>	133	209	163	170	189	253	370	312	331	410
	<i>TN</i>	-	-	-	-	-	-	-	-	-	-
	<i>FP</i>	23	43	38	27	50	76	85	70	92	128
	<i>FN</i>	136	74	151	87	81	306	286	350	343	254
	<i>F1 score (%)</i>	62.59	78.13	63.30	74.89	74.26	56.98	66.61	59.77	60.35	68.22
	<i>Sensitivity (%)</i>	49.44	73.85	51.91	66.15	70.00	45.26	56.40	47.13	49.11	61.75

method yielded an F1 score and a sensitivity that are 15.46% and 27.06% higher, respectively. In order to highlight the impact that detection of touching villi has on the

Table 3.1 Overall F1 score and sensitivity of the previous method, the proposed method without detection of touching villi algorithm, and the proposed method with touching villi detection algorithm. The results are reported as the weighted average of five fold cross validation.

	<i>F1 score</i>	<i>Sensitivity</i>
<i>Previously published method[29]</i>	65.30 %	55.12 %
<i>Proposed method without detection of touching villi</i>	74.58 %	61.28 %
<i>Proposed method with detection of touching villi</i>	80.76 %	82.18 %

overall performance of our proposed method, the overall result of the proposed method without detection of touching villi is also included in Table 3.2.

As it is defined in Equation 3.5, the sensitivity of a method depends on the total number of FNs. In this work, the majority of FN cases are generated when multiple villi are located in close proximity of each other, and the algorithm is not able to discern these as separate villi. As can be observed from the sensitivity column in Table 2, the difference gap between the previous work's sensitivity and the proposed method without detection of touching villi step is 6.68%. The comparison between the sensitivity of the proposed method with and without detection of touching villi shows a considerably large difference of 20.9%. The great improvement in sensitivity over the previous method is primarily achieved by applying the detection of touching villi step in our proposed algorithm.

The processing time required for a complete analysis of each image was approximately 21 min per image on a Core i7-6700 CPU at 3.4 GHz in Matlab environment. We observed that this processing time is quite reasonable as the developed algorithm can run in the background without the need for any user interactions. Although we have used parallel processing in the feature-based analysis stage, the rest of the programming code is not optimized. The villi density in each image and the convexity of the extracted contours were two of the most influential factors in the determination of the processing time required for each analysis. Most of the processing time (approximately 80%) was required for the boundary detection step, which can be substantially improved by parallel (GPU) processing.

3.4 Discussion

An initial step for automated analysis of placental images is segmenting the chorionic villi – the functional unit of the placenta – within the histological specimens. Villi segmentation allows extraction of key features (e.g., villi count, size distribution, shape feature) for an objective assessment of placental images. Due to the complexity of villous structures and the presence of imaging artifacts, segmentation of individual villi in a placental image is a difficult task. However, the proposed method successfully identifies complex villous structures in a large series of images with high accuracy. Our proposed method yielded an F1 score of 80.76% and sensitivity of 82.18% for a dataset comprising of 4946 sample villi, which considerably outperformed a previously published method[29].

Two major contributing factors in obtaining better results in our method compared to the previous method were: 1) comprehensive artifact removal; and 2) the ability to detect and properly segment touching villi. The second factor had a large influence on the results, particularly the sensitivity score. A commonly recognized feature of the placental disease in PE is “villous agglutination” – a condition in which neighboring villi structures adhere to one another [209]. This feature of placental images from pregnancies complicated with PE likely contributed to the decreased performance of the previously published segmentation method when compared to healthy controls. Importantly, our algorithm yielded high and comparable performance when applied to histopathology cases from healthy controls (weighted average F1 score of 83.37%) or pathological cases of preeclampsia (weighted average F1 score of 79.61%) – providing confidence for the utility of this automated process in a clinical setting.

The most important limitation of our work was including connective tissue tears in VC. In our work, the calculation of fetal red blood cells density played a key role in multiple steps of our algorithm. The placental specimens used had been washed, so a considerable amount of red regions (fetal blood cells) inside placental vasculature (within the core of the villi structure) had been removed and replaced by white/light pink regions. Therefore, we also included these white regions in the calculation of VC density. However, some included areas were related to connective tissue tears, which represent no particular tissue.

The long-term goal of our work is to develop an algorithm that has the ability to process the whole placenta scan. This can be achieved by block analysis. With the scans used in our dataset, on average, around 50 extracted images would be required to cover the whole scan. Although the presented pipeline has the potential to analyze the entire scans (i.e., analysis of ~50 images per scan), in our study, we only present the results from three images from each placental scan. Manual segmentation, to establish a ground truth, is a time-consuming process, and it was believed that it was more important for this study to have images that better capture the variability across scans than within a scan.

3.5 Conclusion

In this study, a fully automated method for segmentation of chorionic villi structures in histopathology images of the human placenta was presented. The proposed method has the ability to identify complex villous structures, including touching and overlapping villi, by performing color analysis on the detected concavities of the villi structures. Our proposed method yielded an F1 score of 80.76% and sensitivity of 82.18% for a dataset comprising of nearly 5000 sample villi, considerably higher performance than previously published methods. In addition, the proposed method has a comparable performance on placental

samples from healthy and pathological pregnancies, a critical consideration in determining clinical utility.

With the help of an accurate and robust automated analysis method, pathologists might be presented with a set of otherwise inaccessible morphometric values for each image, such as the distribution of specific structures (physically and class-wise) and presence or absence of certain histological features/findings within identified structures. The pathologist can use these values to help make a final diagnosis. Although we focused only on histopathology images of the placenta, it might be possible to generalize this tool across multiple fields of clinical pathology. This could result in a marketable tool that would substantially reduce the subjectivity that exists in current practice.

Chapter 4: Automated estimation of the mean linear intercept in histopathology images of mouse lung tissue

In this chapter, a pipeline for automated estimation of MLI score in histopathology WSIs of mouse lung tissue is presented. A CNN model based on U-Net architecture is used to extract foreground regions (*i.e.*, diagnostically relevant lung tissue compartments) of the mouse lung WSIs. Conventional image processing and machine learning approaches are applied for the segmentation of the main biological structures of mouse lung tissue (*i.e.*, bronchi, blood vessels, and alveoli), which paves the way for automated estimation of the MLI scores. This chapter has been published as Salsabili, S., *et al.* “Fully automated estimation of the mean linear intercept in histopathology images of mouse lung tissue.” *Journal of Medical Imaging*, vol. 8, 2021, p. 27501, doi: <https://doi.org/10.1117/1.JMI.8.2.027501>

4.1 Introduction

BPD is the most common complication of preterm birth [37]. BPD is a chronic lung disease, characterized by an arrest in alveolar and vascular growth within the lung. BPD is a multifactorial disease, caused by ventilator and oxygen therapy administered for acute respiratory failure, and is commonly associated with ante- and post-natal inflammation [210], [211]. Although, there is currently no effective treatment for BPD, ongoing investigations are in progress to better understand the pathophysiology of this disease. To discover new potential therapies, it is crucial that researchers quantify the lung injury phenotype in an accurate and efficient manner.

A common metric used to quantify lung injury is the MLI, which represents the mean distance between alveolar septa within the lung [36]. Investigators using animal models to

mimic neonatal chronic lung disease, often use the MLI as a parameter to describe the simplification of the lung architecture, characteristic of BPD. The conventional method for MLI quantification of lung tissue specimens usually includes the microscopic assessment of histopathological slides by an expert, which is inefficient, tedious, and time-consuming. Moreover, there is a lack of objective visual gold standards for structures found in microscopic views of tissues, which leads to inter- and intra- expert variation and reproducibility issues [212]–[214]. In recent years, there has been a shift toward the development of automated methods for assessment of histopathology images to address the shortcomings in the conventional approaches [18], [74], [215], [216]. However, due to technical impediments such as object variability, varying straining, and artifacts, the development of robust and comprehensive methods for assessment of histopathology images remains a challenging task.

Automated estimation of the MLI score requires detailed and accurate segmentation of biological structures in lung histopathology images, which makes development of such approaches difficult. In recent years, few studies have been reported on lung histopathology image analysis with a focus of automating the MLI quantification process. However, these methods often have difficulty identifying non-alveolar structures (e.g., blood vessels, bronchi, etc.), which leads to underestimation of the MLI score in comparison to manual measurements[217]. Moreover, the technical details provided for such algorithms are often limited. As a result, there are currently no accessible and reliable automated approach for MLI quantification in the literature and consequently, the most trusted methods remain manual/semi-automated techniques[218]–[220], which can be laborious, time-consuming, and subjective. In this work, our aim is to present a fully automated pipeline for estimation

of the MLI score in histopathology images of mouse lung tissue. The main contributions of this paper are: 1) proposing a novel approach for assessment of digitized histopathology slides to automate the estimation of the MLI scoring; 2) performing accurate segmentation of different biological structures in the histopathology images of mouse lung tissue; and 3) evaluation of the proposed method against human raters.

4.2 Materials and methods

4.2.1 Histopathology images of mouse lung tissue

Our dataset comprises high-resolution WSIs of ten lung histopathology specimens of mice obtained from the Sinclair Centre for Regenerative Medicine (Ottawa Hospital Research Institute, Ottawa, ON). All animal experiments were conducted in accordance with protocols approved by the University of Ottawa Animal Care Committee. The lungs specimens were inflation fixed through the trachea with 10% buffered formalin, under 20cm H₂O pressure, for five minutes. After the trachea was ligated, the lungs were immersion fixed in 10% buffered formalin for 48 hours at room temperature and then immersed in 70% ethanol for 24 hours at room temperature. The Louise Pelletier Histology Core Facility at the University of Ottawa paraffin-embedded, cut (4µm sections), mounted, and stained the lung tissue with H&E.

The slides were scanned using an Aperio CS2 slide scanner (Leica), and high-resolution color images at 20× magnification (i.e., approximately 0.5µm per pixel) were obtained. In total, ten WSIs were generated from two different experimental groups. The WSIs were randomly selected from different mice lung tissue, while blinded to the experimental groups. This procedure was consistent across all animals. The first group contains five WSIs of healthy mouse lung tissue from mice that were housed in room air (*RA* group),

which is used as the control group in this experiment. The second group consists of five WSIs of diseased mouse lung tissue from mice that were exposed to a high concentration of oxygen and lipopolysaccharide (*O2 + LPS* group). The *O2 + LPS* experimental group mimics the conditions that a preterm infant is exposed to (high concentration of oxygen and inflammation), which can lead to a lung injury phenotype seen in BPD. The WSIs generated from the *O2 + LPS* group normally contain fewer and larger alveoli, in comparison to WSIs from the *RA* control group. This is expected to reflect in the MLI score by calculation of a higher value of MLI in the *O2 + LPS* group in comparison to the *RA* control group, which have much more and smaller alveoli.

4.2.2 Conventional MLI quantification

Conventionally, the MLI score is calculated using a semi-automated process[36]. In this process, the microscope software (MetaMorph Software version 7.8, Molecular Devices, LLC) automatically presents the human rater with a field-of-view (FOV) from pre-defined grid points. Each FOV image is a 1072×1388 pixel sub-image from the input WSI, overlaid with guidelines in the middle of the image (see Figure 4.1); the top horizontal guideline, of length $155.34 \mu m$, is used by a human rater in the MLI quantification procedure.

The human rater first decides whether the shown FOV image can be used for MLI quantification or not. The rejection of a FOV image is based on two criteria. First, if a part of the horizontal guideline is in the pleural space (*i.e.*, outside the lung space; Figure 4.1.a), the FOV image is rejected; that is, it is not used in the calculation of the MLI score. Second, if the horizontal line intersects a bronchus (Figure 4.1.b) or a vessel (Figure 4.1.c), the FOV image is rejected. If the FOV image is not rejected, the human rater counts the number of

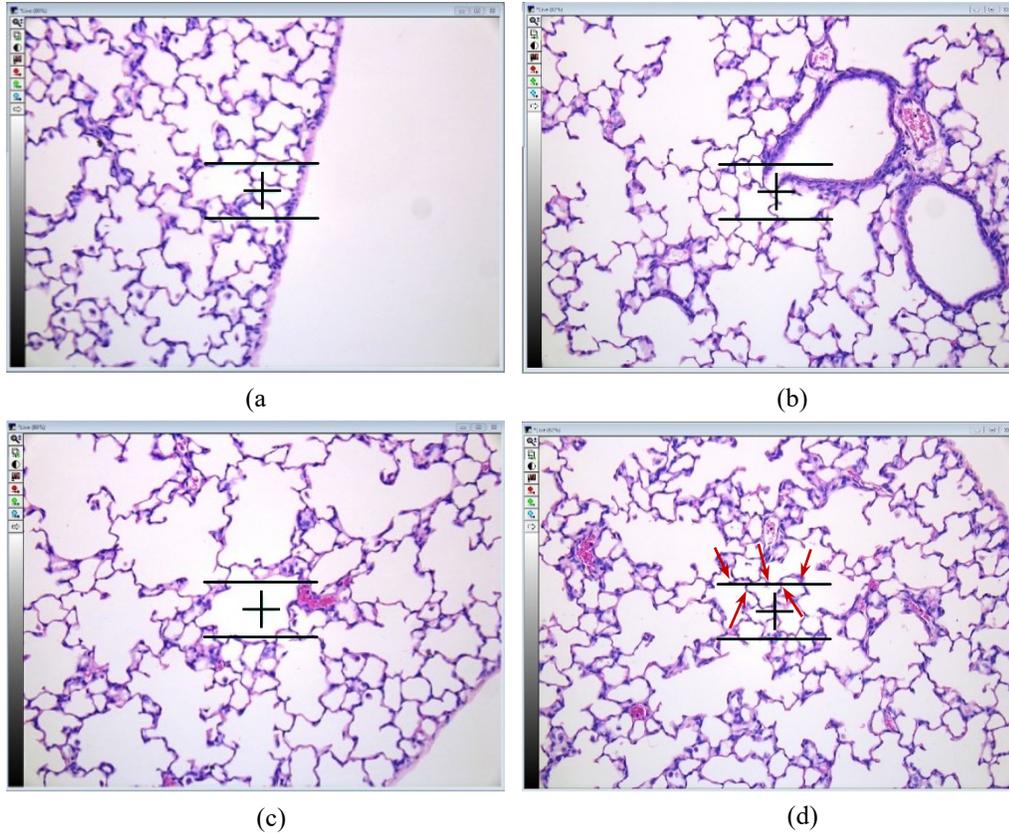


Figure 4.1 Examples of extracted fields of view (FOV) images using the image analysis software. (a) This FOV image is rejected because the guideline is partially outside of the section. (b) The FOV image is rejected due to intersection of the guideline with a bronchus. (c) The FOV image is rejected because the guideline intersects with a blood vessel. (d) An example of accepted FOV image with five intersections. The intersections of the guideline with the septa is shown using red arrows.

intersections. An intersection is when the horizontal guideline fully crosses over the septa, which is the alveolar border wall (Figure 4.1.d).

The MLI score is calculated as:

$$MLI = \frac{N_{accepted\ FOV\ images} \times 155.34}{N_{intersections}} \quad 4.1$$

Where, $N_{accepted\ FOV\ images}$ is the number of accepted FOV images, 155.34 refers to the length of the horizontal line in μm , and $N_{intersections}$ is the total number of intersections counted from all of the accepted FOV images. A minimum of 250 accepted FOV images (i.e., number of FOV images, not including those that were rejected), is desired for the computation of the MLI score[221].

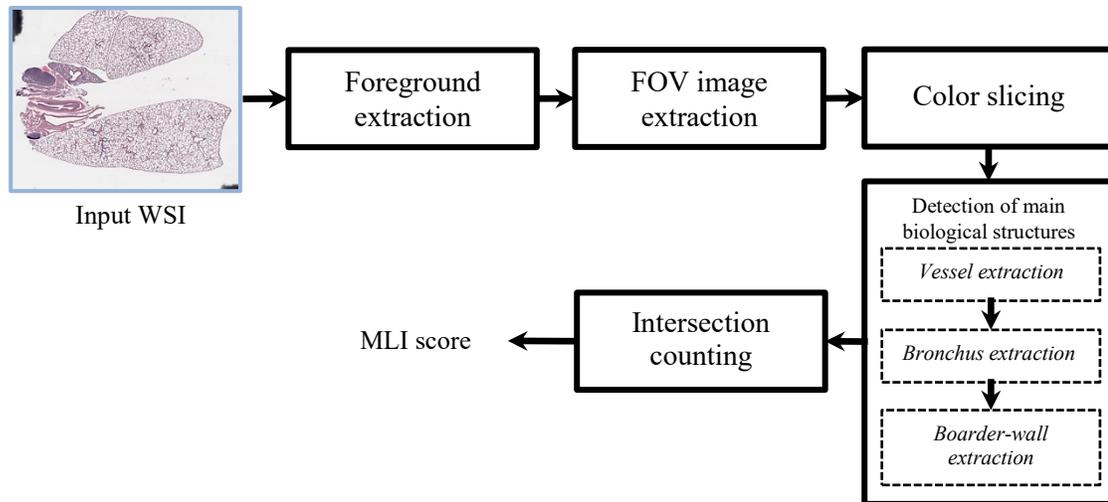


Figure 4.2 Block diagram of the proposed methodology. The abbreviations WSI, FOV, and MLI are referred to whole slide image, field-of-view, and mean linear intercept, respectively.

4.2.3 Automated calculation of MLI score

Our proposed pipeline for the automated estimation of the MLI score consists of five steps, which are shown in Figure 4.2.

4.2.3.1 Foreground extraction

Each WSI may contain various imaging artifacts and undesired biological structures (see Input WSI in Figure 4.2). As an initial step in our pipeline, we segment the lung space from the pleural space and undesired artifacts, which are considered the foreground and background, respectively. The foreground ROIs were segmented using a convolutional neural network (CNN), based on the U-Net[158] architecture.

Since the histopathology slides were scanned at $20\times$ magnification, images' sizes are large (average size of $21,052\times 18,124\times 3$ pixels) and contain a high level of detail that is not required to segment the foreground ROI. As such, images were down-sampled by a factor of 10, greatly reducing the computational cost, while still allowing for accurate ROI segmentation. We conducted five-fold cross-validation in our experiments. Eight of the WSIs were used to form our training dataset (used for training and validation of the CNN),

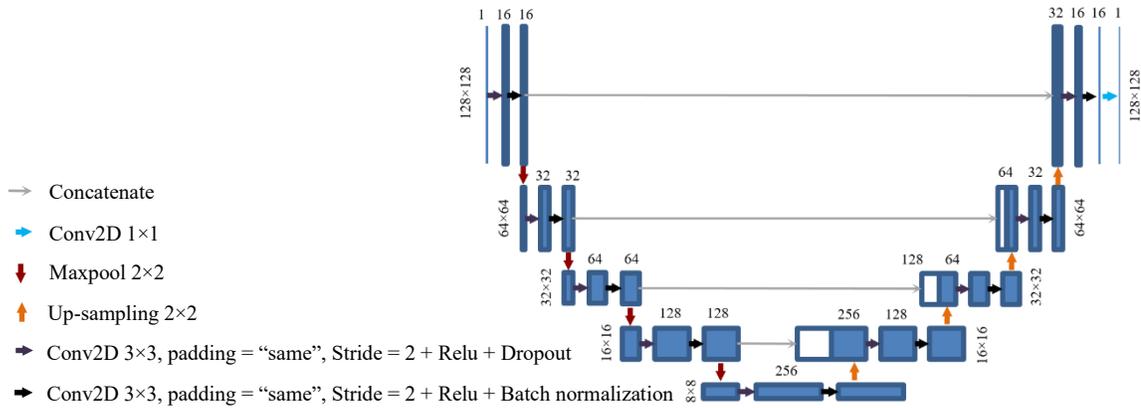


Figure 4.3 The CNN architecture. The blue and white boxes represent the multi-channel feature maps and copied feature maps, respectively. The size of the feature maps is indicated on top of each box and the size of each input layer is denoted on the left-hand side.

and the remaining two WSIs (one from the *RA* group and one image from the *O2 + LPS* group) were used as the test dataset. This was repeated five times such that each WSI was used as the test dataset. From each WSI, image patches (128×128 pixels) were extracted using a sliding window, with 50% overlap in the horizontal and vertical directions. Data augmentation (90° rotations and image flipping) was employed to the training dataset to increase the classifier performance[222]. On average, 28,092 image patches were generated from the eight WSIs at each fold, which were assigned to the training and validation sets with a 9:1 ratio (*i.e.*, 25,283 and 2,809 for training and validation, respectively).

The remaining two WSIs in each fold, used for the test dataset, had the average of 7,023 image patches.

The CNN was trained from scratch, with four convolution layers in the contracting path and four transpose convolution layers in the expansive path. The complete architecture of the network is illustrated in Figure 4.3. The CNN was trained using the Adam optimizer[223], binary cross-entropy loss function, and mean intersection over union evaluation metric. To account for the overfitting problem, we perform dropout by a factor

of 0.5 at each layer and perform cross-validation at each epoch with the ratio of 10:1 (train on 90% of the training data and validation on the remaining 10% at each epoch). Batch normalization was applied to each layer to reduce the training time and prevent diverging gradients. We trained our model for 200 epochs with a batch size of 100. We used the Keras framework for algorithm development on a standard workstation with an Intel Core i7-3770 3.40 GHz CPU, 12 GB of installed RAM, and a single NVIDIA RTX 2060 with 6 GB memory. A total number of 2,161,649 trainable parameters were optimized in our segmentation model.

Manual segmentation of the foreground was performed on each individual WSI image using ImageJ software (<https://imagej.net/>) by the first author (S.S.). The manual segmentation was used as the ground truth for our foreground extraction approach.

4.2.3.2 FOV image extraction

From the foreground region of each original high-resolution WSI, FOV images (sub-images of size 1072×1388 pixels) were extracted using a sliding window, with a 50% and 75% overlap in the horizontal and vertical directions, respectively. For each FOV image, a horizontal guideline (thickness 1 pixel; length 312 pixels, which corresponds to 155.34 μ m) was superimposed at the center of each image and used by human raters to count the intersections. In the automated process the guideline is only virtually superimposed, appearing only for visualization purposes. If the horizontal guideline does not fully reside within the foreground region, the FOV image is rejected.

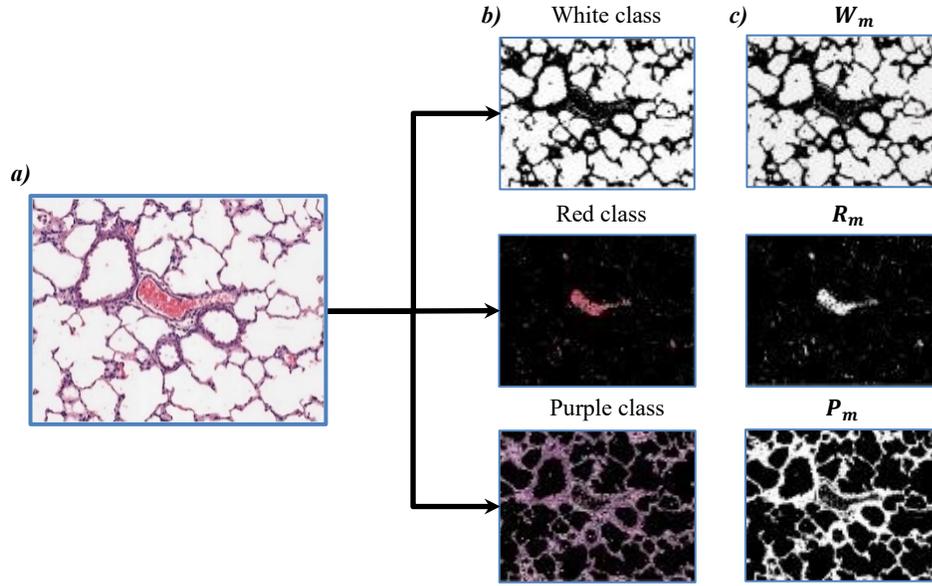


Figure 4.4 The organization of extracted colors in a field-of-view (FOV) image. a) The original input FOV image. b) The RGB representation of each color slice. c) The visualization of each color binary mask, P_m , W_m , and R_m .

4.2.3.3 Color slicing

A number of factors can contribute to variations of the color content in histological images (*e.g.*, histochemical staining time, amount of histology stain used) across different WSIs. We applied color normalization[68] to the input WSIs to mitigate such variations. Next, pixels are classified, using a color slicing algorithm[224], into the three main categories of color in the lung images (see Figure 4.4): 1) white, 2) red, and 3) purple.

The white pixels in each FOV image belong to two main sources: 1) white areas in the pleural space (*i.e.*, background regions), and 2) white areas within the lung space. The white pixels located in the pleural space are segmented in the foreground extraction step. Therefore, only the white pixels corresponding to the lung region need to be identified. The binary mask W_m denotes the white pixels within the lung space and is determined using Equation 4.2:

$$\left\{ \begin{array}{l} bw_1(k, l) = |I_i(k, l) - I_j(k, l)| \leq C \\ bw_2(k, l) = |I_i(k, l)| \geq TH_1 \\ W_m = bw_1 \cap bw_2 \\ \text{where } \left\{ \begin{array}{l} i \neq j \\ i, j \in (R, G, B) \\ k \in (1, M), l \in (1, N) \end{array} \right. \end{array} \right. \quad 4.2$$

In Equation 4.2, I_R , I_G , and I_B are the red, green, and blue channels in RGB color space, respectively. TH_1 is a threshold whose value is determined using Otsu's method[121] at each fold. This threshold ensures that the extracted pixels have high intensities (close to white). The constant C is a fixed threshold for all folds to ensure that the intensity difference between the R, G, and B channels is small. The value assigned to C was 15, which was determined empirically. The mask W_m is generated by the intersection of the bw_1 and bw_2 binary masks, and contains the white pixels within the lung space. The M and N are the corresponding number of rows and columns, respectively.

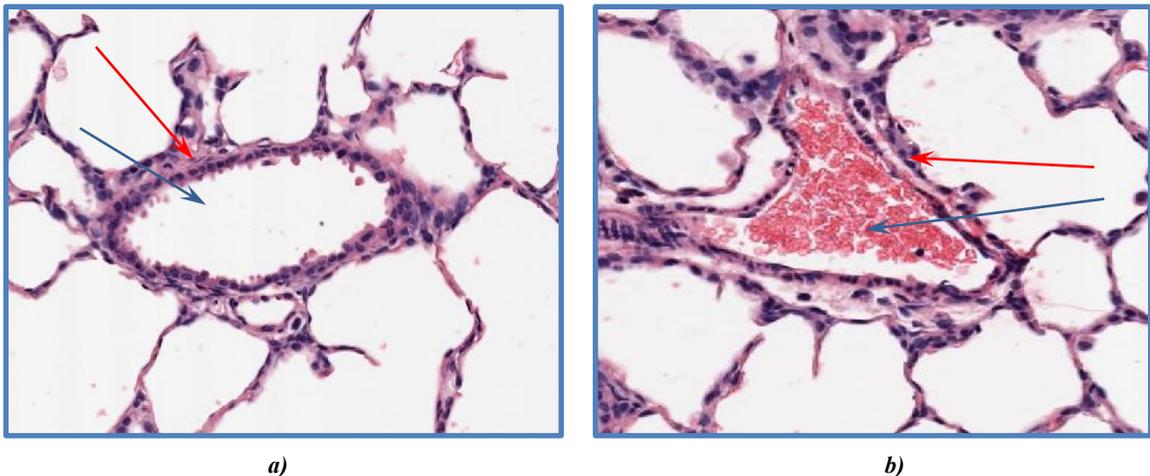


Figure 4.5 The visualization of a bronchus (a) and a vessel (b). The corresponding lumen region (LR) and border wall region of each structure is identified by blue and red arrows, respectively. As it can be seen in (b), the LR of the vessel is densely covered by the red blood cells.

R_m is the binary mask denoting the red pixels in each FOV image. These pixels usually represents the remaining blood cells in the lung tissue, which can indicate the presence of a vessel in the neighboring region. The binary mask R_m is determined using Equation 4.3:

$$\begin{cases} T \triangleq I_R - \text{mean}(I_G + I_B) \\ R_m = (T \geq TH_2) \cap \overline{W_m} \end{cases} \quad 4.3$$

In Equation 4.3, the $\overline{W_m}$ represents all of the pixels that are not in W_m . The value of the threshold TH_2 is determined using Otsu's method[121] at each fold.

The remaining pixels are assigned to the purple color binary mask P_m , which represents several cellular compartments (*e.g.*, pneumocytes cells, glands, smooth muscle, etc.).

4.2.3.4 Detection of main biological structures

Each FOV image is segmented into three biological structures: 1) alveoli, 2) vessel, and 3) bronchi. As stated in section 4.2.2, FOV images whose horizontal guideline intersects with vascular and bronchus regions should be rejected (*i.e.*, excluded from MLI calculation). We first automatically segment the lumen region (LR) of vessels based on a region growing method[225]. We then separate bronchi from alveoli, based on multiple morphological features extracted from each individual LR. Finally, the segmentation approach is completed by identification of the border wall of the vessel and bronchus (see Figure 4.5).

4.2.3.4.1 Vessel extraction

To identify different lung structures in each FOV image, it is desirable to detect all of the LRs using W_m . However, due to high density of blood cells in the LR of the vessels,

accurate extraction of the LR of vessels from W_m is not feasible. Therefore, we develop an alternative approach, where we first extract candidate seed regions for each individual vessel in the FOV image and then use a region growing method to segment all of the LR of the vessels.

Since most vessel structures have a considerable amount of red blood cells within their lumen region (see Figure 4.5.b), it is possible to locate the candidate seed regions by detection of the areas in the red color binary mask R_m , where the density of red pixels is relatively high. In order to detect these regions in R_m , using a 25×25 window, we iteratively sweep R_m with 50% horizontal and vertical overlap. At each iteration, the local density of the red pixels is calculated. If the local density is higher than 0.9, all of the red pixels present in the window will be added to the binary mask BW_{seed} .

Figure 4.6 illustrates different steps in the LR segmentation method. To extract the lumen regions associated with each seed region in BW_{seed} , we apply a region growing

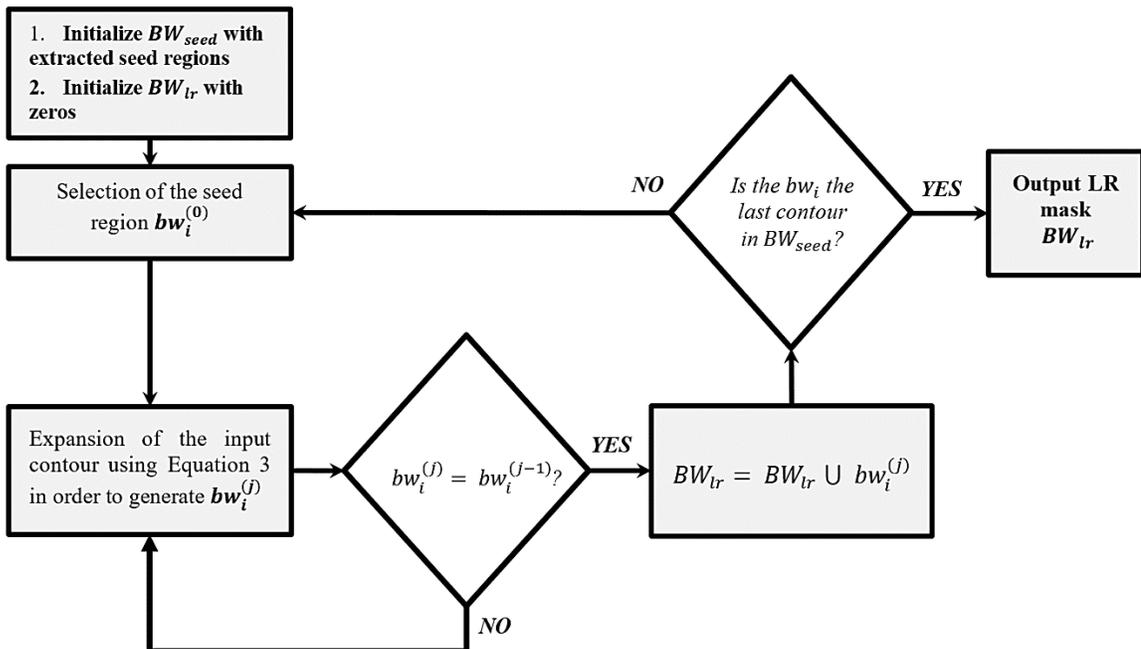


Figure 4.6 The block diagram of the region growing based lumen region (LR) segmentation. The notation $bw_i^{(j)}$ refers to the “ i th” selected seed region and superscript “ (j) ” indicates that bw_i is expanded j times.

segmentation approach to each individual seed region. Let $bw_1^{(0)}, bw_2^{(0)}, \dots, bw_n^{(0)}$ denote n different seed regions in BW_{seed} . In the proposed LR segmentation, the selected seed region $bw_i^{(0)}$ is iteratively expanded using Equation 4.4.

$$bw_i^{(j)} = \{ bw_i^{(j-1)} \oplus SE^{\odot[2]} \} \& W_m \quad 4.4$$

Equation 4.4 is the morphological dilation operation (\oplus) of input binary mask $bw_i^{(j-1)}$ by a circular structuring element with radius of two pixels ($SE^{\odot[2]}$), followed by a bitwise AND operation of the resultant mask with W_m . If the expanded binary mask $bw_i^{(j)}$ does not change after applying Equation 4.4 ($bw_i^{(j)} = bw_i^{(j-1)}$), the algorithm quits the iterative expansion loop and adds the binary mask $bw_i^{(j)}$ to the output LR binary mask BW_{lr} (see Equation 4.5). The operation is continued until all of the seed regions are processed by the algorithm.

$$BW_{lr} = BW_{lr} \cup bw_i^{(j)} \quad 4.5$$

4.2.3.4.2 Bronchus extraction

The LR of the remaining structures (bronchi and alveoli) are mostly white and can be directly extracted using W_m . The next step is to separate the LR of the bronchi from the alveoli's LR. Based on unique visual characteristics of the LR of the bronchi, multiple morphological and textural features are extracted from each individual LR, which is used for classification of the remaining structures. These features are briefly introduced in the following:

1) Area of the LR: One of the salient visual features in classification of the remaining LRs is the *area* of the LR. In room air condition (*i.e.*, RA group), there is a considerable

size difference between the LR area of bronchi compared to the LR of alveoli, and therefore the size feature plays an effective role in classification of bronchi versus alveoli. However, in O2 + LPS experimental group, the alveoli structures are enlarged. As such, LR area may not be as reliable a feature for the O2 + LPS group and the classifier may be more reliant on the other extracted features.

2) Shape of the LR: Based on our visual observation, the shape of the LR is another distinguishing feature to discern bronchi versus alveoli LR. For instance, the LR of bronchus structures are mostly ovular and do not contain many branches. The alveoli LR are often non-ovular and contain many sub-branched (see Figure 4.7.a). We used Equation 4.6 in order to quantify the circularity of the LR.

In Equation 4.6, $area(...)$ and $perimeter(...)$ represent the functions that calculate the area and the perimeter of the input binary mask input contour bw , respectively. The

$$S_{circularity} = \frac{4\pi \text{area}(bw)}{\text{perimeter}(bw)^2} \quad 4.6$$

value of $S_{circularity}$ is an integer between “0” and “1”, where 1 represents a circular contour.

3) Thickness of the boarder wall region: The majority of bronchus structures have a thick border wall region in comparison to other structures in lung tissue (see Figure 4.7.a). Therefore, the thickness of the border-wall region of the remaining structures are extracted. In order to quantify the thickness of the boarder-wall region, the complete segmentation of the boarder-wall of each structure is not required. Instead, we expand the selected LR using Equation 4.7 until the ratio of W_m pixels in the expanded area become larger than 25%.

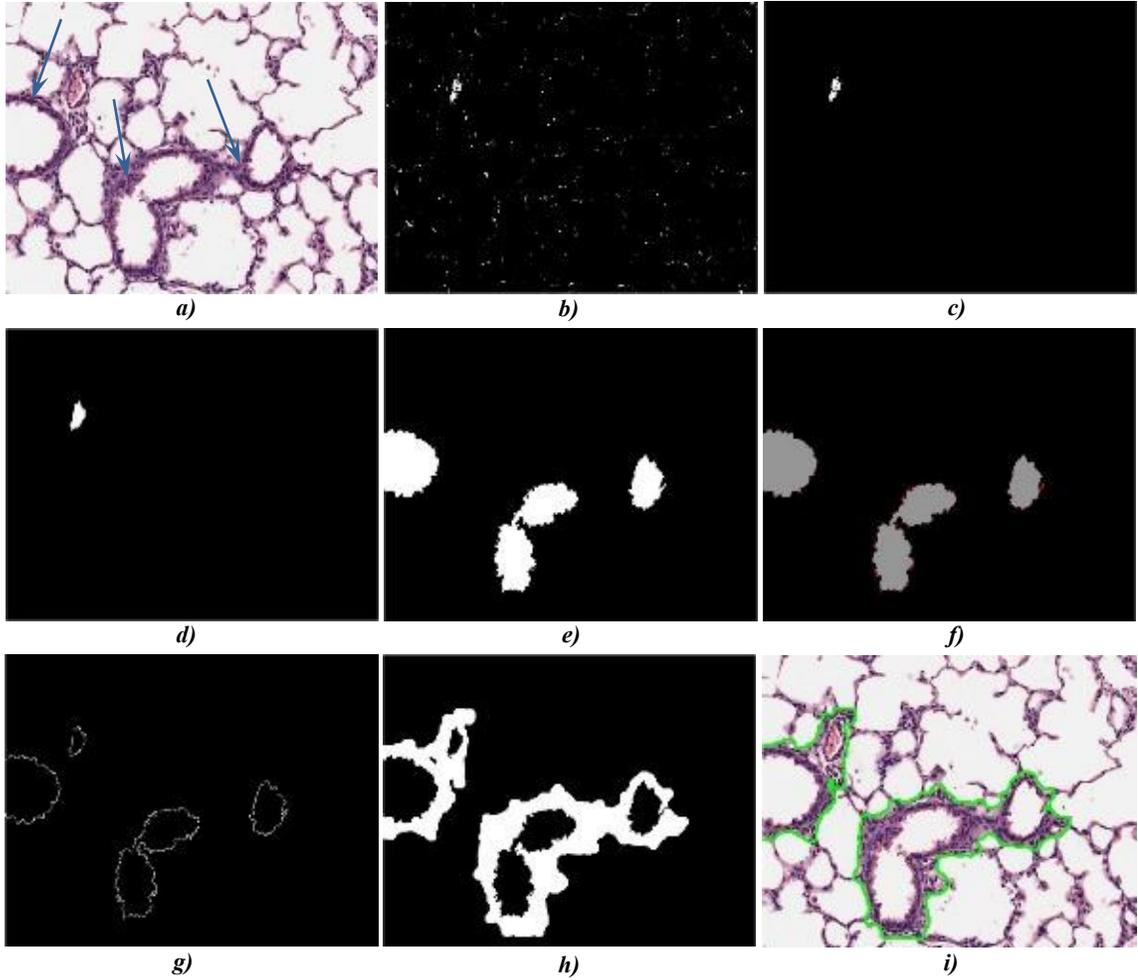


Figure 4.7 a) A sample field-of-view (FOV) image. The blue arrows indicate the border-wall region of the bronchi structures. b) Red color binary mask R_m . c) The extracted vessel's seed region. d) The extracted vessel lumen region (LR). e) The extracted bronchi LR. f) The visualization of the ripple pattern in bronchi structures. The ripple pattern is marked as red in this figure. g) The corresponding seed region in border-wall extraction step. h) The extracted border-wall regions of every vessel and bronchi. i) The segmented bronchi and vessel in the FOV image.

The number of times that Equation 4.7 will be used to expand the LR can be used as an indication of the border-wall region thickness.

$$bw^{(j)} = \{bw^{(j-1)} \oplus SE^{\odot[1]}\} - bw^{(j-1)} \quad 4.7$$

Equation 4.7 is the morphological dilation operation (\oplus) of input binary mask $bw^{(j-1)}$ by a circular structuring element with radius of one pixels ($SE^{\odot[2]}$), followed by removing the pixels of the $bw^{(j-1)}$ from the resultant binary mask ($bw^{(j)}$).

4) LR's perimeter ripple pattern: Another important biological feature is the frequent detection of ripple shape patterns on the perimeter of the LR of bronchi (see Figure 4.7.f). Considering the fact that this ripple pattern is more apparent in the LR of the bronchi, it can be a valuable discerning feature. For each LR, this feature is quantified by calculation of the ratio of the total area of the ripples over the area of the LR (Equation 4.8).

$$\begin{cases} bw_{ripple} = bw \cap (bw \circ SE^{\odot[\alpha]}) \\ R = \frac{area(bw_{ripple})}{area(bw)} \end{cases} \quad 4.8$$

In Equation 4.8, bw is the generated binary mask from the input LR, bw_{ripple} is the binary mask of the ripple pattern, \circ represents the opening morphological operator, $SE^{\odot[\alpha]}$ is a circular structuring element with a radius of α , $area(\dots)$ represents the function that calculates the area of the input binary mask, and R is the *ripple pattern* feature. In this work, we used $\alpha = 3$ as the size of the structuring element.

After the feature extraction step, we classify the remaining structures as bronchi or alveoli using a decision tree based classifier. The classifier is trained on FOV images extracted from eight WSIs and will be tested on the remaining two WSIs at each fold. In our dataset, the population ratio of the bronchus samples to alveoli samples is $\sim 1:100$. Therefore, we perform synthetic oversampling of minority class method (SMOTE)[226] to address the class imbalance in our dataset. In this approach, the synthetic observations are created based on the existing minority observations. For each minority class observation, SMOTE calculates k -nearest neighbors. Depending on the number required oversampled observations, one or more of the k -nearest neighbors are selected to create the synthetic examples.

After identification of the LRs of the bronchi structures, the vessels and the bronchi structures LR will be used in the next step for extraction of the boarder-wall region.

4.2.3.5 Border-wall extraction

As seen in Figure 4.7.a, all of the biological structures in the FOV image are surrounded by a border-wall. From a biological standpoint, these border-wall regions are considered as a part of each structure in the lung tissue. Therefore, the border-wall corresponding to each vessel and bronchus LR, are extracted. Here, we use a region growing approach similar to that of the extraction of the vessels LR. First, we acquire the initial seed pixels using the surrounding neighboring pixels of each structure LR (see Figure 4.7.g). The initial seed pixel is then iteratively expanded using Equation 4.9 until the entire border-wall region of the selected LR is extracted.

$$BW_{border} = \left\{ \left\{ BW_{border} \oplus SE_1^{\odot[\alpha]} \right\} \& P_m \right\} \circ SE_2^{\odot[T]} \quad 4.9$$

In Equation 4.9, BW_{border} is the binary mask containing the border-wall pixels, \oplus is the morphological dilation operator, $SE_1^{\odot[\alpha]}$ is a circular structuring element with radius of α , $\&$ indicates the arithmetic AND operation, P_m is the purple color binary mask, and \circ is the morphological opening operation, $SE_2^{\odot[T]}$ is a circular structuring element with radius of T . In this work, we used $\alpha = 2$ and $T = 3$ as the size of the structuring element.

In order to define the expansion stopping criteria, we take advantage of the fact that the border-wall surrounding each vessel and bronchus LR is relatively thicker than the alveoli's border-wall (*i.e.*, septa). As a result, it is possible to separate the border-wall of the vessel and bronchi from septa by measuring the thickness of the wall (see Figure 4.7.h). The erosion criteria is implemented as the morphological opening term in Equation 4.9,

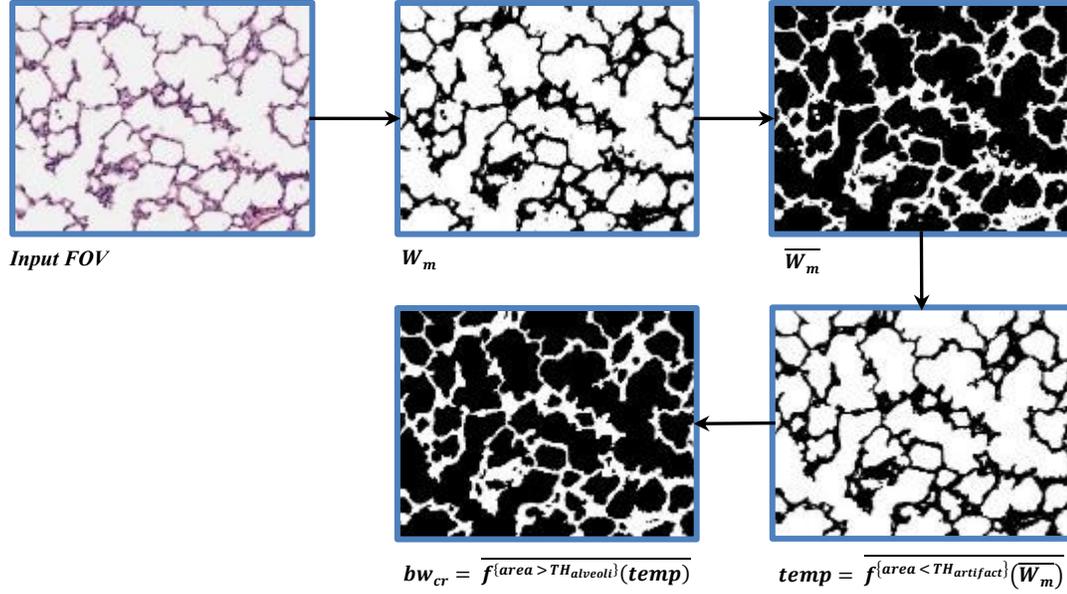


Figure 4.8 An example of septa segmentation in a field-of-view (FOV) image.

using a circular structuring element with radius of T . The optimized value of the T over our training data (extracted FOV images, which were used for training at each fold) was 16 pixels in average.

4.2.3.6 Intersection counting

In order to calculate the MLI score, we are required to count the total number of intersections of the horizontal guideline with septa within the FOV image. First, we reject any FOV image, where the horizontal guideline touches a vessel or bronchus. Then, in the remaining FOV images, we segment the septa regions using Equation 4.10. Figure 4.8 visualizes different steps in segmentation of septa region in a FOV image.

$$\begin{cases} temp = \overline{f^{\{area < TH_{artifact}\}}(\overline{W_m})} \\ bw_{cr} = \overline{f^{\{area > TH_{alveoli}\}}(temp)} \end{cases} \quad 4.10$$

In Equation 4.10, bw_{cr} is the estimated septa region, $f^{\{area < TH_{artifact}\}}(.)$ is an operator that removes the contours with area less than the threshold $TH_{artifact}$, and $f^{\{area < TH_{alveoli}\}}(.)$ is an operator that removes the contours with area larger than the threshold $TH_{alveoli}$. After extraction of septa regions, it is straightforward to count the total number of intersections in each FOV image.

The values of $TH_{alveoli}$ and $TH_{artifact}$ were determined using the grid search hyper-parameter optimization algorithm[227] (see Figure 4.9). In order to acquire the optimum values of the hyper-parameters at each fold, the algorithm was run multiple times (number of grid points used in the grid search) using different hyper-parameters values. The cost value (calculated using Equation 4.11) was observed in order to capture the hyper-parameters associated with the minimum cost values.

$$Loss = \frac{1}{m} \times \sum_{i=1}^m (\widehat{CR}_i - CR_i)^2 \quad (4.11)$$

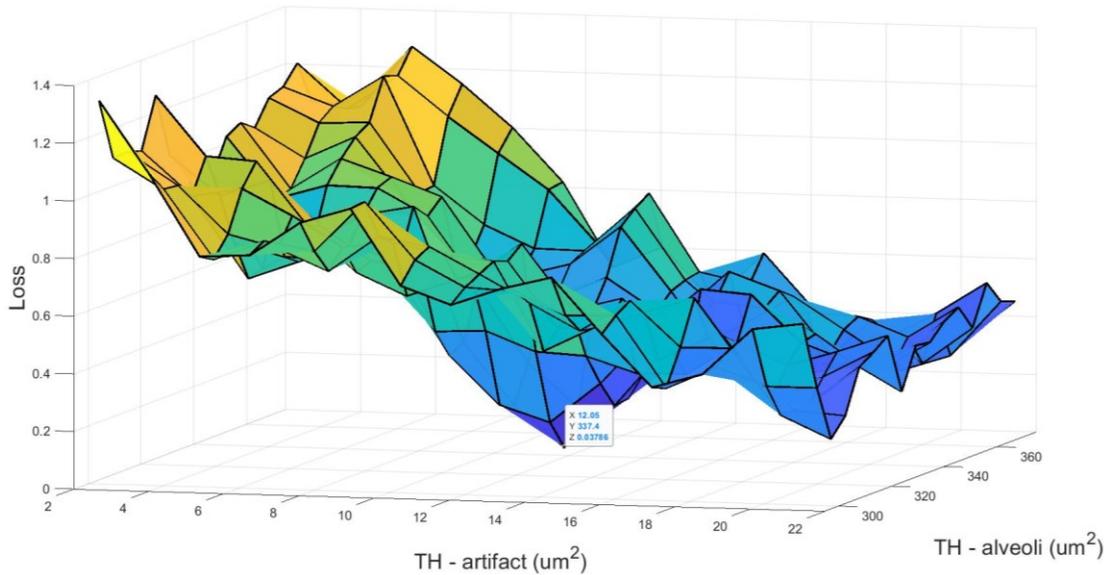


Figure 4.9 An example of hyper-parameters grid search at each fold. At each fold, the values of $TH_{alveoli}$ and $TH_{artifact}$ were determined over eight training WSIs and will be tested on two remaining WSIs in the test set. This procedure continues until all WSIs used as test set.

where, \widehat{CR}_i is the estimated intersections in the i^{th} FOV image, CR_i is the true number of intersections according to the manual assessment, and m is the batch size.

4.2.4 Evaluation of the developed method

The overall performance of the proposed method is affected by three main steps: 1) Foreground extraction, 2) detection of main biological structures, 3) intersection counting. Performance metrics are provided for each step. We performed 5-fold cross validation. Within each fold, we used eight of the WSIs from the dataset (four from *RA* and four from *O2 + LPS*) as the training dataset, and the remaining two WSIs were used as the test dataset in each stage of the proposed pipeline. This was repeated five times such that each WSI was used as the test set.

4.2.4.1 Foreground extraction

In the foreground extraction step, the training dataset (i.e., eight WSIs) was used for training and validation (90% training and 10% validation) of the CNN and the test dataset (i.e., remaining two WSIs) for testing. On average, 28,092 image patches were generated from the eight WSIs at each fold, which were assigned to the training and validation sets with a 9:1 ratio (i.e., 25,283 and 2,809 for training and validation, respectively). The remaining two WSIs in each fold, used for the test dataset, had the average of 7,023 image patches.

The proposed foreground extraction results were compared with the manual segmentation. Manual segmentation of the foreground was performed on each individual WSI image using ImageJ software (<https://imagej.net/>) by the first author (S.S.). The manual segmentation was used as the ground truth for our foreground extraction approach.

To quantitatively evaluate the performance of our segmentation method, we used both region-based and boundary-based metrics. For region-based, we computed the Dice similarity coefficient[228] (DSC) and pixel-wise accuracy (AC).

$$DSC = \frac{2 \times |X \cap Y|}{|X| + |Y|} \quad (4.12)$$

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4.13)$$

Where,

- X: are the set of pixels within the ROI region in the manual segmentation
- Y: are the set of pixels within the ROI region in the CNN segmentation
- $|\cdot|$: is the cardinality of the set
- True positive (TP): ROI region is correctly detected as ROI region (*i.e.*, $|X \cap Y|$).
- True negative (TN): Non-ROI region is correctly detected as Non-ROI region (*i.e.*, $|\overline{X \cup Y}|$)
- False positive (FP): Non-ROI region is incorrectly detected as ROI region (*i.e.*, $|Y| - |X \cap Y|$)
- False negative (FN): ROI region is incorrectly detected as non-ROI region (*i.e.*, $|X| - |X \cap Y|$)

For boundary-based, the Hausdorff distance [229] between the border of the segmented WSI and that of the ground truth was calculated. The Hausdorff distance is used to calculate the maximum distance of a set to the nearest point in the other set:

$$H(A, B) = \max_{a \in A} \{ \min_{b \in B} \{ d(a, b) \} \} \quad 4.14$$

where, H is the Hausdorff distance, a and b represent the points of the objects A and B , and $d(a, b)$ is the distance between a and b .

4.2.4.2 Detection of main biological structures

For our dataset, we extracted 18,321 FOV images of size 1072×1388 pixels (9,009 FOV images from the *RA* group and 9,312 FOV images from the *O2 + LPS* group). At training phase, we used the ground truth foreground extraction as the guideline for extraction of FOV images in order to generate the training data. At each fold, an average of 14,657 FOV images extracted from the eight WSIs, which were used for training of the later steps (*i.e.*, bronchi vs alveoli classification) of our pipeline. At the testing phase, the CNN model was used to extract FOV images from the test dataset (*i.e.*, the remaining two WSIs). An average of 3,664 FOV images were used for testing at each fold.

Manual MLI scoring was conducted independently by co-authors S.S. and S.L. (both are considered novice scorers, who were provided 5 hours of training in B.T.'s laboratory), and co-author M.L. (verified by co-author B.T.; and considered an expert scorer with three years of experience), referred to as rater 1, rater 2, and rater 3, respectively.

The detection of vessels and bronchi are necessary to properly reject FOV images, where the horizontal guideline touches one of these biological structures. In this step, the

$$Precision = \frac{TP}{TP + FP} \quad 4.15$$

$$Recall = \frac{TP}{TP + FN} \quad 4.16$$

$$F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad 4.17$$

metrics in Equation 4.14 – 4.16 were utilized to evaluate the performance of the proposed method in detection of the accepted/rejected FOV images.

We compared the performance of the proposed method against the ground truth, which consists of manual scores for calculation of the MLI for each WSI by three human raters. Our ground truth includes the assessment of accepted/rejected FOV images and the number of intersections with alveoli wall in each accepted FOV image. To evaluate our method in detection of accepted/rejected FOV images, two evaluation approaches were conducted:

1. Acquiring the ground truth for accepted/rejected FOV images based on the majority vote (*i.e.*, correct if it agrees with at least two of the raters).

- True positive (TP): a FOV image is rejected, and at least two raters rejected the FOV image
- False positive (FP): a FOV image is rejected, and at least two raters accepted the FOV image
- False negative (FN): a FOV image is accepted, and at least two raters rejected the FOV image

2. A less restrictive approach, where the automated method is deemed correct if it agrees with at least one of the raters.

- True positive (TP): a FOV image is rejected, and at least one rater rejected the FOV image
- False positive (FP): a FOV image is rejected, and no rater rejected the FOV image
- False negative (FN): a FOV image is accepted, and no rater accepted the FOV image

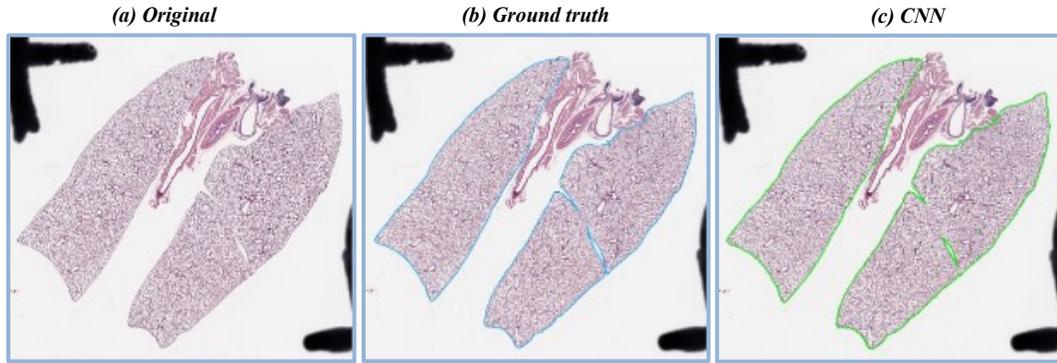


Figure 4.10 An example of the complete foreground extraction approach. a) The original whole slide image. b) The manually segmented ground truth mask. c) The proposed method segmentation mask.

4.2.4.3 Intersection counting

In order to evaluate the reliability of agreement between the manually assessed FOV images and the algorithm-generated assessment, the Fleiss' Kappa[230] statistical measure is calculated in each study group (*i.e.*, *RA* and *O2 + LPS*). The Fleiss' Kappa score reflects the performance of our method for counting the number of intersections in accepted FOV images.

As the final step in the evaluation of the results, the MLI scores were calculated and compared to that of the manually assessed scores.

4.3 Results

Figure 4.10 visualizes an example of foreground extraction over a histopathology WSI and the corresponding manual segmentation. Table 4.1 contains performance metrics for foreground extraction, which indicate that the proposed method performed well for all WSIs.

Table 4.2 Distribution of extracted field-of-view (FOV) images. The “*completely inside pleural space*” column represents the FOV images that contain no specific tissue compartments and therefore, are eliminated in the foreground extraction step. The remaining FOV images are used in detection of main biological structures step to identify the accepted FOV images.

<i>Fold #</i>	<i>FOV images in the dataset</i>	<i>Completely inside pleural space</i>	<i>Number of rejected FOV images</i>		<i>Number of accepted FOV images</i>
			<i>Intersection with pleural space</i>	<i>Intersection with vessel/bronchi</i>	
<i>1</i>	3,219	1,637	249	334	999
<i>2</i>	3,024	1,444	182	300	1,098
<i>3</i>	4,266	2,666	185	276	1,139
<i>4</i>	3,780	2,180	256	302	1,042
<i>5</i>	4,032	2,432	227	291	1,082
<i>Total</i>	18,321	10,359	1,099	1,503	5,360

In order to evaluate the performance of the proposed method in detection of accepted FOV images, the precision, recall, and F1-score metrics are calculated for each fold for the *RA* group and the *O2 + LPS* group. Table 4.2 represents the distribution of rejected and accepted FOV images in our dataset. In Table 4.3, the results for evaluation of our proposed method in detection of FOV images for different evaluation approaches are presented. As it is shown in Table 4.3, the *RA* group as compared to *O2 + LPS* group has higher precision and lower recall in both evaluation approaches, while the F1-scores are similar. Examples

Table 4.1 The evaluation of the foreground extraction step performance using pixel-wise accuracy (AC), Dice coefficient (DSC), and Hausdorff distance.

<i>Image</i>	<i>AC (%)</i>	<i>DSC (%)</i>	<i>Hausdorff distance (μm)</i>
Average score	98.34	98.22	109.68
Maximum	98.95	98.58	182.85
Minimum	97.50	97.49	39.55

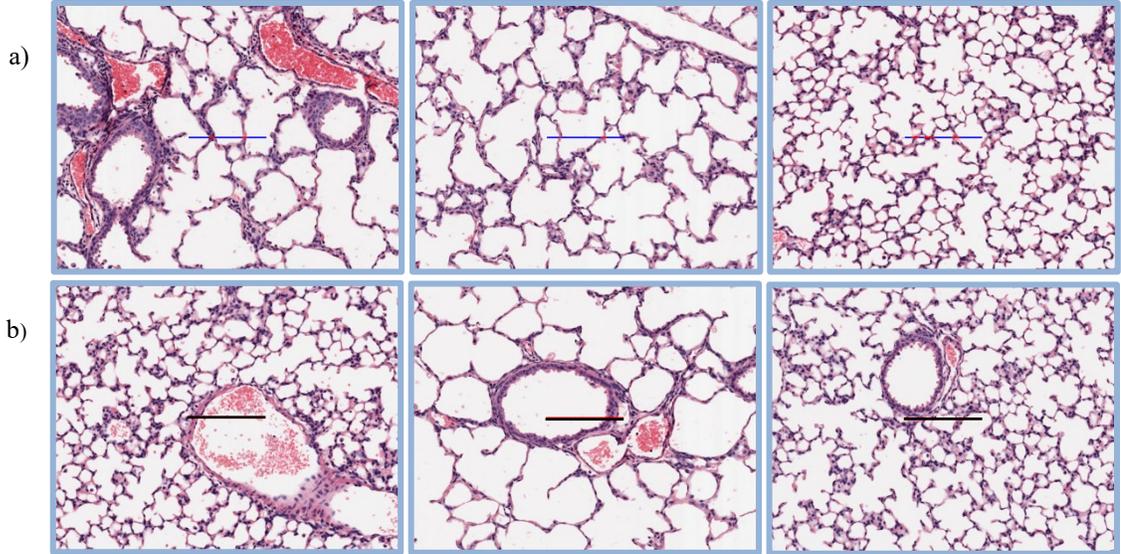


Figure 4.11 Visualization of the automated assessment results. a) Accepted field-of-view (FOV) images. The intersections with the septa are marked as red. The blue color indicates that there are no intersections. b) Rejected FOV images.

of automated FOV image assessment using our proposed method are visualized in Figure 4.11.

Another step in determination of the overall performance is the accuracy by which our method can count the number of intersections in each accepted sample FOV image. Figure **Table 4.3** Evaluation of the proposed method performance in detection of field-of-view (FOV) images. For evaluation approach (1), the decisions are correct if the automated method agrees with at least two out of three raters. For evaluation approach (2), the decisions are correct if the automated method agrees with at least one rater out of three raters. The results are presented for five different folds for each group. The abbreviation STD stands for standard deviation. In this table, the abbreviations *RA* and *O2 + LPS* refers to the mice group housed in room air and mice group were exposed to a high concentration of oxygen and lipopolysaccharide, respectively.

		<i>RA</i>					<i>O2 + LPS</i>					<i>Total</i>		
		<i>Fold #1</i>	<i>Fold #2</i>	<i>Fold #3</i>	<i>Fold #4</i>	<i>Fold #5</i>	<i>Mean ± STD</i>	<i>Fold #1</i>	<i>Fold #2</i>	<i>Fold #3</i>	<i>Fold #4</i>	<i>Fold #5</i>	<i>Mean ± STD</i>	<i>Mean ± STD</i>
<i>Evaluation approach (1)</i>	<i>Precision (%)</i>	89.17	96.50	93.68	97.09	97.42	94.77 ± 3.46	84.25	76.56	88.16	86.64	90.59	85.24 ± 5.37	90.01 ± 6.59
	<i>Recall (%)</i>	68.81	60.79	69.26	76.50	72.14	69.50 ± 5.75	88.17	72.32	83.40	82.72	81.00	81.52 ± 5.79	75.51 ± 8.35
	<i>F1-score (%)</i>	77.68	74.59	79.64	85.58	82.89	80.08 ± 4.31	86.16	74.38	85.71	84.63	85.53	83.28 ± 5.01	81.68 ± 4.72
<i>Evaluation approach (2)</i>	<i>Precision (%)</i>	93.75	97.90	94.74	98.91	98.97	96.85 ± 2.44	90.41	80.22	93.47	92.67	92.68	89.89 ± 5.52	93.37 ± 5.45
	<i>Recall (%)</i>	74.01	73.30	78.95	81.44	80.67	77.67 ± 3.78	92.96	80.81	90.16	92.27	90.17	89.27 ± 4.89	83.47 ± 7.37
	<i>F1-score (%)</i>	82.72	83.83	86.12	89.33	88.89	86.18 ± 2.95	91.67	80.51	91.78	92.47	91.41	89.57 ± 5.08	87.87 ± 4.30

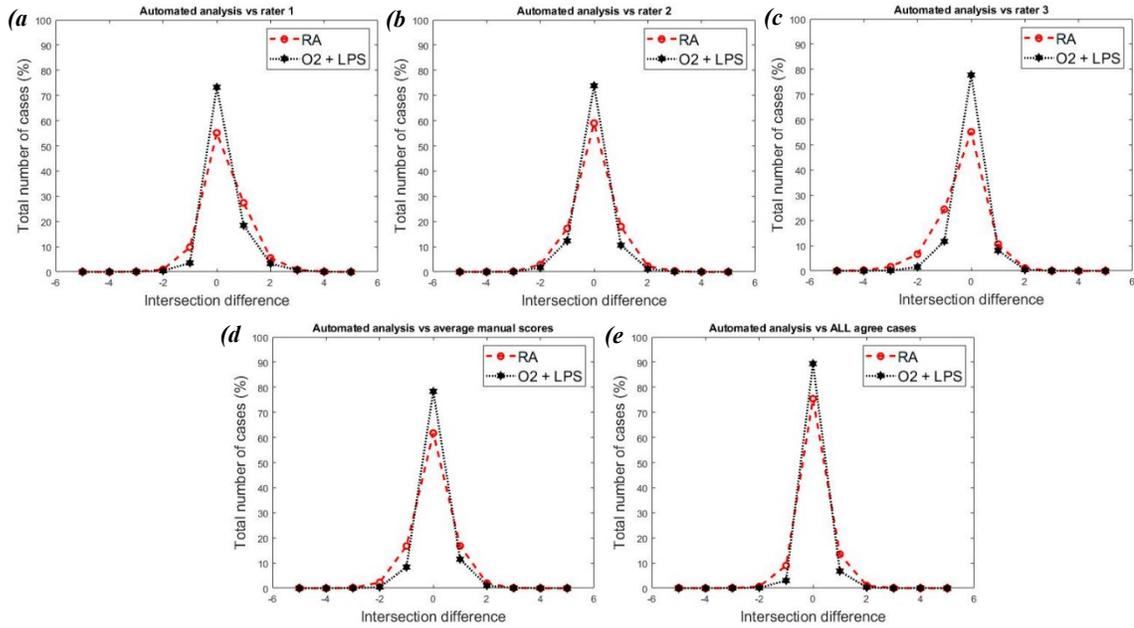


Figure 4.12 The evaluation of the proposed method in detection of the intersections against the human raters. (a), (b), and (c) represent the intersection difference between the automated method and each individual raters scores. (d) shows the comparison between the generated intersections by the proposed method and the average number of intersections reported by human raters. (e) represents the comparison in instances that all human raters agree on number of intersections (*e.g.*, all raters agree that there are five intersections with alveoli septa in the corresponding field-of-view (FOV) image).

4.12 represents the comparison between the proposed method accuracy in detection of the intersections in accepted FOV images against the manually generated scores by human raters. Rater 1 exhibits a slight negative skew (Figure 4.12.a) and rater 3 a slight positive skew (Figure 4.12.c), rater 2 (Figure 4.12.b) and the rater's average (Figure 4.12.d) distributions are quiet symmetric. Figure 4.12.d shows a comparison between the proposed method intersection counting scores and the rater's average score (*i.e.*, the arithmetic mean over the reported intersections by human raters and rounding it to the nearest whole number). In Figure 4.12.e, the intersection difference between the proposed method and the manual scores is visulized, when all three raters were agreed on the number of intersections.

The calculated Fleiss' Kappa scores are presented in Table 4, showing the reliability of agreement in counting the number of intersections in accepted FOV images between

Table 4.4 The evaluation of the reliability of agreement in detection of intersections in accepted field-of-view (FOV) images between the proposed method and the manual assessments using Fleiss’ Kappa score.

	<i>RA</i>					<i>O2 + LPS</i>					
	<i>Rater #1</i>	<i>Rater #2</i>	<i>Rater #3</i>	<i>Rater’s average</i>	<i>All agree cases</i>	<i>Rater #1</i>	<i>Rater #2</i>	<i>Rater #3</i>	<i>Rater’s average</i>	<i>All agree cases</i>	<i>Overall score</i>
<i>Weighted percent agreement (PA)</i>	0.55	0.59	0.55	0.59	0.75	0.78	0.74	0.73	0.77	0.89	0.82
<i>Weighted percent chance agreement (PE)</i>	0.22	0.23	0.24	0.22	0.25	0.30	0.30	0.32	0.31	0.33	0.29
<i>Kappa score</i>	0.42	0.47	0.41	0.46	0.67	0.68	0.63	0.61	0.66	0.84	0.76
<i>Agreement assessment</i>	Moderate	Moderate	Moderate	Moderate	Substantial	Substantial	Substantial	Substantial	Substantial	Perfect	Substantial

manual assessment by human raters and automated algorithm. In Table 4.4, the qualitative interpretation of scores’ agreement in “agreement assessment” section has been derived from Landis and Koch paper[231].

Figure 4.13 shows a comparison between calculated MLI scores by the three raters and the automatically generated MLI scores by the proposed method. As it can be seen in this

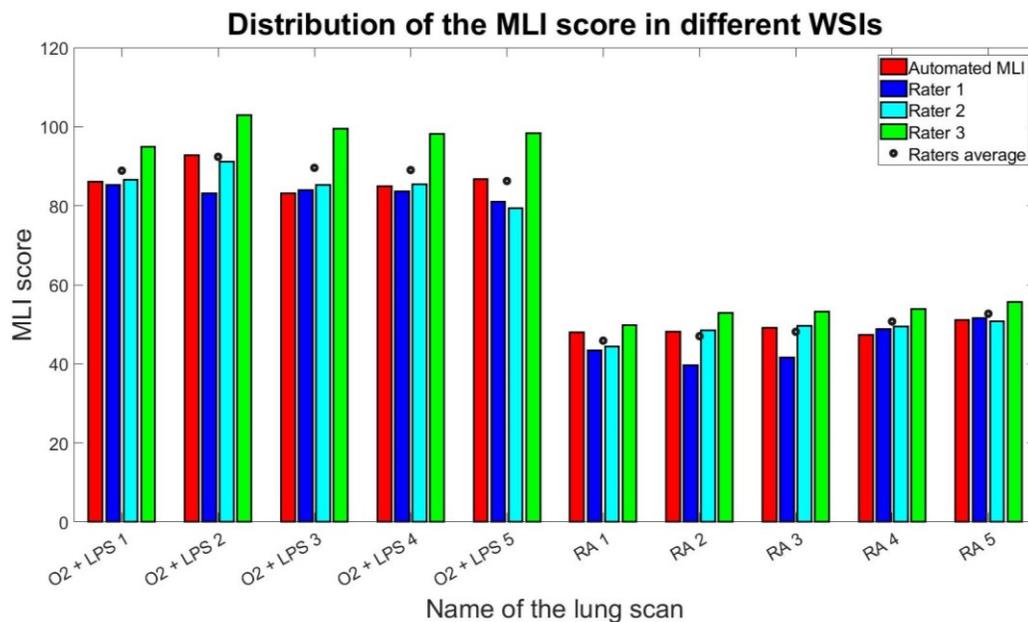


Figure 4.13 The evaluation of the calculated mean linear intercept (MLI) score by the proposed method against human raters.

figure, there is noticeable variability in the MLI scores between individual raters and the automated MLI score falls within this variability. All raters and the automated method successfully discern the *RA* and *O2 + LPS* conditions based on their MLI score.

4.4 Discussion

The current approaches in estimation of MLI score involves manual/semi-automated assessment of histopathology images[219], [220], [232]–[234], which can be time-consuming, subjective, and expensive. In recent years, there has been a growing interest in developing fully automated methods to encounter the inefficiencies of manual/semi-automated MLI scoring but none of these methods addressed the segmentation of lung main biological structures, which plays a vital role in estimation of MLI score. Jacob *et al.* [232] proposed an automated approach for estimation of MLI score in histopathology images of mouse lung tissue. However, their proposed method was not applied to WSIs and the authors used a few selective images to provide their experimental results. Rieger-Fackeldey *et al.* [235] proposed an automated method for estimation of MLI score to study the effects of hyperoxia in histopathology images of newborn mice lung tissue. The authors used a digital image analysis software (Image Pro Plus version 4.0) and a custom macro commands for automated investigations of alveolar morphological characteristics. However, they provided no technical details on their proposed method, nor how well it worked in comparison to manual assessments by human raters. Sallon *et al.* [217] proposed an automated approach for estimation of MLI score in WSIs of mouse lung tissue. The authors used a thresholding approach and closed contour assessment based on size to classify the alveolar structures. However, their proposed algorithm was unable to distinguish between main lung biological structures (alveoli, bronchi, and blood vessels).

To the best of our knowledge, our fully automated pipeline is the first approach capable of comprehensively account for main challenges involved with estimation of MLI score, including 1) taking a histopathology WSIs as input and extract the diagnostically relevant tissue compartments for extraction of FOV images, 2) screening of the FOV images, rejecting images based on presence of certain biological structures (bronchi and blood vessel).

To evaluate the performance of our proposed method, we compared our work against MLI scores from three human raters. Using ten high-resolution WSIs of mouse lung tissue, comprised of two distinct experimental groups (*i.e.*, *RA* and *O2 + LPS*). We independently tested the performance of each step in our pipeline against manual assessment.

In extraction of the foreground regions, the proposed approach showed promising performance in removing the imaging artifacts and undesired biological components and identifying the ROIs, yielding 98.34%, 98.22%, and 109.68 μ m, *AC*, *DSC*, and Hausdorff distance, respectively.

We proposed two different approaches to evaluate the performance of our proposed method in detection of the main biological structures in histopathology images of mouse lung tissue. In evaluation approach (1), the proposed method was able to detect the rejected FOV images with mean precision, recall, and F1-score of 90.01%, 75.51%, and 81.68%, respectively. In evaluation approach (2), the proposed method was able to detect the rejected FOV images with mean precision, recall, and F1-score of 93.37%, 83.47%, and 87.87%, respectively. The higher performance metrics of evaluation approach (2), as compared to evaluation approach (1), can be an indication of the subjectivity involved with the process of calculating the manual MLI scores. Two main observation can be drawn

from the Table 3. First, the mean precision in *RA* group is higher in comparison to the mean precision in *O2 + LPS* group (94.77% versus 85.24%). This indicates that the number of incorrectly rejected FOV images (i.e., FPs) are higher in *O2 + LPS* group. This may be related to fact that the alveoli's LR are enlarged in the *O2 + LPS* group. As we used the area of the LR as a feature in our classification approach, this may result in more misclassifications of alveoli as bronchi (higher number of FPs). This may potentially explain the higher precision in *RA* group compared to *O2 + LPS* group. Second, the mean recall in *RA* group is much lower in comparison to *O2 + LPS* group (69.50% versus 81.52%). This indicates that the total number of incorrectly accepted FOV images (i.e., FNs) are much higher in *RA* group. In our dataset, we noticed that the density of remaining blood cells in *RA* group was lower in comparison to *O2 + LPS* group. Since we used the presence of blood cells as a feature to identify blood vessels, the reduced amount of remaining red blood cells in the *RA* group results in increased number of FNs.

In classification of bronchi versus alveoli structures, as the alveoli structures are enlarged in *O2 + LPS* group, there is a possibility of a bronchi misclassified as a bronchi as alveoli or vice versa. If an alveoli is misclassified as a bronchi, the FOV would be rejected and may result in underestimation of MLI. If a bronchi is misclassified as an alveoli, the FOV would not be successfully rejected and this may result in an overestimation in MLI, as bronchi tend to be larger than alveoli. These types of misclassifications do not appear to be happening with any noticeable frequency in our dataset. We speculate that the occasional misclassifications would also not have a large impact, as the MLI is computed across a large number of FOVs.

We also used Fleiss' Kappa score to measure the reliability of agreement between the human rater's and automatically generated intersections. The mean Fleiss' Kappa scores were 0.46 and 0.66 for the *RA* group and the *O2 + LPS* group, respectively. This shows that the proposed method has a slightly higher agreement with the *O2 + LPS* group than the *RA* group against the average manual. We hypothesize two contributing factors for this difference: 1) The MLI scoring task in *RA* group is a more complex task compared to *O2 + LPS* group. The Fleiss Kappa scores among human raters in *RA* group and *O2 + LPS* group were 0.5949 and 0.6991, respectively. This indicates that the images in the *RA* group were inherently more difficult to analyze. 2) As it is mentioned in section 2.3.6, a contributing factor in preparation of the septa region for automated intersection counting is the value of TH_{alveoli} , which represent the minimum area of the alveoli's lumen region. Using this threshold, the white objects that are smaller than TH_{alveoli} are eliminated from the process of intersection counting. We optimized this value over both study groups (*RA* and *O2 + LPS*). As the alveoli's LR are enlarged in the *O2 + LPS* group, we expect that the calculated TH_{alveoli} would result in more misclassifications of alveoli in *RA* group compared to the *O2 + LPS* group, affecting the agreement between human raters and the proposed method.

In Table 4.4, we observe that the rate of agreement in both experimental groups is increased considerably, when all raters agree on the number intersections (Fleiss' Kappa scores of 0.67 and 0.84 in *RA* group and *O2 + LPS* group). This is also another indication that the subjectivity of the manual analysis can dramatically influence the results. The mean difference between the calculated MLI score between the automated method and average rater's score was 2.33 (4.25%) with standard deviation of 4.13 (5.67%), which shows that

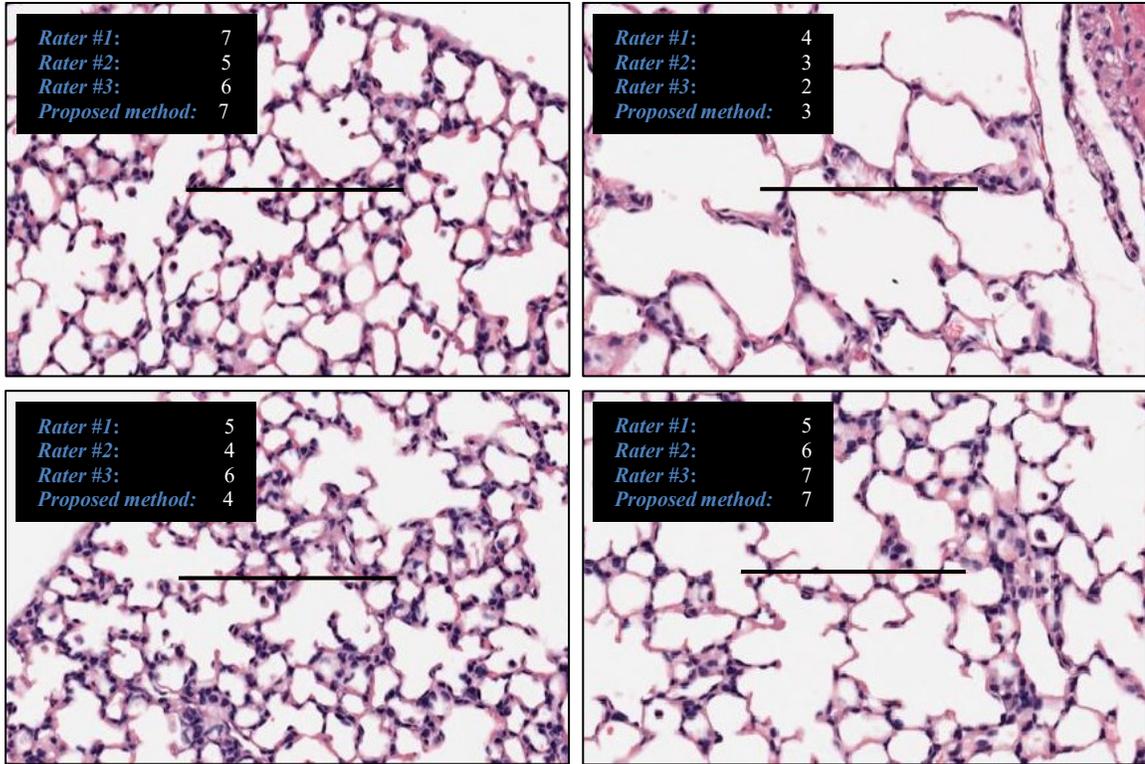


Figure 4.14 The variability in intersection counting by human raters.

our proposed method has the ability to accurately estimate the MLI scores with regards to manual scores calculated by human raters. A Student's t-test was performed to see if the average MLI scores of the human raters and the MLI scores from the proposed method could statistically differentiate the *RA* and *O2 + LPS* groups ($\alpha = 0.05$, with a Bonferroni correction). Results demonstrate a statistically significant difference for both the human raters ($p = 5.65 \times 10^{-9}$) and proposed method ($p = 4.03 \times 10^{-9}$).

There is a strong agreement between the manual assessment and the proposed method, when the average intersections are calculated for all three raters (Figure 4.12.d). Therefore, the proposed method agrees with the average raters' score in 69.79% of cases and the agreement will include 96.77% of cases, if a maximum of one intersection difference is included in the calculations. The agreement between manual scores and the proposed method is stronger, when all raters counted the exact same number of intersections (see

Figure 4.12.e). In the cases where all raters agree on the number of intersections, the proposed method counts the exact same number of intersections with a probability of 83.84% or counts a maximum of one intersection difference with probability of 98.77%. Considering the fact that the proposed method is fully automated and completely reproducible, this result can be an indication of reproducibility issues in detection of intersections in FOV images by human raters. To demonstrate the subjectivity in the manual MLI scoring, a few examples of intersection counting are shown in Figure 4.14. This discrepancy in their counting could imply the raters may have a particular bias in their MLI scoring. The results suggest the relative MLI scorings of the proposed method in comparison to each individual rater's scores are constant (see Figure 4.13). As a result, even if the automated method has a bias itself, it would be constant and highly reproducible.

The processing time for foreground extraction within a single fold was approximately ~3 hours and few minutes for CNN training and testing, respectively. This was performed on a standard workstation with an Intel Core i7-3770 3.40 GHz CPU, 12 GB of installed RAM, and a single NVIDIA RTX 2060 with 6 GB memory. The processing time required for the remaining steps of our proposed pipeline (*i.e.*, color slicing, detection of the main biological structures, and intersection counting) was ~37 hours and ~8 hours for each fold for training and testing, respectively. The vast majority of this time is attributed to detection of main biological structures. The average time required to manually calculate the MLI score for each WSI was approximately ~10 hours. In total, the time required for the proposed method and a human rater to score the entire dataset (10 WSIs) was approximately 40 hours and 100 hours, respectively. Considering the fact that the algorithm can automatically run in the background 24/7 with no supervision, the manual scoring of

the dataset by a human rater that may take up to several weeks for a human rater, can be achieved by the proposed method in less than two days.

To develop our CNN model for foreground segmentation we used Tensorflow and Keras frameworks, and the Python programming language. Other steps are developed using Matlab environment.

It should be noted that optimizing processing time was not a focus of this work. As such, large reductions are likely possible. Reduction of processing time could be easily achieved by leveraging parallel processing capabilities.

In our work, we faced various limitations and challenges that we were not able to address in our proposed pipeline. One of the limitations of our work was detection of the seed regions in segmentation of the blood vessels. We used the remaining blood cells in the perimeter of each vessel as an indication of existing blood vessel. Although, the majority of the blood vessels in our dataset had remaining blood cells in their perimeters, there were some cases with no blood cells present. As a result, in these cases our algorithm fails to correctly identify and segment the vessel regions. The complexity of blood vessels made it difficult for us to find a suitable handcrafted feature, which could be effective in identification of these structures. Another limitation was the way we have evaluated the performance of our proposed method in segmentation of biological structures. Although, the presented evaluation procedure can be a good indication of how detection error propagates into calculation of MLI scores, it does not evaluate the segmentation performance of our proposed approach in the detection of the lung structures. The main barrier is the unavailability of manual segmentation of the various biological structures of lung tissue in our dataset. Our future work is to develop a database of images with manual

segmentation, which will also support the development of more advanced supervised learning segmentation algorithms, such as deep learning, that we anticipate will improve the overall segmentation performance.

In this work, we developed a pipeline to measure the changes in the lung architecture observed in mice with experimental bronchopulmonary dysplasia compared to control, healthy animals. The pathology of bronchopulmonary dysplasia in humans is more complex than in mice. In humans, the disease is variable within the lungs of one patient and there is more variability between patients. However, by accounting a wider variability in disease pathology, it is feasible that in the future, the current pipeline to be translated for use in human lung histopathology for bronchopulmonary dysplasia.

4.5 Conclusion

In this paper, we proposed a new pipeline for automating the estimation of the MLI score. The proposed method uses U-Net architecture for segmentation of diagnostically relevant tissue specimens, which yielded accurate results. Our proposed method utilized color image analysis and region growing for segmentation of the main biological structures (bronchi and vessels) in histopathology images of mouse lung tissue, which showed promising performance. The comparison between the automated method and the manual assessment showed substantial agreement in the calculation of the MLI score. The result demonstrated that the proposed method could replace the manual/semi-automated methods for calculating the MLI score.

Chapter 5: Deep learning based image segmentation in automated histopathology image analysis

5.1 Introduction

As demonstrated in Chapter 3 and Chapter 4, rule-based approaches, using conventional machine learning and image processing techniques, can obtain relative success in AHIA; however, there remain challenges with rule-based approaches, which justifies the exploration of deep learning approaches in AHIA. The main challenges in rule-based approaches are: 1) limited performance, 2) generalizability, and 3) processing time.

One of the main challenges associated with rule-based approaches is that the overall performance of the model is closely linked to the accuracy of the extracted handcrafted features. As such, in many rule-based AHIA approaches, consultation with expert pathologists plays a vital role. However, despite the availability of accurate features, it can be difficult to develop a rule-based approach that can account for the complexity and variability that often exists in histopathological images of tissue. While the segmentation performance can be increased by increasing the complexity of the rule-based approach, the performance increase will quickly plateau. Practically, this will limit the performance of a rule-based approach.

The challenge of generalizability with rules-based approaches refers to the limitation when transferring the pipeline architecture from one application to another. While there may be some shared methods between applications, there is often a considerable amount of time and effort required to develop a rules-based approach for each application. This can be seen when examining the rules-based approach for villi segmentation (Chapter 3) and rules-based approach used for lung segmentation (Chapter 4); while there are shared

methods (*e.g.*, color extraction), the overall method needed to be uniquely developed for each application.

Finally, the rules-based approaches typically take a long time to process. For instance, the average processing time required for our rules-based approach to process a single WSI (on a Core i7-6700 3.4 GHz work station with 16 GB of RAM) for villi segmentation (Chapter 3) and lung segmentation (Chapter 4) was approximately 16 hours and 5 hours, respectively.

In recent years, deep learning approaches – CNNs in particular – have demonstrated the ability to obtain excellent performance in a variety of computer vision tasks, including semantic and instance segmentation [236], object tracking [237], and image classification [238]. This is mainly because of the unique capabilities of the deep learning methods in the learning hierarchy of increasingly complex features [239], which enables them to learn a wide range of complex patterns from the given data. Given the successes of deep learning in image segmentation, often outperforming rules-based approaches [240], a deep learning approach may provide a large increase in performance for image segmentation in AHIA. Deep learning architectures are also generalizable, where the same architecture can be used for different segmentation tasks. While the training phase in deep learning can be time-consuming, the inference phase, where images are segmented, is relatively quick.

In this chapter, we describe a generic deep learning framework for image segmentation in AHIA. To develop our deep learning framework, we experiment on three widely-used state-of-the-art semantic segmentation models: 1) U-Net [158], 2) SegNet [241], and 3) Dilated-Net [242]. Each architecture has unique characteristics, which enables developing a high-performance generic deep learning model. A more detailed discussion on these

characteristics will be provided in section 2.5.2. We will evaluate the performance of these models in terms of segmentation performance and ability to address the histopathology applications, which were addressed in Chapter 3 and Chapter 4 for segmentation placental villi and automated calculation MLI score, respectively.

5.2 Deep learning-based image segmentation

Development of a deep learning segmentation workflow in AHIA comprises of multiple steps (see Figure 5.1): 1) selection of a deep learning based image segmentation model, 2) image preprocessing, 3) image patch extraction, 4) data augmentation, and 5) model training. In this section, we present a brief overview of these steps in AHIA and discuss the methodology used to implement these steps in our generic deep learning approach.

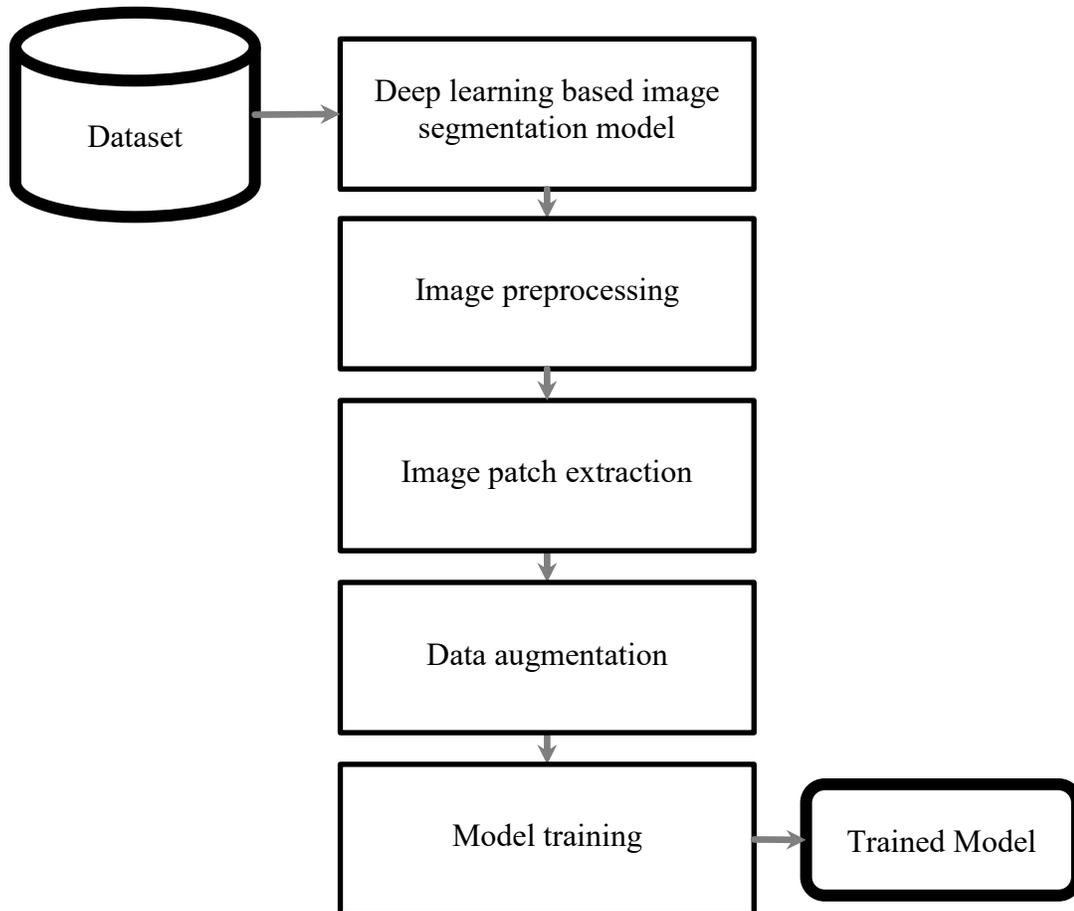


Figure 5.1 A typical deep learning framework in automated histopathology image analysis in the training phase.

5.2.1 Dataset

The histopathology WSIs of human placenta (Chapter 3) and histopathology WSIs of mouse lung tissue (Chapter 4) are utilized to evaluate the performance of the generic deep learning models against conventional rule-based approaches.

5.2.1.1 Histopathology images of human placenta

The placental dataset comprises of high-resolution (scanned at $20\times$ magnification) WSIs of 12 histopathology specimens obtained from the Research Centre for Women's and Infants Health (RCWIH) Biobank (Mount Sinai Hospital, Toronto, ON). The histopathology images of human placenta are taken from two different experimental groups; the healthy control group and the patients diagnosed with PE. All images were taken from different patients and the sizes of the images were up to $39,963\times 31,959$ pixels. From each WSI, three sub-images of size 2740×3964 are extracted, amounting to a total of 36 sub-images (nine images from healthy term pregnancies and 27 from term pregnancies complicated with PE). The single-class placenta dataset consists of two classes; 1) placental villous structures (*i.e.*, Villi), and 2) background components, including imaging artifact, undesired biological components (*i.e.*, non-Villi).

To generate the ground truth segmentation for the placenta dataset, each placental sub-image was segmented using the rule-based approach in Chapter 3 at $20\times$ magnification. Then, the algorithm-generated segmentation masks were manually refined, by the author, using ImageJ software (<https://imagej.net/>). In total, 4,946 villous structures were annotated for the ground truth segmentation of our single-class placenta dataset.

5.2.1.2 Histopathology images of mouse lung tissue

The mouse lung tissue dataset comprises of high-resolution WSIs of 10 lung histopathology specimens of mice obtained from the Sinclair Centre for Regenerative Medicine (Ottawa Hospital Research Institute, Ottawa, ON). The slides were scanned using an Aperio CS2 slide scanner (Leica), and high-resolution color images at 20× magnification (i.e., approximately 0.5µm per pixel) were obtained. The lung dataset consists of five WSIs from healthy mouse lung tissue housed in room air (*RA*) as control group and five WSIs from an experimental group of mice who were exposed to a high concentration of oxygen and lipopolysaccharide (*O2 + LPS*). WSIs were taken from different mice and the size WSIs size were up to 22,430×22,900 pixels. The multi-class mouse lung dataset consists of five classes: 1) Blood Vessel, 2) Bronchus, 3) Alveoli Sack, 4) Alveoli Wall, and 5) Background.

Manual segmentation for these images was generated using a semi-automated approach. As an initial step, the background regions were manually segmented and then, the bronchi and blood vessels were manually annotated using the ImageJ software (<https://imagej.net/>). Using the segmentation pipeline in Chapter 4, the labels for the remaining structures (i.e., alveoli sacs and alveoli walls) are automatically generated at 20× magnification and then manually refined. In total, 4,566 Blood Vessel structures, 510 Bronchus structures, and 298,829 Alveoli Sac structures were annotated for the manual segmentation of our multi-class mouse lung tissue dataset.

Although this approach for establishing the ground truth for placental and mice lung images considerably speeds up the manual segmentation process, it should also be noted that it introduces a bias in evaluation; that is, the performance metrics for the conventional

approaches are anticipated to be higher, as compared to if the ground truth was established completely manually. As such, when evaluating the proposed deep learning method approach, and comparing the performance to the conventional segmentation methods, this bias should be acknowledged.

5.2.2 Deep learning-based image segmentation model

One of the important steps in the development of a deep learning based semantic segmentation pipeline is to find an appropriate segmentation architecture that addresses the specific challenges associated with the application at hand. In recent years, different deep learning segmentation architectures have been proposed for AHIA. Although various factors may contribute to selecting a segmentation architecture, including field-of-view, resolution, size of the input images, and complexity of the given images, there are no fixed rules for determining the optimal segmentation architecture for a specific application. A widely used rule of thumb is to select a previously proposed deep learning model, which performed well on a dataset with similar images. To develop a generic deep learning framework, we initially investigate the characteristics of the state-of-the-art segmentation models based on the unique requirements of our datasets. Then, we compare the performance of these selected models to determine the optimal model to perform segmentation over our datasets.

Here, we briefly introduce three widely used semantic segmentation architectures (*i.e.*, including U-Net, SegNet, and Dilated-Net), which have shown exceptional performance in different computer vision applications and can be considered strong candidates for developing our generic deep learning model.

5.2.2.1 U-Net

As discussed in section 2.5.2, one of the most widely used segmentation architectures is the U-Net. An important characteristic of the U-Net is the use of skip connections, which is referred to as copying and concatenating each layer's feature map between the encoder and the decoder modules. The use of skip connections in the architecture enables the model to compensate for the spatial resolution that is lost in the down-sampling in the encoder module [158]. In addition, studies suggest that the U-Net does not require a very large dataset to obtain high performance and has the capability to receive images with different sizes [158] because of the lack of any dense layers in its architecture. Details of the U-Net architecture can be found in section 2.5.2.

5.2.2.2 SegNet

The SegNet is an efficient pixel-wise semantic segmentation architecture, which is primarily motivated by the road scene understanding applications. The building blocks of the SegNet architecture consist of three main components: 1) the encoder network, 2) the corresponding decoder network, and 3) the pixel-wise classification layer (see Figure 5.2). In the original SegNet paper [241], the encoder network consists of 13 convolutional layers, which corresponds to 13 convolutional layers in VGG16, a powerful object classification CNN. The utilization of VGG16 enables the network to discard the connected layers in favor of retaining higher resolution feature maps at the deepest encoder layer, considerably reducing the number of trainable parameters (*i.e.*, 134 million parameters to 14.7 million parameters [241]). In the encoder network, at each layer, convolution with a filter bank is applied, followed by an element-wise rectified linear non-linearity (ReLU) and max-pooling with the size of 2×2 window and stride 2, which sub-samples the input layer by

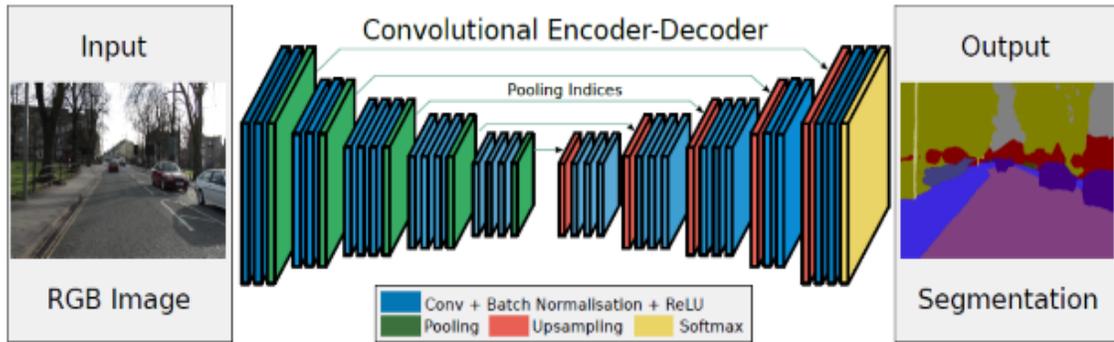


Figure 5.2 The architecture of SegNet [242].

a factor of 2 at the output. While several layers of max-pooling and sub-sampling operations can achieve more translation in-variance for a robust classification, correspondingly, there is a loss of spatial resolution of the feature maps. Therefore, it is necessary to capture and store boundary information in the encoder feature maps before sub-sampling is performed. It involves storing the max-pooling indices, which are the locations of the maximum feature value in each pooling window. As a result, similar to U-Net architecture skip connection procedure, the accurate segmentation of boundary of ROIs using the SegNet architecture can be very memory demanding, specifically in high-resolution images.

In the decoder network, the input feature maps are up-sampled using the memorized max-pooling indices from the corresponding encoder feature maps, generating sparse feature maps. Then, the sparse feature maps are convolved with a trainable decoder filter banks followed by a batch normalization step to produce dense feature maps. At each layer, the batch normalization step is then applied to each of these maps. The final feature map (*i.e.*, output of the decoder) is fed to a trainable softmax classifier, which generates a channel probability map. The predicted segmentation corresponds to the class with maximum probability at each pixel.

In contrast to U-Net, no skip connections are used in SegNet architecture and only the pooling indices from the encoder network are transferred to decoder for pixel-wise classification. This enables the SegNet to have a less computationally expensive training process with a spatial resolution trade-off.

5.2.2.3 Dilated-Net

Many initially proposed CNNs, such as AlexNet [243], VGGs [142], and ResNets [143], were designed for sparse classification applications (*e.g.*, image classification). These CNN architectures rely on max-pooling operations to abstract the data, translating the input image into a few number of classes. As a result, a decrease in the spatial resolution in sparse classification applications is not constraining but considered the objective. However, in dense classification applications (*e.g.*, semantic segmentation), where pixel coordinates of the target objects are required, utilizing such CNNs may be problematic.

The Dilated-Net is a powerful dense classification architecture, which has shown promising performance in different semantic segmentation applications. The architecture of the Dilated-Net is similar to a regular encoder-decoder CNN. The main diverging components of the Dilated-Net architecture are: 1) replacing regular convolutions with dilated convolutions and 2) implementing multi-scale contextual information aggregation module to enhance the final segmentation map resolution.

Unlike regular convolution operations, which collect information from a localized region of the image using a kernel (usually 3×3), the dilated convolutions use kernels of the same size as regular convolutions but spread out over a larger area on the image. A detailed explanation of the dilated convolutions can be found in [242]. The utilization of dilated convolutions enables the networks to efficiently capture global contextual

information without losing resolution in the final segmentation map. The multi-scale contextual information aggregation module is designed to increase the performance of dense prediction architectures, which can potentially increase the accuracy of the feature maps by passing them through multiple layers that expose contextual information.

5.2.3 Image preprocessing

Image segmentation in AHIA may require various steps to preprocess the images based on the segmentation model specifications. One of the most commonly used preprocessing steps is color normalization. The color normalization is implemented to mitigate the dramatic color variations caused in tissue processing and slide digitization steps. Various high-performance and rather complex histopathology color normalization techniques have been proposed over the past few decades [66]. We applied the Reinhard *et al.* [67] approach, which uses a simple logic by using principal component analysis to transfer the color and intensity of a reference image to the target images and performs well to normalize our dataset. A detailed review of recent histopathology color normalization techniques has been provided in section 2.3.1.

Another important preprocessing step in any deep learning approach is to standardize the input images. Although, in most state-of-the-art CNN models, the batch-normalization is used to standardize the inputs of each layer of the neural network, it is a good practice to standardize the input images as a preprocessing step to have a mean of zero and standard deviation of one. As stated in [244], standardization stabilizes the learning process and dramatically reduces the number of training epochs required to train a deep neural network. As such, we standardize our input images to $N(0,1)$.

5.2.4 Image patch extraction

As discussed in section 1.2, high-resolution histopathology WSIs are usually too large to directly be fed into a deep learning model. As such, patch-based deep learning approaches are commonly used in AHIA. When extracting patches, it is vital to consider the trade-off between field-of-view and computational complexity. Patches should be large enough to provide an adequate field-of-view to accurately segment the desired structures, while remaining small enough to enable training without running into memory issues. However, in many applications, where the target objects have dramatic size variations (*e.g.*, bronchi and blood vessels in histopathology WSI of mouse lung tissue), finding an optimum patch size that can provide required information on field-of-view and resolution can be a challenging task. In Chapter 6, we will expand on this limitation and discuss how this limitation can be resolved in patch-based approaches.

To develop our generic model, for each input histopathology image, we extracted patches of size 256×256 with 50% vertical and horizontal overlap at $20\times$ magnification. Using $20\times$ magnification to extract image patches preserves the spatial resolution, which is potentially beneficial in semantic segmentation of complex biological structures in histopathology WSIs. The final heat-maps are constructed by averaging over the intersecting regions in heat-map patches.

5.2.5 Data augmentation

In recent years, one of the main strategies to obtain higher performance with deep learning models was to increase the complexity of the model by adding more layers to the network. However, one of the main challenges with training deep learning models is that these networks require very large datasets to avoid overfitting. Overfitting refers to the

phenomenon when a network, instead of learning specific patterns in an image, learns a function with high variance, which can perfectly model the training data [244]. Access to a large dataset may not always be feasible, particularly in histopathology applications, where the manual assessment of the images requires a certain level of expertise.

Data augmentation is a common strategy in many deep learning based image analysis methods to increase the size of the available dataset, without actually collecting new data. Data augmentation techniques can be broadly categorized into two groups [244]: 1) Basic image manipulation, such as color space transformation, geometric transformation, and random erasing, and 2) Deep learning approaches, such as GAN data augmentation, neural style transfer, and adversarial training. One of the main factors that play an important role in applying different data augmentation techniques is the differences between training data and testing data. For example, in datasets that the conditions of capturing the input images may be subject to dramatic variations, randomly changing the brightness and contrast of the training dataset is a typical data augmentation, which makes the deep learning model robust to such variations. We noticed that in our datasets, due to applying color normalization and standardization of input images, we are not facing such dramatic variations. As a result, for training of our generic deep learning approach, we utilized standard geometric transformations, including images rotation with 90° , 180° , and 270° and vertical and horizontal flipping.

5.2.6 Training model

Deep learning models can be seen as a function that maps the input data into outputs. This function is composed of series of intermediate layers (*e.g.*, convolutional, pooling, fully-connected), which consist of a set of weights. At each layer, mathematical

transformations are applied to the inputs and weights, generating intermediate results (*i.e.*, feature maps) for the downstream layers. The main objective in the training of a deep learning model is to find the optimal weights of each layer to minimize the loss (the sum of logarithmic differences between the inputs and the outputs). A detailed explanation for training process is provided in section 2.5.2. The training process can be a time-consuming task, and similar to any other stochastic processes, may not always lead to an optimum solution. Multiple hyper-parameters are involved in the training of a deep learning model, and play a critical role in the training time and accuracy. Hyper-parameter tuning is usually applied before training a deep learning model. The main hyper-parameters in training a CNN model include: 1) optimization algorithm (*e.g.*, adaptive moment estimation (Adam) [223], stochastic gradient descent (SGD) [245], and Nesterov-accelerated adaptive moment estimation (Nadam) [246], etc.) and optimizer learning rate, 2) weight initialization methods (*e.g.*, random initialization, HE initialization [247], and Xavier initialization [248]), 3) number of training epochs, 4) regularization techniques, and 5) training batch size. Similar to the network architecture, there are no specific written rules to find the optimum values and approaches for the training step and may vary from application to application.

We use Nadam [249] as our main optimizer with an initial learning rate of 10^{-4} . The Nadam is an extension of Adam optimizer, which combines Adam and Nesterov-accelerated gradient, which improves the stability and convergence of the Adam by estimating the future values of the parameters to adjust the momentum. The cross-entropy loss is used to calculate the gradients in the backward pass, optimizing the weights in our deep learning model. Both HE initialization and Xavier initialization methods serve as

good starting points for initialization and mitigate the chances of exploding or vanishing gradients [250]. We utilize the HE initialization as the main weight initialization technique in our deep learning model, which is proven to be a good weight initialization technique with ReLU activation function [250]. For training of our deep learning models, the training dataset is divided with 3:1 ratio, where each group is used for training and validation at each epoch, respectively (25% is used for validation and 75% for training). The model is trained for 300 epochs with a batch size of 50, saving the best model with minimum validation loss.

5.3 Evaluation

The performance of our generic deep learning models are evaluated through three different approaches: 1) comparison against manual segmentation, and 2) comparison of the performance of the deep learning models against the conventional approaches in detection of touching villi structures in single-class placenta dataset (Chapter 3) and estimation of the MLI score in multi-class lung dataset (Chapter 4).

For comparison against the manual segmentation, for the single-class dataset, we perform five-fold cross-validation. We randomly divided our 12 placental images into training data and test data. At each fold, we use nine images for training and three for testing. This is repeated five times, so all placental images are used as test data. A similar approach is implemented to evaluate our models over our multi-class dataset; at each fold, eight lung WSIs are used for training and two for testing. Using five-fold cross-validation, the performance of the models is evaluated on every WSI. We calculate pixel-wise accuracy [251] (PA), mean Dice similarity coefficient [252] (DSC), and mean positive

predictive value (PPV) accuracy [251] performance metrics to evaluate the performance of the deep learning models against manual segmentation.

$$PA = \sum_c \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c} \quad 5.1$$

$$DSC = 1/C \times \sum_c \frac{2 \times TP_c}{2 \times TP_c + FP_c + FN_c} \quad 5.2$$

$$PPV = 1/C \times \sum_c \frac{TP_c}{TP_c + FP_c} \quad 5.3$$

In Equations 5.1 – 5.3, C is the number of classes, and TP_c , TN_c , FP_c , and FN_c are the numbers of true positives, true negatives, false positives, and false negatives for class c , respectively.

For evaluation of the deep learning models in terms of detection of touching villi structures in our single-class dataset, we calculate the sensitivity and F1-score metrics, which are defined in Equation 3.5 and Equation 3.6, and compare against the conventional approach in Chapter 3. For evaluation of the deep learning models in estimation of the MLI score in our multi-class dataset, we calculate the precision, recall, and F1-score using Equations 4.14, Equation 4.15, and Equation 4.16, respectively, to measure the performance of the models in the detection of accepted FOVs. We also compare the calculated MLI scores by the conventional approach against MLI scores calculated by the deep learning models for each WSI.

5.4 Results

Figure 5.3 visualizes examples of semantic segmentation of our single-class placenta dataset at 20× magnification. The generic deep learning models show comparably high accuracy in the segmentation of placental villi structures, specifically in terms of detection of touching villous structures. In Figure 5.3, the objects that have an intersection with the

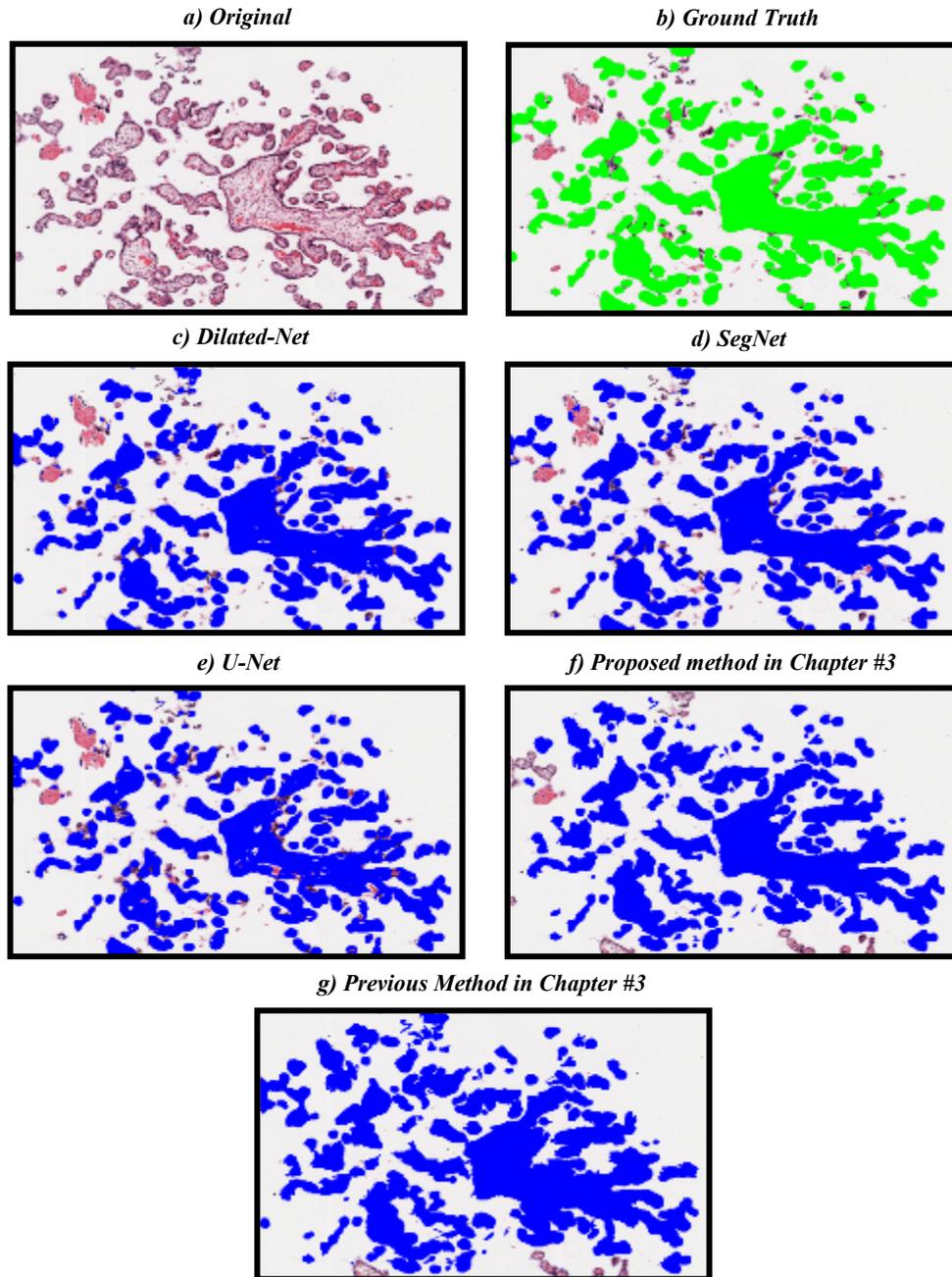


Figure 5.3 An example semantic segmentation in single-class placental villi dataset. The ground truth segmentation and the predictions for villi class are labeled as green and blue, respectively. boundary of the input image are not segmented in Figure 5.3.f and 5.3.g; this is mainly due to the inability of our proposed method in Chapter 3 to detect concavities in boundary structures. However, the deep learning method is able to correctly segment these structures with no limitation. As such, we visualized the boundary objects for deep learning methods

Table 5.1 Overall F1 score and sensitivity metrics for detection of villous structures instances in our single-class placenta dataset. The results are presented for deep learning-based models, the proposed method in Chapter 3 with and without detection of touching villi step, and the previously published work [29].

<i>MODEL</i>	<i>F1 - SCORE</i>	<i>SENSITIVITY</i>
<i>U-Net</i>	<i>0.904</i>	<i>0.919</i>
<i>Dilated-Net</i>	0.861	0.851
<i>SegNet</i>	0.899	0.918
<i>Previous Method [29]</i>	0.653	0.551
<i>Proposed method without detection of touching villi</i>	0.746	0.613
<i>Proposed method with detection of touching villi</i>	0.808	0.821

but not for methods described in Chapter 3. However, we excluded these boundary structures for calculation of the quantitative results in Table 5.1. Table 5.1 summarizes the semantic segmentation performance metrics for each method in the control group healthy

Table 5.2 Pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) accuracy metrics for semantic segmentation in single-class placental dataset.

<i>MODEL</i>	<i>HEALTHY</i>			<i>DIAGNOSED WITH PE</i>			<i>OVERALL</i>		
	<i>PA</i>	<i>DSC</i>	<i>PPV</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>
<i>U-Net</i>	<i>0.968</i>	<i>0.941</i>	<i>0.935</i>	<i>0.943</i>	<i>0.915</i>	<i>0.902</i>	<i>0.956</i>	<i>0.928</i>	<i>0.919</i>
<i>Dilated-Net</i>	0.959	0.921	0.907	0.931	0.895	0.888	0.945	0.908	0.898
<i>SegNet</i>	0.962	0.936	0.922	0.936	0.904	0.896	0.949	0.920	0.909
<i>Proposed method Chapter 3</i>	0.841	0.739	0.786	0.827	0.725	0.762	0.834	0.732	0.774
<i>Previous Method [29]</i>	0.838	0.721	0.737	0.821	0.693	0.713	0.829	0.707	0.725

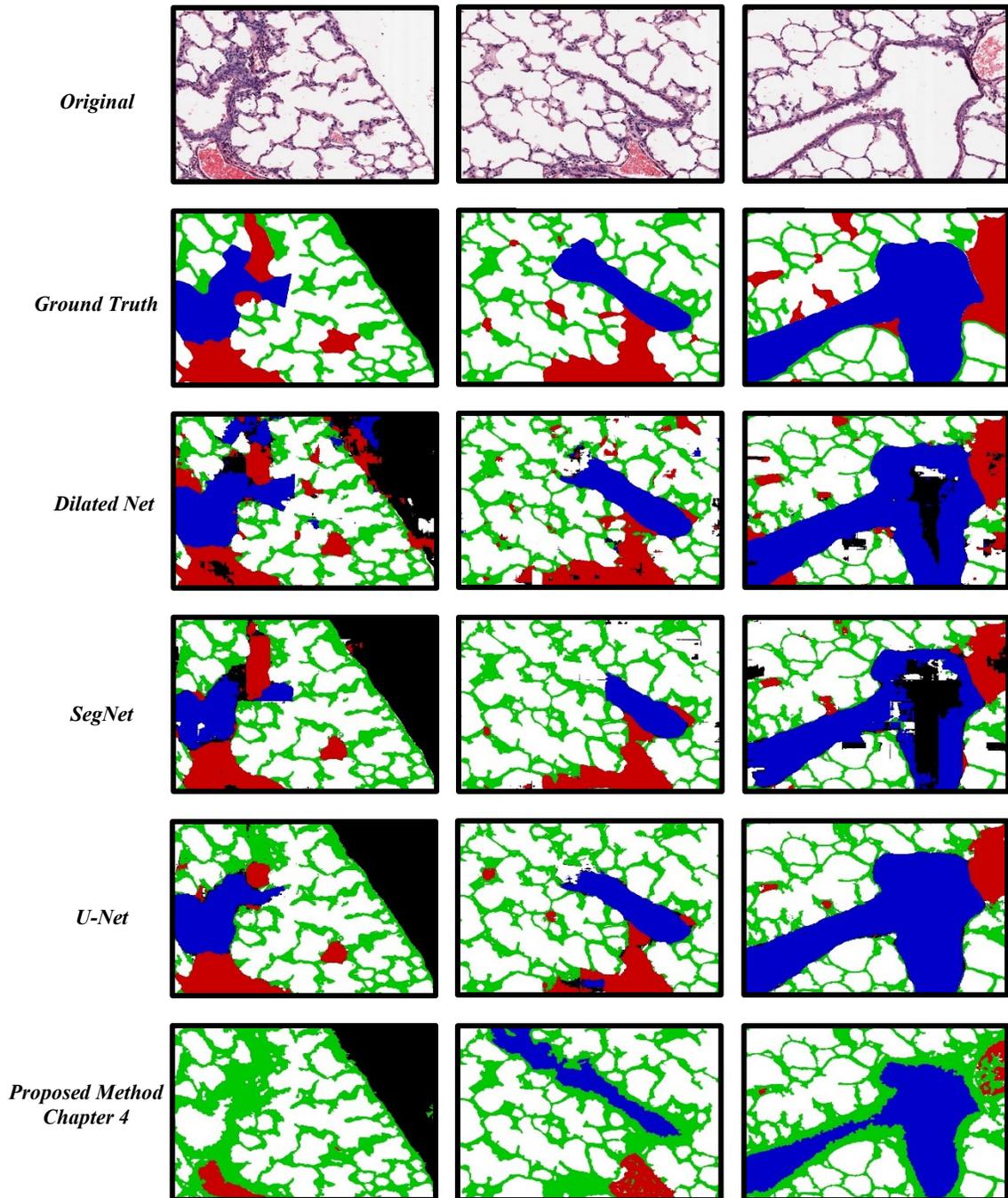


Figure 5.4 Examples of semantic segmentation of histopathology images of mouse lung dataset. The results for Blood Vessel, Bronchi, Alveoli, Alveoli wall, and back ground classes are marked as red, blue, white, green, and black, respectively.

placentas and the group diagnosed with PE group. As demonstrated in this table, the deep learning models outperform the models in Chapter 3 by a considerable margin. A comparison among the generic deep learning models indicates that the U-Net model

Table 5.3 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) accuracy metrics for semantic segmentation in multi-class mouse lung dataset.

<i>MODEL</i>	<i>RA</i>			<i>O2 + LPS</i>			<i>OVERALL</i>		
	<i>PA</i>	<i>DSC</i>	<i>PPV</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>
<i>U-Net</i>	0.962	0.856	0.823	0.946	0.827	0.798	0.954	0.841	0.811
<i>Dilated-Net</i>	0.960	0.803	0.781	0.940	0.812	0.777	0.950	0.808	0.779
<i>SegNet</i>	0.945	0.753	0.720	0.939	0.783	0.755	0.942	0.768	0.738
<i>Proposed method Chapter 4</i>	0.859	0.683	0.610	0.839	0.709	0.651	0.849	0.696	0.631

performs the best with the highest performance metrics in both groups and overall. The superior performance of the U-Net model can also be seen in terms of detection touching villous structures (see Table 5.2), where the villous structures instances can be detected with 9.6% and 9.8% higher F1-score and sensitivity against our proposed method in Chapter 3, respectively.

Figure 5.4 shows multiple examples of semantic segmentation in the histopathology images of multi-class mouse lung dataset for generic deep learning models and the proposed method used in Chapter 4. The semantic segmentation performance metrics are presented in Table 5.3, which demonstrate the superior performance of the generic deep learning models in comparison to the conventional rule-based approach in Chapter 4. The U-Net model slightly outperforms the DilatedNet and SegNet models for all performance metrics.

A comparison between the performance of the deep learning models and the proposed method in Chapter 4 for detection of FOV images for different evaluation approaches (see section 4.2.4) is presented in Table 5.4. As shown in this table, in Evaluation Approach (1)

Table 5.4 Evaluation of the proposed method performance in detection of field-of-view (FOV) images. For evaluation approach (1), the decisions are correct if the automated method agrees with at least two out of three raters. For evaluation approach (2), the decisions are correct if the automated method agrees with at least one rater out of three raters.

		<i>RA</i>			<i>O2 + LPS</i>			<i>Overall</i>		
		<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<i>Evaluation approach (1)</i>	<i>U-Net</i>	0.836	0.957	0.892	0.778	0.957	0.858	0.807	0.957	0.875
	<i>Dilated-Net</i>	0.868	0.920	0.893	0.880	0.870	0.875	0.874	0.895	0.884
	<i>SegNet</i>	0.843	0.914	0.877	0.783	0.896	0.836	0.813	0.905	0.857
	<i>Proposed method Chapter 4</i>	0.948	0.695	0.801	0.852	0.815	0.833	0.900	0.755	0.817
<i>Evaluation approach (2)</i>	<i>U-Net</i>	0.872	0.936	0.903	0.836	0.949	0.889	0.854	0.943	0.896
	<i>Dilated-Net</i>	0.914	0.882	0.897	0.910	0.853	0.881	0.912	0.867	0.889
	<i>SegNet</i>	0.894	0.883	0.889	0.812	0.880	0.845	0.853	0.882	0.867
	<i>Proposed method Chapter 4</i>	0.968	0.777	0.862	0.899	0.893	0.896	0.934	0.835	0.879

and Evaluation approach (2) the overall performance of the deep models is considerably improved in terms of Recall and F1-score performance metrics; however, the proposed method in Chapter 4 scored the highest Precision score. An interesting observation in Table 5.4 is a noticeable performance increase of the deep learning methods in terms of Recall performance metric in comparison to our proposed method in Chapter 4. This noticeable increase in Recall may be an indication for considerable decrease in FNs in deep learning methods, which is a demonstration for higher performance of our deep learning models in detection of lung's biological structures.

As stated in Chapter 4, in *RA* experimental group, the size of the mice lung structures dramatically smaller than those exposed to conditions in *O2 + LPS* group. Based in the results in Table 5.4, the difference in Recall between deep learning models and rule-based approach in Chapter 4 is considerably higher in *RA* group compared to *O2 + LPS* group.

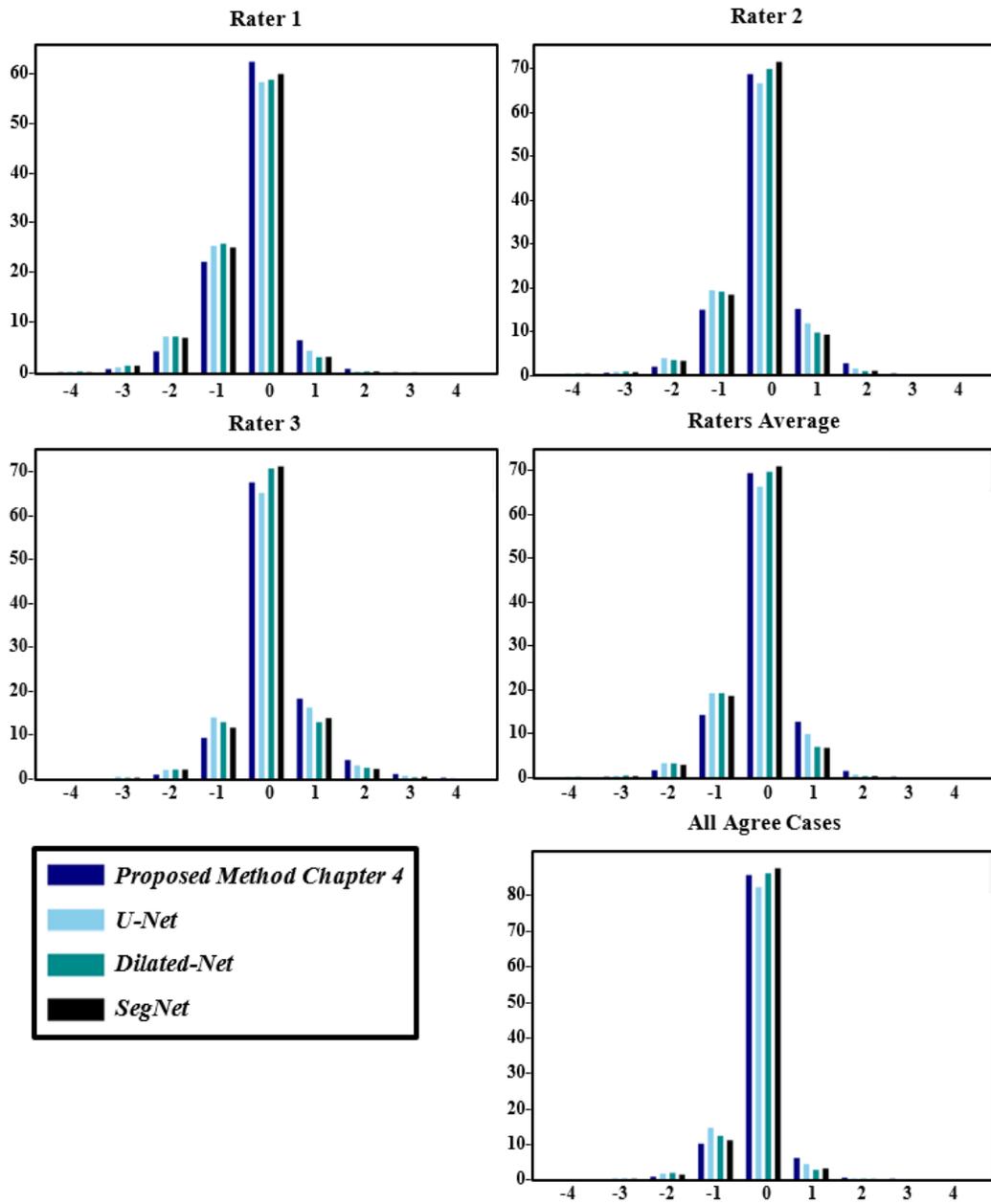
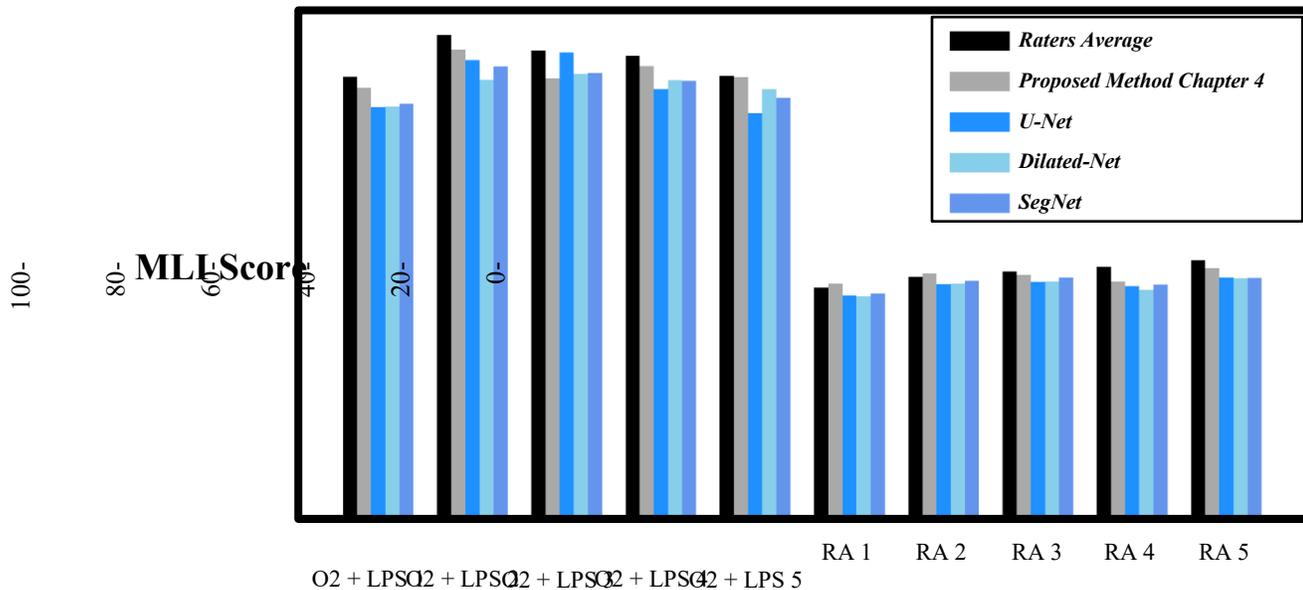


Figure 5.5 The evaluation of the deep learning models and proposed method in Chapter 4 in detection of the intersections against the human raters. The plot for Raters Average and All Agree Cases shows intersection difference between algorithm-generated results and the average number of intersections reported by human raters and counted intersections when all raters were agree on number of intersections.

This indicates that the deep learning models are superior in segmentation of smaller structures in comparison to our rule-based approach in Chapter 4.

The overall performance of the methods has been evaluated in terms of calculating the difference between the algorithm-detected intersections verses the reported number of



Mouse Lung Image Name

Figure 5.6 The Estimated MLI scores in histopathology images of mouse lung dataset

intersections by each rater in accepted FOV images (see Figure 5.5). All methods show a similar rate of agreement in different comparison cases visualized in Figure 5.5; however, the Dilated-Net shows a slightly higher rate of agreement in all comparison cases, except for agreement with the Rater 1, where the proposed method in Chapter 4 shows higher

Table 5.5 The evaluation of the reliability of agreement in detection of intersections in accepted field-of-view (FOV) images between the deep learning methods and the proposed method in Chapter 4 with respect to the manual assessments using Fleiss' Kappa score.

	<i>RA</i>					<i>O2 + LPS</i>					<i>Overall score</i>
	<i>Rater #1</i>	<i>Rater #2</i>	<i>Rater #3</i>	<i>Rater's average</i>	<i>All agree cases</i>	<i>Rater #1</i>	<i>Rater #2</i>	<i>Rater #3</i>	<i>Rater's average</i>	<i>All agree cases</i>	
<i>U-Net</i>	0.57	0.57	0.49	0.57	0.77	0.61	0.62	0.61	0.63	0.80	0.75
<i>Dilated-Net</i>	0.57	0.57	0.49	0.57	0.77	0.64	0.64	0.62	0.65	0.83	0.78
<i>SegNet</i>	0.56	0.56	0.49	0.56	0.76	0.63	0.63	0.62	0.65	0.81	0.78
<i>Proposed Method Chapter 4</i>	0.42	0.47	0.41	0.46	0.67	0.68	0.63	0.61	0.66	0.84	0.76

agreement rate. The Fleiss' Kappa score for evaluating the rate of agreements in different cases is shown in Table 5.5. The overall scores show a minor improvement in the rate of agreement of SegNet and Dilated-Net models in comparison to other methods. The interesting observation in Table 5.5 is the noticeable increase in the rate of agreement of the deep learning methods in *RA* group in comparison with the proposed method in Chapter 4.

Figure 5.6 shows the estimated MLI scores by deep learning models and the proposed method in Chapter 4, in comparison to raters' average MLI score. A comparison between the raters' average score and the automated methods shows that generic deep learning models and rule-based method in Chapter 4 are able to consistently predict the changing patterns of MLI score across different WSIs.

5.5 Discussion

In this chapter, we proposed a generic deep learning pipeline for semantic segmentation of biological structures in high-resolution histopathology images to resolve an important bottleneck in rule-based approaches, where a developed rule-based approach for one histopathology application may not be generalized to other applications. In our experiment, we implemented three state-of-the-art semantic segmentation architectures (*i.e.*, U-Net, Dilated-Net, and SegNet), which have been evaluated in the segmentation of biological structures in single-class human placenta and multi-class mouse lung tissue histopathology images.

By comparing the performance of the deep learning models, the U-Net outperforms the Dilated-Net and SegNet in all performance metrics by scoring 0.956, 0.928, and 0.919 in terms of PA, DSC, and PPV in our single-class placental villi dataset, respectively. In

comparison to the rule-based method in Chapter 3, the U-Net acquired outstanding semantic segmentation performance by scoring 0.122, 0.196, and 0.145 higher performance in terms of PA, DSC, and PPV performance metrics, respectively. The superior performance of the U-Net model can also be seen in the detection of touching villous structures (see Table 5.2) by scoring 0.904 and 0.919 in terms of F1-score and sensitivity performance metrics, which outperforms other segmentation models.

A similar pattern can also be seen in semantic segmentation in our multi-class lung dataset, where U-Net demonstrates higher overall performance in terms of all segmentation performance metrics (Table 5.3). Our rule-based approach shows the highest precision score for both evaluation approaches in the detection of FOV images (0.900 in evaluation approach 1 and 0.934 in evaluation approach 2). However, the U-Net model scores the highest overall RE in both evaluation approaches and highest F1-score in evaluation approach and second-best in evaluation approach 1. Another important observation is the superior performance of the U-Net model in terms of the Recall performance metric in all experimental groups and evaluation approaches. Note that detection of FOV images is considered a recall-oriented task, where FNs (*i.e.*, accepting FOVs that were rejected by raters) may be considered to have a more detrimental effect to MLI scoring than FPs (*i.e.*, rejecting FOVs that were accepted by raters).

The most important limitation of our work was the low performance of our deep learning models in the segmentation of very large biological structures, specifically in the segmentation of very large terminal bronchi and blood vessels in histopathology WSIs of mouse lungs (see Figure 5.4). This may be related to the limited field-of-view in the extracted patches (at 20 \times magnification) used for training of the deep learning models,

where large biological structures are present and the deep learning models fail to acquire the necessary visual cues that is required for accurate segmentation of the structures. This can be addressed by lowering the resolution of the input WSIs; however, this can be problematic because lowering the resolution might affect the performance of the deep learning models in the segmentation of smaller structures. As such, another possible approach to increase the field-of-view can be increasing the patch sizes. However, increasing the patch-size potentially reduces the computational efficiency of the model, limiting the training of the deep learning models in terms of memory and training time. Therefore, the possible solution for this three-fold trade-off between field-of-view, computational efficiency, and spatial resolution is a multiresolution analysis of WSIs, where contextual information from different magnifications are provided for the deep learning models for semantic segmentation of biological structures.

We utilized a standard workstation with an Intel Core i7-3770 3.40 *GHz* CPU, 12 GB of installed RAM, and a single NVIDIA RTX 2060 with 6 GB memory to train and test our deep learning pipeline. We implemented our deep learning models using Tensorflow and Keras frameworks, and the Python programming language. However, to generate our input patches, Matlab environment is used.

5.6 Conclusion

In this chapter, we developed a generic deep learning pipeline using three state-of-the-art semantic segmentation models. The overall comparison between the deep learning models and the rule-based approaches showed that in addition to the generalizability of the deep learning model, a considerable improvement could be achieved by deep learning models. The U-Net model demonstrated superior performance in semantic segmentation in

both multi-class mouse lung dataset and single-class placental villi dataset in comparison to other methods, including the rule-based approaches in Chapter 3 and Chapter 4. It also scored the highest performance in detection of touching villi in placenta dataset and detection of FOV images in mouse lung dataset, which shows that the U-Net may be suitable architecture for semantic segmentation in histopathology images of mouse lung tissue and placental villi histopathology images.

We demonstrated that the three-fold trade-off between field-of-view, computational efficiency, and spatial resolution in extracted patches is an important limitation in deep learning models, which may be resolved by a multiresolution approach. In Chapter 6, we focus on resolving this three-fold trade-off.

Chapter 6: Multi-Resolution Semantic Segmentation of Biological Structures in Digital Histopathology

In this chapter, a pipeline for multi-resolution semantic segmentation of histopathology WSIs is presented, which addresses the three-fold trade-off between field-of-view, computational efficiency, and spatial resolution. We propose a two-stage multiresolution approach for semantic segmentation of histopathology WSIs. In the first stage, we use four different CNNs to extract the contextual information from input patches at four different magnifications (i.e., 2.5 \times , 5 \times , 10 \times , and 20 \times). In the second stage, we use another CNN to aggregate the extracted information in the first stage and generate the final segmentation masks. We evaluated the performance of our proposed method using two histopathology datasets, which consist of structures with large size variations: 1) a single-class placental villi dataset, and 2) a multi-class mouse lung tissue dataset. We also compared the performance of our proposed method with three well-known state-of-the-art patch-based semantic segmentation models (U-Net, Dilated-Net, and SegNet) at different magnifications. This Chapter has been submitted as Salsabili, S., Ukwatta, E., & Chan, A. D. C. (2021). “Multi-Resolution Semantic Segmentation of Biological Structures in Digital Histopathology.” to *Journal of Medical Imaging*.

6.1 Introduction

In modern pathology, analysis of digital histopathology images plays a vital role in assessing the status and function of organisms and in the diagnosis of different diseases. Manual assessment of the digital histopathology images by expert pathologists is expensive, time-consuming, and subject to inter- and intra- rater variability [253]. As such, extensive efforts have been made to automate the various analysis steps, including the

segmentation of regions of interest (ROIs). Digital histopathology whole slide images (WSIs) are very high-resolution in nature, and often we need to segment both the small and large biological structures for subsequent analysis. For example, in lung tissue WSIs, there is a large size difference between the major biological structures, such as alveoli sacs, bronchi, and blood vessels. Segmentation of the various biological structures in histopathology images of lung tissue is a necessary step for automated estimation of mean linear intercept (MLI) score, which is a common metric used for quantification of lung injury in respiratory diseases such as bronchopulmonary dysplasia (BPD) [37], [227]. Another example is placental WSIs, where the size gap between different placental structures can be noticeably large (a few μm^2 in syncytial knots to several mm^2 in large terminal villi). The segmentation of villi structures in placenta histopathology images facilitates the extraction of important features that are used in the analysis of placenta-mediated diseases, such as preeclampsia [254].

In recent years, deep convolutional neural networks (CNN) have shown promising performance in segmenting biological structures in histopathological WSIs [255], [256]. However, there are various challenges associated with the semantic segmentation of WSIs using CNNs. One challenge is the image size; high-resolution WSIs (*e.g.*, 20 \times magnification, corresponding to 0.5 micrometers per pixel) can often have image sizes around 80,000 \times 80,000 pixels. Large images cannot be directly fed into the CNN because of the need for large memory resources and high computational cost of the resulting model. As a result, patch-based approaches are often used [31], [257], [258], where the WSI is divided into small patches (typically around 256 \times 256 pixels), the patches are individually

fed into the CNN, and then the processed patches are combined to create a segmentation of the entire WSI.

Selecting an optimal patch size can be difficult. Small patches have a limited field-of-view, which can be problematic for the model to make accurate predictions, particularly during the segmentation of larger biological structures, because it may lack context from a broader range of texture patterns. Increasing the patch size will resolve the issue with the narrow field-of-view with the cost of reducing the computation efficiency of the model, which can potentially lead to memory issues and a longer model training time. An alternative solution is to down-sample the input WSI before extracting patches, providing a wider field-of-view with a smaller patch size; however, the reduced resolution of the patches may affect the performance of the model, particularly in the segmentation of small biological structures.

In an attempt to overcome the three-fold trade-off between the field-of-view, computational efficiency, and spatial resolution, various multiresolution approaches for WSI analyses have been proposed, which aggregate the contextual information from each resolution. Multiresolution approaches mimic the technique often taken by pathologists, analyzing the WSI at different magnifications. Alsubaie *et al.* [259] performed tumor classification by training a single CNN using image patches at different magnifications and compared the performance of their proposed method against CNNs that were trained on 40 \times , 20 \times , and 10 \times magnifications, which improved the performance by 5%, 1%, and 1% in terms of pixel accuracy, respectively. Kosaraju *et al.* [260] proposed a multi-magnification deep learning model to classify adenocarcinoma in high-resolution histopathology images of the colon and stomach. The authors extracted approximately

160k patches of 256×256 pixels from 20× and 5× magnification WSIs. Due to the high complexity of the model, they were not able to process 40× magnification WSIs. Their proposed method achieved an F1-score of 93.4% in the classification of patches, which showed 3%, 7%, and 6% improvement against well-known deep learning models, CAT-Net [261], MRD-Net [262], and Dense-Net [263], respectively. These works demonstrated the benefits of a multiresolution approach, even though they were focused on histopathology image classification rather than segmentation.

Sirinukunwattana *et al.* [264] implemented different CNN architectures and trained the CNNs using patches of different resolutions, addressing the challenges associated with a narrow field-of-view in patch-based semantic segmentation approaches of prostate and breast histopathology WSIs. The authors applied multiple setups of the long short-term memory (LSTM) units to integrate the results of each CNN to improve the overall segmentation performance. Their proposed bidirectional LSTM setup was able to achieve class-average F1-score of 0.789 and 0.523 for prostate and breast histopathology WSIs. Tokunaga *et al.* [163] proposed a multiresolution approach for semantic segmentation of adenocarcinoma in lung WSIs, which extracted contextual information from different magnifications using multiple expert CNNs, and generated the results by adaptively weighting the heat-maps from the expert CNNs and aggregating the extracted information using a shallow CNN. Their proposed method achieved 82.1% and 53.6% mean intersection-over-union, which showed 3% and 19% improvement against Hard-Switch-CNN [265] method for single-class and multi-class segmentation tasks, respectively. Rijthoven *et al.* [165] proposed HookNet, which attempted to address the trade-off between resolution and field-of-view in lung and breast histopathology WSIs, which utilized skip

connections between two parallel CNNs to aggregate the contextual information from different magnification. Their proposed method achieved 0.91 and 0.72 F1-score for semantic segmentation in breast and lung datasets, respectively. While the multiresolution schemes of these methods enable them to improve the performance of the segmentation model by resolving the trade-off between field-of-view and resolution, the segmentation of biological structures with large size differences was neither the objective of these works nor evaluated in histopathology WSIs.

In this work, we propose a two-stage multiresolution approach for the semantic segmentation of biological structures in histopathological WSIs. In the first stage, we use multiple Expert CNNs (ECNNs) to extract contextual information at multiple magnifications of the WSIs. In the second stage, we use another CNN to aggregate the extracted contextual information, using structure size to weight the ECNN heat-maps. The proposed multiresolution approach addresses the trade-off between field-of-view, computational efficiency, and spatial resolution in conventional patch-based CNN models. The computationally efficient schema of the proposed method utilizes small image patch sizes (256×256 pixels) for training, which enables the model to run on small GPUs. The performance of the proposed method is evaluated over two different histopathology image datasets: 1) the histopathology WSIs of human placenta (single-class dataset), and 2) histopathology WSIs of mouse lung tissue (multi-class dataset). Both of these datasets consist of biological structures with drastic size variability, which can potentially demonstrate the performance of our proposed method in the segmentation of large and small biological structures. The proposed method is compared against different state-of-the-art CNN baselines at different magnifications.

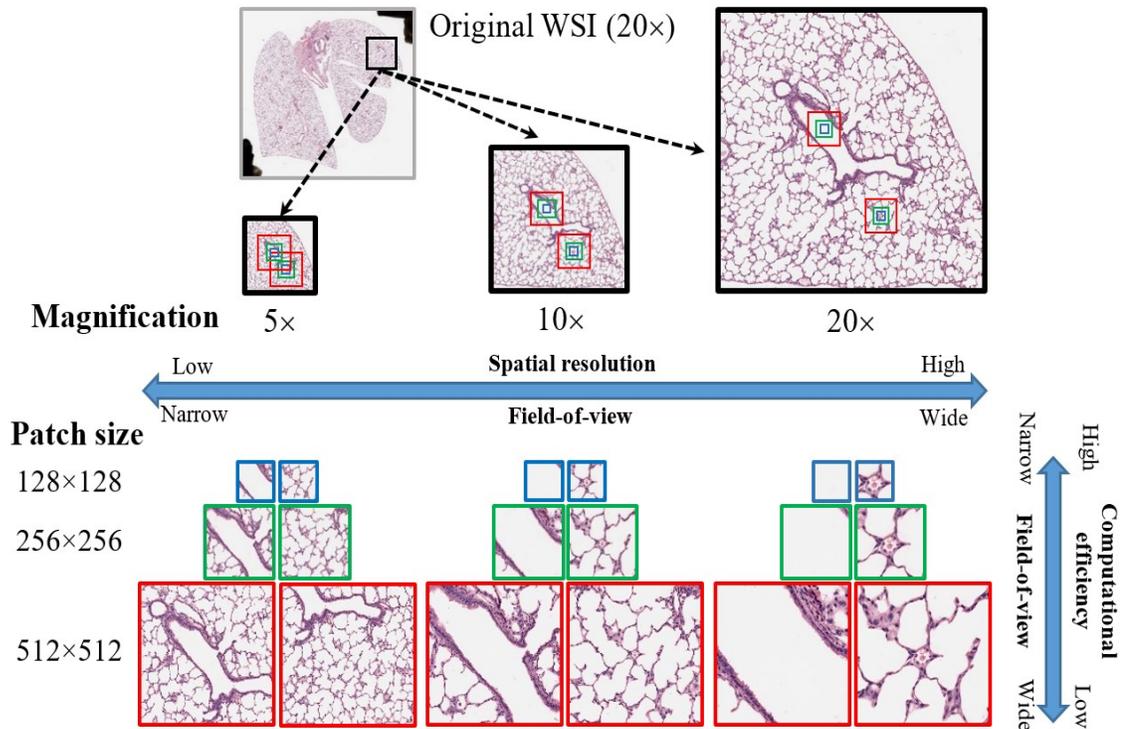


Figure 6.1 The trade-off between trade-off between the field-of-view, computational efficiency, and spatial resolution. By increasing the magnification (moving from left to right), the field-of-view of the patches narrows, while the spatial resolution increases. By increasing the patch size (moving from top to bottom), the field-of-view of the patches increases, while the computation efficiency decreases.

6.2 The multi-resolution semantic segmentation

There is a three-fold trade-off between field-of-view, computational efficiency, and resolution in patch-based CNNs. Figure 6.1 illustrates image patch selection with respect to patch size and magnification. Changing the patch size (moving in the vertical axis in Figure 6.1) changes the field-of-view, with the trade-off of computational efficiency. Small patches may have a limited field-of-view, lacking sufficient contextual information for the CNN model to make accurate decisions. Large patches, however, reduces the computational efficiency of the CNN model, which may result in memory issues and slow training speed. To resolve this trade-off between field-of-view and computational efficiency, one can reduce the magnification of the image in order to obtain an adequately large field-of-view, while maintaining a computationally efficient patch size. Changing the

magnification of the input image (moving in the horizontal axis in Figure 6.1) changes the field-of-view, with the trade-off of spatial resolution. The lower spatial resolution can reduce the segmentation performance, particularly with smaller biological structures.

We propose a two-stage, multiresolution semantic segmentation method to address the trade-off between the field-of-view, computational efficiency, and spatial resolution (Figure 6.2). The first stage is comprised of a number of parallel CNNs, each using the same fixed patch size but at different spatial resolutions, which provides different field-of-view sizes for each CNN. Each of these CNNs extracts the contextual information from the input WSI, generating a segmentation heat-map (probability map) for its particular spatial resolution. In the second stage, we weigh the structures in each heat-map based on the size of the structure. The weighting process enables the efficient inclusion of required contextual information for accurate segmentation of a structure based on its size. The weighted heat-maps are concatenated and an aggregating CNN is used to generate the final segmentation map (binary images associated with each class).

6.2.1 Contextual information extraction

In this stage, contextual information is extracted using parallel CNNs, each using a different magnification of the WSI; these CNNs will be referred to as expert CNNs (ECNNs). The number of ECNNs (N) can be adjusted to vary the range of magnifications appropriate for the given application. In this work, we use $N = 4$ magnifications that are separated by a factor of $2\times$ ($2.5\times$, $5\times$, $10\times$, and $20\times$), obtained by down-sampling the original WSI, which had a $20\times$ magnification. For a particular magnification, the WSI is divided into non-overlapping image patches of a fixed size (in this work we used a patch size of 256×256), and the associated ECNN is used to generate heat-maps for each image

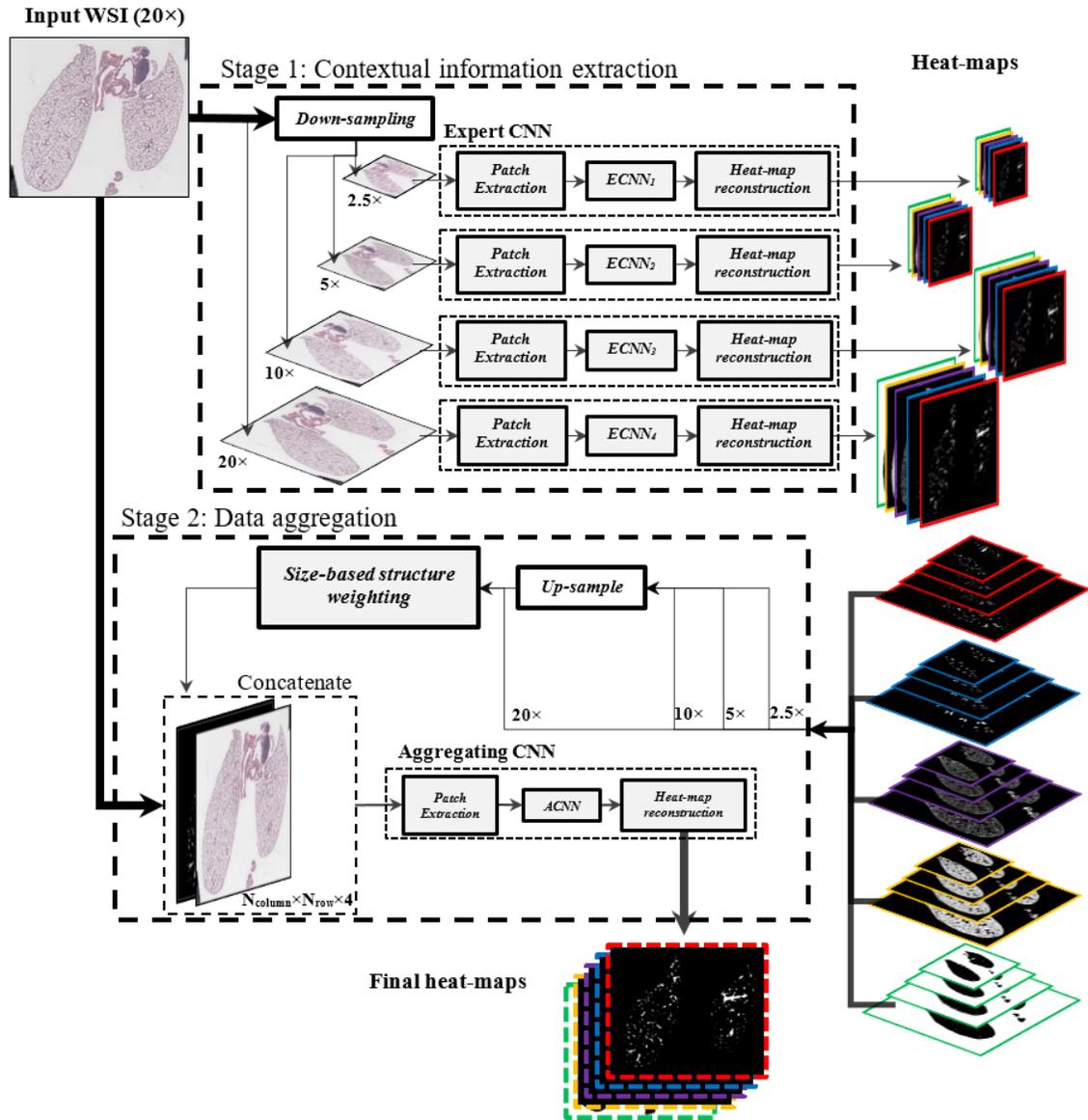


Figure 6.2 The overview of the proposed multiresolution segmentation pipeline. In contextual information extraction stage, four different expert CNNs (ECNNs) are used to generate the heat-maps associated with each class at different resolutions as shown in green, yellow, purple, blue, and red for background, alveoli lumen, alveoli border wall, bronchi, and blood vessel, respectively. In data aggregation stage, the size-based structure weighting step is applied to the heat-maps associated with each class at each magnification. The generated weighted heat-maps associated with all classes are processed using the aggregating CNN (ACNN) to generate the final heat-maps for each class.

patch. Each generated heat-map patches corresponds to a class in the dataset. The generated heat-map patches of each class from a particular ECNN (*i.e.*, the output of the ECNN for that class) are put together to construct the heat-maps of the classes for that magnification (see Figure 6.2).

The vanilla U-Net [266] architecture is used for the ECNN models. The effective use of skip connections between down-sampling layers to up-sampling layers in U-Net architecture enables the model to preserve the spatial resolution. This is particularly important for our application because our aim is to segment the large biological structures as well as small biological structures, where resolution loss may result in misclassification of the smaller structures. The U-Nets were trained from scratch, with four convolution layers in the contracting path and four transpose convolution layers in the expansive path. Hyper-parameter tuning was performed over the training data. The U-Net was trained using the Nadam optimizer [249] with an initial learning rate of 10^{-3} . The Nadam offers adaptive learning rate adjustment, which simplifies the hyper-parameter tuning and accelerates gradient calculations, which moderately increases convergence speed. The weighted categorical cross-entropy was used for calculating the loss:

$$\left\{ \begin{array}{l} L^{wcce} = -\frac{1}{M} \sum_{c=1}^{N_c} \sum_{n=1}^{N_p} \sum_{m=1}^M w_c \times y_m^c(n) \log(\hat{y}_m^c(n)) \\ w_c = \frac{N_p}{M \times N_{pc}} \end{array} \right. \quad 6.1$$

where, L^{wcce} is the weighted categorical cross-entropy loss function; $y_m^c(n)$ and $\hat{y}_m^c(n)$ are the n^{th} pixels of the m^{th} training patch from c^{th} class of the ground truth and predicted heat-map, respectively; M is the number of training patches; N_p is the number of pixels in the training patch; N_c is the number of classes; w_c is the weight for c^{th} class; and N_{pc} is the number of pixels in the c^{th} class.

To account for the overfitting problem, we perform dropout by a factor of 0.25 at each layer and perform cross-validation at each epoch with the ratio of 10:1 (train on 90% of the training data and validation on the remaining 10% at each epoch). Batch normalization is

applied to each layer - to reduce the training time and prevent diverging gradients - followed by a rectified linear unit activation function. We train our model for 50 epochs, with a batch size of 20, and we save the best model in terms of minimum validation loss. A total number of 8,643,458 trainable parameters are optimized in each ECNN network.

6.2.2 Data aggregation

In this stage, we aggregate the extracted contextual information from different magnifications to produce the final heat-maps, which is performed in three steps: 1) up-sampling, 2) size-based structure weighting, and 3) aggregating using a CNN.

6.2.2.1 Up-sampling

The heat-maps, generated by the ECNNs from the contextual information extraction stage, have different sizes. Heat-maps of lower magnifications are up-sampled to the same size as the largest heat-map, which corresponds to the original WSI with the highest magnification (*i.e.*, 20×). The bi-cubic interpolation method with 4×4 kernel size is used to up-sample the heat-maps.

6.2.2.2 Size-based structure weighting

As previously discussed, the segmentation of the larger structures benefits from a larger field-of-view, while the segmentation of smaller structures benefits from a higher spatial resolution. As such, we weight the heat-maps to favor larger structures from lower magnifications and smaller structures from higher magnifications, generating a weighted heat-map for each class (see Figure 6.3). In addition, the size-based structure weighting step enables to efficiently integrate the extracted contextual information from different magnifications to a single weighted heat-map, which improves the scalability of inputs in

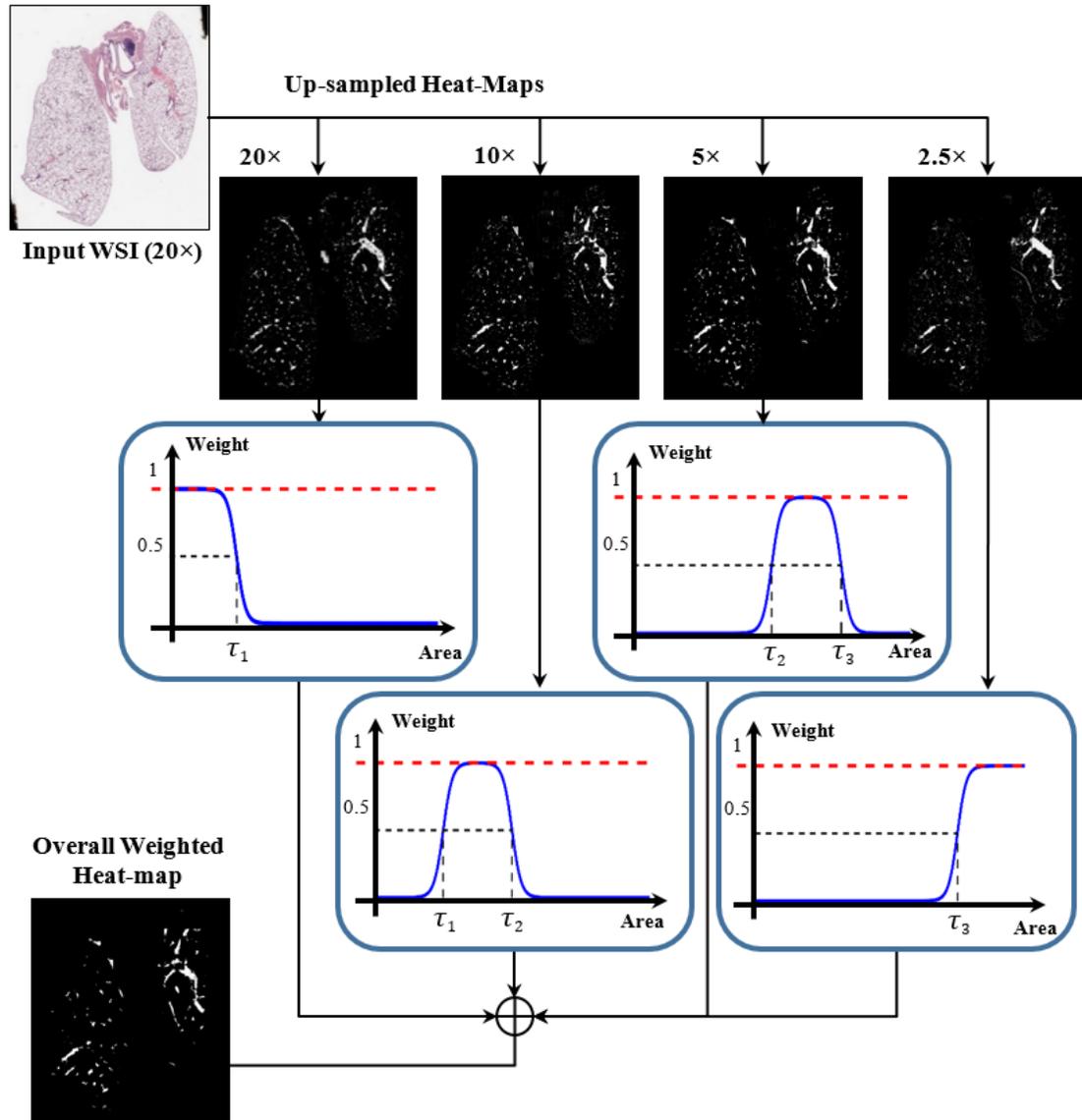


Figure 6.3 Size-based structure weighting of the heat-maps of different magnification for each class. In this figure, an example WSI of lung dataset is presented, which contains five different classes. For simplicity, only the heat-map weighting process for one class (Blood Vessel) is visualized.

second stage. Although, the main objective for the size-based weighting step is to integrate the heat-maps containing structures with large size variations, it can still be effective if the size of the structures does not change drastically. As such, the Equation 6.2 is used to generate the weights for the structures of different size at each magnification:

$$\left\{ \begin{array}{l} W_{M_i}^j = \frac{1}{1 + e^{-\Delta(A_{M_i}^j - \tau_i)}} - \frac{1}{1 + e^{-\Delta(A_{M_i}^j - \tau_{i+1})}}, \quad i \in \{0, 1, 2, 3\} \\ M_i \in \{20 \times, 10 \times, 5 \times, 2.5 \times\} \\ \tau_i \in \{\tau_0 = -\infty, \tau_1, \tau_2, \tau_3, \tau_4 = +\infty\} \end{array} \right. \quad 6.2$$

where, $W_{M_i}^j$ is the weight that applies to the j^{th} structure at magnification M_i , which has the size of A . τ_1, τ_2 , and τ_3 are hyper-parameters used to adjust the weighting at each magnification. Δ is a constant value, representing the slope of attenuation at threshold τ_i .

Equation 6.2 ensures that the weighted heat-maps maintain the same range of values (between 0 and 1). We heuristically optimized the values of τ_1, τ_2 , and τ_3 as part of the hyper-parameter tuning process in the second stage.

At each magnification, the calculated weights in Equation 6.2 are multiplied by the pixel values of the corresponding structures to weight the heat-map of that magnification. The weighted heat-maps, from each magnification, are superimposed to generate the overall weighted heat-map (see Equation 6.3).

$$\left\{ \begin{array}{l} H^{\text{weighted}}(x, y) = \sum_{i=0}^3 \sum_{j=1}^{N_{M_i}} W_{M_i}^j \times H_{M_i}(x, y) \\ (x, y) \in \text{pixel coordinates of the } j^{\text{th}} \text{ structure in } H_{M_i} \end{array} \right. \quad 6.3$$

where, H_{M_i} is up-sampled heat-map outputted by the ECNN at M_i magnification, H^{weighted} is the weighted heat-map, $W_{M_i}^j$ is the weight for j^{th} structure at M_i magnification (Equation 6.2), and N_{M_i} is the number of structures in H_{M_i} .

6.2.2.3 Aggregating CNN

In the size-based structure weighting step, we integrated the contextual information on resolution and field-of-view of four different magnifications (*i.e.*, 20 \times , 10 \times , 5 \times , and 2.5 \times) to generate an overall weighted heat-map at 20 \times magnification for each class. Here, our aim is to improve the overall segmentation performance of our proposed method by developing an aggregating CNN (ACNN), which utilizes the corresponding overall weighted heat-maps of each class and the original WSI and generates the final heat-map for that class. As such, for each class, we train a separate ACNN model, which is expert in segmentation of the biological structures in that class. For our single-class placenta dataset only one ACNN is used. For our multi-class lung dataset, which consists of five classes (*i.e.*, Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background), five ACNNs are used.

To prepare the input data for the ACNNs, for each class, the WSIs at 20 \times magnification and the corresponding overall weighted heat-map of that class are concatenated to form a four-dimensional matrix (*i.e.*, one dimension corresponds to the weighted heat-map and three dimensions correspond to the color channels of the input WSI). ACNN inputs are patches of size 256 \times 256 pixels from the 4D matrix (*i.e.*, the patch size is 256 \times 256 \times 4). Although, 256 \times 256 patches at 20 \times magnification offer a limited field-of-view for the segmentation of larger biological structures, the limited field-of-view is mitigated by the information provided by the overall weight heat-maps.

To construct the final heat-map for each class, the outputs of the corresponding ACNN model (*i.e.*, the heat-map patches) are put together. For our single-class placenta dataset and multi-class lung dataset, one and five final heat-maps are constructed for each input

WSI, respectively. To encounter the discontinuities in the output heat-map patches, the input patches are extracted with 50% overlap in vertical and horizontal axis, in the testing phase. The final heat-maps are constructed by averaging over the intersecting regions in heat-map patches.

To generate the final segmentation maps (*i.e.*, binary masks) for each input sub-image in our single-class placenta dataset, we use 0.5 threshold to capture Villi and Non-Villi pixels from the output of ACNN (probability heat-map). For segmentation maps in our multi-class lung dataset, we use score voting among the outputs of ACNNs of each class (*i.e.*, classifying a pixel by which class has the highest $H^{weighted}(x, y)$ value).

In our experiments, we realized that due to the difficulty of the segmentation task in this step, the vanilla U-Net is not a suitable architecture for the ACNN models. As such, we implemented the ACNNs with U-Net as the core segmentation architecture and EfficientNetB0 backbone as the encoder of our U-Net model. The EfficientNet[267] variants are the state-of-the-art classification models, which demonstrated the best classification performance on ImageNet dataset[268], while maintaining a small number of parameters in comparison to other state-of-the-art classification models, such as ResNet [269] variants. Although EfficientNetB0 is the simplest variant within the EfficientNet family, EfficientNetB0 is a suitable architecture for our ACNN models due to the lower complexity of the segmentation task in our datasets in comparison to ImageNet (*i.e.*, 1000 class classification on ImageNet in contrast to the ≤ 5 class classification in our dataset). To implement the EfficientNetB0-U-Net model, the fully connected layers of the EfficientNetB0 are removed, and the output of the EfficientNetB0 tower is connected to the input of the up-sampling units in the contracting path of the U-Net model. The ACNN

architecture consists of five skip connections, which corresponds to the five main convolutional layers in EfficientNetB0 architecture.

The ACNNs are trained from scratch using the Nadam optimizer with an adaptive learning rate with an initial learning rate of 10^{-3} . We noticed that our datasets were large enough for optimizing the ACNN network. We implemented the weighted binary cross-entropy loss (see Equation 6.4) for the training of our ACNN model, which enables us to address the class imbalance in our datasets.

$$\left\{ \begin{array}{l} L^{wbce} = -\frac{1}{M} \sum_{n=1}^N \sum_{m=1}^M [w \times y_m(n) \times \log(\hat{y}_m(n)) + (1 - y_m(n)) \times \log(1 - \hat{y}_m(n))] \\ w = \frac{N}{M \times N_p} \end{array} \right. \quad 6.4$$

where, L^{wbce} is the weighted binary cross-entropy loss function. $y_m(n)$ and $\hat{y}_m(n)$ are the n^{th} pixels of the m^{th} training patch of the ground truth and predicted heat-map, respectively. The M , N , N_p , and w are the number of training examples, the total number pixels, the number of pixels of the positive class, and the weight for villi class, respectively.

We used intersection over union, Dice coefficient, and pixel accuracy as the evaluation metrics for hyper-parameter tuning. The cross-validation is performed at each epoch with the ratio of 4:1 (training on 75% of the training data and validation on the remaining 25%) and the best model was saved with minimum cross-validation loss. Batch normalization was applied to each layer to reduce the training time and prevent diverging gradients. We trained our model for 100 epochs with a batch size of 10. A total number of 22,851,229 trainable parameters were optimized in our ACNN network.

6.3 Experiment

6.3.1 Dataset

The performance of the proposed multiresolution semantic segmentation method is evaluated on two different datasets: 1) high-resolution histopathology WSIs of human placenta, and 2) high-resolution histopathology WSIs of mouse lung tissue. The human placenta images are used to evaluate the method in single-class semantic segmentation (*i.e.*, Villi and Non-Villous). The mouse lung images are used to evaluate the method in multi-class semantic segmentation (*i.e.*, Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background).

6.3.1.1 Histopathology images of human placenta

The single-class dataset comprises high-resolution digital scans of 10 placental histopathology specimens obtained from the Research Centre for Women's and Infants Health (RCWIH) Biobank (Mount Sinai Hospital, Toronto, ON). The ethics approval to perform sub-analyses on the Biobank samples was obtained from the Carleton University Research Ethics Board, Ottawa Health Science Network Research Ethics Board, and the Children's Hospital of Eastern Ontario (CHEO) Research Ethics Board. The placental specimens were fixed in paraffin wax, stained with hematoxylin, washed in a 0.3% acid alcohol solution, and counterstained with eosin following the standard protocol for hematoxylin and eosin (H&E) staining at the Department of Pathology and Laboratory Medicine at the CHEO. The slides were scanned using an Aperio CS2 slide scanner (Leica), and high-resolution color images at $20\times$ magnification (*i.e.*, approximately $0.5\mu\text{m}$ per pixel) were obtained. All images were taken from different patients and the sizes of the images were up to $39,963\times 31,959$ pixels. From each placental WSI, five sub-images of

size 2740×3964 pixels were extracted, amounting to a total of 50 sub-images, which were used as our single-class dataset.

To generate the ground truth segmentation for the placenta dataset, each placental sub-image was segmented using a previously published algorithm [270] at 20× magnification. Then, the algorithm-generated segmentation maps were refined by the first author S.S. using ImageJ software (<https://imagej.net/>). In total, 10,809 Villi structures were annotated for the ground truth segmentation of our single-class placenta dataset.

6.3.1.2 Histopathology images of mice lung tissue

The multi-class semantic segmentation dataset comprises high-resolution WSIs of 20 lung histopathology specimens of mice obtained from the Sinclair Centre for Regenerative Medicine (Ottawa Hospital Research Institute, Ottawa, ON). All animal experiments were conducted in accordance with protocols approved by the University of Ottawa animal care committee. The lungs specimens were inflation fixed through the trachea with 10% buffered formalin, under 20 cm H₂O pressure, for five minutes. After the trachea was ligated, the lungs were immersion fixed in 10% buffered formalin for 48 hours at room temperature and then immersed in 70% ethanol for 24 hours at room temperature. Lungs were then paraffin-embedded, cut into 4 μ m sections, and stained with hematoxylin and eosin (H&E). The slides were scanned using an Aperio CS2 slide scanner (Leica), and high-resolution color images at 20× magnification (*i.e.*, approximately 0.5 μ m per pixel) were obtained. All WSIs were taken from different mice, and the WSI size was as large as 22,430×22,900 pixels.

Manual segmentation for these images was generated using a semi-automated approach by the first author S.S. To generate the labels, as an initial step, the background regions

were manually segmented and then, the bronchi and blood vessels were manually annotated using the ImageJ software (<https://imagej.net/>). Using the segmentation pipeline in a previously published algorithm [271], the labels for the remaining structures (*i.e.*, alveoli sacs and alveoli walls) are automatically generated at 20× magnification and then manually refined. In total, 8,447 Blood Vessel structures, 1,071 Bronchus structures, and 620,646 Alveoli Sac structures were annotated for the manual segmentation of our multi-class mouse lung tissue dataset.

6.3.2 Preprocessing

We apply a preprocessing step to address color variations and imaging artifacts, which may negatively affect segmentation performance. A number of factors can contribute to variations of the color content in histological images (e.g., histochemical staining time, amount of histology stain used) across different WSIs. We apply color normalization [272] to the input WSIs to mitigate such variations. In our dataset, there are imaging artifacts exist that appear as darkened areas of the image, which result in misclassification of the region. Examples of these regions in our placental villi dataset and the approach taken to suppress these darkened regions are based on a previously published work [270].

6.3.3 Evaluation

Semantic segmentation was performed using the proposed multiresolution approach (section 2) and compared against three state-of-the-art segmentation models: 1) U-Net [266], 2) SegNet [241], and 3) Dilated-Net [273]. We trained our proposed methods and each of these models on our single-class and multi-class datasets using 5-fold cross-validation, which enables the models to be tested on our entire dataset.

For single-class segmentation, within each fold, we used sub-images of eight WSIs from the dataset (*i.e.*, 40 sub-images) as the training dataset, and the remaining two WSIs (*i.e.*, 10 sub-images) were used as the test dataset. This was repeated five times such that the sub-images of each WSI was used as the test set. As our proposed method consists of two main training stages, at each fold, the training dataset is randomly divided into two sub-training datasets (*i.e.*, each sub-training dataset contains 20 sub-images). The first sub-training dataset is used for training of ECNNs and the second sub-training dataset is used for training of the ACNN. Similar datasets are utilized to train the evaluation models (*i.e.*, U-Net, SegNet, and DilatedNet) at each fold; however, we used both sub-training datasets

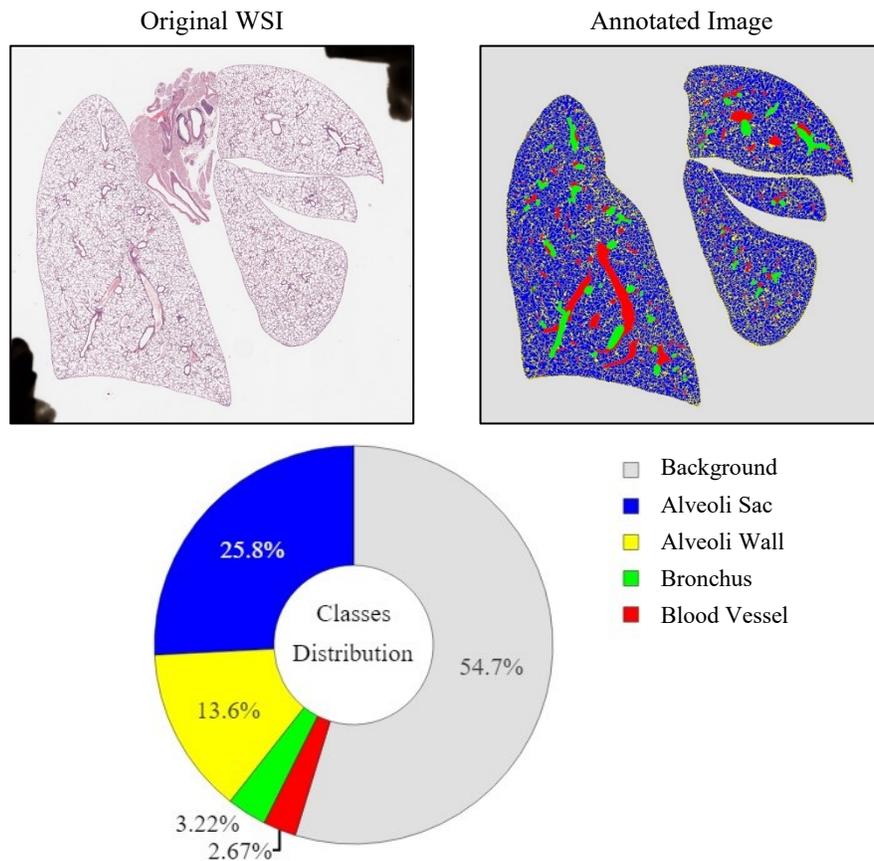


Figure 6.4 An example of histopathology WSI of mouse lung tissue and distribution of different classes.

(40 sub-images) in the training phase. The distribution of the classes in our single-class dataset was [Villi, Non-Villous] = [0.38, 0.62].

For multi-class segmentation, within each fold, we used 16 WSIs from the dataset as the training dataset, and the remaining four WSIs were used as the test dataset. This was repeated five times such that each WSI was used in the test set. For the proposed method, at each fold, the training dataset is randomly divided into two sub-training datasets (*i.e.*, each sub-training dataset contains eight WSIs) and each sub-training dataset is used for training of ECNN and ACNN stages. For evaluation of U-Net, SegNet, and Dilated-Net models, both sub-training datasets are used for training. The distribution of the classes in our multi-class dataset was [Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background] = [0.02, 0.03, 0.26, 0.14, 0.55] (see Figure 6.4). Due to the large class imbalance in our multi-class dataset, multiple steps had been undertaken. First, we excluded patches that were mainly extracted from the background class by removing patches with a mean pixel value close to white ($[R, G, B] = [255, 255, 255]$, where R, G, and B are the red, green, and blue components in RGB color mode). Second, we used data augmentation (90° rotations and image flipping) to increase the number of patches of minority classes (*i.e.*, Blood Vessel and Bronchus classes). The distribution of the classes used for training was [Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background] = [0.11, 0.17, 0.31, 0.16, 0.25].

$$PA = \sum_c \frac{TP_c + TN_c}{TP_c + TN_c + FP_c + FN_c} \quad 6.5$$

$$DSC = 1/C \times \sum_c \frac{2 \times TP_c}{2 \times TP_c + FP_c + FN_c} \quad 6.6$$

$$PPV = 1/C \times \sum_c \frac{TP_c}{TP_c + FP_c} \quad 6.7$$

6.3.4 Performance metrics

Segmentation performance is evaluated in terms of pixel-wise accuracy [251] (PA), mean Dice similarity coefficient [252] (DSC), and mean of the positive predictive value (PPV).

where, C is the number of classes, and TP_c , TN_c , FP_c , and FN_c are the numbers of true positives, true negatives, false positives, and false negatives for class c , respectively.

We also evaluated the performance of the proposed method in terms of its ability to segment the structures based on their size, on a natural log scale. To evaluate our proposed in this setting, we initially define N different size bins, where each bin represents the structures in the ground truth masks that their size falls in the range of the bin. The number

$$\begin{cases} SE_c^{B_i} = \frac{TP_c^{B_i}}{TP_c^{B_i} + FN_c^{B_i}}, i \in \{1, 2, \dots, N\} \\ B_i \in \{S_j \mid b_{i-1} \leq A_j < b_i\}, S_j \in \{S_1, S_2, \dots, S_M\}, b_i \in \{b_0, b_1, \dots, b_N\} \end{cases} \quad 6.8$$

of bins may vary between our single-class dataset and different classes in our multi-class dataset because the distribution of the structures' size are different in these classes. Then, for the final segmentation map of each class, the average sensitivity (SE) of the class for each bin is calculated using Equation 6.8. We report the average SE with respect to the structures' size:

where, B_i represents i^{th} bin, which contains the structures S_j whose size A_j is larger or equal to b_{i-1} and smaller than b_i , N denote the total number of bins, and M the total number of structures, respectively. $SE_c^{B_i}$, $TP_c^{B_i}$, and $FN_c^{B_i}$ are the average sensitivity, the number of true positives, and false negatives for class c for the structures in B_i , respectively.

6.4 Results

Figure 6.5 visualizes examples of semantic segmentation of our single-class dataset (*i.e.*, histopathology sub-images of placental villi) at 20 \times magnification. All of the segmentation models show comparably high accuracy in the segmentation of placental villi structures. This behavior is also confirmed by the performance metrics calculated over our single-class placenta dataset in Table 6.1. Based on the results in Table 6.1, our proposed method shows the best performance in terms of PA and DSC metrics, and the PPV metric is 0.017 lower than the best performing model (U-Net at 10 \times). Figure 6.6 visualizes the average sensitivity of each method with respect to the size of the placental villi structures. In Figure

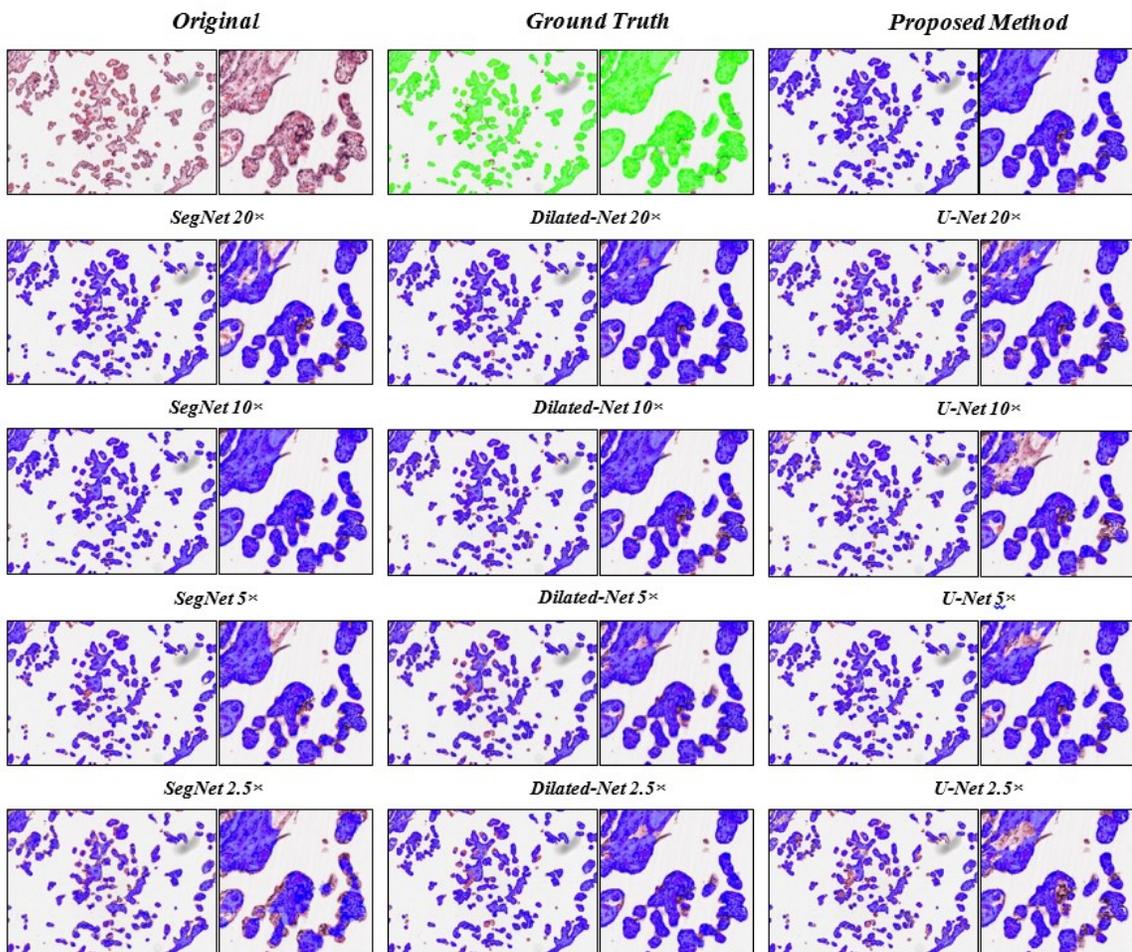


Figure 6.5 An example semantic segmentation result in single-class placenta villi dataset. The ground truth segmentation and the predictions for villi class are labeled as green and blue, respectively.

Table 6.1 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) metrics for our single-class placenta dataset. The metrics are reported for mean \pm standard deviation over each sample.

<i>MODEL</i>	<i>MAGNIFICATION</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>
U-Net	20 \times	0.917 \pm 0.028	0.846 \pm 0.016	0.986 \pm 0.011
U-Net	10 \times	0.925 \pm 0.024	0.861 \pm 0.020	0.988 \pm 0.080
U-Net	5 \times	0.941 \pm 0.018	0.895 \pm 0.013	0.982 \pm 0.012
U-Net	2.5 \times	0.901 \pm 0.032	0.810 \pm 0.035	0.981 \pm 0.011
Dilated-Net	20 \times	0.948 \pm 0.016	0.910 \pm 0.014	0.982 \pm 0.011
Dilated-Net	10 \times	0.930 \pm 0.019	0.874 \pm 0.023	0.965 \pm 0.045
Dilated-Net	5 \times	0.945 \pm 0.021	0.906 \pm 0.013	0.970 \pm 0.025
Dilated-Net	2.5 \times	0.911 \pm 0.036	0.840 \pm 0.041	0.955 \pm 0.031
SegNet	20 \times	0.934 \pm 0.026	0.883 \pm 0.015	0.983 \pm 0.010
SegNet	10 \times	0.936 \pm 0.023	0.892 \pm 0.021	0.936 \pm 0.047
SegNet	5 \times	0.910 \pm 0.032	0.833 \pm 0.022	0.970 \pm 0.022
SegNet	2.5 \times	0.841 \pm 0.053	0.681 \pm 0.031	0.903 \pm 0.045
Proposed method	(20 \times , 10 \times , 5 \times , 2.5 \times)	0.956 \pm 0.021	0.925 \pm 0.025	0.9713 \pm 0.018

6.6, our proposed method shows a consistently high performance in the segmentation of large structures as well as small structures. However, the performance of the baseline

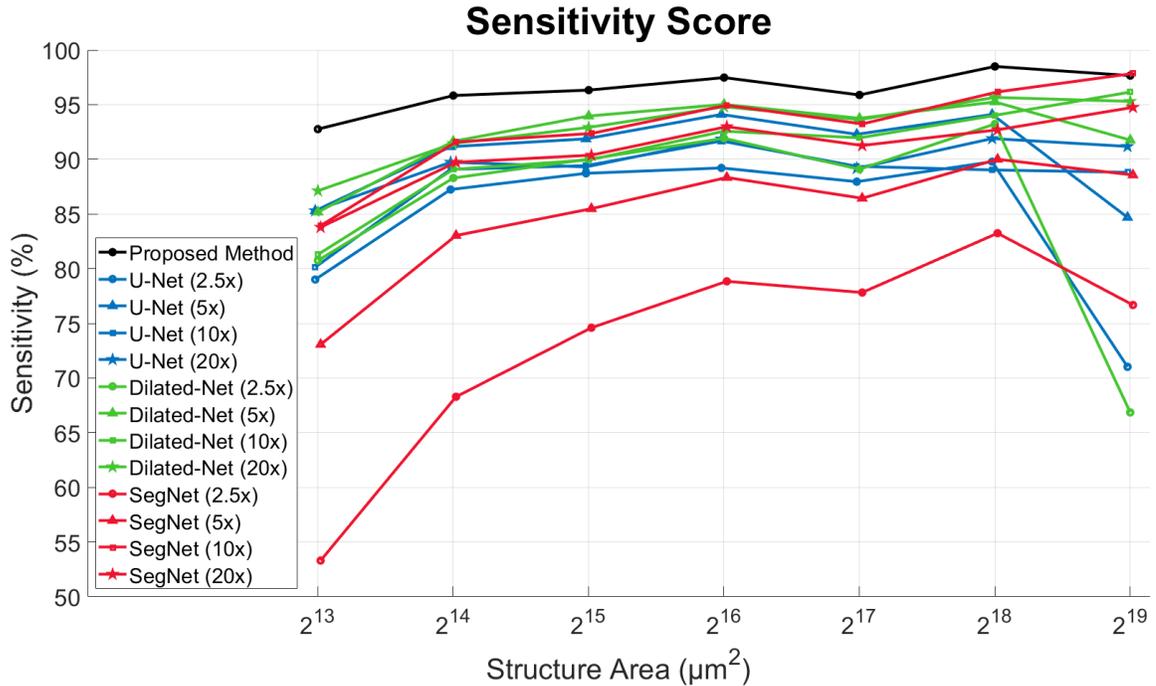


Figure 6.6 The average sensitivity of each method with respect to the structures' size in single-class placental villi dataset.

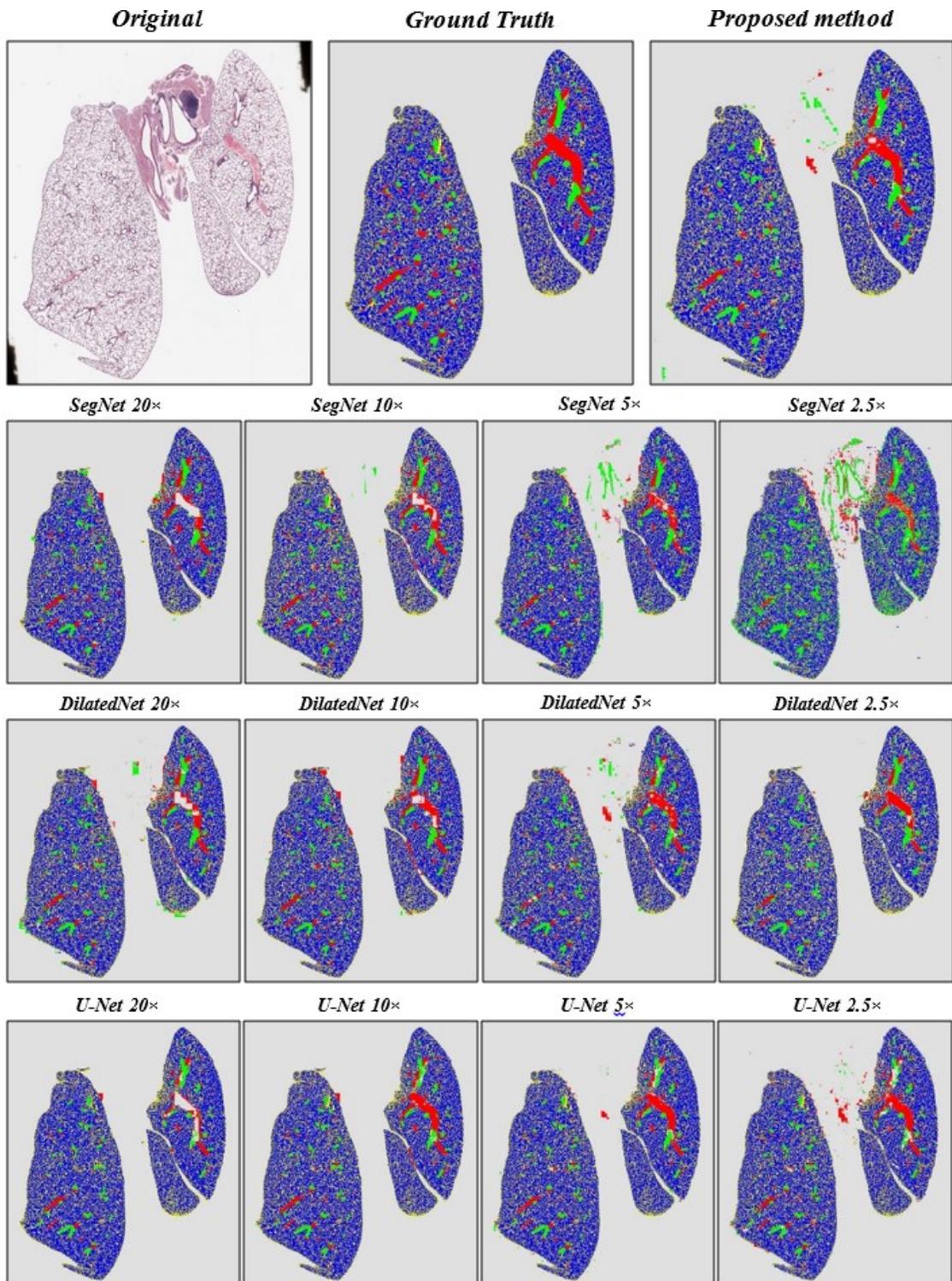


Figure 6.7 An example semantic segmentation of a lung histopathology WSI using proposed method and SegNet²⁷, DilatedNet²⁸, and U-Net¹⁹ at different magnifications. The Blood Vessel, Bronchus, Alveoli Sac, Alveoli Wall, and Background classes are labeled as red, green, blue, yellow, and gray, respectively.

methods is subject to a high variability as the size of the structures changes, specifically in the segmentation of small structures at lower magnifications.

Table 6.2 The of pixel-wise accuracy (PA), mean Dice similarity coefficient (DSC), and mean positive predictive value (PPV) metrics for our multi-class lung dataset. The metrics are reported for mean \pm standard deviation over each sample.

<i>MODEL</i>	<i>MAGNIFICATION</i>	<i>PA</i>	<i>DSC</i>	<i>PPV</i>
U-Net	20 \times	0.959 \pm 0.034	0.786 \pm 0.044	0.783 \pm 0.037
U-Net	10 \times	0.962 \pm 0.028	0.815 \pm 0.031	0.829 \pm 0.029
U-Net	5 \times	0.962 \pm 0.021	0.804 \pm 0.017	0.826 \pm 0.019
U-Net	2.5 \times	0.955 \pm 0.043	0.764 \pm 0.038	0.806 \pm 0.031
Dilated-net	20 \times	0.959 \pm 0.028	0.793 \pm 0.026	0.751 \pm 0.048
Dilated-net	10 \times	0.961 \pm 0.022	0.804 \pm 0.024	0.775 \pm 0.033
Dilated-net	5 \times	0.965 \pm 0.017	0.831 \pm 0.019	0.806 \pm 0.024
Dilated-net	2.5 \times	0.958 \pm 0.038	0.789 \pm 0.031	0.773 \pm 0.051
SegNet	20 \times	0.945 \pm 0.027	0.723 \pm 0.033	0.670 \pm 0.057
SegNet	10 \times	0.960 \pm 0.037	0.796 \pm 0.027	0.767 \pm 0.047
SegNet	5 \times	0.963 \pm 0.025	0.817 \pm 0.028	0.787 \pm 0.039
SegNet	2.5 \times	0.961 \pm 0.054	0.811 \pm 0.051	0.782 \pm 0.061
Proposed method	(20 \times , 10 \times , 5 \times , 2.5 \times)	0.971 \pm 0.024	0.873 \pm 0.033	0.833 \pm 0.028

Figure 6.7 visualizes an example of multi-class semantic segmentation of a lung histopathology WSI at 20 \times magnification, in comparison to baseline methods at 20 \times , 10 \times , 5 \times , and 2.5 \times magnification. As shown in Figure 6.7, unlike the baseline methods that perform well in segmentation of large structures at lower magnifications and small structures at higher magnifications, our proposed method is capable of accurate segmentation of both large and small structures. Table 6.2 contains the performance metrics for semantic segmentation of our multi-class lung dataset, which confirms the superior performance of our proposed method against state-of-the-art segmentation methods at different magnifications over PA, DSC, and PPV performance metrics. The average sensitivity of the Bronchus, the Blood Vessel, and the Alveoli Sac classes with respect to the structure size are visualized in Figure 6.8. In Figure 6.8.a, large differences can be observed between the performances of the baseline methods, particularly at lower magnifications (*i.e.*, 2.5 \times and 5 \times) when segmenting small structures in the Alveoli class.

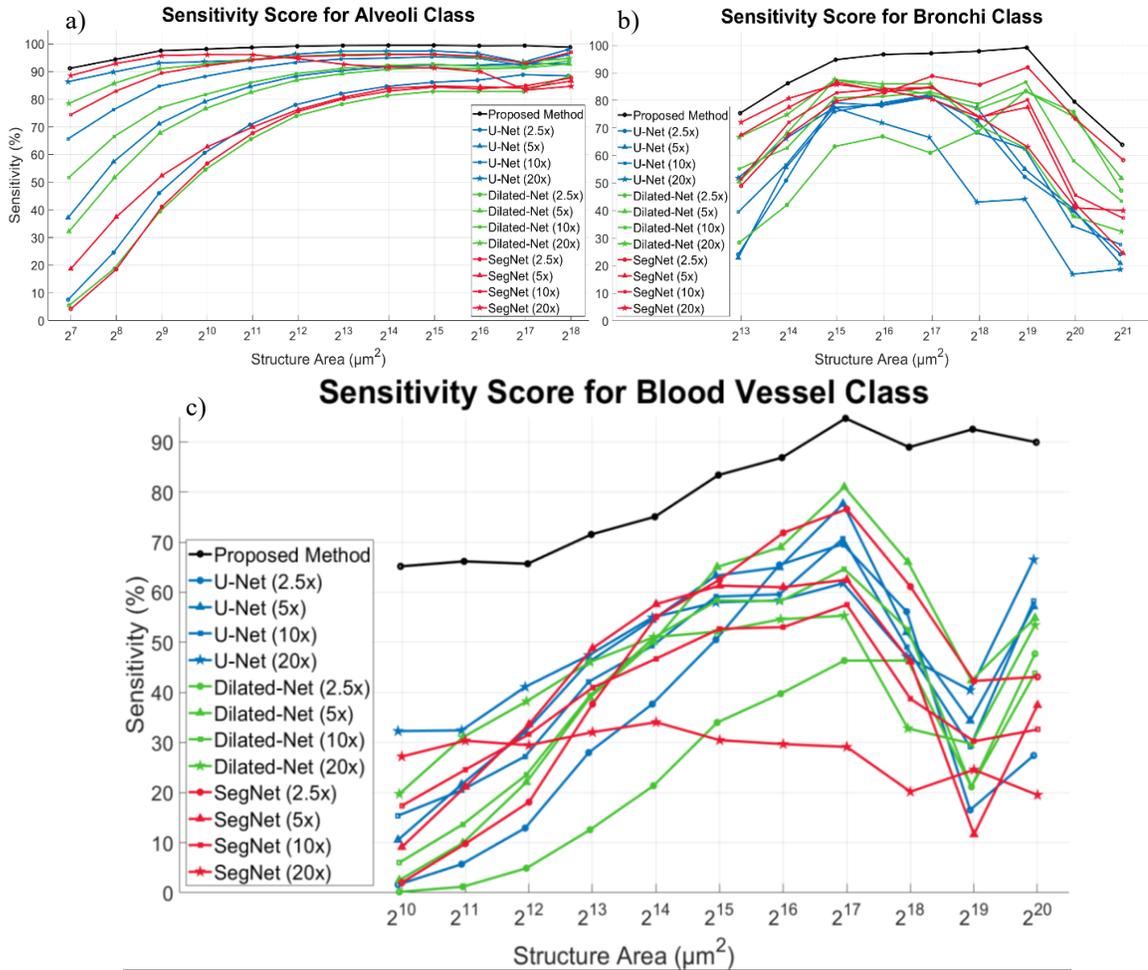


Figure 6.8 The average sensitivity of each method with respect to the structures' size for a) Alveoli class, b) Bronchus class, and c) Blood Vessel class in multi-class lung dataset.

The proposed method shows consistently high performance for the segmentation of small and large structures in this class. As it can be seen in Figure 6.8.b and Figure 6.8.c, by moving along size axis the performance of the baseline models drastically changes for segmentation of structures in Bronchus and Blood Vessel classes. The proposed method demonstrates a comparably consistent and superior performance in the segmentation across different size structures.

6.5 Discussion

In this work, we aim to resolve the three-fold trade-off between field-of-view, computational efficiency, and spatial resolution in semantic segmentation of histopathology

WSIs, which contains biological structures with drastic size variations. We proposed a two-stage multiresolution semantic segmentation approach, which extracts the contextual information from different magnifications in the first stage, and aggregates the extracted information in the second stage to improve the overall segmentation performance. Implementation of the proposed method in two separate stages enables us to independently use all possible contextual information at different magnifications for training of our model. A common approach in end-to-end multiresolution approaches is to match the field-of-view by center cropping the heat-maps at lower magnifications [163], [275], which makes the usage of the contextual information at lower magnifications inefficient and may limit the overall performance [276], [277].

For the single-class dataset, our proposed method outperformed the baseline methods in terms of DSC and PA metrics. The PPV score of our proposed method was slightly lower than the highest PPV score (1.7% lower than U-Net at 10 \times). Due to the low complexity of the segmentation task in our single-class dataset (*i.e.*, distinct visual differences between villi structures and the background regions, and relatively low size variability of the villi structures), the baseline methods perform comparably well when using patches extracted at 5 \times magnification. However, by analyzing the performance of the models based on the size of the structures in Figure 6.6, it can be observed that the size of the structures influences the performance of the baseline methods. This behaviour is more evident in the segmentation of small structures, where lowering the resolution of the input patches caused the performance of the SegNet and U-Net models at 2.5 \times and 5 \times magnifications to drastically decline. The sensitivity of most baseline methods steadily increases with increasing structure's size, and SegNet and U-Net at 5 \times magnification shows a comparable

performance against our proposed method in the segmentation of large structures (approximately 0.5 mm^2). Overall, our proposed method was able to demonstrate a slight improvement against the baseline methods, by offering a consistently high performance in segmentation of different size structures in the single-class dataset.

The overall performance of the semantic segmentation in our multi-class lung dataset indicates that our proposed method considerably outperforms other baseline methods, obtaining the highest scores in PA, DSC, and PPV performance metrics (Table 6.2). The size-based sensitivity scores showing similar trends as of the segmentation in our single-class dataset (Figure 6.8.a). This can be justified based on similar segmentation complexities in our single-class dataset and Alveoli class, where the structures' size distribution and visual distinguishability are similar. However, in the segmentation of structures in the Blood Vessel (Figure 6.8.b) and Bronchus class (Figure 6.8.c), the baseline methods show high variability in the segmentation performance of structures of different sizes. In figure 6.8.b, the baseline methods tend to segment small Bronchus structures better at higher magnifications, and segment larger Bronchus structures better at lower magnifications. This confirms the sensitivity of the segmentation task in Bronchus class to trade-off between resolution and field-of-view of the extracted patches. Our proposed method shows a consistently higher performance against the baseline methods across all structure sizes. For very large structures (*i.e.*, approximately 2 mm^2), the performance increase is smaller. This can be explained by complexity of the segmentation task in the Bronchus class, where large terminal bronchi are being segmented and not enough visual cues are present even for human interpretation. This can also be confirmed by the sharp

decline in all of our baseline methods' performance in segmentation of very large Bronchus structures.

As shown in Figure 6.8.c, similar to segmentation in Bronchus class, the baseline methods at high magnifications tend to show comparably higher performance in segmentation of small structures. However, different trends can be observed in average sensitivity of the methods as the size of the structures increase. By increasing the size in Blood Vessel class, all of the methods demonstrating a steady increase followed by a sharp decline for structures of larger than 0.2 mm^2 . However, for structures larger than 0.2 mm^2 , the performance of most baseline methods drastically increases. This may be related to the complexity of the segmentation task in Blood Vessel class, where resolution plays a key role to capture complex patterns of the structures in this class. This can be confirmed by the controversial behaviour of baseline methods at lower magnification, where U-Net at $20\times$ magnification scores the best average sensitivity among other baseline methods. Despite following a similar trend, our proposed method outperforms baseline methods in segmentation of the Blood Vessel class across all structure sizes by a large margin, and exhibits lower variability as the size of the structures change. This remarkable performance is influenced by two main factors: 1) accessing contextual information from different magnifications, and 2) using EfficientNetB0 encoder in our ACNN model.

We implemented our multiresolution approach using four level of magnification ($2.5\times$, $5\times$, $10\times$, and $20\times$). We chose the number of magnifications based on the structures' size variation in our datasets; however, this parameter was not optimized. Future work will examine the impact of the number of magnifications (e.g., ablation study). We do anticipate that having too few levels of magnifications will reduce performance and hypothesize that

having more levels of magnifications than optimal will have little to no impact, noting that since magnifications levels are separate exponentially, if there were too many magnifications it would likely only be one or perhaps two more levels.

In our experiment to develop the ACNN stage, the EfficientNetB0 was used as the encoder. We noticed that in comparison to the vanilla U-Net architecture, the EfficientNetB0-UNet model offers a more convex objective function, which makes the optimization process faster and prevents diverging gradients. We specifically chose the EfficientNetB0 encoder for our U-Net segmentation model, which has the smallest network size in EfficientNet family because 1) it has the closest input size (*i.e.*, 224×224) to the architecture used to train in ImageNet dataset, 2) our segmentation task (with five and one classes in our multi-class and single-class datasets, respectively) was not as complex as classification with over 1,000 classes on ImageNet, and 3) we had limited computation resources. The training of our models was performed on a standard workstation with an Intel Core i7-3770 3.40 GHz CPU, 12 GB of installed RAM, and a single NVIDIA RTX 2060 with 6 GB memory using Tensorflow and Keras frameworks, which is based on Python programming language. The code to generate our input image patches and performing size-based structure weighting step are developed using Matlab environment.

An important limitation of our work was to implement our proposed method in two separate stages, which prevents optimization of our model and makes the training process inefficient and time consuming. We expect further improvements in terms of training speed and performance can be achieved by re-structuring our model into a single stage. Another limitation of this work was using the sample images as input to perform segmentation in our single-class placenta dataset instead of using the entire WSIs. Each of our placental

histopathology WSIs consisted of thousands villous structures, which could take months to be completely annotated.

6.6 Conclusion

The proposed multiresolution semantic segmentation method is capable of addressing the three-fold trade-off between field-of-view, computational efficiency, and spatial resolution in analysis of high-resolution histopathology WSIs. We demonstrated that the performance of the patch-based CNNs in segmentation of histopathological structures with drastic size variations is limited by the trade-off between field-of-view and resolution of the extracted patches. Our method outperformed the state-of-the-art patch-based CNNs at different magnifications on two histopathology datasets. We demonstrated that our proposed method is able to segment histopathological structures with drastic size variations with high accuracy and consistency. In our future work, we will investigate the effect of number of magnification parameter in training of the ACNN models. This can pave the way to develop an end-to-end deep learning model that can aggregate the contextual information from different magnifications in a single stage.

Chapter 7: Summary and Future Work

Analysis of histopathology WSIs is a vital step for research, diagnosis, and treatment purposes in many clinical applications. The most accepted approaches for analysis of histopathology WSIs are manual/semi-automated techniques, which can be expensive, time-consuming, and subject to inter- and intra-rater variability. In this thesis, we proposed multiple automated pipelines for the segmentation of complex biological structures in high-resolution histopathology images, which is a fundamental step in many AHIA applications.

The developed methods had been evaluated in two clinical research applications: 1) segmentation of placental villi in histopathology images of human placenta, and 2) segmentation of main biological structures in histopathology images of mouse lung tissue. Through comprehensive evaluation against manual assessments, we demonstrated that each proposed method addresses a specific bottleneck in AHIA, which can be a potential replacement for conventional manual/semi-automated practices in the field.

In this research, the following contributions have been made:

1. Developed a fully automated pipeline for rules-based segmentation of villi in histopathology WSIs of human placenta. To acquire the foreground region (*i.e.*, villous structures), color image-processing techniques were applied. We developed a novel concavity detection method to efficiently propose potential overlapping regions in the foreground masks, which were classified using conventional machine learning approaches. The proposed segmentation method yielded high accuracy and consistency in the detection of villous structures for a dataset comprising of nearly 5000 villi, considerably outperforming the previously published methods.

2. Developed a fully automated pipeline for rules-based segmentation of biological structures in histopathology WSIs of mouse lung tissue. In our approach, we implemented a CNN model based on U-Net architecture to extract foreground regions (*i.e.*, diagnostically relevant tissue) of histopathology WSIs of the mouse lung tissue. Color image processing were used to acquire different structures of the mouse lung (*i.e.*, blood vessel, bronchi, and alveoli) and a conventional machine learning model was used to detect each structure. The comparison between the proposed method and the manual assessment showed substantial agreement in the calculation of the MLI score. The result demonstrated that the proposed method could replace the manual/semi-automated methods for calculating the MLI score.
3. Developed a generic deep learning approach for automated segmentation of complex biological structures in histopathology WSIs of human placenta and mouse lung tissue. We developed a generic deep learning pipeline using three state-of-the-art semantic segmentation models. We demonstrated that despite evident visual differences between histopathology WSIs of human placenta and mouse lung tissue, the deep learning models could be easily used for semantic segmentation in both applications with minor modifications. We also showed that the deep learning models outperform the conventional approaches in terms of detection of touching villi in placenta WSIs and estimation of MLI score in mouse lung WSIs.
4. Propose a novel multi-resolution semantic segmentation method for segmentation of complex biological structures in histopathology images, which utilize the contextual information of histopathology images at different spatial resolutions and aggregate this information to generate higher overall performance. The proposed algorithm was

capable of improving the state-of-the-art segmentation accuracy by addressing the dramatic size difference (*i.e.*, approximately 1,000,000 times difference). We proposed a multi-resolution semantic segmentation method to address the three-fold trade-off between field-of-view, computational efficiency, and spatial resolution in the analysis of high-resolution histopathology WSIs. We demonstrated that the performance of the patch-based CNNs in the segmentation of histopathological structures with drastic size variations is limited by the trade-off between field-of-view and resolution of the extracted patches. Our method outperformed the state-of-the-art patch-based CNNs at different magnifications on histopathology WSIs of placental villi and mouse lung tissue. We demonstrated that our proposed method is able to segment histopathological structures with drastic size variations with high accuracy and consistency.

An important limitation in this thesis was to acquire the ground truth segmentation of the main biological structures in our histopathology datasets. Accessing a large training dataset can have a significant role in the performance of the segmentation models, specifically in deep learning-based approaches. Manual segmentation of high-resolution histopathology WSIs is a complex and time-consuming task, which requires pathology expertise to perform the annotations. As such, by the time of completion of this thesis, we were able to manually segment 50 sub-images of size 2740×3964 from 12 human placenta WSIs and 20 WSIs of mouse lung tissue. The performance of our proposed methods could be improved by simply providing more manually segmented images. The potential solutions to overcome this challenge in semantic segmentation in AHIA can be utilizing semi-supervised [278] and weakly supervised [279] algorithms. In semi-supervised

learning methods, a small labeled dataset and a large unlabeled dataset are used to train the deep learning model. The labeled dataset is used to train a CNN model. Then, the trained CNN is used to make predictions on the unlabeled dataset; the pseudo labels (*i.e.*, the labels with the highest prediction confidence) and the labeled data are linked together to train a new CNN, which can potentially enhance the performance of the resultant model. This approach is specifically applicable in our datasets, where large unlabeled WSIs are available.

Weakly supervised learning methods, on the other hand, benefit from decreasing the cost and increasing the efficiency of human effort to provide manually segmented data. However, due to the complexity of biological structures in high-resolution histopathology WSIs, the weak labeling can potentially decrease the performance of the resultant model.

In preparation for our single-class placenta dataset, we randomly extracted multiple sample images from 12 different placenta WSIs of healthy and diagnosed with PE. As such, although, the images contained large variability in terms of imaging artifacts, undesired biological components, and diverse ranges of villous structures, due to complexity and difficulty of manual segmentation of the placenta WSIs, not all regions in the WSIs were included in development of our proposed methods. Moreover, we realized that in order to develop a robust and generalizable automated pipeline, we needed more placenta WSIs than only 12, which can limit the performance of our proposed method in processing unseen data. In our multi-class lung dataset, on the other hand, we were able to use the entire regions of our lung histopathology WSIs (*i.e.*, 20 WSIs) from four different experimental groups. However, similar to placenta dataset, the number of WSIs used in developing of our proposed methods required to include more histopathology WSIs of

mouse lung tissue to improve the robustness and generalizability. As a result, in our future work, we will include a larger and more diverse datasets to address the above mentioned limitations.

Another limitation of our work was the implementation of the proposed multi-resolution semantic segmentation method in two separate stages, which prevents global optimization of our model and makes the training process inefficient and time-consuming. Further improvements in terms of training speed and performance may be achieved by restructuring our model into a single stage, where contextual information from different magnifications is extracted in the contracting path of a single CNN model. However, to maintain a suitable field-of-view for extracting image patches, which contain required contextual information for accurate segmentation of large biological structures in histopathology WSIs, the size of the extracted patches will be very large. As such, due to the computational complexity of the resultant model, extensive computational resources will be required for the training stage.

References

- [1] Katsuragawa, S., & Doi, K. (2007). Computer-aided diagnosis in chest radiography. *Computerized Medical Imaging and Graphics*, 31(4-5), 212-223.
- [2] MacMahon, H., Doi, K., Chan, H. P., Giger, M. L., Katsuragawa, S., & Nakamori, N. (1990). Computer-aided diagnosis in chest radiology. *J Thorac Imaging*, 5(1), 67-76.
- [3] Van Ginneken, B., Hogeweg, L., & Prokop, M. (2009). Computer-aided diagnosis in chest radiography: Beyond nodules. *European journal of radiology*, 72(2), 226-230.
- [4] Abe, Y., Hanai, K., Nakano, M., Ohkubo, Y., Hasizume, T., Kakizaki, T., ... & Moriyama, N. (2005). A computer-aided diagnosis (CAD) system in lung cancer screening with computed tomography. *Anticancer research*, 25(1B), 483-488.
- [5] Petrick, N., Haider, M., Summers, R. M., Yeshwant, S. C., Brown, L., Iuliano, E. M., & Pickhardt, P. J. (2008). CT colonography with computer-aided detection as a second reader: observer performance study. *Radiology*, 246(1), 148-156.
- [6] Yoshida, H., & Dachman, A. H. (2004, October). Computer-aided diagnosis for CT colonography. In *Seminars in Ultrasound, CT and MRI* (Vol. 25, No. 5, pp. 419-431). WB Saunders.
- [7] Elizabeth, D. S., Nehemiah, H. K., Raj, C. R., & Kannan, A. (2012). Computer-aided diagnosis of lung cancer based on analysis of the significant slice of chest computed tomography image. *IET image processing*, 6(6), 697-705.
- [8] Kumar, S. S., Moni, R. S., & Rajeesh, J. (2013). An automatic computer-aided diagnosis system for liver tumours on computed tomography images. *Computers & Electrical Engineering*, 39(5), 1516-1526.
- [9] Hua, K. L., Hsu, C. H., Hidayati, S. C., Cheng, W. H., & Chen, Y. J. (2015). Computer-aided classification of lung nodules on computed tomography images via deep learning technique. *OncoTargets and therapy*, 8.
- [10] Christe, A., Leidolt, L., Huber, A., Steiger, P., Szucs-Farkas, Z., Roos, J. E., ... & Ebner, L. (2013). Lung cancer screening with CT: evaluation of radiologists and different computer assisted detection software (CAD) as first and second readers for lung nodule detection at different dose levels. *European journal of radiology*, 82(12), e873-e878.
- [11] Niaf, E., Rouvière, O., Mège-Lechevallier, F., Bratan, F., & Lartizien, C. (2012). Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Physics in Medicine & Biology*, 57(12), 3833.
- [12] El-Dahshan, E. S. A., Mohsen, H. M., Revett, K., & Salem, A. B. M. (2014). Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm. *Expert systems with Applications*, 41(11), 5526-5545.
- [13] Lartizien, C., Rogez, M., Niaf, E., & Ricard, F. (2013). Computer-aided staging of lymphoma patients with FDG PET/CT imaging based on textural information. *IEEE journal of biomedical and health informatics*, 18(3), 946-955.
- [14] Martínez-Murcia, F. J., Górriz, J. M., Ramirez, J., Puntonet, C. G., Salas-Gonzalez, D., & Alzheimer's Disease Neuroimaging Initiative. (2012). Computer aided diagnosis tool for Alzheimer's disease based on Mann-Whitney-Wilcoxon U-test. *Expert Systems with Applications*, 39(10), 9676-9685.
- [15] Moon, W. K., Lo, C. M., Chang, J. M., Huang, C. S., Chen, J. H., & Chang, R. F. (2012). Computer-aided classification of breast masses using speckle features of automated breast ultrasound images. *Medical physics*, 39(10), 6465-6473.
- [16] Cui, J., Sahiner, B., Chan, H. P., Nees, A., Paramagul, C., Hadjiiski, L. M., & Shi, J. (2009). A new automated method for the segmentation and characterization of breast masses on ultrasound images. *Medical physics*, 36(5), 1553-1565.
- [17] Drukker, K., Sennett, C. A., & Giger, M. L. (2008). Automated method for improving

- system performance of computer-aided diagnosis in breast ultrasound. *IEEE Transactions on medical imaging*, 28(1), 122-128.
- [18] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B. (2009). Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2, 147-171.
- [19] Komura, D., & Ishikawa, S. (2018). Machine learning methods for histopathological image analysis. *Computational and structural biotechnology journal*, 16, 34-42.
- [20] Srinidhi, C. L., Ciga, O., & Martel, A. L. (2021). Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67, 101813.
- [21] Jimenez-del-Toro, O., Otálora, S., Andersson, M., Eurén, K., Hedlund, M., Rousson, M., & Atzori, M. (2017). Analysis of histopathology images: From traditional machine learning to deep learning. In *Biomedical Texture Analysis* (pp. 281-314). Academic Press.
- [22] Pal, N. R., & Pal, S. K. (1993). A review on image segmentation techniques. *Pattern recognition*, 26(9), 1277-1294.
- [23] Dabass, M., Vig, R., & Vashisth, S. (2018, December). Review of histopathological image segmentation via current deep learning approaches. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1-6). IEEE.
- [24] Tosta, T. A. A., Neves, L. A., & do Nascimento, M. Z. (2017). Segmentation methods of H&E-stained histological images of lymphoma: a review. *Informatics in medicine unlocked*, 9, 35-43.
- [25] Xing, F., & Yang, L. (2016). Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: a comprehensive review. *IEEE reviews in biomedical engineering*, 9, 234-263.
- [26] Mahmood, F., Borders, D., Chen, R. J., McKay, G. N., Salimian, K. J., Baras, A., & Durr, N. J. (2019). Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE transactions on medical imaging*, 39(11), 3257-3267.
- [27] Kannan, S., Morgan, L. A., Liang, B., Cheung, M. G., Lin, C. Q., Mun, D., & Kolachalama, V. B. (2019). Segmentation of glomeruli within trichrome images using deep learning. *Kidney international reports*, 4(7), 955-962.
- [28] Sirinukunwattana, K., Pluim, J. P., Chen, H., Qi, X., Heng, P. A., Guo, Y. B., & Rajpoot, N. M. (2017). Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35, 489-502.
- [29] Kidron, D., Vainer, I., Fisher, Y., & Sharony, R. (2017). Automated image analysis of placental villi and syncytial knots in histological sections. *Placenta*, 53, 113-118.
- [30] Swiderska-Chadaj, Z., Markiewicz, T., Koktysz, R., & Cierniak, S. (2018). Image processing methods for the structural detection and gradation of placental villi. *Computers in biology and medicine*, 100, 259-269.
- [31] Qaiser, T., Tsang, Y. W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D., & Rajpoot, N. (2019). Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. *Medical image analysis*, 55, 1-14.
- [32] Chan, L., Hosseini, M. S., Rowsell, C., Plataniotis, K. N., & Damaskinos, S. (2019). Histosegnet: Semantic segmentation of histological tissue type in whole slide images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10662-10671).
- [33] Fraser, A., Nelson, S. M., Macdonald-Wallis, C., Cherry, L., Butler, E., Sattar, N., & Lawlor, D. A. (2012). Associations of pregnancy complications with calculated cardiovascular disease risk and cardiovascular risk factors in middle age: the Avon Longitudinal Study of Parents and Children. *Circulation*, 125(11), 1367-1380.
- [34] Swiderska-Chadaj, Z., Markiewicz, T., Koktysz, R., & Cierniak, S. (2018). Image processing methods for the structural detection and gradation of placental villi. *Computers in biology and medicine*, 100, 259-269.

- [35] Kidron, D., Vainer, I., Fisher, Y., & Sharony, R. (2017). Automated image analysis of placental villi and syncytial knots in histological sections. *Placenta*, 53, 113-118.
- [36] Knudsen, L., Weibel, E. R., Gundersen, H. J. G., Weinstein, F. V., & Ochs, M. (2010). Assessment of air space size characteristics by intercept (chord) measurement: an accurate and efficient stereological approach. *Journal of applied physiology*, 108(2), 412-421.
- [37] Thébaud, B., Goss, K. N., Laughon, M., Whitsett, J. A., Abman, S. H., Steinhorn, R. H., ... & Jobe, A. H. (2019). Bronchopulmonary dysplasia. *Nature reviews Disease primers*, 5(1), 1-23.
- [38] Mahesh, M. (2013). The essential physics of medical imaging. *Medical physics*, 40(7), 077301.
- [39] Smith-Bindman, R., Kwan, M. L., Marlow, E. C., Theis, M. K., Bolch, W., Cheng, S. Y., & Miglioretti, D. L. (2019). Trends in use of medical imaging in US health care systems and in Ontario, Canada, 2000-2016. *Jama*, 322(9), 843-856.
- [40] Dos Reis, F. J. J., de Barros e Silva, V., de Lucena, R. N., Mendes Cardoso, B. A., & Nogueira, L. C. (2016). Measuring the pain area: an intra-and inter-rater reliability study using image analysis software. *Pain Practice*, 16(1), 24-30.
- [41] Sommers, M., Beacham, B., Baker, R., & Fargo, J. (2013). Intra-and inter-rater reliability of digital image analysis for skin color measurement. *Skin Research and Technology*, 19(4), 484-491.
- [42] Costa, A. C., Dibai Filho, A. V., Packer, A. C., & Rodrigues-Bigaton, D. (2013). Intra and inter-rater reliability of infrared image analysis of masticatory and upper trapezius muscles in women with and without temporomandibular disorder. *Brazilian journal of physical therapy*, 17, 24-31.
- [43] Waite, S., Scott, J., Gale, B., Fuchs, T., Kolla, S., & Reede, D. (2017). Interpretive error in radiology. *American Journal of Roentgenology*, 208(4), 739-749.
- [44] Bruno, M. A., Walker, E. A., & Abujudeh, H. H. (2015). Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. *Radiographics*, 35(6), 1668-1676.
- [45] Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise differences in the comprehension of visualizations: A meta-analysis of eye-tracking research in professional domains. *Educational psychology review*, 23(4), 523-552.
- [46] Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, 72(5), 1205-1217.
- [47] Brady, A. P. (2017). Error and discrepancy in radiology: inevitable or avoidable?. *Insights into imaging*, 8(1), 171-182.
- [48] Giger, M. L., & Suzuki, K. (2008). Computer-aided diagnosis. In *Biomedical information technology* (pp. 359-XXII). Academic Press.
- [49] Doi, K. (2005). Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology*, 78(suppl_1), s3-s19.
- [50] Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5), 198-211.
- [51] Giger, M. L., Chan, H. P., & Boone, J. (2008). Anniversary paper: history and status of CAD and quantitative image analysis: the role of medical physics and AAPM. *Medical physics*, 35(12), 5799-5820.
- [52] Lusted, L. B. (1955). Medical electronics. *New England Journal of Medicine*, 252(14), 580-585.
- [53] Lodwick, G. S., Keats, T. E., & Dorst, J. P. (1963). The coding of roentgen images for computer analysis as applied to lung cancer. *Radiology*, 81(2), 185-200.
- [54] Becker, H. C., Nettleton, W. J., Meyers, P. H., Sweeney, J. W., & Nice, C. M. (1964). Digital computer determination of a medical diagnostic index directly from chest X-ray images. *IEEE Transactions on Biomedical Engineering*, (3), 67-72.

- [55] Toriwaki, J. I., Suenaga, Y., Negoro, T., & Fukumura, T. (1973). Pattern recognition of chest X-ray images. *Computer Graphics and Image Processing*, 2(3-4), 252-271.
- [56] Roellinger Jr, F. X., Kahveci, A. E., Chang, J. K., Harlow, C. A., Dwyer III, S. J., & Lodwick, G. S. (1973). Computer analysis of chest radiographs. *Computer Graphics and Image Processing*, 2(3-4), 232-251.
- [57] Bağcı, U., Bray, M., Caban, J., Yao, J., & Mollura, D. J. (2012). Computer-assisted detection of infectious lung diseases: a review. *Computerized Medical Imaging and Graphics*, 36(1), 72-84.
- [58] Lemaître, G., Martí, R., Freixenet, J., Vilanova, J. C., Walker, P. M., & Meriaudeau, F. (2015). Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Computers in biology and medicine*, 60, 8-31.
- [59] Filippou, V., & Tsoumpas, C. (2018). Recent advances on the development of phantoms using 3D printing for imaging with CT, MRI, PET, SPECT, and ultrasound. *Medical physics*, 45(9), e740-e760.
- [60] Jalalian, A., Mashohor, S. B., Mahmud, H. R., Saripan, M. I. B., Ramli, A. R. B., & Karasfi, B. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, 37(3), 420-426.
- [61] G. Rolls, "An Introduction to Specimen Processing." [Online]. Available: <https://www.leicabiosystems.com/knowledge-pathway/an-introduction-to-specimen-processing/>.
- [62] Mescher, A. L. (2013). *Junqueira's basic histology: text and atlas* (Vol. 12). 13th ed. New York: McGraw-Hill Medical.
- [63] Y. Rivenson *et al.*, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," *Nat. Biomed. Eng.*, 2019.
- [64] C. Laurent *et al.*, "Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., & Ozcan, A. (2019). Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nature biomedical engineering*, 3(6), 466-477.," *Hum. Pathol.*, 2013.
- [65] Lauro, G. R., Cable, W., Lesniak, A., Tseytlin, E., McHugh, J., Parwani, A., & Pantanowitz, L. (2013). Digital pathology consultations—a new era in digital imaging, challenges and practical applications. *Journal of digital imaging*, 26(4), 668-677.
- [66] Roy, S., kumar Jain, A., Lal, S., & Kini, J. (2018). A study about color normalization methods for histopathology images. *Micron*, 114, 42-61.
- [67] Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer graphics and applications*, 21(5), 34-41.
- [68] Macenko, M., Niethammer, M., Marron, J. S., Borland, D., Woosley, J. T., Guan, X., & Thomas, N. E. (2009, June). A method for normalizing histology slides for quantitative analysis. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1107-1110). IEEE.
- [69] Khan, A. M., Rajpoot, N., Treanor, D., & Magee, D. (2014). A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Transactions on Biomedical Engineering*, 61(6), 1729-1738.
- [70] Li, X., & Plataniotis, K. N. (2015). A complete color normalization approach to histopathology images using color cues computed from saturation-weighted statistics. *IEEE Transactions on Biomedical Engineering*, 62(7), 1862-1873.
- [71] Aljuhani, A., Srivastava, A., Cronin, J. P., Chan, J., Machiraju, R., & Parwani, A. V. (2022). *Whole Slide Imaging: Deep Learning and Artificial Intelligence*. In *Whole Slide Imaging* (pp. 223-236). Springer, Cham.
- [72] Aswathy, M. A., & Jagannath, M. (2017). Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked*, 8, 74-79.
- [73] Gurcan, M. N., Boucheron, L. E., Can, A., Madabhushi, A., Rajpoot, N. M., & Yener, B.

- (2009). Histopathological image analysis: A review. *IEEE reviews in biomedical engineering*, 2, 147-171.
- [74] Veta, M., Pluim, J. P., Van Diest, P. J., & Viergever, M. A. (2014). Breast cancer histopathology image analysis: A review. *IEEE transactions on biomedical engineering*, 61(5), 1400-1411.
- [75] Tokuoka, Y., Yamada, T. G., Mashiko, D., Ikeda, Z., Hiroi, N. F., Kobayashi, T. J., & Funahashi, A. (2020). 3D convolutional neural networks-based segmentation to acquire quantitative criteria of the nucleus during mouse embryogenesis. *NPJ systems biology and applications*, 6(1), 1-12.
- [76] Irshad, H., Veillard, A., Roux, L., & Racoceanu, D. (2013). Methods for nuclei detection, segmentation, and classification in digital histopathology: a review—current status and future potential. *IEEE reviews in biomedical engineering*, 7, 97-114.
- [77] Belsare, A. D., & Mushrif, M. M. (2012). Histopathological image analysis using image processing techniques: An overview. *Signal & Image Processing*, 3(4), 23.
- [78] Kumar, R., Srivastava, R., & Srivastava, S. (2015). Detection and classification of cancer from microscopic biopsy images using clinically significant and biologically interpretable features. *Journal of medical engineering*, 2015.
- [79] Basavanthally, A., Agner, S., Alexe, G., Bhanot, G., Ganesan, S., & Madabhushi, A. (2008). Manifold learning with graph-based features for identifying extent of lymphocytic infiltration from high grade, her2+ breast cancer histology. *Image Anal. Appl. Biol.(in Conjunction MICCAI)*, New York [Online]. Available: <http://www.miaab.org/miaab-2008-papers/27-miaab-2008-paper-21.pdf>.
- [80] Ladekarl, M., Bæk-Hansen, T., Henrik-Nielsen, R., Mouritzen, C., Henriques, U., & Sørensen, F. B. (1995). Objective malignancy grading of squamous cell carcinoma of the lung. Stereologic estimates of mean nuclear size are of prognostic value, independent of clinical stage of disease. *Cancer*, 76(5), 797-802.
- [81] Adiga, U., Malladi, R., Fernandez-Gonzalez, R., & de Solorzano, C. O. (2006). High-throughput analysis of multispectral images of breast cancer tissue. *IEEE Transactions on Image Processing*, 15(8), 2259-2268.
- [82] Monaco, J., Tomaszewski, J. E., Feldman, M. D., Moradi, M., Mousavi, P., Boag, A., ... & Madabhushi, A. (2009, March). Probabilistic pairwise markov models: Application to prostate cancer detection. In *Medical Imaging 2009: Image Processing* (Vol. 7259, p. 725903). International Society for Optics and Photonics.
- [83] Gil, J., Wu, H., & Wang, B. Y. (2002). Image analysis and morphometry in the diagnosis of breast cancer. *Microscopy research and technique*, 59(2), 109-118.
- [84] Naik, S., Doyle, S., Feldman, M., Tomaszewski, J., & Madabhushi, A. (2007, September). Gland segmentation and computerized gleason grading of prostate histology by integrating low-, high-level and domain specific information. In *MIAAB workshop* (pp. 1-8).
- [85] Naik, S., Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008, May). Automated gland and nuclei segmentation for grading of prostate and breast cancer histopathology. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 284-287). IEEE.
- [86] Doyle, S., Agner, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2008, May). Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 496-499). IEEE.
- [87] Jütting, U., Gais, P., Rodenacker, K., Böhm, J., Koch, S., Präuer, H. W., & Höfler, H. (1999). Diagnosis and prognosis of neuroendocrine tumours of the lung by means of high resolution image analysis. *Analytical Cellular Pathology*, 18(2), 109-119.
- [88] Doyle, S., Madabhushi, A., Feldman, M., & Tomaszewski, J. (2006, October). A boosting cascade for automated detection of prostate cancer from digitized histology. In *International*

- conference on medical image computing and computer-assisted intervention (pp. 504-511). Springer, Berlin, Heidelberg.
- [89] Petushi, S., Garcia, F. U., Haber, M. M., Katsinis, C., & Tozeren, A. (2006). Large-scale computations on histology images reveal grade-differentiating parameters for breast cancer. *BMC medical imaging*, 6(1), 1-11.
- [90] Kayser, G., Riede, U., Werner, M., Hufnagl, P., & Kayser, K. (2002). Towards an automated morphological classification of histological images of common lung carcinomas. *Elec J Pathol Histol*, 8, 022-03.
- [91] Saha, M., Chakraborty, C., & Racoceanu, D. (2018). Efficient deep learning model for mitosis detection using breast histopathology images. *Computerized Medical Imaging and Graphics*, 64, 29-40.
- [92] Cruz-Roa, A. A., Ovalle, J. E. A., Madabhushi, A., & Osorio, F. A. G. (2013, September). A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 403-410). Springer, Berlin, Heidelberg.
- [93] Demir, C., & Yener, B. (2005). Automated cancer diagnosis based on histopathological images: a systematic survey. Rensselaer Polytechnic Institute, Tech. Rep.
- [94] Demir, C., & Yener, B. (2005). Automated cancer diagnosis based on histopathological images: a systematic survey. Rensselaer Polytechnic Institute, Tech. Rep.
- [95] Loukas, C. G., & Linney, A. (2004). A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. *Computer Methods and Programs in Biomedicine*, 74(3), 183-199.
- [96] Xing, F., Xie, Y., Su, H., Liu, F., & Yang, L. (2017). Deep learning in microscopy image analysis: A survey. *IEEE transactions on neural networks and learning systems*, 29(10), 4550-4568.
- [97] Li, C., Chen, H., Li, X., Xu, N., Hu, Z., Xue, D., ... & Sun, H. (2020). A review for cervical histopathology image analysis using machine vision approaches. *Artificial Intelligence Review*, 53(7), 4821-4862.
- [98] Mosquera-Lopez, C., Agaian, S., Velez-Hoyos, A., & Thompson, I. (2014). Computer-aided prostate cancer diagnosis from digitized histopathology: a review on texture-based systems. *IEEE reviews in biomedical engineering*, 8, 98-113.
- [99] Bándi, P., van de Loo, R., Intezar, M., Geijs, D., Ciompi, F., van Ginneken, B., & Litjens, G. (2017, April). Comparison of different methods for tissue segmentation in histopathological whole-slide images. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 591-595). IEEE.
- [100] Veta, M., Van Diest, P. J., Kornegoor, R., Huisman, A., Viergever, M. A., & Pluim, J. P. (2013). Automatic nuclei segmentation in H&E stained breast cancer histopathology images. *PloS one*, 8(7), e70221.
- [101] Wang, P., Hu, X., Li, Y., Liu, Q., & Zhu, X. (2016). Automatic cell nuclei segmentation and classification of breast cancer histopathology images. *Signal Processing*, 122, 1-13.
- [102] Singh, M., Kalaw, E. M., Giron, D. M., Chong, K. T., Tan, C. L., & Lee, H. K. (2017). Gland segmentation in prostate histopathological images. *Journal of medical imaging*, 4(2), 027501.
- [103] Khan, A. M., & Ravi, S. (2013). Image segmentation methods: A comparative study.
- [104] Kaur, D., & Kaur, Y. (2014). Various image segmentation techniques: a review. *International Journal of Computer Science and Mobile Computing*, 3(5), 809-814.
- [105] Seerha, G. K., & Kaur, R. (2013). Review on recent image segmentation techniques. *International Journal on Computer Science and Engineering*, 5(2), 109.
- [106] Kumar, R., & Arthanarice, A. M. (2013). A comparative study of image segmentation using edge-based approach. *International Journal of Mathematical and Computational Sciences*,

- 7(3), 510-514.
- [107] Kanopoulos, N., Vasanthavada, N., & Baker, R. L. (1988). Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits*, 23(2), 358-367.
 - [108] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698.
 - [109] Muthukrishnan, R., & Radha, M. (2011). Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology*, 3(6), 259.
 - [110] Haider, W., Malik, M. S., Raza, M., Wahab, A., Khan, I. A., Zia, U., & Bashir, H. (2012). A hybrid method for edge continuity based on Pixel Neighbors Pattern Analysis (PNPA) for remote sensing satellite images. *Int'l J. of Communications, Network and System Sciences*, 5(29), 624-630.
 - [111] Bali, A., & Singh, S. N. (2015, February). A review on the strategies and techniques of image segmentation. In *2015 Fifth International Conference on Advanced Computing & Communication Technologies* (pp. 113-120). IEEE.
 - [112] Nazeran, H., Rice, F., Moran, W., & Skinner, J. (1995). Biomedical image processing in pathology: a review. *Australas Phys Eng Sci Med*, 18(1), 26-38.
 - [113] Senthilnath, J., Omkar, S. N., Mani, V., Tejovanth, N., Diwakar, P. G., & Shenoy, A. B. (2012). Hierarchical clustering algorithm for land cover mapping using satellite images. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(3), 762-768.
 - [114] Huang, Z., Wang, X., Wang, J., Liu, W., & Wang, J. (2018). Weakly-supervised semantic segmentation network with deep seeded region growing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7014-7023).
 - [115] Rouhi, R., Jafari, M., Kasaei, S., & Keshavarzian, P. (2015). Benign and malignant breast tumors classification based on region growing and CNN segmentation. *Expert Systems with Applications*, 42(3), 990-1002.
 - [116] Kang, W. X., Yang, Q. Q., & Liang, R. P. (2009, March). The comparative research on image segmentation algorithms. In *2009 First International Workshop on Education Technology and Computer Science* (Vol. 2, pp. 703-707). IEEE.
 - [117] Lee, S. U., Chung, S. Y., & Park, R. H. (1990). A comparative performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics, and Image Processing*, 52(2), 171-190.
 - [118] Suetens, P., Bellon, E., Vandermeulen, D., Smet, M., Marchal, G., Nuyts, J., & Mortelmans, L. (1993). Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *European journal of radiology*, 17(1), 14-21.
 - [119] Lindeberg, T., & Li, M. X. (1997). Segmentation and classification of edges using minimum description length approximation and complementary junction cues. *Computer Vision and Image Understanding*, 67(1), 88-98.
 - [120] Zaitoun, N. M., & Aqel, M. J. (2015). Survey on image segmentation techniques. *Procedia Computer Science*, 65, 797-806.
 - [121] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.
 - [122] Thilagamani, S., & Shanthi, N. (2011). A survey on image segmentation through clustering. *International Journal of Research and Reviews in Information Sciences*, 1(1), 14-17.
 - [123] Hassan, T. M., Elmogy, M., & Sallam, E. (2015). Medical image segmentation for liver diseases: a survey. *International Journal of Computer Applications*, 118(19).
 - [124] Singh, A., Yadav, A., & Rana, A. (2013). K-means with Three different Distance Metrics. *International Journal of Computer Applications*, 67(10).
 - [125] Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., & Chen, T. J. (2006). Fuzzy c-means

- clustering with spatial information for image segmentation. *computerized medical imaging and graphics*, 30(1), 9-15.
- [126] Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.
- [127] Arevalo, J., Cruz-Roa, A., & GONZÁLEZ O, F. A. (2014). Histopathology image representation for automatic analysis: A state-of-the-art review. *Revista Med*, 22(2), 79-91.
- [128] Jain, A. K., Zhong, Y., & Dubuisson-Jolly, M. P. (1998). Deformable template models: A review. *Signal processing*, 71(2), 109-129.
- [129] Kass, M., Witkin, A., & Terzopoulos, D. (1988). Snakes: Active contour models. *International journal of computer vision*, 1(4), 321-331.
- [130] Khunteta, A., & Ghosh, D. (2016). Object Boundary Detection Using Active Contour Model via Multiswarm PSO with Fuzzy-Rule Based Adaptation of Inertia Factor. *Advances in Fuzzy Systems*, 2016.
- [131] Yan, P., Zhou, X., Shah, M., & Wong, S. T. (2008). Automatic segmentation of high-throughput RNAi fluorescent cellular images. *IEEE transactions on information technology in biomedicine*, 12(1), 109-117.
- [132] Qi, X., Xing, F., Foran, D. J., & Yang, L. (2011). Robust segmentation of overlapping cells in histopathology specimens using parallel seed detection and repulsive level set. *IEEE Transactions on Biomedical Engineering*, 59(3), 754-765.
- [133] Ali, S., & Madabhushi, A. (2012). An integrated region-, boundary-, shape-based active contour for multiple object overlap resolution in histological imagery. *IEEE transactions on medical imaging*, 31(7), 1448-1460.
- [134] Chang, H., Wen, Q., & Parvin, B. (2015). Coupled segmentation of nuclear and membrane-bound macromolecules through voting and multiphase level set. *Pattern recognition*, 48(3), 882-893.
- [135] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*.
- [136] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.
- [137] Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267-285). Springer, Berlin, Heidelberg.
- [138] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- [139] Mahmood, Z., Muhammad, N., Bibi, N., & Ali, T. (2017). A review on state-of-the-art face recognition approaches. *Fractals*, 25(02), 1750025.
- [140] Shen, D., Wu, G., & Suk, H. I. (2017). Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19, 221-248.
- [141] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [142] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [143] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [144] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).
- [145] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., ... & Adam,

- H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
- [146] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).
- [147] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
- [148] Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2017, September). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In International MICCAI brainlesion workshop (pp. 178-190). Springer, Cham.
- [149] Li, Y., Qi, H., Dai, J., Ji, X., & Wei, Y. (2017). Fully convolutional instance-aware semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2359-2367).
- [150] Yuan, Y., Chao, M., & Lo, Y. C. (2017). Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance. IEEE transactions on medical imaging, 36(9), 1876-1886.
- [151] Xiao, R., Goldklang, M. P., Stearns, K. N., & D'Armiento, J. M. (2019). Deep Learning in Computer Vision Provides Valuable Insights and Improved Efficiency for Lung Cytology and Histopathology. In C107. GOING VIRAL TO REPAIR (pp. A5761-A5761). American Thoracic Society.
- [152] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).
- [153] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 1440-1448).
- [154] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99.
- [155] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
- [156] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.
- [157] Ciresan, D., Giusti, A., Gambardella, L., & Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. Advances in neural information processing systems, 25, 2843-2851.
- [158] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
- [159] Dosovitskiy, A., Fischer, P., Springenberg, J. T., Riedmiller, M., & Brox, T. (2015). Discriminative unsupervised feature learning with exemplar convolutional neural networks. IEEE transactions on pattern analysis and machine intelligence, 38(9), 1734-1747.
- [160] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016, October). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In International conference on medical image computing and computer-assisted intervention (pp. 424-432). Springer, Cham.
- [161] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham.
- [162] Zhang, Z., Liu, Q., & Wang, Y. (2018). Road extraction by deep residual u-net. IEEE

- Geoscience and Remote Sensing Letters, 15(5), 749-753.
- [163] Tokunaga, H., Teramoto, Y., Yoshizawa, A., & Bise, R. (2019). Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12597-12606).
 - [164] Babu Sam, D., Surya, S., & Venkatesh Babu, R. (2017). Switching convolutional neural network for crowd counting. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5744-5752).
 - [165] van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., & Ciompi, F. (2021). HookNet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. *Medical Image Analysis*, 68, 101890.
 - [166] Ho, D. J., Yarlagadda, D. V., D'Alfonso, T. M., Hanna, M. G., Grabenstetter, A., Ntiamoah, P., & Fuchs, T. J. (2021). Deep multi-magnification networks for multi-class breast cancer image segmentation. *Computerized Medical Imaging and Graphics*, 88, 101866.
 - [167] Can, A., Bello, M., Cline, H. E., Tao, X., Ginty, F., Sood, A., ... & Montalto, M. (2008, May). Multi-modal imaging of histological tissue sections. In 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro (pp. 288-291). IEEE.
 - [168] Yang, L., Meer, P., & Foran, D. J. (2005). Unsupervised segmentation based on robust estimation and color active contour models. *IEEE Transactions on Information Technology in Biomedicine*, 9(3), 475-486.
 - [169] BenTaieb, A., & Hamarneh, G. (2017). Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging*, 37(3), 792-802.
 - [170] Xu, Y., Zhu, J. Y., Eric, I., Chang, C., Lai, M., & Tu, Z. (2014). Weakly supervised histopathology cancer image segmentation and classification. *Medical image analysis*, 18(3), 591-604.
 - [171] Wen, S., Kurc, T. M., Gao, Y., Zhao, T., Saltz, J. H., & Zhu, W. (2017). A methodology for texture feature-based quality assessment in nucleus segmentation of histopathology image. *Journal of pathology informatics*, 8.
 - [172] Jia, Z., Huang, X., Eric, I., Chang, C., & Xu, Y. (2017). Constrained deep weak supervision for histopathology image segmentation. *IEEE transactions on medical imaging*, 36(11), 2376-2388.
 - [173] Song, Y., Chang, H., Gao, Y., Liu, S., Zhang, D., Yao, J., ... & Cai, W. (2018, April). Feature learning with component selective encoding for histopathology image classification. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 257-260). IEEE.
 - [174] Xu, Y., Jia, Z., Wang, L. B., Ai, Y., Zhang, F., Lai, M., ... & Chang, C. (2017). Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features. *BMC bioinformatics*, 18(1), 1-17.
 - [175] Caicedo, J. C., Cruz, A., & Gonzalez, F. A. (2009, July). Histopathology image classification using bag of features and kernel functions. In Conference on Artificial Intelligence in Medicine in Europe (pp. 126-135). Springer, Berlin, Heidelberg.
 - [176] Bayramoglu, N., Kannala, J., & Heikkilä, J. (2016, December). Deep learning for magnification independent breast cancer histopathology image classification. In 2016 23rd International conference on pattern recognition (ICPR) (pp. 2440-2445). IEEE.
 - [177] Kieffer, B., Babaie, M., Kalra, S., & Tizhoosh, H. R. (2017, November). Convolutional neural networks for histopathology image classification: Training vs. using pre-trained networks. In 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA) (pp. 1-6). IEEE.
 - [178] Paules, C., Youssef, L., Rovira, C., Crovetto, F., Nadal, A., Peguero, A., & Gratacós, E. (2019). Distinctive patterns of placental lesions in pre-eclampsia vs small-for-gestational age and their association with fetoplacental Doppler. *Ultrasound in Obstetrics & Gynecology*, 54(5), 609-616.

- [179] Huppertz, B. (2011). Placental pathology in pregnancy complications. *Thrombosis research*, 127, S96-S99.
- [180] Sava, R. I., March, K. L., & Pepine, C. J. (2018). Hypertension in pregnancy: Taking cues from pathophysiology for clinical practice. *Clinical cardiology*, 41(2), 220-227.
- [181] Kovo, M., Schreiber, L., Ben-Haroush, A., Gold, E., Golan, A., & Bar, J. (2012). The placental component in early-onset and late-onset preeclampsia in relation to fetal growth restriction. *Prenatal diagnosis*, 32(7), 632-637.
- [182] Roberts, J. M., & Escudero, C. (2012). The placenta in preeclampsia. *Pregnancy Hypertension: An International Journal of Women's Cardiovascular Health*, 2(2), 72-83.
- [183] Morgan, T. K. (2016). Role of the placenta in preterm birth: a review. *American journal of perinatology*, 33(03), 258-266.
- [184] Stillbirth Collaborative Research Network Writing Group. (2011). Causes of death among stillbirths. *Jama*, 306(22), 2459.
- [185] Bellamy, L., Casas, J. P., Hingorani, A. D., & Williams, D. J. (2007). Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. *Bmj*, 335(7627), 974.
- [186] De Boo, H. A., & Harding, J. E. (2006). The developmental origins of adult disease (Barker) hypothesis. *Australian and New Zealand Journal of Obstetrics and Gynaecology*, 46(1), 4-14.
- [187] Redline, R. W. (2015). Classification of placental lesions. *American journal of obstetrics and gynecology*, 213(4), S21-S28.
- [188] Chen, A., & Roberts, D. J. (2018). Placental pathologic lesions with a significant recurrence risk—what not to miss!. *Apmis*, 126(7), 589-601.
- [189] Catov, J. M., Scifres, C. M., Caritis, S. N., Bertolet, M., Larkin, J., & Parks, W. T. (2017). Neonatal outcomes following preterm birth classified according to placental features. *American journal of obstetrics and gynecology*, 216(4), 411-e1.
- [190] Vinnars, M. T., Nasiell, J., Holmström, G., Norman, M., Westgren, M., & Papadogiannakis, N. (2014). Association between placental pathology and neonatal outcome in preeclampsia: a large cohort study. *Hypertension in pregnancy*, 33(2), 145-158.
- [191] Roescher, A. M., Timmer, A., Erwich, J. J. H., & Bos, A. F. (2014). Placental pathology, perinatal death, neonatal outcome, and neurological development: a systematic review. *PloS one*, 9(2), e89419.
- [192] Chisholm, K. M., & Heerema-McKenney, A. (2015). Fetal thrombotic vasculopathy: significance in liveborn children using proposed society for pediatric pathology diagnostic criteria. *The American journal of surgical pathology*, 39(2), 274-280.
- [193] Redline, R. W. (2015, February). The clinical implications of placental diagnoses. In *Seminars in perinatology* (Vol. 39, No. 1, pp. 2-8). WB Saunders.
- [194] Khong, T. Y., Mooney, E. E., Ariel, I., Balmus, N. C., Boyd, T. K., Brundler, M. A., ... & Gordijn, S. J. (2016). Sampling and definitions of placental lesions: Amsterdam placental workshop group consensus statement. *Archives of pathology & laboratory medicine*, 140(7), 698-713.
- [195] Redline, R. W. (2008). Placental pathology: a systematic approach with clinical correlations. *Placenta*, 29, 86-91.
- [196] Redline, R. W. (2007). Villitis of unknown etiology: noninfectious chronic villitis in the placenta. *Human pathology*, 38(10), 1439-1446.
- [197] Turowski, G., Berge, L. N., Helgadottir, L. B., Jacobsen, E. M., & Roald, B. (2012). A new, clinically oriented, unifying and simple placental classification system. *Placenta*, 33(12), 1026-1035.
- [198] Hargitai, B., Marton, T., & Cox, P. M. (2004). Best practice no 178: Examination of the human placenta. *Journal of clinical pathology*, 57(8), 785-792.
- [199] Warsof, S. L., Gohari, P., Berkowitz, R. L., & Hobbins, J. C. (1977). The estimation of fetal

- weight by computer-assisted analysis. *American journal of obstetrics and gynecology*, 128(8), 881-892.
- [200] Almoussa, N., Dutra, B., Lampe, B., Getreuer, P., Wittman, T., Salafia, C., & Vese, L. (2011, March). Automated vasculature extraction from placenta images. In *Medical Imaging 2011: Image Processing* (Vol. 7962, p. 79621L). International Society for Optics and Photonics.
- [201] Chang, J. M., Huynh, N., Vazquez, M., & Salafia, C. (2013, July). Vessel enhancement with multiscale and curvilinear filter matching for placenta images. In *2013 20th international conference on systems, signals and image processing (IWSSIP)* (pp. 125-128). IEEE.
- [202] Palee, P., Sharp, B., Noriega, L., Sebire, N. J., & Platt, C. (2013). Image analysis of histological features in molar pregnancies. *Expert systems with applications*, 40(17), 7151-7158.
- [203] Finlayson, G. D., Schiele, B., & Crowley, J. L. (1998, June). Comprehensive colour image normalization. In *European conference on computer vision* (pp. 475-490). Springer, Berlin, Heidelberg.
- [204] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.
- [205] Chan, T. F., & Vese, L. A. (2001). Active contours without edges. *IEEE Transactions on image processing*, 10(2), 266-277.
- [206] Stoz, F., Schuhmann, R. A., & Schmid, A. (1987). Morphometric investigations of terminal villi of diabetic placentas in relation to the White classification of diabetes mellitus.
- [207] Chaikitgosiyakul, S., Rijken, M. J., Muehlenbachs, A., Lee, S. J., Chaisri, U., Viriyavejakul, P., & McGready, R. (2014). A morphometric and histological study of placental malaria shows significant changes to villous architecture in both *Plasmodium falciparum* and *Plasmodium vivax* infection. *Malaria journal*, 13(1), 1-13.
- [208] Haralick, R. M., Sternberg, S. R., & Zhuang, X. (1987). Image analysis using mathematical morphology. *IEEE transactions on pattern analysis and machine intelligence*, (4), 532-550.
- [209] Stark, M. W., Clark, L., & Craver, R. D. (2014). Histologic differences in placentas of preeclamptic/eclamptic gestations by birthweight, placental weight, and time of onset. *Pediatric and Developmental Pathology*, 17(3), 181-189.
- [210] Jobe, A. H., & Bancalari, E. (2001). Bronchopulmonary dysplasia. *American journal of respiratory and critical care medicine*, 163(7), 1723-1729.
- [211] Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. A., Shankaran, S., ... & Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. (2015). Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *Jama*, 314(10), 1039-1051.
- [212] Rizzardi, A. E., Johnson, A. T., Vogel, R. I., Pambuccian, S. E., Henriksen, J., Skubitz, A. P., ... & Schmechel, S. C. (2012). Quantitative comparison of immunohistochemical staining measured by digital image analysis versus pathologist visual scoring. *Diagnostic pathology*, 7(1), 1-10.
- [213] Poggi, S. H., Salafia, C., Paiva, S., Leak, N. J., Pezzullo, J. C., & Ghidini, A. (2009). Variability in pathologists' detection of placental meconium uptake. *American journal of perinatology*, 26(03), 207-210.
- [214] Meyer, J. S., Alvarez, C., Milikowski, C., Olson, N., Russo, I., Russo, J., & Parwaresch, R. (2005). Breast carcinoma malignancy grading by Bloom–Richardson system vs proliferation index: reproducibility of grade and advantages of proliferation index. *Modern pathology*, 18(8), 1067-1078.
- [215] Jothi, J. A. A., & Rajam, V. M. A. (2017). A survey on automated cancer diagnosis from histopathology images. *Artificial Intelligence Review*, 48(1), 31-81.
- [216] AlZubaidi, A. K., Sideseq, F. B., Faeq, A., & Basil, M. (2017, March). Computer aided diagnosis in digital pathology application: Review and perspective approach in lung cancer

- classification. In 2017 annual conference on new trends in information & Communications technology applications (NTICT) (pp. 219-224). IEEE.
- [217] Sallon, C., Soulet, D., Provost, P. R., & Tremblay, Y. (2015). Automated high-performance analysis of lung morphometry. *American journal of respiratory cell and molecular biology*, 53(2), 149-158.
- [218] Crowley, G., Kwon, S., Haider, S., Caraher, E. J., & Nolan, A. (2017). C80-C Imaging methodology and application to lung disease: quantitative lung morphology: semiautomated method of mean chord length measurements. *American Journal of Respiratory and Critical Care Medicine*, 195.
- [219] Crowley, G., Kwon, S., Caraher, E. J., Haider, S. H., Lam, R., Batra, P., & Nolan, A. (2019). Quantitative lung morphology: semi-automated measurement of mean linear intercept. *BMC pulmonary medicine*, 19(1), 1-9.
- [220] Horai, Y., Kakimoto, T., Takemoto, K., & Tanaka, M. (2017). Quantitative analysis of histopathological findings using image processing software. *Journal of toxicologic pathology*, 30(4), 351-358.
- [221] Tschanz, S. A., Burri, P. H., & Weibel, E. R. (2011). A simple tool for stereological assessment of digital images: the STEPanizer. *Journal of microscopy*, 243(1), 47-59.
- [222] Mikołajczyk, A., & Grochowski, M. (2018, May). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117-122). IEEE.
- [223] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- [224] Salsabili, S., Mukherjee, A., Ukwatta, E., Chan, A. D., Bainbridge, S., & Gynspan, D. (2019). Automated segmentation of villi in histopathology images of placenta. *Computers in biology and medicine*, 113, 103420.
- [225] Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on pattern analysis and machine intelligence*, 16(6), 641-647.
- [226] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
- [227] Jobe, A. H. (2011). The new bronchopulmonary dysplasia. *Current opinion in pediatrics*, 23(2), 167.
- [228] Crum, W. R., Camara, O., & Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging*, 25(11), 1451-1461.
- [229] Beauchemin, M., Thomson, K. P., & Edwards, G. (1998). On the Hausdorff distance used for the evaluation of segmentation results. *Canadian journal of remote sensing*, 24(1), 3-8.
- [230] Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5), 378.
- [231] Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- [232] Jacob, R. E., Carson, J. P., Gideon, K. M., Amidan, B. G., Smith, C. L., & Lee, K. M. (2009). Comparison of two quantitative methods of discerning airspace enlargement in smoke-exposed mice. *PloS one*, 4(8), e6670.
- [233] Schittny, J. C., Mund, S. I., & Stampanoni, M. (2008). Evidence and structural mechanism for late lung alveolarization. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 294(2), L246-L254.
- [234] Mund, S. I., Stampanoni, M., & Schittny, J. C. (2008). Developmental alveolarization of the mouse lung. *Developmental dynamics: an official publication of the American Association of Anatomists*, 237(8), 2108-2116.
- [235] Rieger-Fackeldey, E., Park, M. S., Schanbacher, B. L., Joshi, M. S., Chicoine, L. G., Nelin, L. D., & Smith, C. V. (2014). Lung development alterations in newborn mice after recovery

- from exposure to sublethal hyperoxia. *The American journal of pathology*, 184(4), 1010-1016.
- [236] Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International journal of multimedia information retrieval*, 7(2), 87-93.
- [237] Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., & Kim, T. K. (2021). Multiple object tracking: A literature review. *Artificial Intelligence*, 293, 103448.
- [238] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, 29(9), 2352-2449.
- [239] Makantasis, K., Karantzalos, K., Doulamis, A., & Doulamis, N. (2015, July). Deep supervised learning for hyperspectral data classification through convolutional neural networks. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 4959-4962). IEEE.
- [240] van Ginneken, B. (2017). Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiological physics and technology*, 10(1), 23-32.
- [241] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12), 2481-2495.
- [242] Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.
- [243] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 1097-1105.
- [244] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (pp. 448-456). PMLR.
- [245] Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.
- [246] T. Dozat, "Incorporating Nesterov Momentum into Adam," *ICLR Work.*, 2016.
- [247] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
- [248] Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249-256). *JMLR Workshop and Conference Proceedings*.
- [249] Tato, A., & Nkambou, R. (2018). Improving adam optimizer.
- [250] Koturwar, S., & Merchant, S. (2017). Weight initialization of deep neural networks (DNNs) using data statistics. *arXiv preprint arXiv:1710.10570*.
- [251] Csurka, G., Larlus, D., Perronnin, F., & Meylan, F. (2013, September). What is a good evaluation measure for semantic segmentation?. In *Bmvc* (Vol. 27, No. 2013, pp. 10-5244).
- [252] Crum, W. R., Camara, O., & Hill, D. L. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE transactions on medical imaging*, 25(11), 1451-1461.
- [253] Mosli, M. H., Feagan, B. G., Zou, G., Sandborn, W. J., D'Haens, G., Khanna, R., ... & Levesque, B. G. (2015). Reproducibility of histological assessments of disease activity in UC. *Gut*, 64(11), 1765-1773.
- [254] Longtine, M. S., Chen, B., Odibo, A. O., Zhong, Y., & Nelson, D. (2012). Villous trophoblast apoptosis is elevated and restricted to cytotrophoblasts in pregnancies complicated by preeclampsia, IUGR, or preeclampsia with IUGR. *Placenta*, 33(5), 352-359.
- [255] Li, C., Chen, H., Li, X., Xu, N., Hu, Z., Xue, D., ... & Sun, H. (2020). A review for cervical

- histopathology image analysis using machine vision approaches. *Artificial Intelligence Review*, 53(7), 4821-4862.
- [256] Zhou, X., Li, C., Rahaman, M. M., Yao, Y., Ai, S., Sun, C., ... & Teng, Y. (2020). A comprehensive review for breast histopathology image analysis using classical and deep neural networks. *IEEE Access*, 8, 90931-90956.
- [257] Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. arXiv preprint arXiv:1606.05718.
- [258] Xing, F., Xie, Y., & Yang, L. (2015). An automatic learning-based framework for robust nucleus segmentation. *IEEE transactions on medical imaging*, 35(2), 550-566.
- [259] Alsubaie, N., Shaban, M., Snead, D., Khurram, A., & Rajpoot, N. (2018, July). A multi-resolution deep learning framework for lung adenocarcinoma growth pattern classification. In *Annual Conference on Medical Image Understanding and Analysis* (pp. 3-11). Springer, Cham.
- [260] Kosaraju, S. C., Hao, J., Koh, H. M., & Kang, M. (2020). Deep-Hipo: Multi-scale receptive field deep learning for histopathological image analysis. *Methods*, 179, 3-13.
- [261] Tsaku, N. Z., Kosaraju, S. C., Aqila, T., Masum, M., Song, D. H., Mondal, A. M., & Kang, M. (2019, November). Texture-based deep learning for effective histopathological cancer image classification. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 973-977). IEEE.
- [262] Alsubaie, N., Shaban, M., Snead, D., Khurram, A., & Rajpoot, N. (2018, July). A multi-resolution deep learning framework for lung adenocarcinoma growth pattern classification. In *Annual Conference on Medical Image Understanding and Analysis* (pp. 3-11). Springer, Cham.
- [263] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).
- [264] Sirinukunwattana, K., Alham, N. K., Verrill, C., & Rittscher, J. (2018, September). Improving whole slide segmentation through visual context—a systematic study. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 192-200). Springer, Cham.
- [265] Babu Sam, D., Surya, S., & Venkatesh Babu, R. (2017). Switching convolutional neural network for crowd counting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5744-5752).
- [266] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.
- [267] Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114). PMLR.
- [268] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE.
- [269] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).
- [270] S. Salsabili, A. Mukherjee, E. Ukwatta, A. D. C. Chan, S. Bainbridge, and D. Grynspan, “Automated segmentation of villi in histopathology images of placenta,” *Comput. Biol. Med.*, vol. 113, p. 103420, Oct. 2019.
- [271] Salsabili, S., Lithopoulos, M., Sreeraman, S., Vadivel, A., Thébaud, B., Chan, A. D., & Ukwatta, E. (2021). Fully automated estimation of the mean linear intercept in histopathology images of mouse lung tissue. *Journal of Medical Imaging*, 8(2), 027501.

- [272] Reinhard, E., Adhikhmin, M., Gooch, B., & Shirley, P. (2001). Color transfer between images. *IEEE Computer graphics and applications*, 21(5), 34-41.
- [273] Yu, F., Koltun, V., & Funkhouser, T. (2017). Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 472-480).
- [274] Tokunaga, H., Teramoto, Y., Yoshizawa, A., & Bise, R. (2019). Adaptive weighting multi-field-of-view CNN for semantic segmentation in pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 12597-12606).
- [275] Gu, F., Burlutskiy, N., Andersson, M., & Wilén, L. K. (2018). Multi-resolution networks for semantic segmentation in whole slide images. In *Computational Pathology and Ophthalmic Medical Image Analysis* (pp. 11-18). Springer, Cham.
- [276] Ho, D. J., Yarlagadda, D. V., D'Alfonso, T. M., Hanna, M. G., Grabenstetter, A., Ntiamoah, P., & Fuchs, T. J. (2021). Deep multi-magnification networks for multi-class breast cancer image segmentation. *Computerized Medical Imaging and Graphics*, 88, 101866.
- [277] Takahama, S., Kurose, Y., Mukuta, Y., Abe, H., Fukayama, M., Yoshizawa, A., & Harada, T. (2019). Multi-stage pathological image classification using semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10702-10711).
- [278] Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, 3(1), 1-130.
- [279] Zhou, Z. H. (2018). A brief introduction to weakly supervised learning. *National science review*, 5(1), 44-53.