Summer 2018

Measurements of retinal microvasculature in mice and humans with deep learning

Wenxiang Deng
University of Iowa

Recommended Citation

Deng, Wenxiang. "Measurements of retinal microvasculature in mice and humans with deep learning." PhD (Doctor of Philosophy) thesis, University of Iowa, 2018.
https://doi.org/10.17077/etd.84ques5w

MEASUREMENTS OF RETINAL MICROVASCULATURE IN MICE AND HUMANS WITH DEEP LEARNING

by

Wenxiang Deng

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy degree
in Biomedical Engineering
in the Graduate College of
The University of Iowa

August 2018

Thesis Supervisor: Associate Professor Mona K. Garvin

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Wenxiang Deng

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Biomedical Engineering at the August 2018 graduation.

Thesis Committee:

Mona K. Garvin, Thesis Supervisor

David G. Wilder

Terry A. Braun

Punam K. Saha

Todd E. Scheetz

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Mona Garvin for being my research advisor in the past years. Her guidance and patience helped me go through the twists and turns in this project. I would also like to thank my academic advisor Prof. David Wilder for all the advice and support. I've learned a lot from this PhD period, not just methodologies for doing research, but the mindsets and critical thinkings towards general new areas.

A special thanks to Dr. Randy Kardon, Dr. Isabella Grumbach, and all of the collaborators I have had during my work. I received a lot of advice and assistance during my work here. I'm glad to have finished some useful work for the group. Most of my PhD work would have been impossible to finish without their help. Also thanks to Michelle Tamplin for helping me acquire and organize OCTA scans and reports used in the study. Without her help, a great portion of the study could not haved finished so smoothly.

I would like also to thank Dr. Punam Saha, Dr. Todd Scheetz, and Dr. Terry Braun for serving on my committee.

In addition, I would like to thank all my friends and lab-mates Junjie, Qiao, Ray, Sam, Shafkat, Yashila, Qingyang, Yue, Leixin and other friends who have also helped me over these years. Thanks to them for their help and for sharing their experience of research and life.

Finally, I would also like to especially thank my family for their support through my PhD period.

This thesis was supported, in part, by the Iowa City VA Center for The Prevention and Treatment of Visual Loss, grant C9251-C, Rehabilitation Research & Development (RR&D), VA-ORD; and I21 RX001561A RR&D, VA-ORD.

ABSTRACT

Retinal microvascular changes can occur in retinal diseases including glaucoma, diabetic retinopathies (DR), and radiation retinopathies (RR). For instance, some of these diseases show sparser microvasculature around the foveal area as an early symptom. Therefore, automatically segmenting and analyzing retinal microvasculature may help diagnose and understand the underlying mechanisms of ocular diseases. However, due to the limitations of the image quality and the difficulties of manual tracing, this is not a trivial problem to solve and very few studies address automated segmentations of microvascular networks.

In this thesis, automated approaches with deep learning are developed to analyze microvasculature in fluorescein angiography (FA) and recently developed optical coherence tomography angiography (OCTA). These approaches accurately quantify retinal microvasculature in mice and humans. First, we develop an automated approach to simultaneously segment retinal structures including the optic nerve head (ONH), main vessels, and the microvascular network in mouse FA images. By combining model-based approaches with deep learning, we can automatically segment these structures and require fewer manual tracings for training. Then, multiple methods are developed to automatically analyze the microvasculature in human OCTA *en-face* images. The key novelties in these approaches involve the use of deep learning to not only segment the microvasculature but also directly find regions with different microvascular status. Furthermore, we investigated and analyzed the clinical significance of our region-based OCTA *en-face* segmentations. In order to achieve this goal, we first analyzed the effectiveness of regional vascular measurements as an indication of RR severity. Additionally, the microvascular differences between affected and unaffected eyes of the RR patients were also studied and compared to healthy subjects. These studies show that the usage of deep-learning networks provides accurate results for the analysis of retinal vasculature.

Overall, the main contributions of this thesis include: 1) combining model-based

semi-automated segmentations and a fully automated deep-learning network and developing an approach to simultaneously segment multiple retinal structures in mouse FA images, 2) developing and evaluating three automated approaches to directly segment microvasculature in human OCTA *en-face* images, 3) developing an alternative deep-learning method to directly segment regions with different vascular states from the OCTA *en-face* images, and 4) using developed results from automated segmentations to measure the disease severity in radiation retinopathy patients.

PUBLIC ABSTRACT

Retinal diseases such as diabetic retinopathies (DR) and radiation retinopathies (RR) show changes in the microvasculature as early symptoms. Therefore, automatically segmenting and analyzing retinal microvasculature may help diagnose and understand the underlying mechanisms of ocular diseases. However, due to the limitations of the image quality, as well as the difficulties of evaluation, very few studies address automated segmentations of microvascular networks.

A commonly used imaging modality to record retinal vasculature is fluorescein angiography (FA). It is an invasive technique and uses fluorescein dyes to provide information of blood flow in the retina. Another relevant technology is optical coherence tomography angiography (OCTA). It is a newly developed non-invasive technique that captures regions with blood flow in the retina in 3D. Better ways to generate automated quantification for these modalities could potentially enhance clinical studies and help in the diagnosis and understanding of retinal diseases.

In this work, automated approaches are developed to accurately quantify retinal microvasculature in mouse FA and human OCTA images. First, we develop an automated approach to simultaneously segment retinal structures including the optic nerve head (ONH), main vessels, and the microvascular network in mouse FA images. By combining traditional approaches and deep learning, we can automatically segment these structures with minimal human effort. Then, multiple methods are developed to automatically analyze the microvasculature in human OCTA projection images. The key novelties in these approaches are by using deep-learning based approaches, we can not only segment the microvasculature, but also directly find regions with different microvascular status. Furthermore, we investigated and analyzed the clinical significance of our region-based OCTA projection segmentations. In order to achieve this goal, we first analyzed the effectiveness of regional vascular measurements as an indication of RR disease. Additionally, the microvascular differences between affected and unaffected eyes of the RR patients were also studied and compared to

healthy subjects. The proposed approaches are expected to provide more accurate and reliable automated measurements to analyze retinal microvasculature.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

Table

LIST OF FIGURES

Figure

CHAPTER 1
INTRODUCTION

Traditionally, retinal diseases are often diagnosed and evaluated based on fundus photography. However, other imaging modalities such as fluorescein angiography (FA) and recently developed optical coherence tomography angiography (OCTA) have shown effectiveness to find early signs of retinal vascular changes in many diseases including glaucoma [3,4], age-related macular degeneration (AMD) [5], diabetic retinopathies (DR) [6–8], and radiation retinopathies (RR) [9,10]. Automated measurements of vasculature network can help study and diagnose the early stage of such diseases.

In this project, we mainly focus on automated analyses of two modalities: FA images from diabetic mice retina scans, and OCTA from human radiation retinopathy patients. Introduced in 1961 [11], FA is a traditional technique to analyze blood flow in the retina by injecting dyes. On the other hand, OCTA is a relatively new noninvasive imaging modality. As a functional extension of OCT, OCTA was first introduced recently in 2012 by Jia et al. [12]. Compared to the structural OCT capturing the 3D structure of the retina and providing volumetric information, OCTA is more of an improvement to FA technique and focuses on measurement of blood flow and vascular status. Compared to FA, it is noninvasive and capable of visualizing blood flow in multiple retinal layers.

There are many studies for analyzing the vascular network in FA and OCTA, human and mice images [13–16]. Through our observations with mouse FA images, the main vessels and microvasculature should be analyzed separately so that the occurrence of these large vessels would not affect the measurement of microvascular density. However, currently, there is no approach to simultaneously segment or analyze the main vessels and the microvascular network separately. For the OCTA modality, only limited studies with automated segmentation of microvasculature are presented. Commercially available scanners such as AngioVue (Optovue, Inc., Fremont, CA) provides a vascular density measurement in software. However, it tends to fail in cases of large vessels in the retinal region. Moreover, only limited information is provided by the scanner.

Overall, in this work, our goal is to develop approaches for quantification of retinal vascular changes in mice and humans. We first develop an automated approach to simultaneously segment retinal structures including the optic nerve head (ONH), main vessels, and the microvascular network in mouse FA images. Then, multiple methods are developed to automatically analyze the microvasculature in human OCTA *en-face* images. The key novelties in these approaches are by using deep-learning based approaches, we can very accurately segment the microvasculature and also directly find regions with different microvascular status. At last, we study the microvascular changes between affected and unaffected eyes of radiation retinopathy patients.

In particular, this thesis is composed of three specific aims as follows:

- **Aim 1: Segmentation of multiple structures in fluorescein angiography images of mice.** Here we develop approaches measuring different retinal structures in mouse FA images. We first develop a model-based approach to roughly segment main vessels and the microvascular network. Then the results are interactively corrected to train a deep learning based approach to segment the optic nerve head (ONH), main vessels, and the microvascular network simultaneously.

- **Aim 2: Segmentation of the microvasculature with optical coherence tomography angiography in humans.** We also develop deep-learning based approaches to measure the microvasculature in human optical coherence tomography angiography (OCTA) data. The retinal surfaces in corresponding structural OCT volumes are first segmented. Then three automated approaches are developed and evaluated for directly segmenting the microvasculature in OCTA *en-face* images. Finally, we also develop and evaluate an alternative approach to simultaneously find avascular regions, regions with capillary dropout, and capillary dense regions.

- **Aim 3: Region analysis for OCTA and study of differences between affected and unaffected eyes.** Lastly, we analyze the regions segmented in Aim 2 and compare them to experts' analysis of severity of radiation retinopathy. We also compare the differences of affected and unaffected eyes in radiation retinopathy patients and normal subjects as a clinical-related application.

1.1 Thesis organization

The rest of this thesis is organized as follows:

- **Chapter 2** introduces the clinical background of this thesis. This chapter covers the image modalities like FA, SD-OCT, and OCTA we use in the aims. It also includes the diseases we are interested in for this thesis.

- **Chapter 3** discusses some prior approaches used for analyzing retinal vessels and layers. The background of the approaches we use in this thesis is also introduced.

- **Chapter 4** provides the methods and validations for segmenting multiple structures, including microvascular network, main vessels, and ONH region in mouse FA images.

- **Chapter 5** develops three approaches to directly analyze the microvascular status in OCTA *en-face* images of superficial layers in RR patients. It also uses a new way to evaluate the effectiveness of each approach, instead of only comparing statistical significance of the analyses between abnormal and normal subjects.

- **Chapter 6** develops an alternative approach to directly segment the normal and abnormal regions from the OCTA *en-face* images using deep learning method. This improves the algorithms ability to analyze the low signal-to-noise ratio (SNR) images.
- **Chapter 7** presents the clinical application of the methods developed in this and previous chapter for measuring the microvascular loss in radiation retinopathy patients.
- **Chapter 8** summarizes all the chapters and concludes this thesis.

CHAPTER 2
CLINICAL BACKGROUND

2.1 Retinal vasculature

The retina in the inner posterior globe of the eye is a light-sensitive layered tissue. Photoreceptor cells in the retina convert light into the eye to electrical neural signals and the signals are transmitted into the brain to form vision.

Blood vessels in the retina are part of the circulation system and supply neurons and retinal tissues with nutrition and oxygen. Around 65% of the blood supply in the retina is through the choroid and 35% by retinal vasculature on the top of the retina [6]. Retinal vasculature originates and converges at the optic nerve head as arterioles and venules and gradually diverges into retinal capillaries. Anomalies in the retinal vasculature are often strongly related to ocular diseases such as glaucoma, diabetic retinopathy, and age-related macular degeneration (AMD). For instance, diabetic retinopathy may cause ischemia and result in new blood vessels growing. This neovascularization is usually a sign of proliferative diabetic retinopathy (PDR) [17]. On the other hand, retinal vascular changes can also indicate cardiovascular diseases such as hypertension and stroke [18, 19]. Diseases like hypertension can lead to pathological changes to retinal vessels mainly shown as the narrowing of retinal arterioles [20].

Automated analyses of retinal vasculature [16, 21, 22] may help in the diagnosis of diseases, as well as help understand the underlying mechanisms of ocular diseases.

2.2 Fluorescein angiography

Fluorescein angiography (FA) is a commonly used imaging modality since its introduction in 1961 [11]. It is closely related to fundus photographs which take images of the interior surface (fundus) of the eye with cameras. While the FA modality on the other hand, also uses fundus cameras to capture retinal images, it is an invasive imaging approach in that it involves injecting sodium fluorescein dye into the

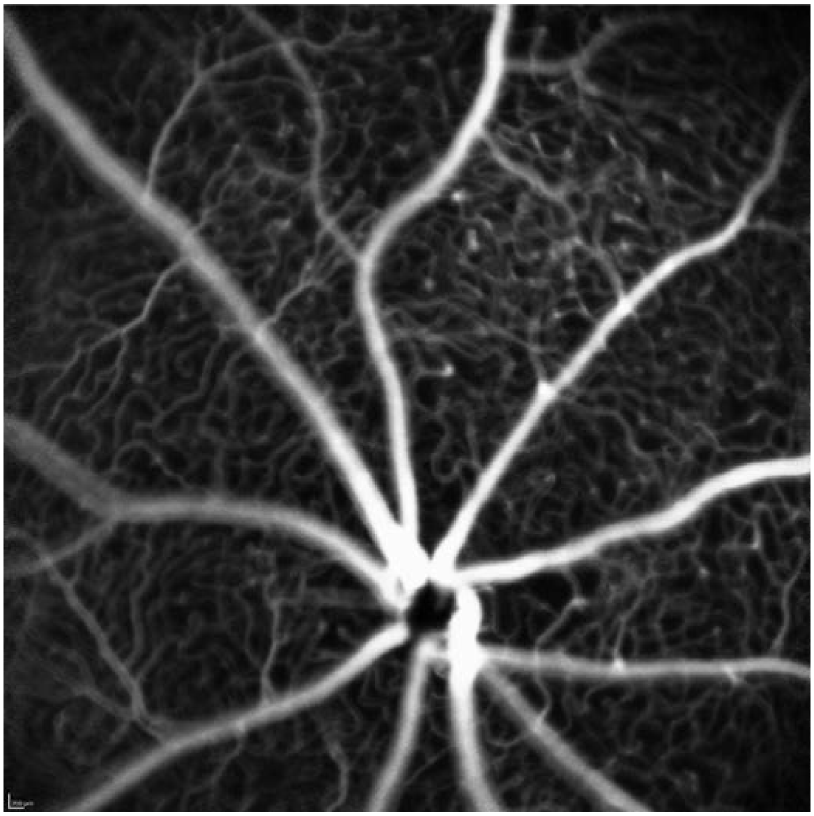


Figure 2.1: An example image of mouse retina (as further described in Section 4.6) in FA modality.

circulation system. The dye binds to leukocytes in the bloodstream and can emit fluorescence after illumination to the retina. An angiogram can then be captured by fundus cameras with narrow-band filters. The FA image records the status of blood flow in the retina, and can provide morphologic as well as circulative information for the retinal vascular plexus. Fig. 2.1 shows an example of mouse FA images with main vessels and capillary plexus.

FA is currently an effective way to measure the functional state of the retina. However, in contrast to fundus photographs, FA is an invasive method, and the injected fluorescent dye could cause side effects like nausea. It also only captures a 2D representation of retina and could mix up superficial and deeper retinal plexuses. Newer technologies like OCTA could potentially avoid these drawbacks.

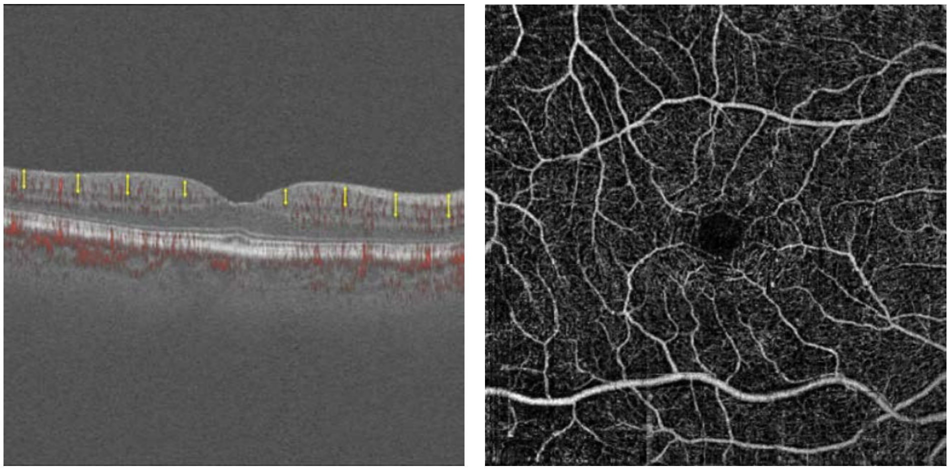
2.3 Optical coherence tomography angiography

Optical coherence tomography (OCT) technology is widely used in ophthalmology because of its ability to visualize 3D context in the retina. It is a noninvasive imaging technique using a relatively long wavelength light for imaging of biological tissue in vivo in micrometer-resolutions [6,23]. OCT is an important modality to assess ocular diseases.

As a functional extension of OCT technology, optical coherence tomography angiography (OCTA) is a relatively new imaging method which combines split-spectrum amplitude-decorrelation angiography (SSADA) data with volumetric scanning, first proposed by Jia et al. [12] in 2012. Some studies have shown OCTA effectiveness in the analysis of macular degeneration, glaucoma, and radiation retinopathy [4,5,24,25]. Briefly, by repeatedly scanning the same region and analyzing multiple scans, OCTA detects regions with blood flow versus no flow. It allows blood vessel visualization in living tissue in a noninvasive way.

Scanners that acquire OCTA scans also provide SD-OCT scans as structural OCT showing the retinal layers. Fig. 2.2 shows a human structural OCT scan acquired from AngioVue (Optovue, Inc., Fremont, CA) and the corresponding projection angiogram of superficial layers. These projection *en-face* angiogram images are generated by segmenting retinal layers and selecting the maximum value in the corresponding layers in the SSADA volumes.

As a result, OCTA is increasingly used for retinal vasculature pathologies in retinal studies [26–28]. Comparing to traditional FA modality, OCTA has some advantages. First is that it is noninvasive and contrast agents will not be needed to generate images. In case of diseases, the dye used for visualizing the vessels could also leak and cause a change of contrast in the image. Such leakage can affect the result of the diagnosis. Also, FA images usually can only show the vascular network clearly in superficial layers. However, studies have shown that blood flow can be measured in



(a) (b)

Figure 2.2: An example OCTA image from the dataset as further described in Section 5.6. (a) SD-OCT b-scan with SSADA as the overlay. (b) The corresponding *en-face* projection angiogram for superficial layers indicated by yellow arrows.

at least three layers in the retina using OCTA [29,30]. Thus, a significant advantage with OCTA is it can give a visualization of the vascular network in multiple layers.

2.4 Diabetic retinopathy

Diabetes mellitus is a prevalent disease in the United States, with more than 29 million patients nation-wide [31]. Elevated blood glucose in the body is known to cause damage to blood vessels and nerve cells. It therefore causes damage to the heart, brain, kidneys, and eyes. Diabetic retinopathy (DR) is a complication of diabetes that causes damage to the retina. DR is the second leading cause of blindness and the primary cause of visual loss in working-age people in the US [6]. For long-term diabetes, nearly all patients have retinal capillary changes. In both humans and rodents, hyperglycemia causes capillary vessel wall damage and occluded vessels [6,32,33]. Eventually, retinal detachment or diabetic macular edema (DME) could occur. DME is the main cause of severe vision loss and blindness in diabetic patients. It breakdowns the blood-retinal barrier, which leads to leakage into retinal tissue and fluid accumulation. However, with early diagnosis and annual screening, vision loss and blindness are preventable [34].

Accompanied by retinal vascular changes, another complication of diabetes in the retina is nerve cell damage. In early-stage diabetes mellitus patients with minimal diabetic retinopathies, change of retinal layers thickness can be observed in OCT images [35,36]. Even in the earliest stage of DR, there is a measurable loss of intraretinal neurons in the retina.

2.5 Radiation retinopathy

Radiation for the eye or the orbit of the eye causes radiation retinopathies (RR), first described by Stallard in 1933 [37]. RR is usually a delayed condition and occurs months or years after the radiation. Anatomically, it is thought to be caused by loss of endothelial cells with sparing of pericytes in the retinal vascular network due to exposure to radiation. This loss would then lead to capillary loss and microaneurysms, and eventually causes neovascularization, macular edema, and retinal detachment [38,39]. Patients after radiotherapy treatments for the eye or head are likely to get RR and have irreversible visual impairment [38, 40–42]. Traditionally, the significance of RR is rated into stages by doctors using ophthalmoscopy and fluorescein angiography, and higher stages of RR will have increased risk of vision loss in the patient. Fig. 2.3 shows fundus and fluorescein images of an RR patient before treatments.

Finger tests [43] using fundus images mainly analyzes macular edema in the RR affected eye. It gives a general outline of the symptoms of RR with ophthalmoscopy, including retinal hemorrhage, microaneurysms, and vitreous hemorrhage. Alternatively, Horgan grading [42] with OCT volume is focused more on macular edema and showed that OCT images could help detect earlier symptoms of RR before shown in fundus images. The detection of early symptoms allows earlier diagnosis and could help prevent further visual loss in the affected eye in the patient.

However, only using SD-OCT could miss some early vascular changes in RR,

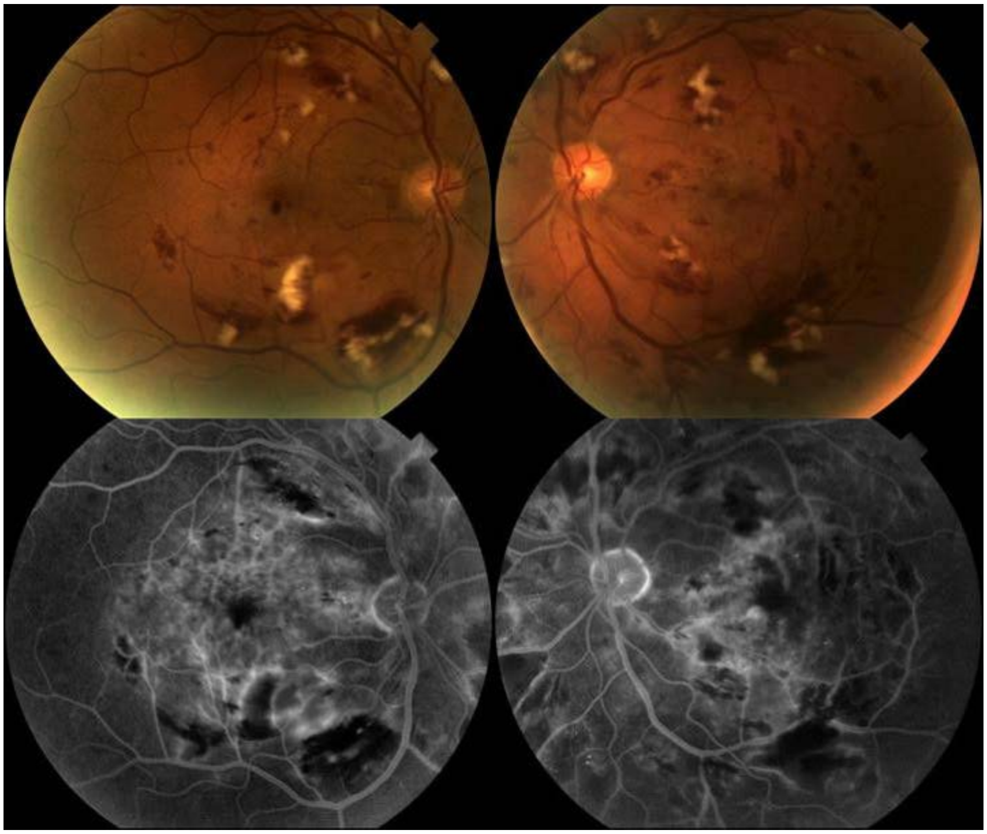
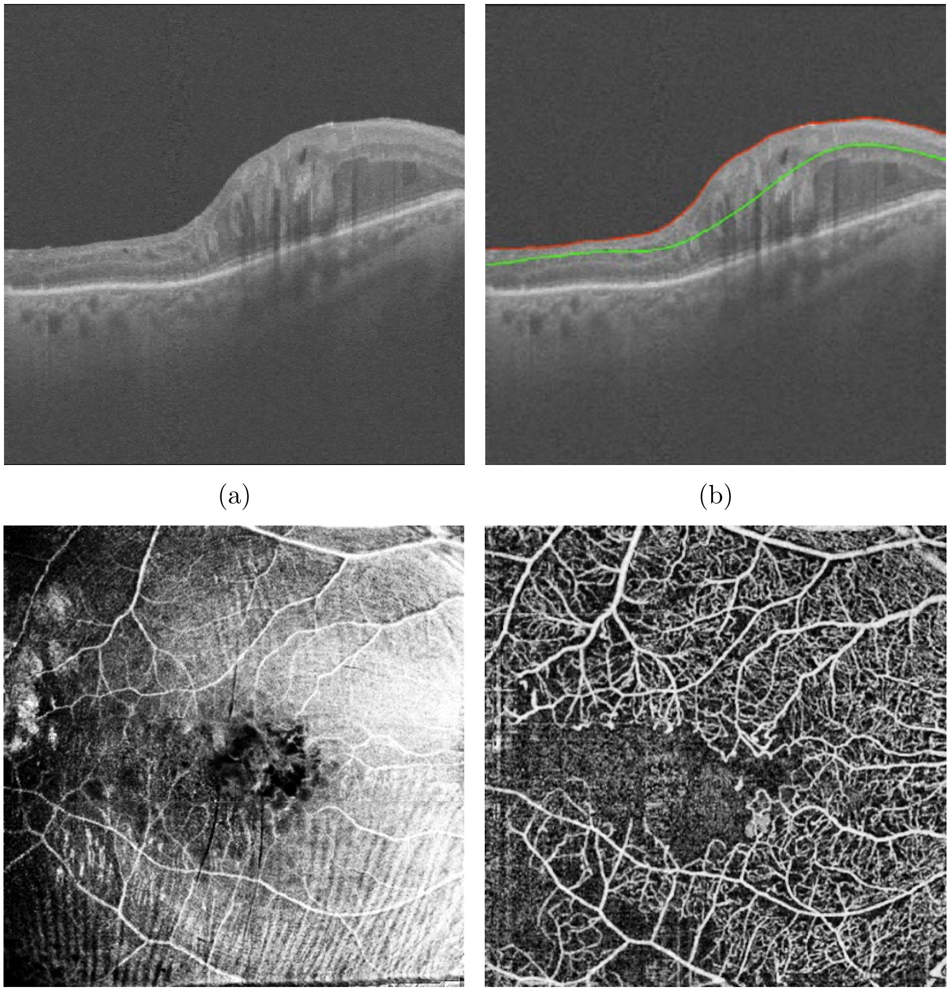


Figure 2.3: Fundus and fluorescein images of a radiation retinopathy patient. The images are obtained under Creative Commons Attribution License from [1].

such as microaneurysms and capillary dropouts. With the help of recently developed OCTA technology, a better analysis of these vascular changes can be more precisely measured. Structural SD-OCT b-scan (Fig. 2.4(a)) and the *en-face* OCTA projection image (Fig. 2.4(d)) of an RR patient can be seen in Fig. 2.4. The case study by Veverka et al. [24] shows that OCTA can accurately analyze changes in retinal vasculature in RR patients, and it can detect RR symptoms before there are visible changes in SD-OCT and fundus images. However, because OCTA is a newer technology, more studies are still needed for better analysis of RR vascular changes with this modality.



(a) (b)

(c) (d)

Figure 2.4: Examples of structural SD-OCT scan, *en-face* projection image, and OCTA *en-face* image of superficial layers from RR patient (as further described in Section 5.3 and Section 5.6). (a) SD-OCT b-scan. The corresponding *en-face* projection image of the SD-OCT (c) and angiogram (d) for superficial layers (b).

CHAPTER 3
TECHNICAL BACKGROUND

3.1 Retinal vessel segmentations

Segmentation of blood vessels separates vessel pixels from non-vessel pixels in the image. There are a large number of related works published for segmenting the retinal vessels. Most of them are automated approaches for segmenting all vessels in human fundus images. All these methods can be divided mainly into model-based and machine learning-based approaches.

For the model-based approaches, there are methods using region growing [44], adaptive thresholding [45], or morphological operations [46] techniques. However, the machine learning-based approaches tend to have better results. These approaches usually use classifications to solve the problem. Niemeijer et al. [21] used k-NN classifiers by using 31 Gaussian-based filters. Soares et al. [47] used Bayesian classifier with Gaussian mixture model (GMM) and used Gabor features. There are also a number of AdaBoost based methods [48], neural network [49, 50], and support vector machine (SVM) [51] based approaches. However, all these earlier methods require specifically designed features as training data. More recent deep-learning based approaches, such as [22, 52] use convolution neural networks (CNNs) for the segmentations. There is also work by Prentašić et al. [16] segmenting OCTA vessels using CNNs.

However, very few studies address automated segmentations of microvascular networks. And none of them are focusing on analyzing main and microvascular vessels separately.

3.2 Graph theoretic algorithms for image segmentations

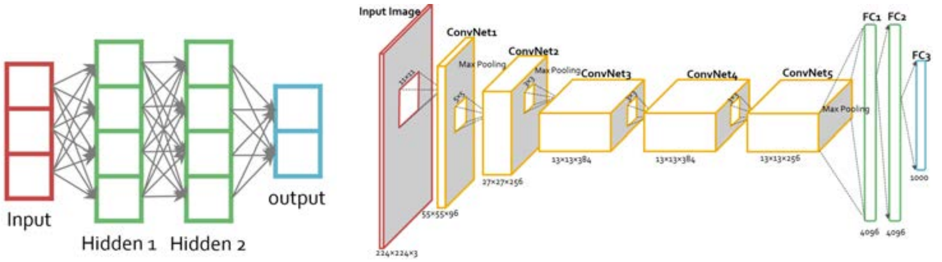
The graph-theoretic algorithms closely related to this study are graph-search algorithms [53–56]. They are multi-dimensional optimal surface segmentation approaches first developed by Li et al. [53]. Li's work creates weighted graph $G = (V, E)$ from the input image, where $v \in V$ set are for pixels (or voxels) $(v_i, v_j) \in E$ set for connected edges between pixels. Weights designed for the data are applied to the graph as costs. By solving the s/t cut problem, this approach can compute minimum closed sets as optimal surfaces in polynomial time complexity.

As an extension of the original graph-search algorithms, multi-layer surface segmentation in SD-OCT volumes with 3D graph-based approach was proposed by Garvin et al. [54] to segment up to 7 layers in retinal SD-OCT volumes. With the approach, multiple retinal surfaces in 3D are simultaneously segmented. It forms a globally optimal solution to the cost function and constraints designed for the data. This work also shows that graph-search can incorporate both surface and region cost functions. Within these frameworks, outer layers are segmented simultaneously, then the inner layers in the retina are segmented. Later multiscale and cascaded approaches under the same framework were developed for human [57] or animal retinal studies [58].

Other than direct surface segmentations, graph-search can be modeled to segment more complex shapes with inter-object interactions [59]. For our specific interests, Xu et al. [55] proposed a graph-based approach to more accurately measure blood vessel boundaries in fundus photographs. By modeling vessels into segments along centerlines, this method segments vessel edges by finding two optimal boundaries simultaneously with weighted graphs.

Another modification to the graph-search is made by Song et al. [60] for segmenting bladder and prostate. It incorporates context and shape-prior terms with weighted edges in the graph. As a result, after taking into account the prior information, smoother and more robust surfaces can be segmented.

In this work, we use the graph-theoretic algorithm in two applications. In the first case, it is used to delineate boundaries of blood vessels in mouse FA images. We also use the approach to segment the superficial layers in the structural OCT data of



(a) (b)

Figure 3.1: Example of NNs. (a) An example of MLP with two hidden layers, (b) AlexNet using CNN for ImageNet. It consists of 5 layers of CNN and 3 fully connected layers.

OCTA scans.

3.3 Deep learning and convolution neural network

Machine learning (ML) systems are often used for image processing tasks, such as object recognition and pixel classification. However, conventional ML methods are usually limited by complex preprocessing and manually designed or tuned features. In many cases, we need to create features to transform image data to some feature space to be suitable for the method.

Deep learning is a class of ML methods using neural networks (NNs) with multiple hidden layers [61]. They are inspired by the biological neural system and use simple non-linear transformations that are determined by training data. It has been very successful in computer vision and medical imaging [62, 63].

One commonly used model of NN is multi-layer perceptrons (MLPs). As shown in Fig. 3.1(a), it comprises multiple layers of fully-connected (FC) layers of neurons. Within such structure, neurons between two neighboring layers are fully-connected in a cascading pairwise fashion. However, for solving image based problems, which could have millions of pixels as input, MLP does not work well.

Deep convolution neural networks (CNNs) is another deep-learning framework designed for image problems. It is a deep network with stacks of non-linear filters that can learn and unveil high-level features from low-level features. It is first proposed by LeCun et al. [64] inspired by the structure of primary visual cortex, and then modified by Krizhevsky et al. [62] in 2012. It has proven to be great at image classifications, with 16% error compared to best 26% with traditional methods in the ImageNet challenge in 2012. Fig. 3.1(b) shows the underlying structure of AlexNet by Krizhevsky. Since AlexNet, CNN has outperformed the state-of-the-art algorithms in different image processing tasks.

3.3.1 Fine-tuning

In some cases, we do not have enough data to train a deep network, especially in cases of medical image analysis. Fine-tuning can help train deep networks with fewer data and time. It uses a pre-trained network, then freezes the learned variables in most of the layers, and uses new data to train only a small part of the network.

There is some work in medical image analyses using the fine-tuning technique. [65] used pre-trained CNN on ImageNet data without tuning the CNN network, to label regions in chest CT volumes. [66] analyzed and recommended that by using a pre-trained network in most cases gives a better or at least equivalent result compared to training from scratch. [67] compared pre-trained networks on CIFAR and ImageNet for abdominal lymph node and lung disease detection. Results show higher accuracies with fine-tuning than training from scratch.

As a result, we can use the fine-tuning technique if the network itself is hard to train or needs a large dataset.

3.3.2 Fully convolutional network and U-Net

In traditional CNN approaches, different networks are often developed to classify a whole image into different classes. On the other hand, tasks we are frequently interested in medical image analysis are localization and segmentation.

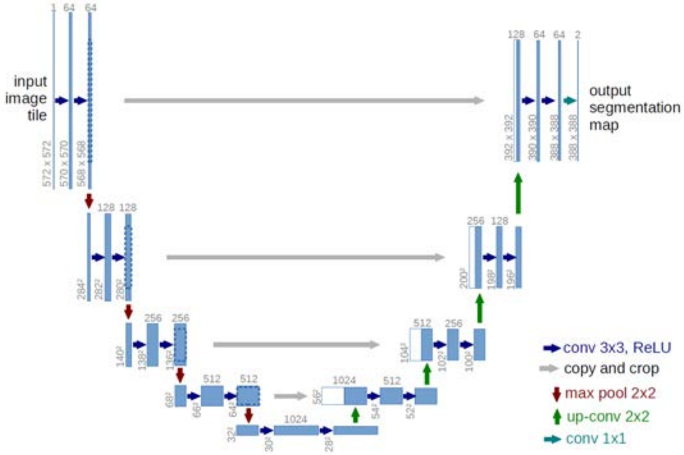


Figure 3.2: The basic structure of U-Net from [2].

Earlier CNN based approaches [68–70] solve such problems with batch-based approach to go through the whole image and get labels for each cropped region. However, these batch based approaches would take much longer time to run, since the network needs to run on each pixel. To solve this problem, Sermanet et al. [70] developed OverFeat technique in 2013. It shows that by converting FC layers into convolutional layers, we can use CNNs for fast localizations. However, this approach is limited by the resolution of output, which is much lower than the original input.

Fully convolutional network (FCN) is then developed by Long et al. [71] for this problem. The main contribution of this work is the introduction of deconvolution layers that is a transposed process of convolution. These parametric layers can be appended to existing CNNs and act as an upsampling mechanism and could combine the information from multiple layers in different resolutions. As an extension of FCN network, U-Net [2] is fully-connected with asymmetric structure with output resolution similar to the input image. Fig. 3.2 shows the basic structure of such network. This framework is quite ideal for our purposes since medical applications usually require higher precision.

A minor problem with U-Net is that it is not a direct derivative of CNNs and can be hard to get pre-trained models. It needs to be trained from scratch instead. Luckily, such structure requires a lot less training data as mentioned in the original paper. For our purposes, we can train the network on smaller, cropped and augmented images with rotations, distortions, and contrast changes. It enables us to use much fewer images to get a better-trained network. Also, because of the nature of FCN networks [70], we can use full resolution images when testing for much faster speed.

3.4 Deep learning for medical image analyses

As a subtask in computer vision, medical image analyses are very different from the traditional tasks. The data size is usually much smaller. Also, the images can be grayscale and don't contain a complex scene or many complex objects of interest. On the other hand, in medical imaging, instead of labeling the whole image, in many cases we need a localized measurement of the image, such as labeling an anomaly or segmenting biological structures.

Some works exploring medical uses focus on finding cropped regions for image-based classification. For example, [63] uses fundus images of diabetic retinopathy patients graded by ophthalmologists for automated grading of diabetic retinopathies. As another example, [72] trains a CNN for classification type of skin cancer. However, these whole-image based classifications require huge datasets (more than 100,000 training images for both approaches) to train. Also, from a clinical standpoint, such CNNs are black boxes with little explanation of its reasoning.

Some other works are based on an FCN. Since the development of FCN [71], some approaches use different forms of FCNs to directly segment structures in human images [52,73–76]. In the field of ophthalmology, there are some existing deep-learning based approaches doing vessel segmentations in DRIVE dataset for fundus images [22,52]. One relevant work in ophthalmology is to segment microvasculature network in OCTA images from Prentašić et al. [16]. They follow a naive implementation of a five layer (three convolutions and two fully connected) CNN to identify each pixel as vessel or background. In another related work, [76] uses a fully convolutional network (U-Net) to segment edema in macular region in SD-OCT volume.

CHAPTER 4
SEGMENTATION OF MAIN VESSELS, MICROVASCULATURE, AND OPTICAL NERVE HEAD REGION IN MOUSE FLUORESCEIN ANGIOGRAPHY IMAGES (AIM 1)

4.1 Introduction

The goal of this section is to develop approaches to segment main vessels, microvasculature, and optical nerve head (ONH) in mouse FA images. The flowchart for this work is shown in Fig. 4.1. This section can generally be divided into two parts.

In the first part, we develop semi-automated approaches to get the retinal structures combining model-based segmentation and manual tracings. We first combine multiscale vessel filters [77], morphologic operations, graph-based approach by Xu et al. [55], and manual corrections to build a framework for the main vessel and microvascular segmentations. The ONH regions are then manually traced.

Next, we use the results from the semi-automated segmentations to develop a deep-learning based approach. We will use a full-convolutional U-Net [2] to simultaneously segment all the structures as pixel classification.

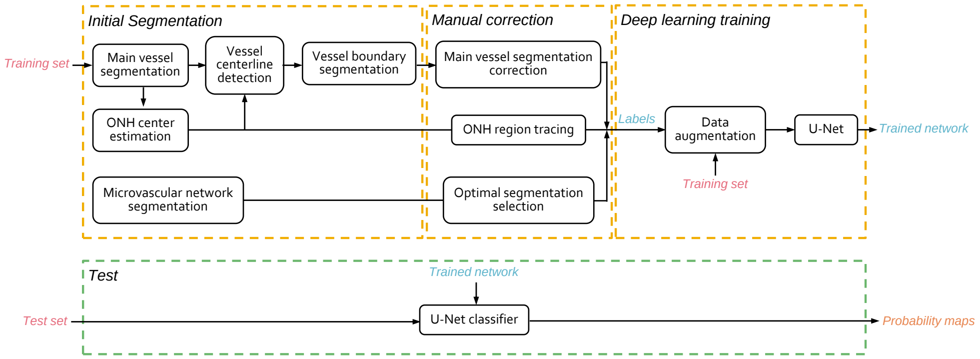


Figure 4.1: The framework for segmenting structures in mouse FA images.

4.2 Segment main vessels with model-based approach

Because the main vessels take a prominent region in FA images, segmentations not distinguishing main vessels and microvasculature could result in inaccurate measurements for microvascular networks. An automated method to segment the main vessels not only excludes their influences on the microvascular network measurements, the main vessels can also act as natural tools to form compartments for localization measurements. We can easily match vessel compartments to other images for multi-modality and longitudinal studies in the future. In this task, we develop a semi-automatic approach to segment main vessels.

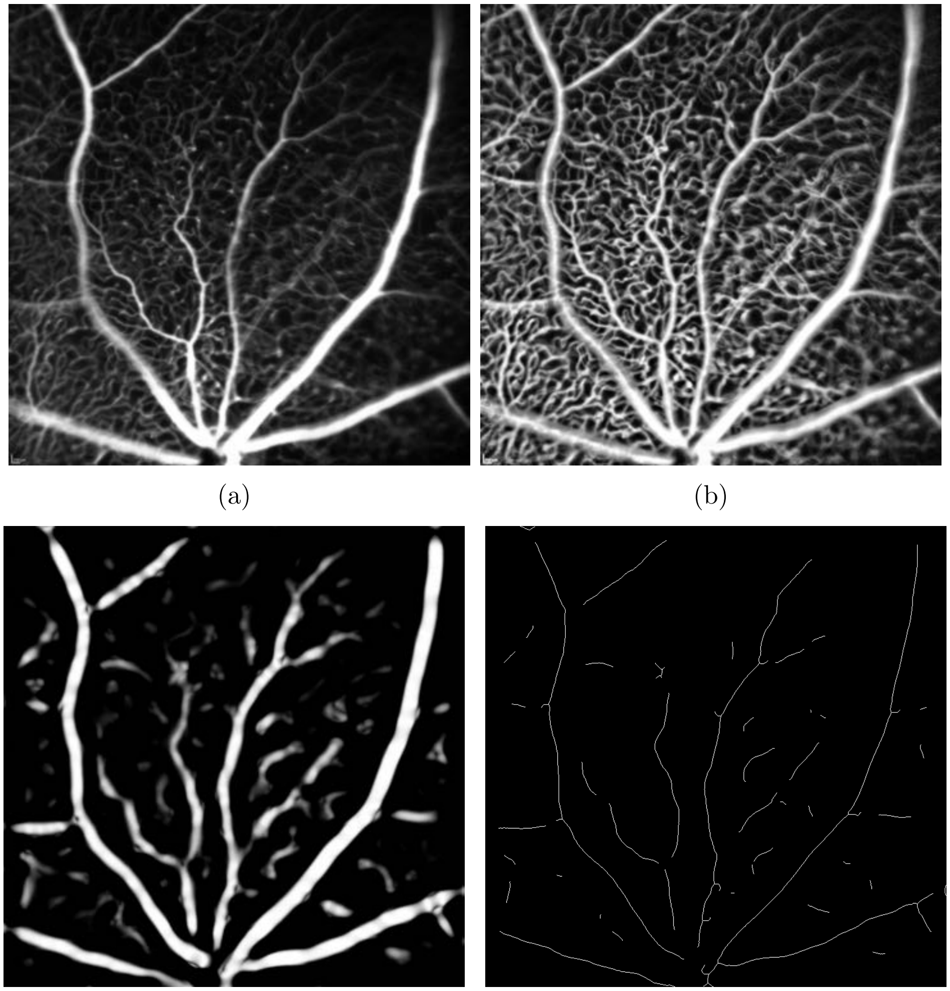
4.2.1 Initial segmentation

As will be described further in Section 4.6, in this study, there are two resolutions of the images in the dataset: 768 × 768 and 1536 × 1536 pixels covering a similar region of 1.5 mm × 1.5 mm. First the resolutions are unified to 768 × 768 pixels. Then contrast limited adaptive histogram equalization (CLAHE) [78] is applied to enhance the vessels in the image (Fig. 4.2(b)). We use Hessian-based multiscale filters by Frangi et al. [77] to segment the main vessels, since this algorithm shows good performance in detecting "tubular" structures in 2D or 3D images. To measure these vesselness structures for each pixel in 2D images, the vesselness function can be written as:

$$\mathcal{F}_{ves} = \begin{cases} 0, & \text{if } \lambda_2 \geq 0, \\ e^{-\frac{{\mathcal{R}_{\mathcal{B}}}^2}{2b^2}}(1 - e^{-\frac{\mathcal{S}^2}{2c^2}}), & \text{otherwise} \end{cases} \tag{4.1}$$

where $\mathcal{R}_{\mathcal{B}} = \frac{\lambda_1}{\lambda_2}$ $(\lambda_1 \leq \lambda_2)$ measures the blob-like versus line-like structures in 2D images. And $\mathcal{S}$ is the local contrast level (structureness) defined by:

$$\mathcal{S} = \sqrt{\sum_{i=1}^{dim} {\lambda_i}^2} \tag{4.2}$$



(a) (b)

(c) (d)

Figure 4.2: The initial segmentation of the main vessels. (a) An example original FA image. (b) The enhanced image of (a). (c) The corresponding vesselness map using Hessian-based filters. (d) The skeletonized vesselness image of (a).

where λ_i are the eigenvalues of the of the Hessian matrix, and *dim* is the dimension of the image. In our cases, λ_is are λ_1 and λ_2 for 2D image analysis. The b and c are the thresholds for the sensitivity of $\mathcal{R}_\mathcal{B}$ and $\mathcal{S}$. Here these thresholds are empirically determined with $b = 0.5$ and c as half of the maximum Hessian norm value.

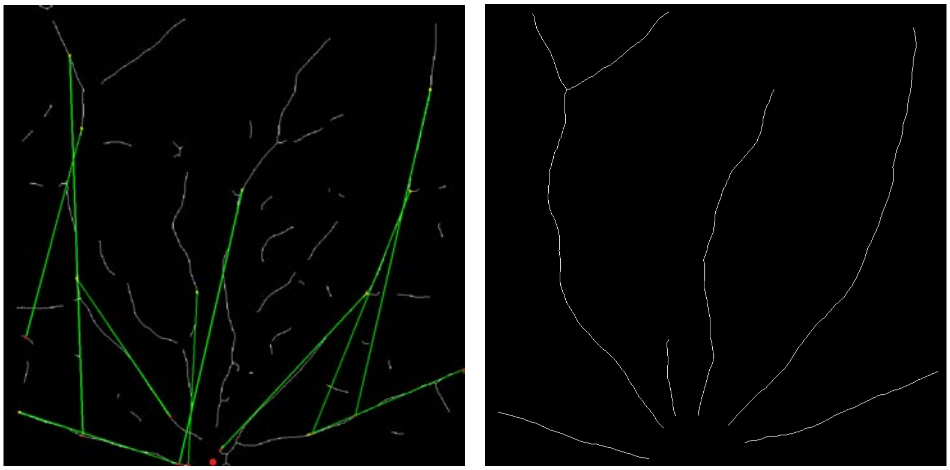
More specifically, Equation 4.1 measures bright curvilinear regions in a darker background. While the results as seen in Fig. 4.2(c) show inaccurate boundaries for the main vessels, for initial segmentations, we only need a rough approximation of these vessels.

After applying thresholding to the filtered image to make it binary, we form skeletonized vessels (Fig. 4.2(d)), by iteratively removing the pixels on the borders [79].

4.2.2 ONH center identification

After the initial process, we find the approximate ONH center for a better assessment of main vessels. From observations, we can see the main vessels surrounding the ONH are straight and pointing towards the ONH center. If there are enough (and there usually are) main vessels shown in the image, we can use a rough position of the main vessels to find the ONH center.

Here we use Hough transform to estimate the positions and directions of the vessels and use these estimated straight lines to predict the center. The skeletonized binary vessel image from the previous section is first transformed into a parameter space (r, θ), for all lines $r = x\cos\theta + y\sin\theta$ in the image [80]. Then the strongest signal points in the space are selected and projected back to the skeletonized image as lines. We then get all the intersection points for the lines and use the median in x y directions as the approximate ONH center. Fig. 4.3(a) shows an example of the approximation.



(a) (b)

Figure 4.3: The segmentation for main vessel centerlines. (a) The fitted lines from Hough transform and approximated ONH center as red dot. (b) Result of skeletonized main vessels.

4.2.3 Skeletonized main vessel identification

Although Hessian-based multiscale filters have already roughly excluded thin vessels, analyzing and excluding irrelevant segments is still needed to construct the main vessel network. In order to remove the effects from the complex pattern around the ONH region, we first remove the points with a distance less than 0.15 mm (approximately 80 pixels) from the approximated ONH center. Next, we detect all the bifurcation and crossing points. The bifurcation points are defined as the location where one vessel centerline diverges into two, and crossing points as the location where two or multiple centerlines overlap. Here we get the centerline pixels with three eight-connected neighbors or more, similar to Xu et al. [55]. Then we use an iterative method based on analyzing the position of each vessel segment. The steps are applied as follows:

1. Find and remove all the bifurcations. The endpoint in a vessel segment away from the bifurcation is annotated as a freepoint.

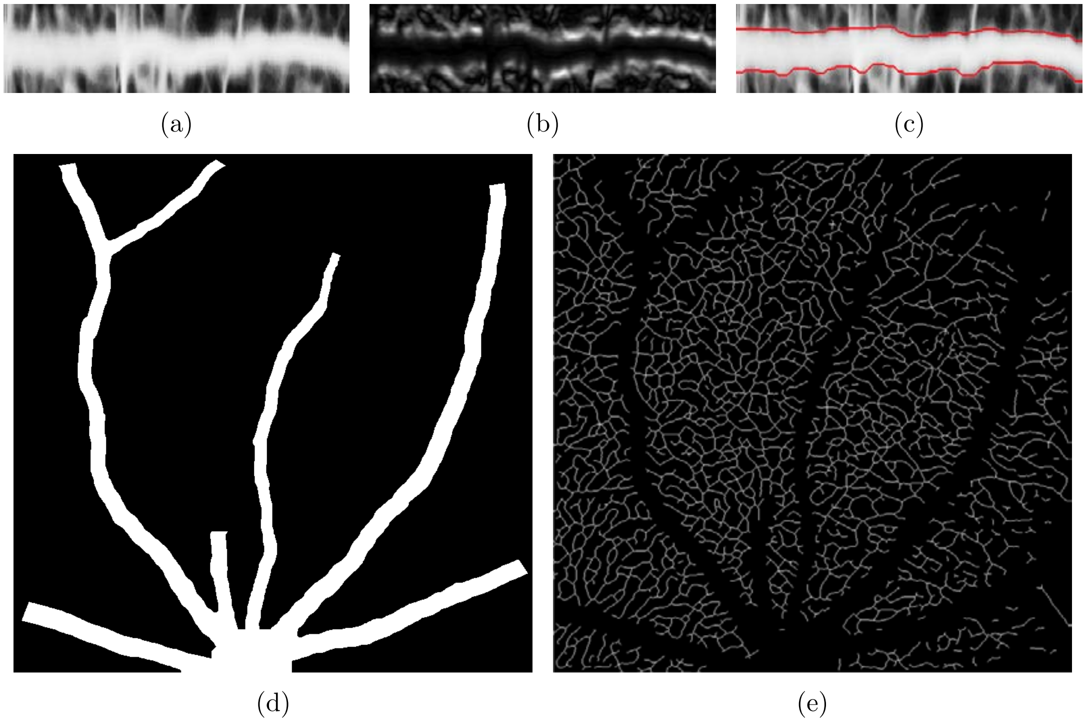
2. Iterate through each vessel segment:
   (a) If the endpoint closer to the ONH center is a freepoint: remove the segment if the freepoint is more than 0.2 mm (100 pixels) away from the ONH center.
   (b) If the further endpoint is a freepoint: remove if the segment is short.
3. Add back the bifurcations and form a new network.
4. Repeat the whole process until no vessel segment is removed.

The result is the main vessel network as seen in Fig. 4.3(b).

4.2.4 Main vessel boundaries segmentation

In this step, we re-segment the main vessels more accurately. A similar approach to Xu et al. [55] is applied to each main vessel segment with the skeletonized main vessel network from the last section. On each vessel, we first apply principal component analysis (PCA) on each segmented vessel centerline pixel and its closest neighbors. We use the first principal component as the direction and the second as the normal. Columns of pixels along the normal direction of a vessel centerline in the original FA image are used to form vessel images. An image with the size of $SegmentLength \times SurroundingPixels$ is generated for every segment, as shown in Fig. 4.4(a).

We use the graph-based method with soft constraints by Song et al. [81] to segment the main vessels smoothly. The boundaries of the vessels in the image are simultaneously segmented with the first-order derivative of Gaussian filters as cost functions, an example can be seen in Fig. 4.4(b). The segmented lines are then projected back to the original image (Fig. 4.4(c)) to form the main vessel mask, as shown in Fig. 4.4(d).


(a) (b) (c)

(d) (e)

Figure 4.4: Segmentation for the vessel networks. (a) An example main vessel piece to be segmented. (b) The corresponding cost function. (c) Segmentation result for the vessel segment. (d) The main vessel network after aligning the segmentations back. (e) The microvascular network.

4.3 Segment microvascular network

To segment the microvascular network, we use the same multiscale Hessian filters [77] as used for segmenting the main vessels to get a vesselness map. Otsu's thresholding is then applied to get the binary vesselness image. We also exclude the main vessel regions with main vessel segmentations as the mask. We then detect and remove small holes and small unconnected components (smaller than 10 pixels) in the thresholded image. The thresholded image is then skeletonized to fixed width form microvascular centerlines. Here we use skeletonized microvascular network because microvasculature occupies little space, but FA images show different widths of the microvasculature dependent on luminance. For the convenience of evaluation, the calculated single-pixel-width skeletonized microvasculature is dilated to fixed 3 pixels. Because the original image quality varies, we also try different scale combinations and thresholding strategies on each image and manually pick the most accurate one. Fig. 4.4(e) show the image we get after the process.

4.4 Visualize and refine segmentations with human interactions

At this stage, we gather all the information from previous sections and use an interactive viewer to generate the labeled data to train the deep network in next section.

We first use the main vessel segmentations from section 4.2.4 and project them back to the original image coordinate space. But instead of directly making the segmentations into pixels masks, we make each segmentation into a vessel centerline plus the predicted thickness around the centerline. This way the segmentations are much easier to interact with than pixel-based prediction.

Also, the graph-based segmentation creates a thickness prediction on each point on the vessel centerline and this will be very hard to interact with. Hence we first use thin-plate-spline [82] to fit a more smooth curve to the vessels boundaries and

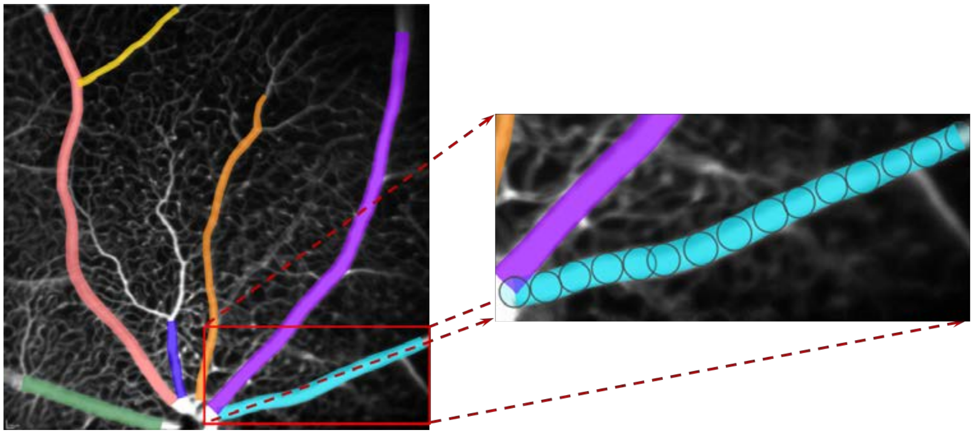


Figure 4.5: The simplified main vessel segmentations projected in an interactive way.

sample points from the smoothed boundaries. We then use Ramer Douglas Peucker (RDP) algorithm [83, 84] to generate a keypoint representation of the vessels. This greedy algorithm simplifies a multi-dimensional sequence into fewer keypoints. We run the algorithm on 3 axes of (x, y, r) where (x, y) are the vessel centerline and r is the thickness measurement. We use an ϵ of 5.0, meaning the maximum distance between simplified form and truth is less than 5 unit distance.

The keypoint representation of the vessels is then visualized and modified manually with a web-based viewer we developed. An illustration can be seen in Fig. 4.5. Users can refine the location and thickness of each keypoint by mouse or keyboard. The refined results (Fig. 4.6) are used in the later stages to train the U-Net.

Then, we use the ONH information from section 4.2.2 and trace the boundary of the ONH region. Next, we combine the corrected main vessel, microvascular segmentation, and ONH tracing to create 4-channel label maps. Channel 0 is the background, and the remaining three are the segmented retinal structures. All the overlaps between the three segmentations are also removed. Fig.4.7 shows an example result overlaying on the corresponding FA image.

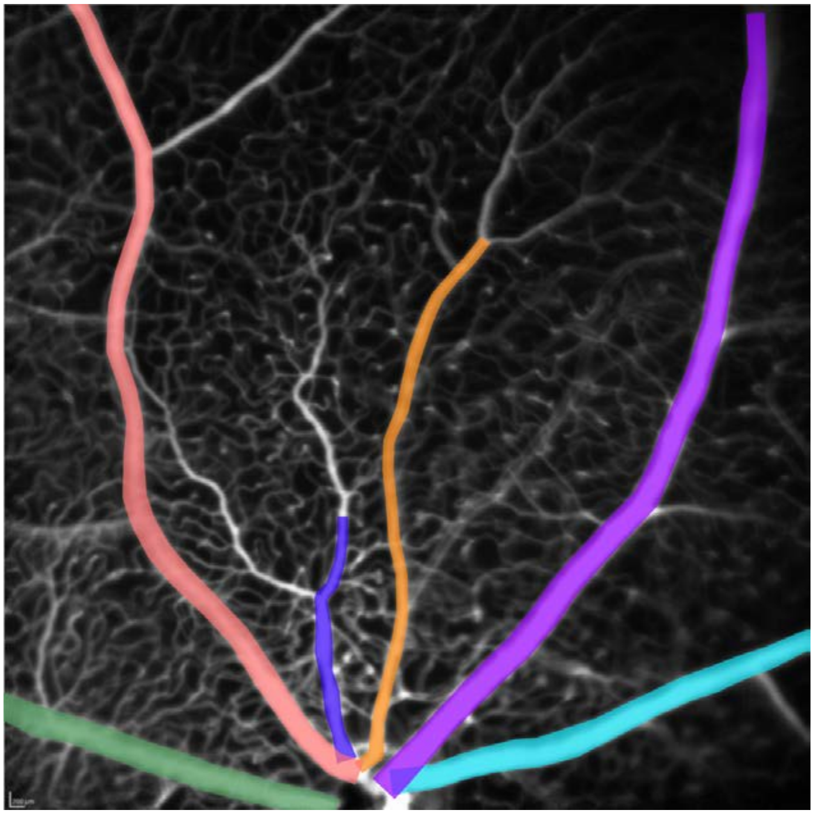


Figure 4.6: The interactively corrected main vessel mask.

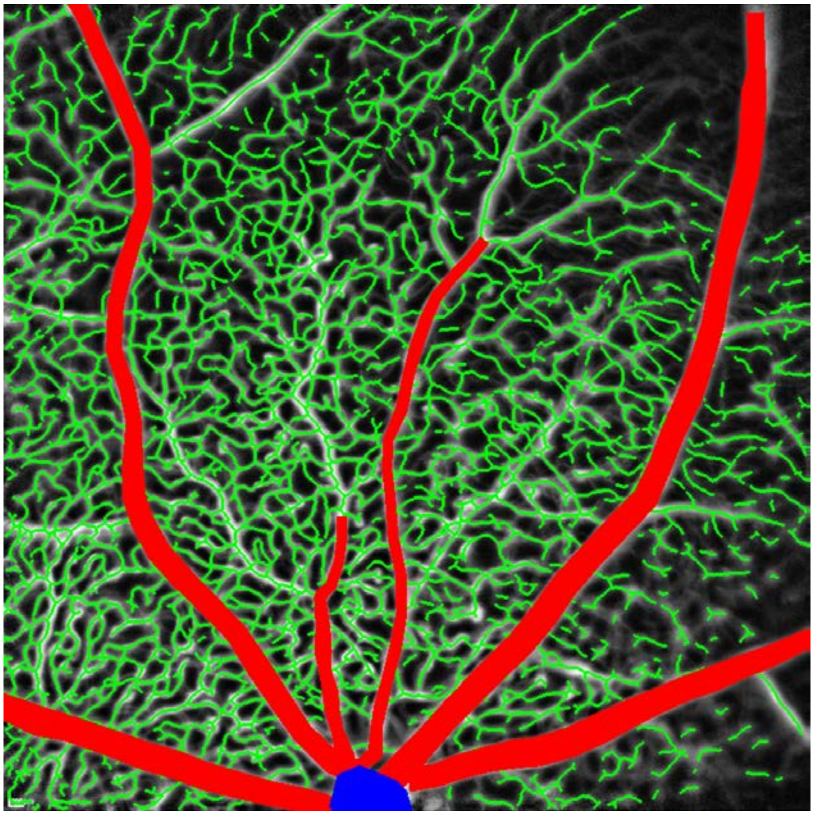


Figure 4.7: An example tracing to be used for training the deep network.

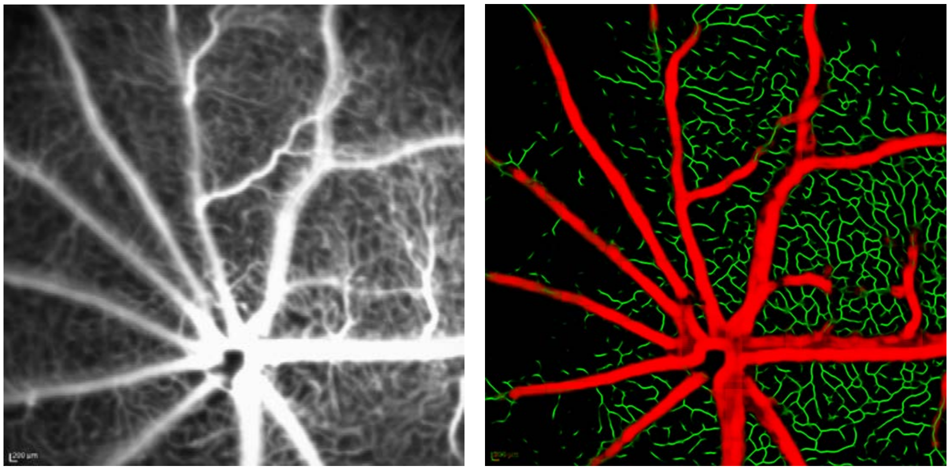
4.5 Segment multiple structures simultaneously with U-Net

In the previous sections, we used semi-automated approaches to get the main vessels and microvascular segmentations. We also traced the ONH regions. Here, we develop a fully automated deep-learning based approach and use the previous results as input to train a deep neural network. U-Net [2] is applied to do a pixel-based classification as retinal structures and background.

4.5.1 Dynamic data augmentation

Because CNNs are a large set of filters with learnable parameters stacked together, we do not apply denoising or image enhancement filters to the images as a preprocessing step. Instead, we use data augmentation so that the network can learn the necessary filters automatically. Here data augmentation means transforming and modifying instances in the dataset, so that we can get a larger and more varied dataset. Traditionally, it is done statically before training stage. Such a static approach to augment data shortens the preprocessing time compared to dynamic modifications before each training iteration. However, it causes a more complex data preprocessing pipeline and also has less data variety, especially when combining more types of augmentation (such as cropping, rotations, and color jittering).

In practice, we notice that compared to some initial results we get from statically augmenting training data by 10 times before training (as seen in Fig. 4.8), dynamic augmentations by modifying input images before each training iteration greatly improve the consistency of the pixel classification results. As a result, for this dataset, cropping, rotation, flipping, and contrast changes are applied dynamically before each training iteration. We first unify the resolutions to 768×768 and normalize the intensities for each image. Then, the training images are cropped at random locations with the resolution of 256×256 pixels. Then the data is augmented with random rotations in range $(-180°, 180°)$. A random horizontal flip is then applied. We also



(a) (b)

Figure 4.8: An example result (b) of (a) by using static augmentation.

do contrast changes from full to half intensities, since in many cases some vessels are very dim and such data augmentation could avoid preprocessing to enhance the image. The whole process is parallelized to multiple threads so that the training speed is not affected.

After the data augmentation, the processed images are directly used to train the U-Net.

4.5.2 Network setup

For the training of the network, here we use a modified version of U-Net [2] as seen in Fig. 4.9. This network has a symmetric structure consisting of an encoder with 8 3×3 convolutional layers to extract information at different scales and a decoder with 7 layers to reconstruct the information back into prediction maps of the same resolutions as inputs.

More specifically, we use an established VGG11 [85] with batch normalization (BN) layers as the encoder part. Although a number of approaches show that by using residual blocks [86–89] may increase precision, these approaches have to train from scratch because pretrained ResNet cannot be directly adapt as the encoder of the U-Net. Recent work by Iglovikov et al. [90] shows by using an established network (VGG11 in this case) as the encoder part of U-Net increase the training speed and segmentation result compared to original U-Net.

Different from [90], we also add batch normalization layers to improve the numerical stability. Batch normalization (BN) [91] is proven to be a very effective way to increase the robustness of training, and widely adopted in newer approaches [87–90]. Briefly, a BN layer first normalizes the output in each channel with the following:

$$\widehat{x}^{(k)} = \frac{x^{(k)} - E[x^{(k)}]}{\sqrt{Var[x^{(k)}]}}, \tag{4.3}$$

where E is the mean of the mini-batch and Var is the variance.

Then the BN layer uses additional learnable parameters to compensate the normalization operation, as shown in following:

$$y^{(k)} = \gamma^{(k)}\widehat{x}^{(k)} + \beta^{(k)}, \tag{4.4}$$

where $\gamma^{(k)}$ and $\beta^{(k)}$ are the parameters to be learned in the training phase.

Also, padding is added to every layer to preserve the size of the output segmentation. An initialization scheme by Kaiming et al. [92] is used to initialize every layer. For the loss functions, cross entropy loss is used on a pixel level to classify four classes as background, microvasculature, the main vessels, or ONH region. More specifically, the loss combines per-pixel softmax function and cross entropy, similar to [2] and [71]. It can be described as:

$$L(x, class) = -\log \frac{e^{x_{class}}}{\sum_j e^{x_j}}, \tag{4.5}$$

where $L(x, class)$ stands for the loss for a specific class, x_{class} is the predicted probability (range from 0 to 1) for this class, and x_j is the predicted probability for each

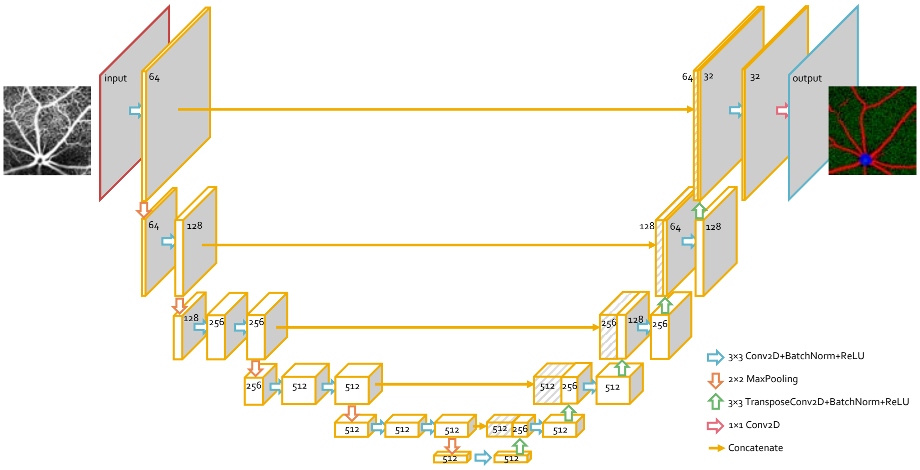


Figure 4.9: The modified version of U-Net used in this work.

possible class j.

Another modification from the original U-Net we do is to use different scales for training batches and testing. Ordinarily, instead of full resolutions, the U-Nets are trained and tested on the same scales using cropped tile images. However, through our initial experiments with this approach, the network tends to generate less confident classifications around the edges. One solution is to use the same full resolution without cropping for training as well as test. The problem with this approach is that it requires much more GPU memory to train each image. This also deteriorates the performance of BN layers since such layers perform better with larger batch sizes [93]. On the other hand, because the network structure is fully convolutional, here we do data augmentation and network training on the cropped images and directly evaluate the original higher resolution images. If the dimension does not fit the U-Net, we expand the size by padding to the next fitted size. For instance, our implementation requires images with pixel resolutions divisible to 32. If a test image has a resolution of 129×130 pixels, zero intensities will be padded to the edges of the image to make it to the next fitted size of 160×160 pixels.

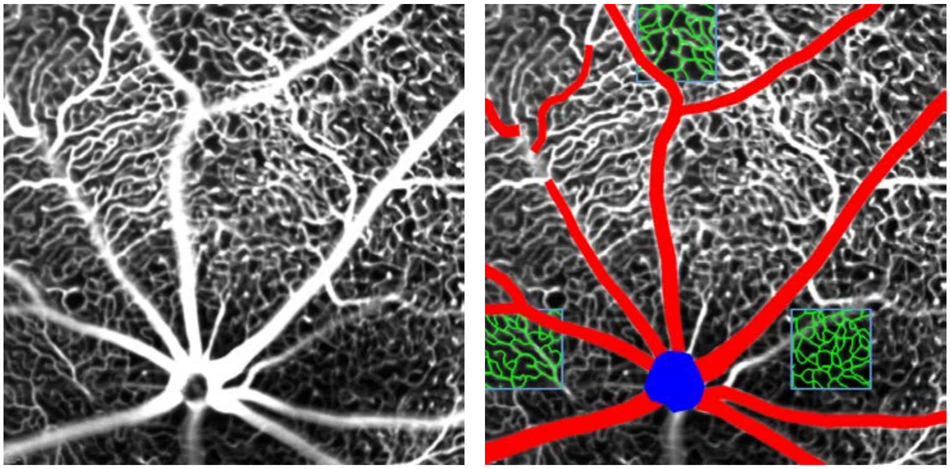

(a) (b)

Figure 4.10: An example test image and the manual tracings (b).

4.6 Experimental methods and results

A total of 73 FA retinal images from 47 diabetic mice acquired using Heidelberg Spectralis scanners (Heidelberg Engineering, Germany) are included in this study. All procedures involving mouse subjects were in compliance with the ARVO statement for the Use of Animals in Ophthalmic and Vision Research, and approved by the Office of Research & Development at the Department of Veterans Affairs. These images are either from OS or OD acquired at one or two timepoints, showing both main vessels and the microvascular network. Within these images, 52 images have dimensions of 768 × 768 pixels and 21 have dimensions of 1536 × 1536 pixels covering a similar region of 1.5 mm × 1.5 mm. For convenience, we uniform the resolution to 768 × 768 before any processing. These images are randomly separated into three categories: a training set, validation set, and test set. Each set doesn't have images from the same mouse. The images are separated as 10 as the test data, 5 for the validation set, and the remaining 58 images are for the training set.

For the test data, the ONH region and the main vessels are manually traced, and the results generated by the trained network are compared to these manual tracings.

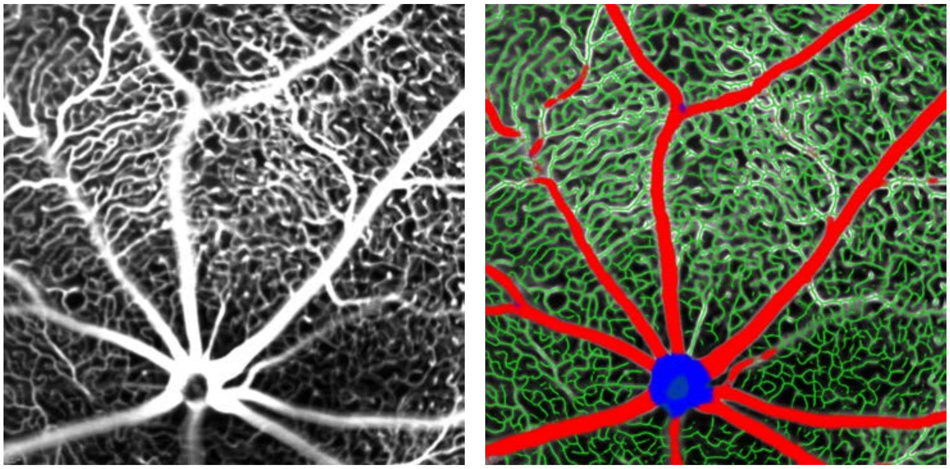
As for the microvascular network, it is hard to trace the whole image, since there are thousands of pieces of microvasculature shown. Here we separate each test image into 6×6 mosaics and randomly select 3 regions with 128×128 pixels. Fig. 4.10(b) shows an example random selection of the regions to be evaluated on, and the figure also shows the manually traced main vessels and ONH region.

For the training part, we use mini-batch stochastic gradient descent (Adam) optimizer [94]. We use the patch size of 256×256 resolution random crops from original training images. Then the data is augmented with random rotations and random contrast changes. A starting learning rate is set to 1e-4 and by monitoring the validation loss, we decrease the learning rate to 1e-6 at epoch 360 and 1e-7 at epoch 450. All training and validation are performed with a single Nvidia GeForce 1080 Ti GPU. Here we use PyTorch to do the training and testing. It has a simple and flexible interface with dynamic computation graphs.

The whole network takes about 2 hours to train, and afterward, the network starts to overfit. We monitor the validation loss and picked the lowest loss one as the final network.

Fig. 4.11(b) shows an example probability map of each of the segmented classes for a FA image (Fig. 4.11(a)). Visually, our approach gives clean and high confidence predictions even before any postprocessing to eliminate small isolated noise. It also is very similar to the manual tracings shown in Fig. 4.10(b) and only have a few small mis-classified branches for main vessels and small false positive areas for microvasculature and ONH regions.

To evaluate the results, we use pixel-based receiver operating characteristic (ROC) curves. ROC is performed by using different intensity thresholds on the probability maps for each predicted class from the neural network to generate binary prediction images. We then compare each predicted pixel to the ground truth labels to calculate true positive rates (TPR) versus false positive rates (FPR). Specifically, for



(a) (b)

Figure 4.11: An example segmentation result (b) of retinal structures for a mouse FA image (a).

the evaluation of microvasculature, we follow a similar evaluation method mentioned in [95,96]. Dilation with disk structuring element of radius 1 is used create a buffered truth and TPR vs FPR is calculated. This is because microvasculature is traced as skeletons and cannot be directly compared to probability maps on a pixel-level. By adding buffer pixels to the skeletonized image, we can compare the two with limited biases. The area under the curve of the ROC (AUC) is also calculated to evaluate each of the segmentations.

The ROC curves of the segmentations can be seen in Fig. 4.12. The AUCs shows very high performance for segmenting main vessels and ONH regions. The performance is slightly lower for microvascular networks. To further show the results, we also calculated the AUC for each test image as seen in Table 4.1. Here the overall AUCs are the same as in Fig. 4.12. It shows that the predictions for main vessels and ONH regions are consistently very high (> 0.99) for all results. The results for microvascular networks are less consistent and lower than the other two in all cases.

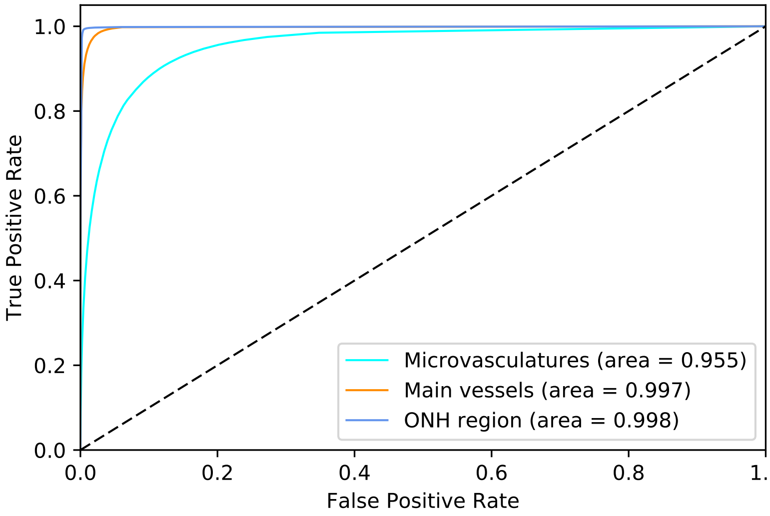


Figure 4.12: The evaluation for fully-automated deep-learning approach. The ROC curves for segmentations of microvascular network, main vessels, and ONH regions.

Table 4.1: Area under ROC curves (AUC) for each test image (10 images in total).

	Microvasculature	Main vessels	ONH region
Image 1	0.951	0.994	0.995
Image 2	0.941	0.998	NA[†]
Image 3	0.944	0.997	1
Image 4	0.93	0.996	0.997
Image 5	0.969	0.997	0.99
Image 6	0.965	0.997	0.997
Image 7	0.964	0.999	1
Image 8	0.962	0.998	1
Image 9	0.962	0.996	1
Image 10	0.973	0.998	0.999
Overall	0.955	0.997	0.998

[†] ROC curve cannot be calculated because Image 2 doesn't show ONH region in neither prediction nor manual tracing.

4.7 Discussion, limitations & conclusions

In this chapter, we develop a method to automatically segment multiple structures in mouse FA images. A semi-automated process is first implemented to minimize the efforts for manual tracing. Then a deep neural network is used to simultaneously find the structures. This approach is trained on 58 mouse FA images and tested on 10 images. The result shows very high AUCs for ONH and main vessel segmentations. The AUCs are lower for segmentation of microvascular networks. First, this is because the microvascular network is harder to precisely evaluate than the other two structures. Our current evaluation method only compares probability maps versus a dilated skeletonized manual tracings in pixel-space instead of 'true' skeletons. Secondly, because the training references are directly from automatic microvascular segmentation without direct corrections, a lower performance can be expected.

We also analyzed the training data and found that about 14% of pixels are marked as microvasculature, about 13% of pixels as main vessels, while 0.9% of them are marked as ONH center. This is another reason for the high AUC for ONH segmentation.

As for main vessel segmentations, although it has a very high AUC, we notice that some of the smaller branches are mis-segmented, as seen in Fig. 4.13. One reason is that the U-Net works as a pixel-based adaptive filter and it will only segment based on the local intensity related information. Also, since the thicker vessels take more than 95% of the pixels and the sometimes mis-segmented small vessels branches only occupy a very small amount of space. Luckily, with the information of ONH center and larger vessels, these branches can be removed through post-processing.

To our knowledge, it is also the only approach to simultaneously segment both main vessel and microvasculature in retinal images. It also to some extent addressed the difficulty to do manual tracings by using traditional approaches to decrease the needs to do all the tracings.

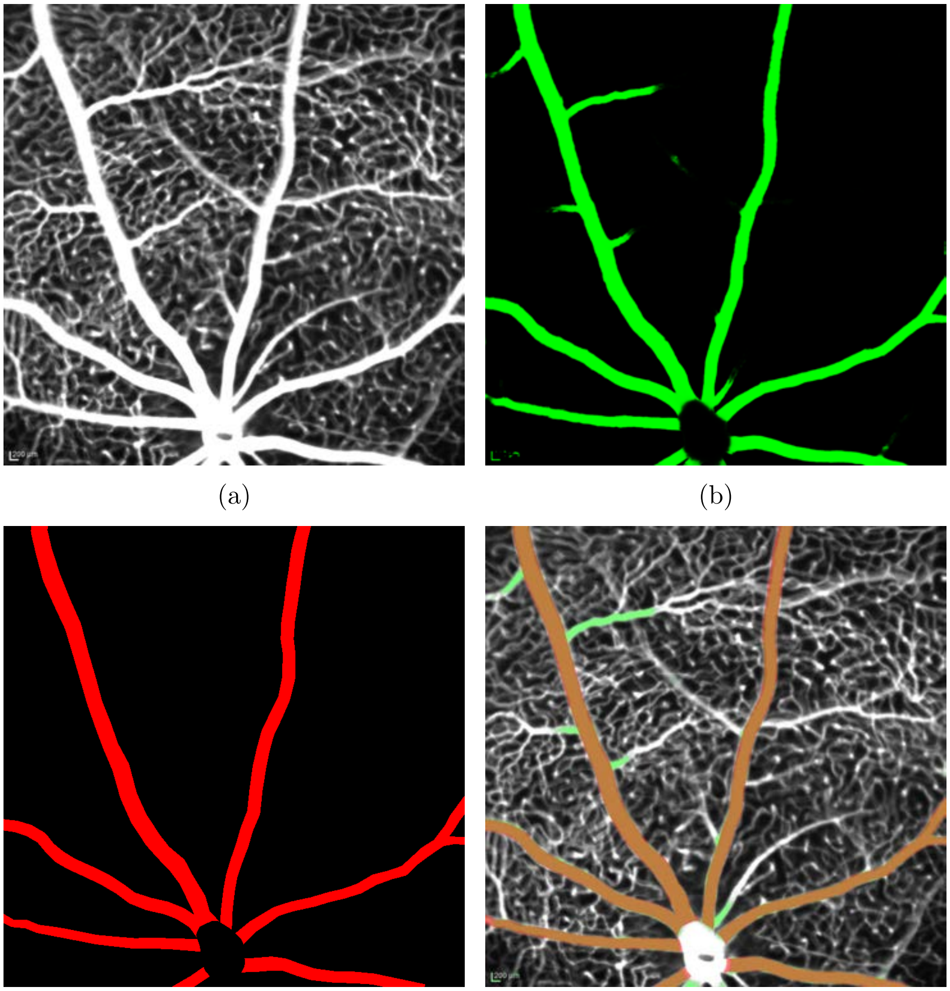

(a)
(b)
(c)
(d)

Figure 4.13: An example image with mis-segmented small branches. (a) The FA image. (b) Pixel segmentation from U-Net. (c) Manual tracing. (d) The overlay shows the main vessels are accurately segmented but some smaller branches are not.

On the other hand, we find using the automated segmentation to train a neural network limits the performance of the deep-learning network. To further improve the performance of the approach, a better way to generate tracing labels for the microvascular network could potentially improve the segmentation. Alternative loss functions (such as Dice coefficients) and replacement of some of the layers may also help.

In conclusion, comparing to the traditional multi-stage model-based approaches with hard-coded parameters and filters, our approach is fully automated and shows the ability to simultaneously segmenting multiple structures reliably in FA images. Also, comparing to approaches only using deep-learning networks in other medical imaging applications, by using an interactive approach to generate training data, we can more easily create references as training data.

CHAPTER 5
ANALYSIS OF MICROVASCULAR DENSITY IN LOW SIGNAL HUMAN OPTICAL COHERENCE TOMOGRAPHY ANGIOGRAPHY IMAGES (AIM 2)

5.1 Introduction

Traditionally, fluorescein angiography (FA) is widely used for assessing retinal micro-circulations. However, this technique is invasive and doesn't distinguish microvasculature in superficial retinal layers from the deeper plexus. OCT angiography (OCTA) is a modality increasingly used for assessing retinal vasculature pathologies in retinal studies [26–28].

Case studies of radiation retinopathy (RR) such as [9, 10, 24] have shown the effectiveness of OCTA for early diagnosis. There are also studies [7,8,15,97,98] that use automated or semi-automated approaches to quantify vesselness. However, there are two problems with these approaches:

1. All the studies related to RR only use 3 mm × 3 mm images. Fig 5.1 shows an illustration of differences between a 3 mm × 3 mm and 6 mm × 6 mm field of view. 3 mm × 3 mm images sometimes are too small to give an overall judgment for macular status.
2. None of the approaches mentioned above are evaluated against manual measurements. The validations are based on statistical significance between different subjects or different eyes.

To address the problems mentioned above, the goal of this chapter is to develop and evaluate methods to quantify microvasculature in 6 mm × 6 mm OCTA *en-face* images. There are three approaches developed and compared for this purpose:

- **Baseline approach**: This is a slightly modified version of the AngioVue software. By using contrast limited adaptive histogram equalization (CLAHE) to enhance the images, intensities in each image are more evenly distributed than

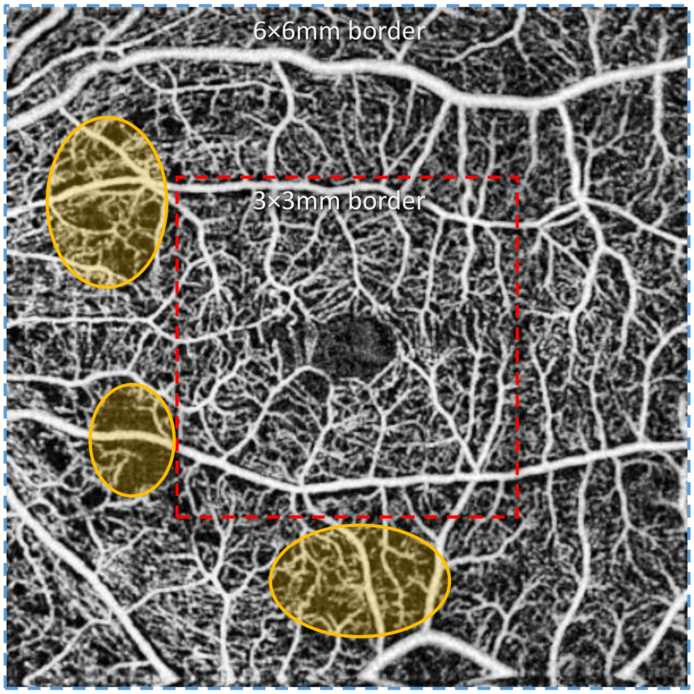


Figure 5.1: An illustration of 6 mm × 6 mm (in blue box) versus 3 mm × 3 mm (in red box) field of view in a RR patient. The yellow circles show example regions of interest outside the 3 mm × 3 mm region where there may be capillary loss.

the original image. This is a representative method that directly enhances the pixels in the OCTA *en-face* image for analysis.

- **Hessian-based approach**: Hessian filters are used to detect vessel-like structures in the OCTA images, and then skeletonization is applied for a unified vasculature appearance. This is a method for detection of vessel-like structures in the images.
- **Deep-learning approach**: Based on the Hessian-based approach, some of the false positives are cleaned up, followed by training a pixel-based classifier to find the regions with vessels. This approach shows the ability to detect vasculature with machine learning.

On the application side, there are two main distinctions in this study. Because the smaller field of view with 3 mm × 3 mm (and 3 mm × 3 mm × 2 mm for volumetric data) may not show enough information of a patient's macular status, the input images we use for the baseline approach are 6 mm × 6 mm 2D *en-face* images, and 6 mm × 6 mm × 2 mm volumes for the Hessian-based and deep-learning approaches. Also, in contrast to many other approaches focusing on assessment of the statistically significant differences between diseased and control images, we evaluate these approaches by how good they predict avascular regions in each macular *en-face* image.

5.2 Baseline approach

For the baseline approach, we use a simple approach based on the commercially available AngioVue analysis with slight improvement. To our best knowledge, AngioVue or AngioVueHD (a higher resolution software update from AngioVue) do density measurements by directly apply blurring filters for the vessel density. However, as seen in Fig. 5.2(a), the large vessels take a prominent place in the image and tend to be clearer and brighter than the microvascular networks. Therefore, the commercial software tends to overestimate the larger vessels and show a higher measurement

(Fig. 5.2(b)). Also, because some parts of the microvascular plexus are dimmer than others, a direct measure will also overlook such regions and classify them as low density.

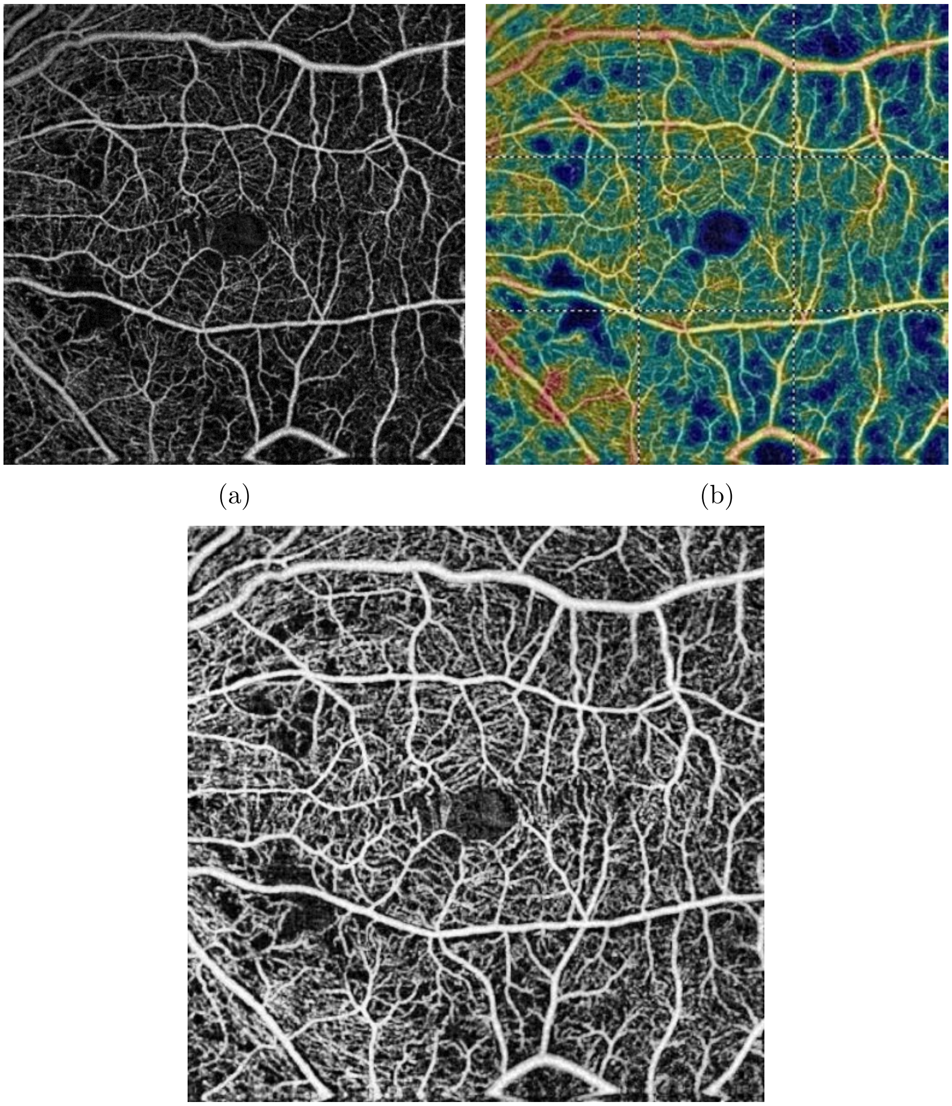
As a result, instead of directly applying filters on *en-face* images to generate density maps, we enhance the images to more consistent intensities. For this approach, we use CLAHE [78] for a more uniformed contrast in the different region of the image. To be more consistent with analysis from AngioVue, we use the *en-face* images generated directly generated from the software as inputs. The result is directly used as the baseline for the effectiveness of microvascular measurements. Fig. 5.3(d) shows an example result of the process. This is also consistent with studies such as [7,15,99] that use some variations of the baseline approach with intensity-based enhancements for the analyses.

5.3 Preprocessing from raw data for the Hessian-based and deep-learning approaches

In each OCTA scan, there is a volumetric structural SD-OCT scan showing the macular tissue and a corresponding split spectrum amplitude-decorrelation angiography (SSADA) scan showing the vesselness. For the other two approaches, we use the raw SD-OCT and SSADA data to segment and form the *en-face* projection angiogram of superficial layers.

More specifically, we first decode the raw SSADA data into OCTA projection image form. To get the layer information, here we segment surfaces in the corresponding SD-OCT volumes in the dataset with graph-based layer segmentation algorithms [54, 57]. The outer surfaces are first segmented, and inner surfaces are then segmented. We can segment up to 10 layers in the SD-OCT volume with the graph-based approach.

Since the raw SSADA volumes precisely correspond to SD-OCT volumes, we use



(a) (b)

(c)

Figure 5.2: OCTA superficial *en-face* images and analysis. (a) The *en-face* image directly from AngioVue software. (b) The density heatmap of (a) generated from AngioVue. (c) The enhanced *en-face* image with CLAHE.

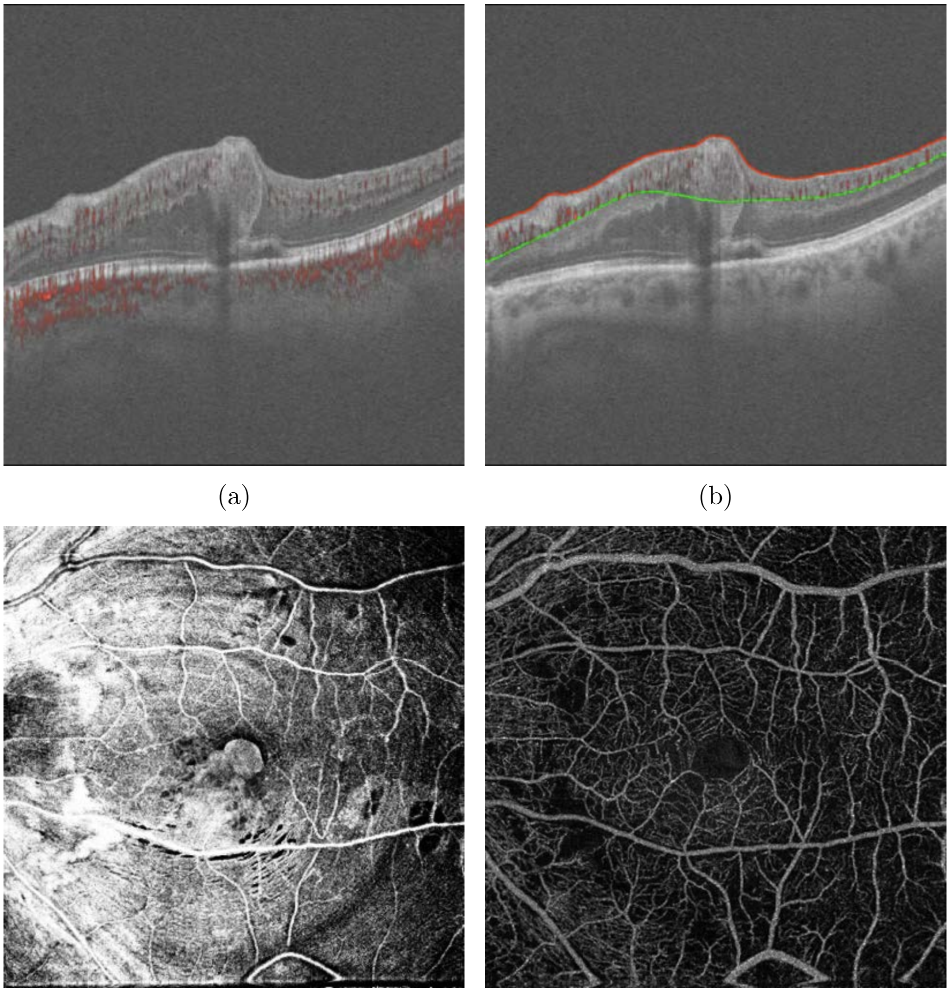
the layer segmentations to measure vessels in different retinal layers. In this approach, we follow the definition from Camino et al. [97] and use the superficial layers as the combination of retinal nerve fiber layer (RNFL) to inner nuclear layer (INL). Fig. 5.3(b) shows the SSADA information overlaid on a b-scan showing the vessel responses in the corresponding layers. To generate the projection images, we use the maximum intensity in each column in the layers. An example result can be seen in Fig. 5.3(d).

We notice that the projection images directly generated from SSADA are not very balanced in the intensity distribution. We also see that this is also correlated with the intensity of the corresponding structural OCT *en-face* projection images. As a result, we use the projection image (Fig. 5.3(c)) of the corresponding layers of the structural OCT to compensate the intensities. Median filters are first applied to the projection image to avoid introducing noise to the OCTA *en-face* image. Histogram equalization is also adopted to enhance the contrast for the image. The OCTA *en-face* image is then multiplied by the processed image with additional constants:

$$proj(x,y) = (2 - 0.9 \cdot proj_{oct}(x,y)) \cdot proj_{octa}(x,y), \tag{5.1}$$

where $proj_{oct}(x,y)$ is the structural OCT projection image, $proj_{octa}(x,y)$ is the OCTA *en-face* image, and $proj(x,y)$ is the compensated image from OCTA *en-face* $proj_{octa}(x,y)$ at each pixel location (x,y). Both $proj_{oct}(x,y)$ and $proj_{octa}(x,y)$ are normalized to $(0,1.0)$ intensity range before calculation. The constants for the process are experimentally determined by testing different combinations. Fig. 5.4(a) shows an example result of the process.

Because the deep learning approach we use prefers a resolution divisible by 32, the generated *en-face* images are unified to 480 × 480 resolutions by upsampling with cubic spline interpolation to preserve more details.



(a) (b) (c) (d)

Figure 5.3: An example of OCTA image. (a) SD-OCT b-scan with SSADA as an overlay in red scatters. (b) The same b-scan showing the superficial layers used in this study. The superficial layers are shown between red and green boundaries, and only SSADA information between the boundaries are used. (c) The projection image of the superficial layers of the structural SD-OCT volume indicated in (b). (d) The corresponding *en-face* projection angiogram.

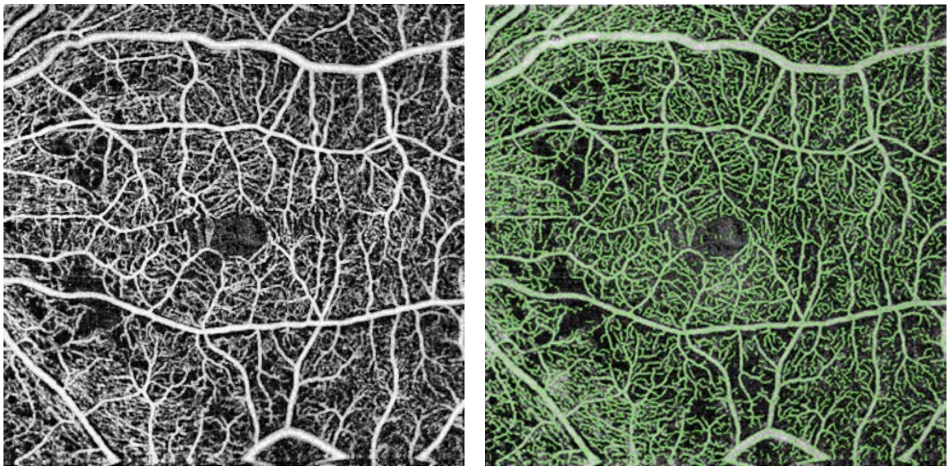


Figure 5.4: Measurement of microvasculature in OCTA images. (a) An OCTA *en-face* image generated with preprocessing steps from Section 5.3. (b) The microvascular network shown in green as overlay on the *en-face* image.

5.4 Hessian-based segmentation

The Hessian-based analysis is adopted in a number of approaches [97,98] for analysis of vesselness and has been proven to be effective detecting "tubular" structures. As a result, similar to what is described in Section 4.2.1, in the second model-based approach we use filters by Frangi et al. [77] to segment the microvasculature in the *en-face* images.

Because the microvasculature should have a similar thickness, we also perform skeletonization and additional morphological post-processing to unify the thickness to one pixel; a similar approach is also used in work like [10]. This also minimizes the effect of large vessels because all vessels become one-pixel thick. Fig. 5.4(b) shows the microvascular network in the projection image.

5.5 Deep-learning segmentation

The Hessian-based segmentation works well in many of the images. However, in some of the more noisy images, such analysis shows regions with false positives as shown in Fig. 5.5(b). Directly removing these segments will also greatly reduce the true positives in some low SNR images.

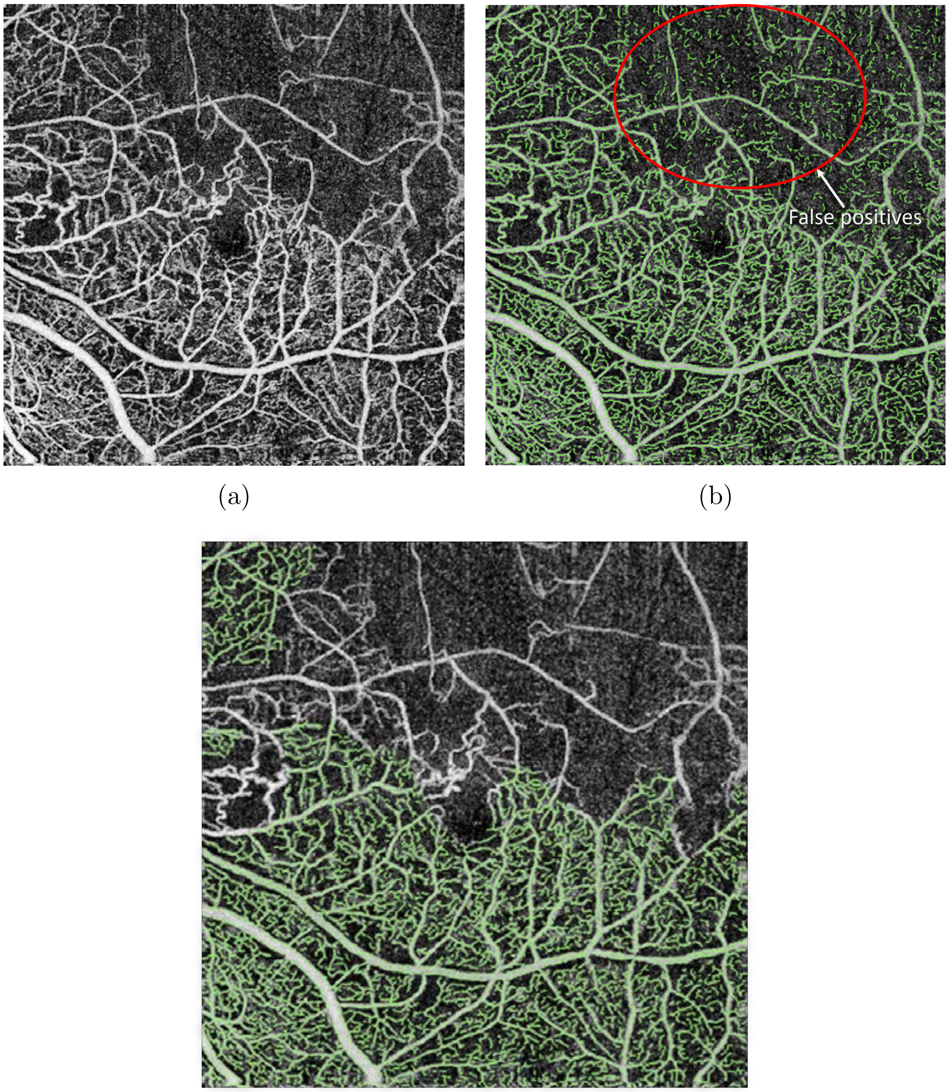
From our experience, the model-based approaches have a difficult time balancing sensitivity and specificity, because the microvasculature in many of the images we use is very similar to speckle noise. Also, the results are highly dependent on the noise level in the pre-processed data. For instance, a change in the histogram equalization or compensation function gives a completely different analysis.

As a result, in this section, we train a machine-learning classifier to automatically segment the microvasculature. We follow the approach from the previous chapter and train a deep network with the manually corrected Hessian-based segmentation.

An illustrative graph of the approach can be seen in Fig. 5.6. The U-Net basic structure here is similar to the network used in Chapter 4. The projection images directly generated from SSADA without enhancement (as shown in Fig. 5.3(d)) from Section 5.3 are used as the inputs to the network. For the training reference data, the results from Hessian-based segmentation with manual corrections are used. As seen in Fig. 5.5, we only manually remove the apparent false positives because it is impossible to trace the vessels accurately in a large number of training images.

The main differences from the network in the last chapter are the output format and the loss function when training. The loss function used in the previous chapter is similar to U-Net and FCN [2, 71]. With the output layer as a Softmax nonlinear function, the output has to be at least two channels, with the channel-wise summation of 1.0. In this work, we use a Sigmoid function as the final layer to output a probability map. This simplifies the process and outputs a single class, instead of background and foreground by Softmax as in the last chapter. As a result, the loss function is as follows:

$$L_x = -\text{mean}(y \cdot \log \sigma(x) + (1 - y) \cdot \log(1 - \sigma(x))), \tag{5.2}$$



(a) (b)

(c)

Figure 5.5: Manual corrections of false positives in Hessian-based segmentations. (a) An example OCTA *en-face* image. (b) The microvascular network is not correctly segmented in some of the regions. (c) The new microvascular measurement with manually removed false positives.

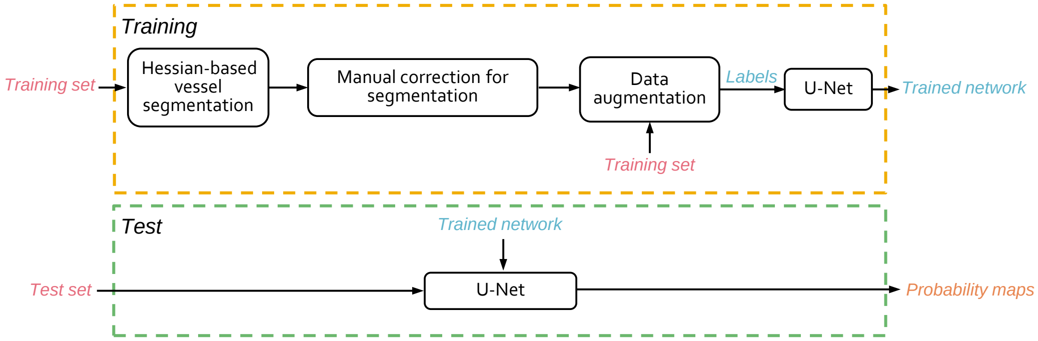


Figure 5.6: The flowchart for segmenting microvasculature with deep learning. The results from the Hessian-based approach and corrected and used as labels for training.

where $\sigma(x) = \frac{1}{1+e^{-x}}$ is the Sigmoid function and the y is the truth.

Also, instead of using the enhanced *en-face* image with CLAHE, we use the original *en-face* image. This is because although the contrast between vessel and background is increased due to the operations, there is a loss of information that may be needed. An example probability map from this process can be seen in Fig. 5.7.

5.6 Experimental methods

Eighty-six macular scan sets with AngioVue generated *en-face* images of superficial layers, structural SD-OCT volumes, and the corresponding SSADA data from OD and OS from 43 human subjects are involved in the study. Within the subjects, 32 are RR patients and 11 are control cases. All the images are acquired with AngioVue (Optovue, Inc., Fremont, CA). All procedures involving human subjects in this study were approved by Institutional Review Boards (IRB) and the Human Subjects Office at the University of Iowa.

Different from previous studies [9, 10, 24] with $3\,\mathrm{mm} \times 3\,\mathrm{mm}$ physical dimensions, all of the images used in this study all have $6\,\mathrm{mm} \times 6\,\mathrm{mm}$ dimensions. The SD-OCT volumes have $400 \times 400 \times 640$ or $304 \times 304 \times 640$ resolutions and the SSADA scans are $400 \times 400 \times 160$ or $304 \times 304 \times 160$ covering the same $6\,\mathrm{mm} \times 6\,\mathrm{mm} \times 2\,\mathrm{mm}$ retinal

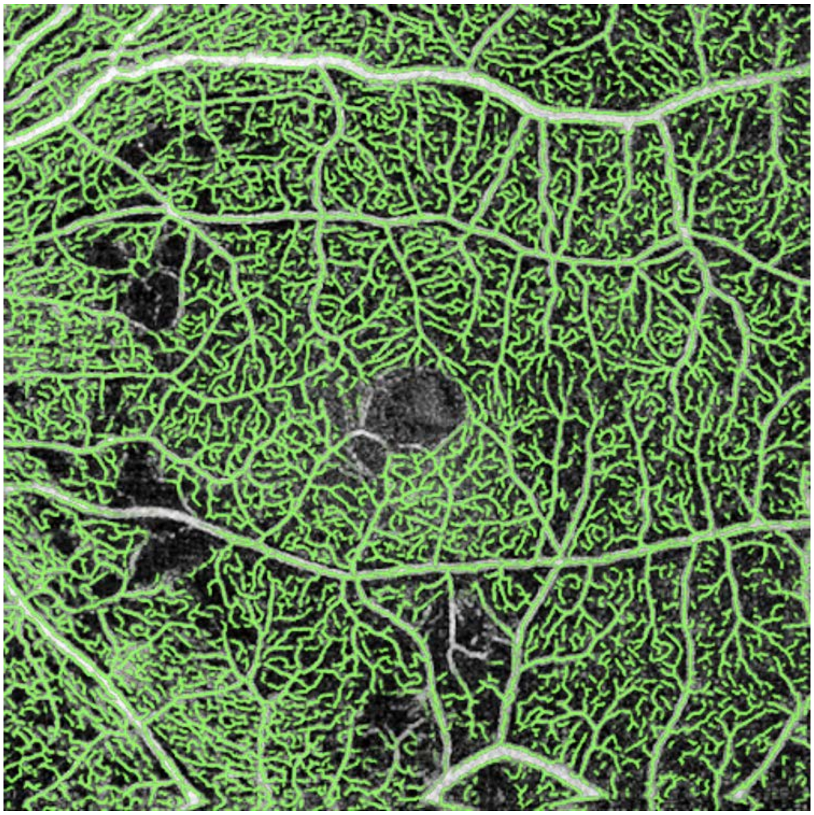


Figure 5.7: An example result of the deep-learning segmentation with predicted map overlaid on the enhanced *en-face* image.

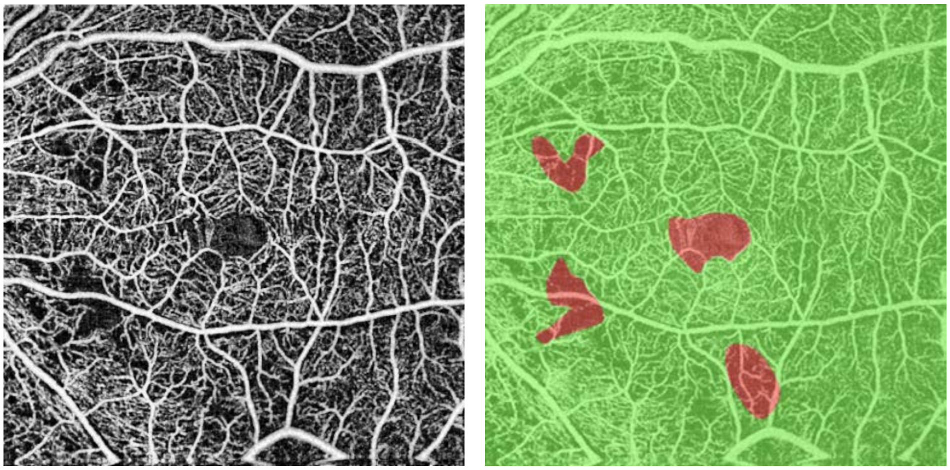
regions. The control scans used in the study are all acquired by AngioVueHD with 400 × 400 resolution. For the RR patient, images from 6 patients are with AngioVue with 304 × 304 and the rest are AngioVueHD.

For evaluation, traditionally, the segmentations are directly compared to the corresponding manual tracings. In the last chapter in our evaluation of microvasculature segmentation, since it was difficult to directly trace the whole images, we compared our results to randomly traced tiles. However, with this dataset, direct delineations of microvasculature are not achievable in many of the *en-face* images because of very low quality *en-face* images. Also, alternative evaluations by comparison of the significant difference between patients and controls are too 'weak' to differentiate the three approaches we developed.

As a result, we evaluate the effectiveness by assessing the approaches' ability to predict the avascular zones in the *en-face* images. These avascular regions in each image are traced and corrected by experts. Fig. 5.8(b) shows an example tracing of these regions. Based on the vessel segmentations from each approach, we generate a probability map of capillary loss regions, as discussed in more details below. The generated predictions are then compared to the manual tracings.

Specifically, in order to do the comparison with quantitative measurements, here we apply a simple unified super-size averaging filter to the resulting vesselness maps from each approach to blur the detailed microvasculature and generate the predictions. Then we apply different thresholds to the predictions so that regions with no vessels can be distinguished from regions with microvasculature. Pixel-wise differences between the results and manual tracing are used to generate ROC curves. The results are compared using the area under the curves of the ROC (AUCs). Statistical significance between different ROC curves are also computed using bootstrap statistical tests [100] using pROC package [101] in R.

For the dataset separation, because the approaches include traditional model-



(a) (b)

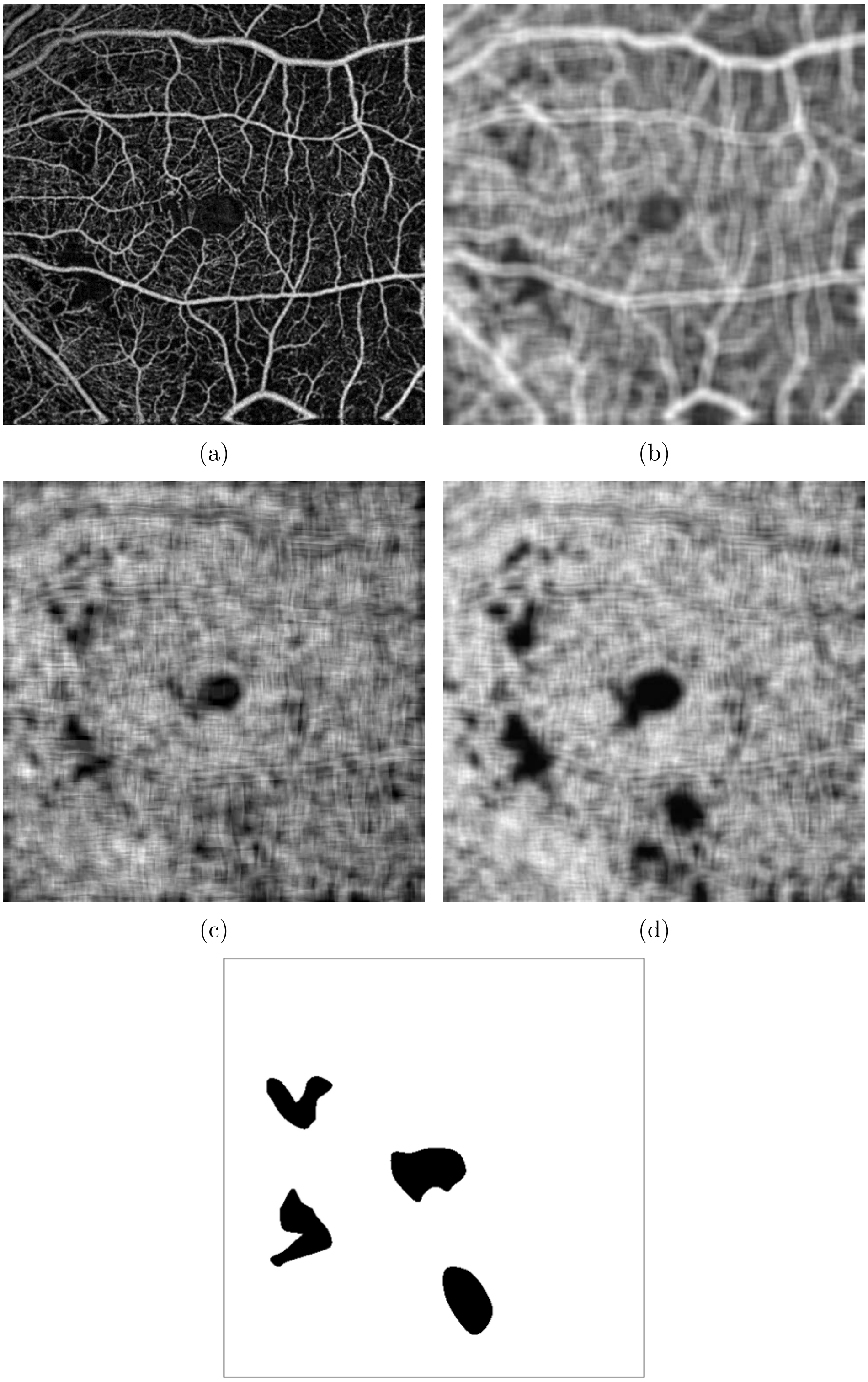
Figure 5.8: The tracing of avascular region (b) in the OCTA superficial *en-face* image (a). The avascular regions are marked in red.

based methods and a machine learning approach, we divide the dataset into training and test sets. We evaluate all of the approaches with the test set. A total of 20 OS and OD images from 10 randomly chosen subjects are used as the test data. The baseline, Hessian-based, and deep-learning approaches are run and compared with manual tracings on the test dataset. The remaining 66 images are used as the training set to train the deep-learning approach.

Also worth mentioning, to train the network in the deep-learning approach (Section 5.5), Adam optimizer is applied with a learning rate of 1e-4. A total of 1200 epochs are used for the total training on a Nvidia GeForce Titan GPU. It takes around 1.5 hours to train the network.

5.7 Results

Example result images for evaluation and the manual tracing of avascular regions can be seen in Fig. 5.9. Qualitatively, the method with deep learning generates the best overall result and is most consistent with the manual tracing. It has a clean background and fewer false positives. For the baseline approach, although we use



(a) (b) (c) (d) (e)

Figure 5.9: Example evaluation images for each approach and the manual tracing reference. (a) An example OCTA *en-face* image to be evaluated. (b) The filtered *en-face* image with the baseline approach. (c) The corresponding result image for the Hessian-based segmentation. (d) The result for the deep-learning approach. (e) The manual tracing.

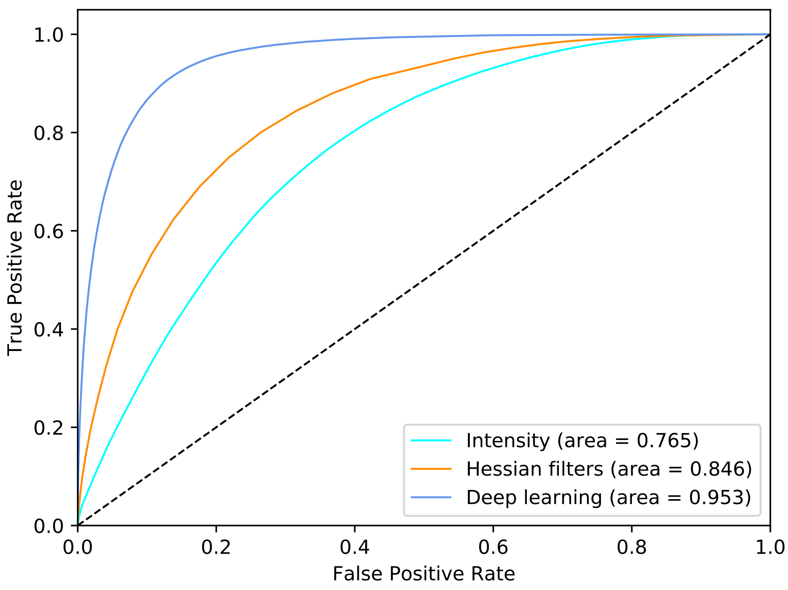


Figure 5.10: The ROC curves for different approaches.

adaptive histogram equalization to balance the intensities, there are regions that have vessels with dim and background regions with bright appearances. This makes it harder to generate a precise prediction of avascular areas. For the Hessian-based approach, it generates better results than baseline, but still has bright false positive regions in the manually traced avascular regions. The problem is that it is difficult to balance the false positives and true positives with the Hessian-based approach, and in some cases (Fig. 5.5(b)) false positives are inevitable.

Quantitative evaluation of ROC curves of these methods is as shown in Fig. 5.10. Each pair of the approaches are significantly different with $p < 0.05$ with bootstrap statistical tests. This confirms the qualitative observations. The deep-learning method gives the best prediction of avascular regions with 0.95 AUC, much better than the other two analysis. It also shows that pure intensity-based approaches (similar to the baseline approach) may not fit very well for analysis of vascular densities with AUC of 0.77.

5.8 Discussion, limitations & conclusions

With the current OCTA technology, microvascular patterns are not very visible in the *en-face* images. This makes direct segmentation and evaluation of microvascular challenging. In this chapter, we develop and compare three approaches including two model-based approaches and a new deep-learning approach to segment microvasculature. Although studies such as [7, 99] use similar methods to the baseline approach, our results show that this approach performs poorly in distinguishing avascular regions from regions with microvasculature (AUC = 0.77). For the Hessian-based approach, to our best knowledge, there is no work that uses similar approaches to segment microvascular networks in RR patients. From the ROC analysis, we see this method gives better performance than the baseline approach in distinguishing vasculature states, and is simpler to implement than the deep-learning approach. However, OCTA *en-face* images from RR patients are often a lower quality compared to control subjects (such as shown in Fig. 5.11). This makes it hard to separate blood vessels from noise with model-based approaches.

We also show that, with the deep-learning based approach, the problem with the previous approaches can be solved. This machine-learning approach can directly learn from manually corrected vesselness maps without any prior knowledge and gives an accurate prediction of the microvascular network. The results show that for detection of avascular regions, the deep-learning based approach gives a significantly better result than the other two approaches.

On the other hand, this work shows limitations of current frameworks. First, a main problem we have is that some of the images have very low SNR ratio such as the example shown in Fig. 5.12. It is impossible to give meaningful predictions of actual microvasculature in such scenarios. For instance, even after manual corrections, the Hessian-based results can only give a vascular prediction map that shows a strong correlation to the microvascular density, instead of direct vesselness predictions, as

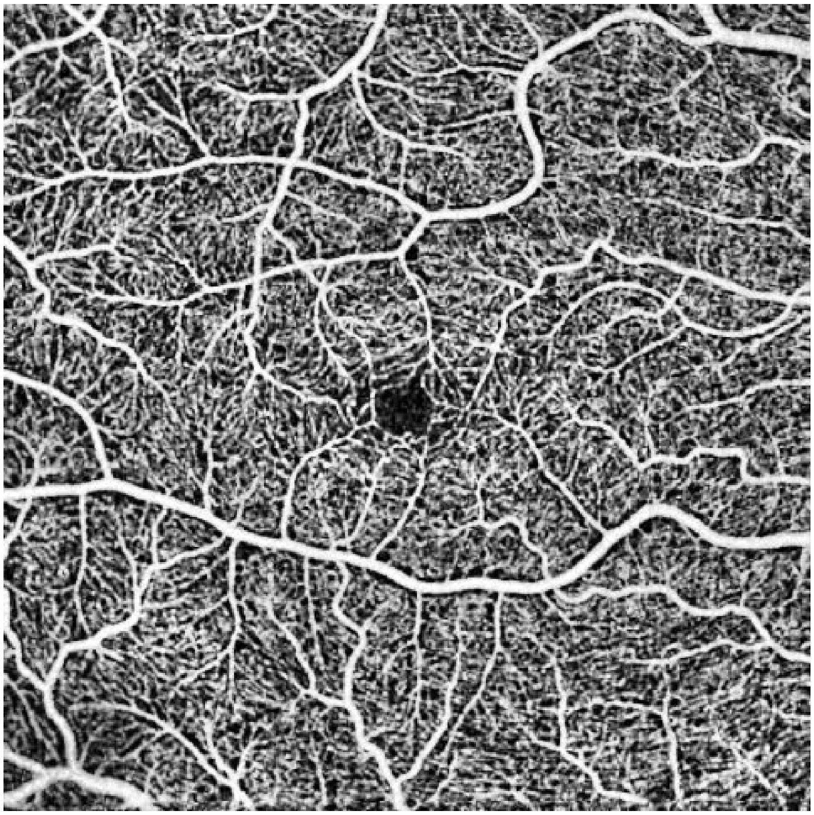


Figure 5.11: An example OCTA *en-face* image from a control subject.

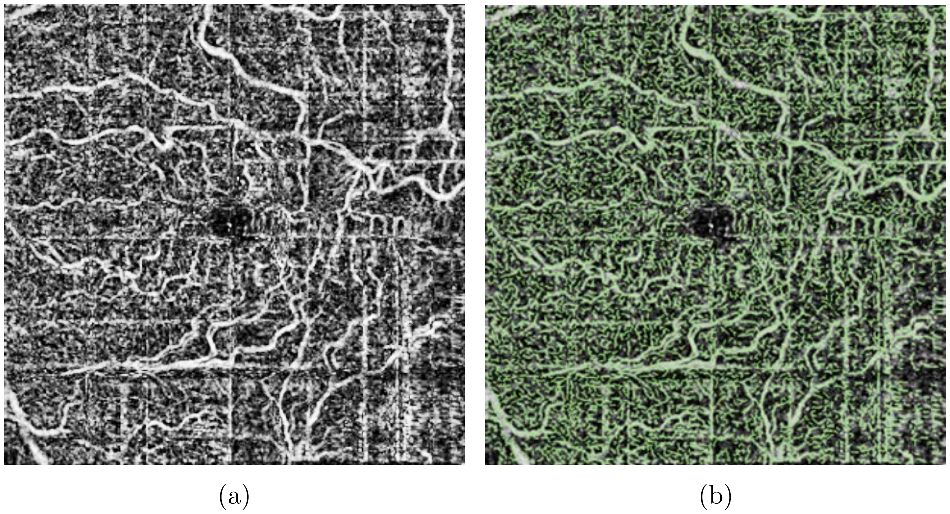

(a) (b)

Figure 5.12: An example of a bad quality OCTA *en-face* image (a) and the corresponding detected vessels with Hessian filters (b).

shown in Fig. 5.12(b). Also, with the deep-learning method, the quality of training data is a major limitation. Because it is very hard to give an accurate depiction of the microvascular network, for the training data, we are only able to remove the false positives mainly in the avascular regions according to the tracings of these regions. This would result in inaccuracy especially in low SNR ratio images.

As a result, to address these problems, we will develop a new approach to automatically segment the regions with a deep learning in next chapter as a follow up. Instead of predicting vascular networks, the future approach will focus on the direct analysis of regions of different microvascular states.

In summary, we have developed and compared three automated ways to predict vesselness in OCTA *en-face* images of RR patients. Our best result shows the vesselness prediction can accurately predict avascular regions against regions with microvasculature. This is also the first study to compare multiple automated approaches and their effectiveness to predict microvasculature in RR patients. Our approaches and validations can potentially enhance the clinical studies and help diagnosis of radiation retinopathies.

CHAPTER 6
REGION ANALYSIS FOR OCTA *EN-FACE* IMAGES (AIM 2)

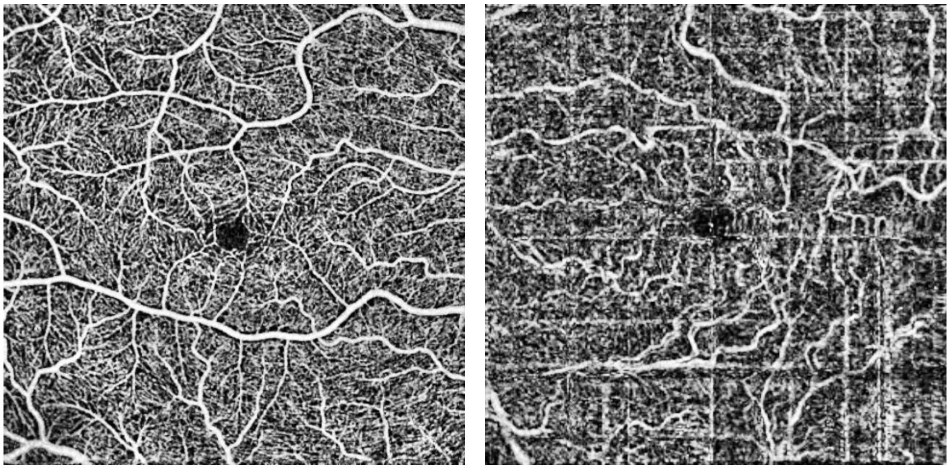
6.1 Introduction

In the previous chapter, we followed the idea to directly segment microvascular network in RR patients. However, the quality of images varies especially in RR patients, and this makes a direct depiction of vessels not feasible. Fig. 6.1 shows an example comparison of a good quality *en-face* image (Fig. 6.1(a)) and an image with an indistinguishable microvascular network (Fig. 6.1(b)). These differences are not isolated phenomena in our dataset. Images in a case study of RR patients such as [24] also shows varying of image qualities. This also makes it difficult to balance sensitivity versus specificity when using the Hessian-based approach. Because the results from Hessian-based approach is used as the training data, this can also affect the results of the deep-learning approach.

On the other hand, to analyze the regions of capillary loss and the overall severity of vascular dropout, we used additional postprocessing such as supersized filters to generate density maps and predicted avascular regions from the maps. This is a more complex pipeline and possibly less accurate. Therefore, direct measurements of regional microvascular state may make more sense.

We also noticed some analysis from OCTA case studies that experts' evaluation of RR severities [24] focuses on the regions with capillary loss or no capillaries. As a result, here we introduce a new deep-learning based approach for region analysis. More specifically, an approach to directly segment the normal and abnormal regions from the OCTA *en-face* images using deep learning is developed.

The purpose of this chapter is to train a single deep-learning network to simultaneously segment avascular regions, regions with capillary dropout, and capillary-dense regions. To achieve this goal, some modifications have been made to the network for this purpose.



(a) (b)

Figure 6.1: A comparison of *en-face* images in the dataset. (a) An example image from control subject showing vessels clearly. (b) An image from RR patient with motion artifact and not directly visible capillaries.

6.2 Network overview

The overall network structure of this work is shown in Fig. 6.2. This approach takes the OCTA *en-face* image with no enhancement and trains the U-Net for separating avascular regions, regions with the capillary dropout, and capillary-dense regions.

Here we use the basic structure of VGG11 [85] with corresponding deconvolution layers, similar to Chapter 4. The main modifications are changing the convolutions in some of the layers to be dilated convolutions introduced by Yu et al. [102]. This approach enables larger receptive field for the network without additional parameters to be learned. We also use a different loss function for training.

6.3 Data augmentation

For preprocessing, we use similar input as in Section 5.5. In addition to the data augmentation methods used in Chapter 4, through experiments, we find that the elastic transformations in augmentations give better results for training than only

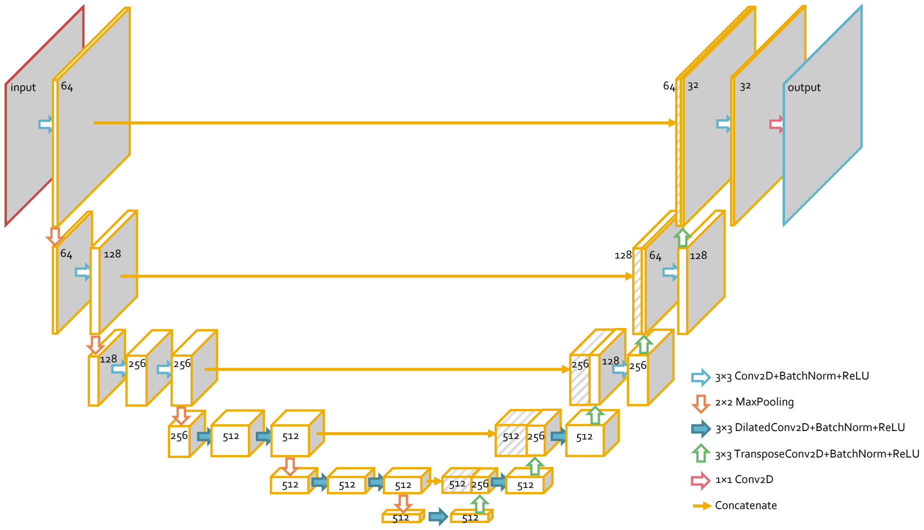


Figure 6.2: The modified version of U-Net with dilated convolutions used in this work.

random flipping and cropping.

Specifically, we use the elastic distortion method mentioned in a 2003 ConvNet paper by Simard et al. [103]. Random displacement fields are first generated with $\Delta_x(x, y) = \delta_x$ and $\Delta_y(x, y) = \delta_y$. For each pixel location (x, y) in the original image, the corresponding pixel in the warped image is shifted by (δ_x, δ_y) from its original location. A Gaussian filter with random σ is then applied to the random displacement fields to create elastic deformations to the training images. An example of this process is shown in Fig. 6.3.

This elastic distortion method is incorporated after the data augmentation techniques we developed in the previous chapter to generate the input training data.

6.4 Dilated convolution

In previous chapters, the segmentations are mainly using local information to find vessel-like structures. In this section, however, more global information is needed for determining if a region only has sparse or dense vessel-like structures around it.

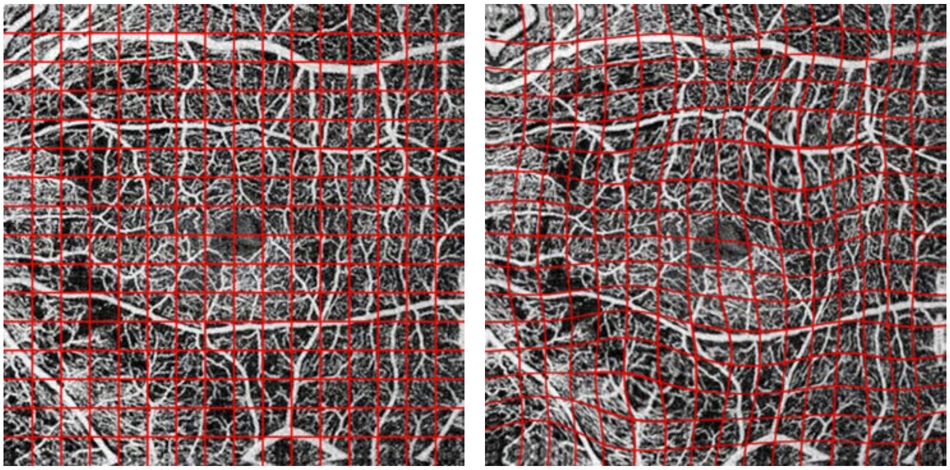
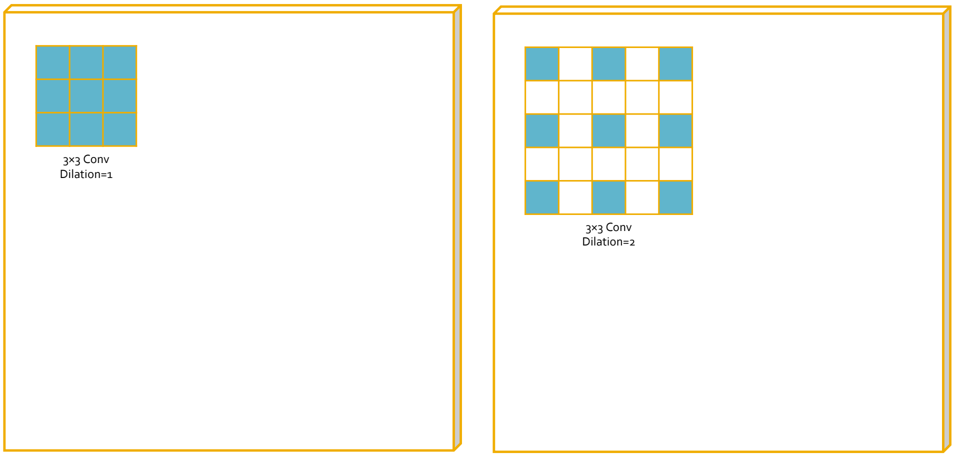


Figure 6.3: An example process of elastic distortion. (a) The image before distortion, with displacement field overlay as red grids. (b) The distorted image with the corresponding displacement field.

This requires a larger effective receptive field [104] to let the network perceive more information around each pixel. As a result, we modify the U-Net to have a larger receptive field.

However, simply increasing the depth of network or expanding the size of convolutions (to, for instance, 5 × 5 convolutions) will increase the memory usage, and the number of parameters to train. Dilated convolution (a.k.a Atrous convolution) [102] are designed to solve such problems. Fig. 6.4 shows an example of dilated convolution with dilation factor of 2 (Fig. 6.4(b)) compared to the normally used convolution (Fig. 6.4(a)). The two convolutions have the same number of learned parameters, but the dilated convolutions can cover more area around each pixel by sampling with skips. This convolution model gives very good segmentation results and is widely adopted in semantic segmentation or scene parsing tasks in natural images [102, 105–107]. Therefore, we change the conventional 3×3 convolutions in our previously developed U-Net to 3×3 with 2 dilation. The convolutions shown in blue arrows in Fig. 6.2 are now dilated convolutions with kernel dilation of 2.



(a) (b)

Figure 6.4: A normal 3 × 3 convolution (a) and 3 × 3 convolution with dilation=2 (b).

6.5 Loss function

In this study, we also extend the loss function for the training of the network. Normally, the output of the probabilities is determined by a combination of Softmax function and multinomial cross-entropy, as seen in Chapter 4.

In the last chapter, we already used Sigmoid + BCE loss for segmenting a single class. In this section, we use this for additional reasons. Softmax works as mutual exclusion, meaning for each pixel, the classes will compete against each other and the network will only favor one type of class. More details are discussed by He et al. [108] in his Mask R-CNN work. In our study, we want to allow the network to simultaneously pick multiple classes when necessary, since the classifications are sometimes not absolute for each pixel. Here we apply Sigmoid layer for probability generations, and BCE for loss, as follows:

$$L_{BCE} = -\sum_i (y_i \cdot \log \sigma(x_i) + (1 - y_i) \cdot \log(1 - \sigma(x_i))), \tag{6.1}$$

where x_i is the predicted value for each pixel and y_i is the manual label. Also, our initial experiments show the necessity of assigning a weight for the loss of each class, otherwise part of the avascular region will sometimes be overlooked. In practice, pixels classified as avascular regions are weighted three times larger than the other two classes.

In addition, some experiments by Iglovikov et al. [90] and Milletari et al. [86] has shown the effectiveness of Dice coefficient loss. It can be defined as intersection-over-union (IoU) and measures similarity between a number of sets. Because the discrete Dice coefficient is not differentiable, the version we use as loss function is a soft Dice loss coefficient:

$$L_{Dice} = \frac{\text{Intersect}}{\text{Union}} = \frac{2 \cdot \sum y_i \cdot x_i + \alpha}{\sum y_i + \sum x_i + \alpha}, \tag{6.2}$$

where $\alpha = 1$ is a Laplace smoothing constant that could prevent division by zero and avoid overfitting.

For the final loss, we join the two losses as:

$$L_{all} = L_{BCE} - \ln\left(L_{Dice}\right). \tag{6.3}$$

With a combination of BCE loss and Dice loss, we can simultaneously maximize the per-pixel prediction as well as the overall intersection between the predicted probability map and the manual tracing.

6.6 Experimental methods and results

Similar to the previous chapter, the same 86 macular *en-face* images of superficial layers generated from SSADA data are used in this study. All the *en-face* images used in this experiment are also unified to 480 × 480 resolutions with cubic spline interpolation upsampling.

In addition to the tracings we did in the last chapter, we separate the *en-face* images into capillary-dense regions, dropped capillary regions, and avascular regions.

All three regions in each image are manually traced. These tracings are used as both references for training as well as the truth to compare with using ROC curves. AUCs are also calculated.

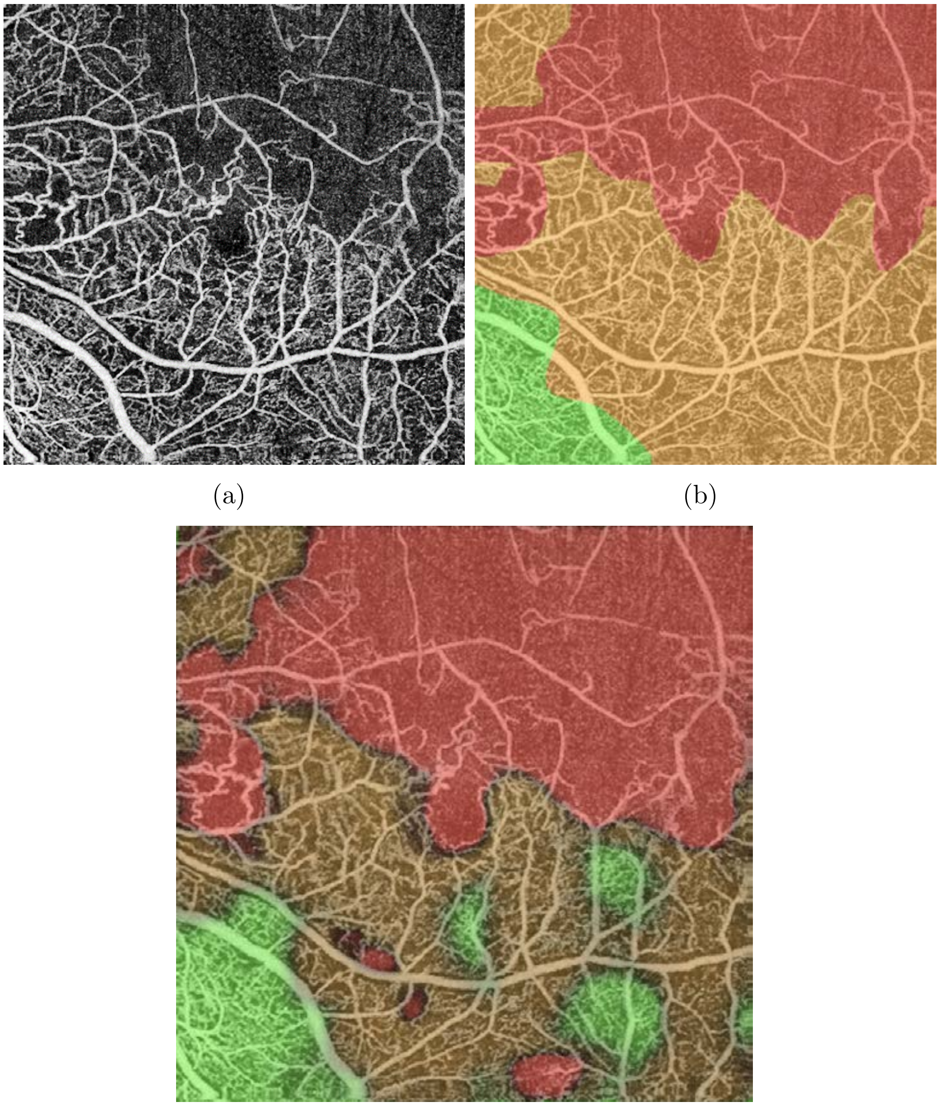
Because we are also generating results to be used in next chapter for analysis for all the subjects in the study, a 10-fold analysis is applied to get predictions on all the images we have. The first 7 folds have 78 images from 39 randomly chosen subjects as training and 8 images from 4 subjects as the test set. The remaining 3 folds have 76 images and 10 images as training and test set.

To train the network, we use Adam optimizer to train the network with an initial learning rate of 1e-3. The learning rate changed to 1e-4 after 200 epochs and 1e-5 around 800 epochs. A total of 1000 epochs are used for the total training on a single Nvidia GeForce 1080 Ti GPU. Additionally, we don't see much overfitting, and thus we do not use a validation set when training. It takes around 2 hours to train one network, and approximately a day needed to train all the U-Nets for the experiment.

On the testing phase, each image takes about 0.1s to run on GPU and 5s to run on a CPU.

An example of the original image, the manual tracing, and the corresponding result can be seen in Fig. 6.5. Visually, this approach offers clean predictions for avascular regions marked in red. The regions are also slightly more consistent and detailed than manual tracings. On the other hand, the predicted capillary-dropped regions in orange are less confident and have multiple misclassified regions.

The ROC curves of the segmentations are shown in Fig. 6.6. Among the ROCs, the prediction of avascular regions shows the best AUC of 0.97. This is also higher than the best result we obtained from the last chapter. The capillary dropout prediction generally gives a worse performance of 0.93 AUC.



(a) (b)

(c)

Figure 6.5: The results from the proposed deep-learning approach. (a) The enhanced *en-face* image. (b) The manual tracing for the image. (c) The segmentation result of the three classes.

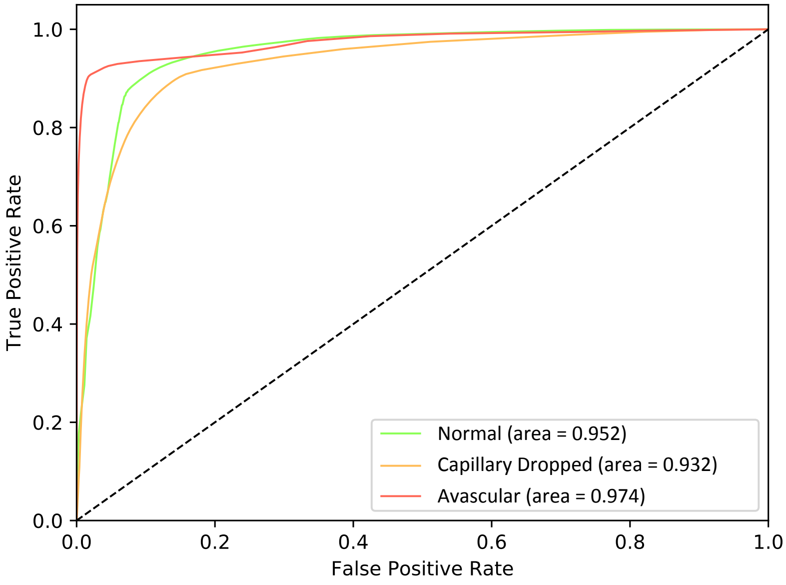


Figure 6.6: The ROC curves for normal, capillary dropout, and avascular regions.

6.7 Discussion, limitations & conclusions

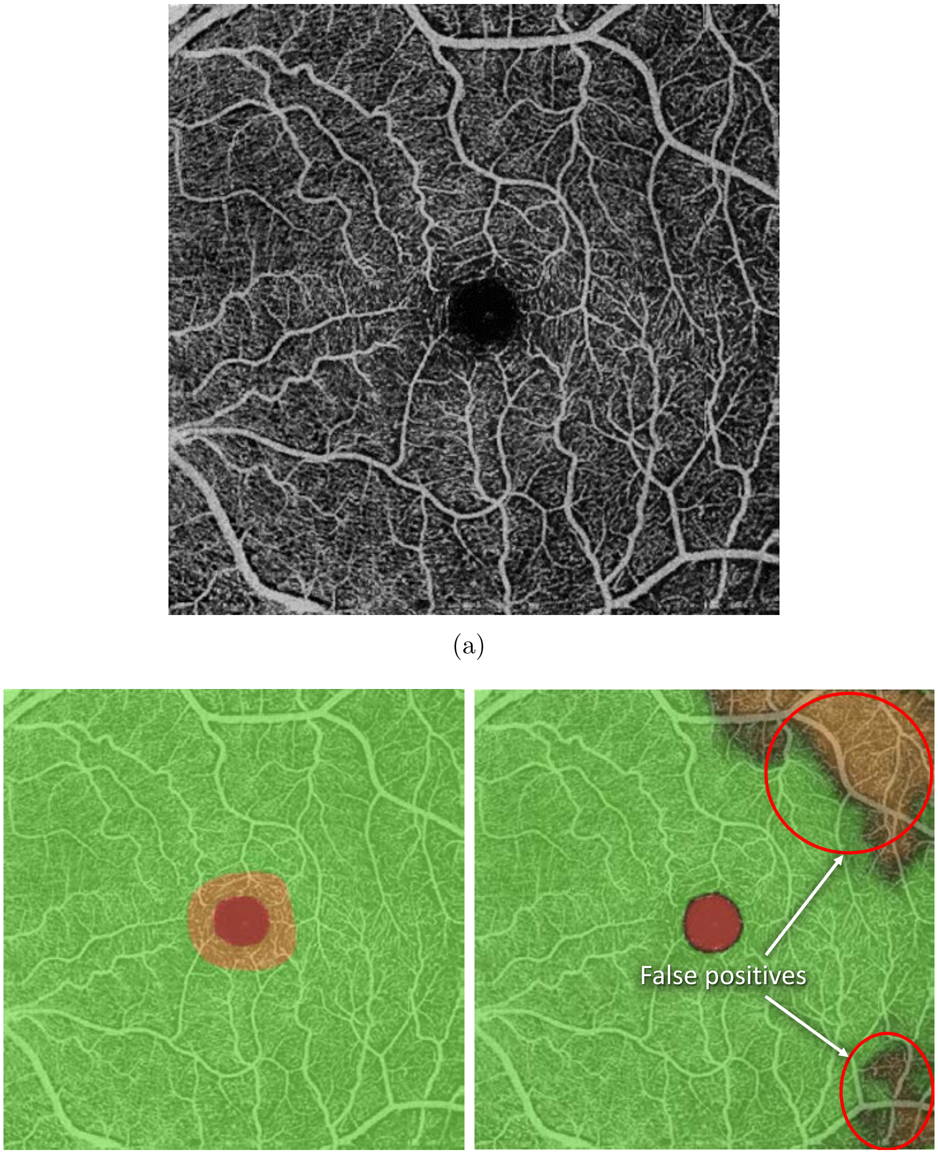
An approach to directly find regions of different microvascular states is proposed in this chapter. This gives a more promising result compared to the last chapter. Moreover, less postprocessing will be needed to get the regions of abnormals. For instance, enlargement of the foveal avascular zone is an early indicator of RR, and with microvascular network directly segmented using the approach from the prior chapter, the delineation of the foveal avascular zone is still needed. On the other hand, the region-based approach can directly generate the area and the shape for a clinician's further analysis.

For the U-Net structure used in this study, we also considered adding residual blocks by He et al. [109]. Although some of the segmentation approaches in medical imaging [86–88] adopted residual blocks into U-Net, such a structure is normally used to train very deep networks (ResNet34, ResNet101, etc.) and is effective for avoiding severe overfitting. In cases of U-Net, it is first not as deep as networks for image classifications. It is also not easy to overfit, which would limit generality. Moreover, none of the approaches mentioned above specifically compared the performance of similar structures with and without residual blocks. As a result, we didn't deploy residual layers as part of the network we develop.

According to the ROC curves in Fig. 6.6, the simultaneous segmentations gives the best prediction for avascular regions and the worst prediction for segmenting capillary-dropped regions. By qualitatively comparing predictions with manual tracings, we find that the segmentation results tend to have more details and predicted regions make better sense than our manual tracings in some cases. Also the confidence level for these predictions (shown as lower intensity in Fig. 6.5(c) and Fig. 6.7(c)) are lower than the prediction of the other two regions for the most part. This is because the tracing is more subjective than the tracing of avascular areas and sometimes hard to distinguish capillary-dropped regions. We think the main bottleneck is obtaining a more formal and consistent definition for the three region classes and tracing them more accurately. A better tracing could help the network get better results.

On the other hand, since there is still a limited number of training images, some cases may not be properly trained. Fig. 6.7 shows an example of such cases. The probability prediction map shows a very different result than manual tracing, especially for the capillary-dropped regions. It may be because that the image is dimmer than most other images from the study.

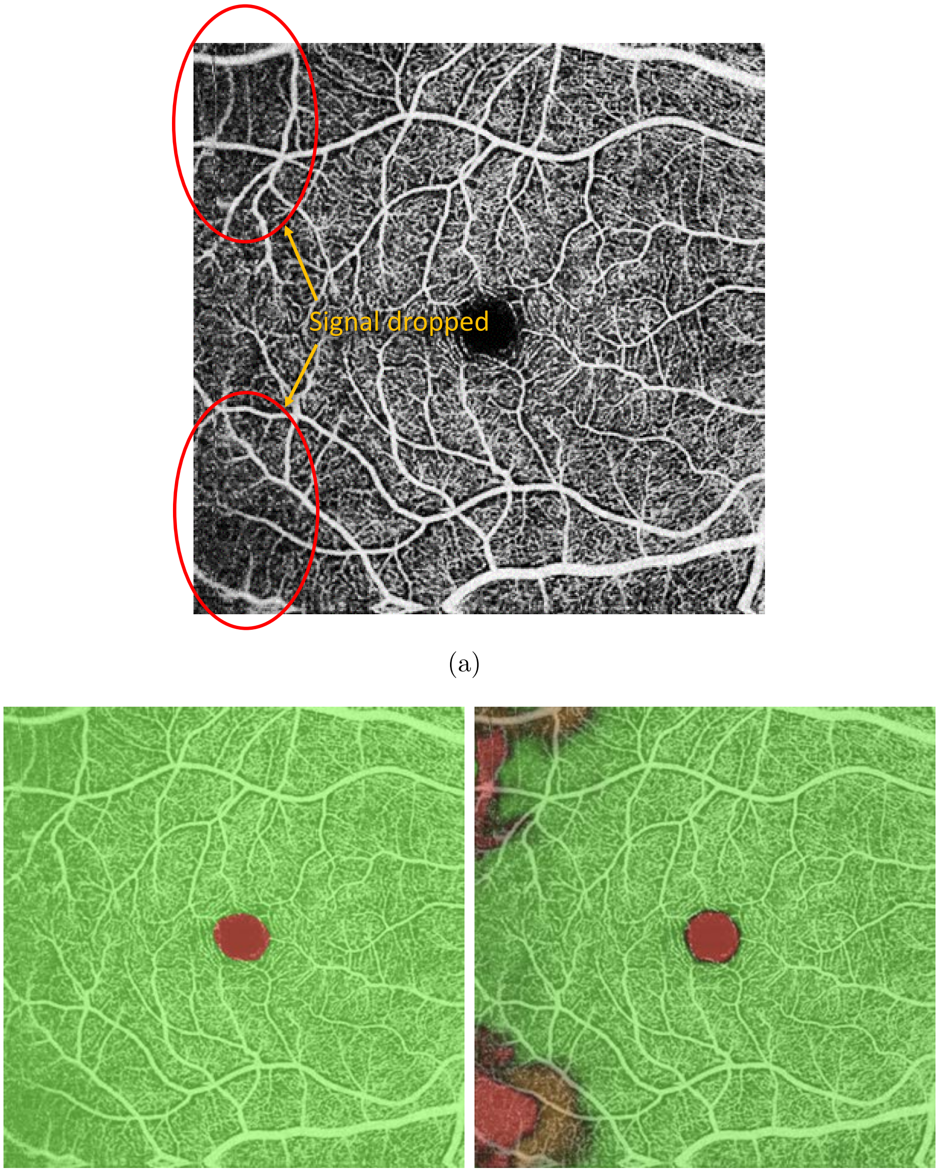
When examining the *en-face* images from healthy control subjects, we also notice sometimes there are regions with deteriorated signal and quality loss. This can be hard to distinguish from capillary dropout and sometimes avascular regions on a local level. An example can be seen as Fig. 6.8. For this specific case, it is hard even for human experts to classify the *en-face* image as normal or diseased. If we can take these phenomena into account, a better approach to automatically analyze these data may be developed.



(a)

(b) (c)

Figure 6.7: An example of inaccurate segmentation of capillary-dropped region. (a) The OCTA *en-face* images from a healthy control subject. (b) The manual tracing of capillary-dense regions in green, capillary-dropped regions in yellow, and avascular regions in red. (c) The corresponding segmentation from the automated approach showing false positives on the corners.



(a)

(b) (c)

Figure 6.8: An illustration of local signal drop. (a) shows an OCTA *en-face* image from a control subject. (b) and (c) shows the manual tracing and segmentation result.

In conclusion, the region-based segmentation method developed in this section gives very accurate predictions for the avascular region in RR study and slightly worse results in predicting capillary-dropped regions. Compared to the approaches we developed in the last chapter, it not only has better performance, but the predictions generated by this approach can potentially be directly used as indications of RR disease. We will explore the relationship between vascular regions and the RR disease in the next chapter.

CHAPTER 7
REGION ANALYSIS FOR OCTA AND STUDY OF DIFFERENCES BETWEEN AFFECTED AND UNAFFECTED EYES (AIM 3)

7.1 Introduction

Radiation retinopathy (RR) is a condition that usually occurs months or years after radiation therapy for the eye or the orbit of the eye. Without proper treatment, it could cause irreversible visual impairment [38,40]. Hence, an accurate measurement of the severity of RR can help better enable early intervention and assess the effectiveness of the treatments.

Traditionally, radiation retinopathy is graded with Finger [43] tests using fundus photographs and Horgan [42] grading using SD-OCT scans. Although OCTA modality could give an earlier diagnosis of the disease [24], recent automated quantification studies on RR only show statistical significance of OCTA analysis in affected and unaffected eyes or the diseased versus healthy subjects [7, 10]. For instance, Zahid et al. [7] uses fractal dimensional analysis to automatically analyze vascular density in retinal OCTA *en-face* images in diabetic retinopathy patients. In this work, only statistical analyses between patients and control subjects are studied, and the results show statistical significance in terms of fractal dimensions (FD). A better quantitative analysis with comparison with expert knowledge is still needed to help understand the usefulness of the modality.

A direct segmentation of regions with different vascular states is usually not easy with traditional automated methods. This is because the quality of the scans changes from scan to scan and it could be non-trivial to find consistent patterns for vesselness. With the help of deep-learning methods, we are able to identify such regions with various complex local patterns more easily. In the previous chapter, we developed an approach to directly measure the regional changes in microvasculature in human OCTA macular *en-face* scans.

Because we only evaluated the approach from Chapter 6 on its ability to predict the regions with different microvascular states, investigations are still needed for the correlations between microvascular loss and RR. In this study, we evaluate the relationships between avasculature area and RR disease based on the algorithm developed in Chapter 6.

The objective of this chapter is to first analyze the effectiveness of regional vascular measurements as an indication of RR severity by comparing it to severity rankings of all the subjects from clinical experts. Then, we study the differences of the microvascular network in the affected and unaffected eye of the RR patients.

7.2 Comparisons between vascular region tracing and automated detection with expert ranking

7.2.1 Methods

The dataset used for this study (previously described in Section 5.6) consists of *en-face* images of superficial layers centered at the macula. There are 86 OS and OD images from 43 subjects acquired with AngioVue or AngioVueHD (Optovue, Inc., Fremont, CA) scanner. The subjects include 11 control cases and 32 uveal melanoma patients with normal or abnormal OCTAs. All the images used in the experiment have unified 480×480 pixel resolutions covering regions of $6\,\mathrm{mm} \times 6\,\mathrm{mm}$ centered at the macula.

The hypothesis we are testing in this study is that there is a strong correlation between the size of microvascular loss regions and RR severities. Hence the manual tracing, as well as automated predictions of avasculature and capillary dropout, are used for the study. The automated predictions are obtained using 10-fold cross-validation so all the images are used in this study. Manual tracings and automated approach include the mappings of avascular, capillary dropped, and capillary dense regions in each of the 86 images. To calculate the size of the microvascular loss regions, the automated probability prediction maps are thresholded to binary predictions with

0.5 as thresholds. The areas of avasculature and capillary dropout are then directly calculated with the binary images from the automated approach as well as the manual tracings. These areas are used as the indicators of RR severities from our region-based analysis.

In addition, the OS and OD pairs of all subjects in the study are classified by two clinical experts into normal eyes, where there is no perceivable anomaly, and abnormals with different level of symptoms related to RR. Furthermore, the experts also rank both normal and abnormals. For the normals, the rankings are mainly based on the size of the foveal avascular zone. And for abnormals, the rankings are based on the diagnosed severity. The images used for ranking are *en-face* views of superficial layers directly generated from AngioVue.

There are two analyses done in this part. First, we measure the consistency between results from our approach and the severity rankings from one expert (expert 1). The manual tracing, the automated predictions for avascular area and capillary dropped area are compared to expert 1's ranking. We also compare an alternative ranking from another expert (expert 2) to expert 1. These rankings are not linearly correlated with the avascular and capillary dropped areas. Hence, we use Spearman's ranking correlation as the metric to compare the different measurements. Second, we also compare the avascular area in the eyes of abnormals and normals based on expert 1's diagnosis. One-way analysis of variance (ANOVA) is applied to quantify the comparison.

7.2.2 Results

The comparisons of the microvasculature areas and expert's ranking are shown in Table 7.1. Specifically, the two columns are the correlation coefficients between expert's ranking and areas from manual tracing, as well as areas from automated segmentations, respectively. And the rows are for the correlations using avascular and capillary-dropout areas as indicators in abnormal, normal, and all cases.

Table 7.1: The Spearman's rank correlations between the areas of manual tracings and expert 1 ranking of severity results; and between automated segmentation results and expert 1.

		Man. vs. Exp.1	Algo. vs. Exp.1
Avasc.	Normal	0.82	0.73
	Abnormal	0.71	0.76
	All	0.70	0.71
Dropped	Normal	$-0.07^{\dagger}$	$-0.05^{\ddagger}$
	Abnormal	0.49	0.53
	All	0.53	0.58

† Doesn't show significant correlation with p-value $= 0.73$.
‡ Doesn't show significant correlation with p-value $= 0.6$.

We see a high correlation between the expert's ranking to the size of the predicted avascular area with $r = 0.76$ in abnormal cases. It shows that the avascular region size can be potentially used as an indication of the severity of RR disease. Capillary dropped area also plays a minor role in the analysis since it shows $r = 0.49$ and $r = 0.53$ correlations for manual tracing and automated detection, respectively. However, the capillary dropout regions have weaker relationships to the severity of RR than avascular area. For this reason, we will only explore the avascular area in next section. The automated approach also shows slightly higher correlations than manual tracings in both avascular and capillary dropout regions. This is qualitatively consistent with our observation that the prediction maps are slightly better than the tracing.

For the images determined as normal, experts rank the images based on the size of the avascular area (specifically the size of the foveal avascular zone). As a result, the size of the avascular region is highly correlated to expert analysis. This also means that the area measurements of capillary dropout regions are not related to the ranking. A non-significant correlation (p-value > 0.05) between the capillary dropped area with the ranking is consistent with this theory.

Table 7.2: The Spearman's rank correlations between expert 1 and expert 2 severity rankings. The labels of normal and abnormal eyes are according to expert 1 classification based on OCTA *en-face* images.

	Correlation
Normal	0.84
Abnormal	0.84
All	0.86

On the other hand, as shown in Table 7.2, the correlation coefficients between two experts are higher than both manual tracing and automated segmentation of avascular regions. There are a few possible reasons for the lower correlation based on region-based analyses. One main reason is that we currently don't incorporate both avascular and capillary dropped areas to form a single measurement. Also, the shapes and locations of vascular changes could play a part in human decisions. Another reason could be portions of manual tracings are inaccurate or inconsistent, especially for the partial capillary dropped regions. Future improved manual tracing and better algorithms, as well as methods to incorporate both avascular regions and capillary dropped regions, could potentially improve the approaches to achieve results closer to the human level.

Fig. 7.1 and Table 7.3 show the size and distribution of the avascular regions. With ANOVA test, the sizes of avascular regions in abnormal and normal cases are significantly different with p-value< 0.05 ($p = 4.3e - 6$). This result also shows that the area of avasculature can act as an indication of RR disease.

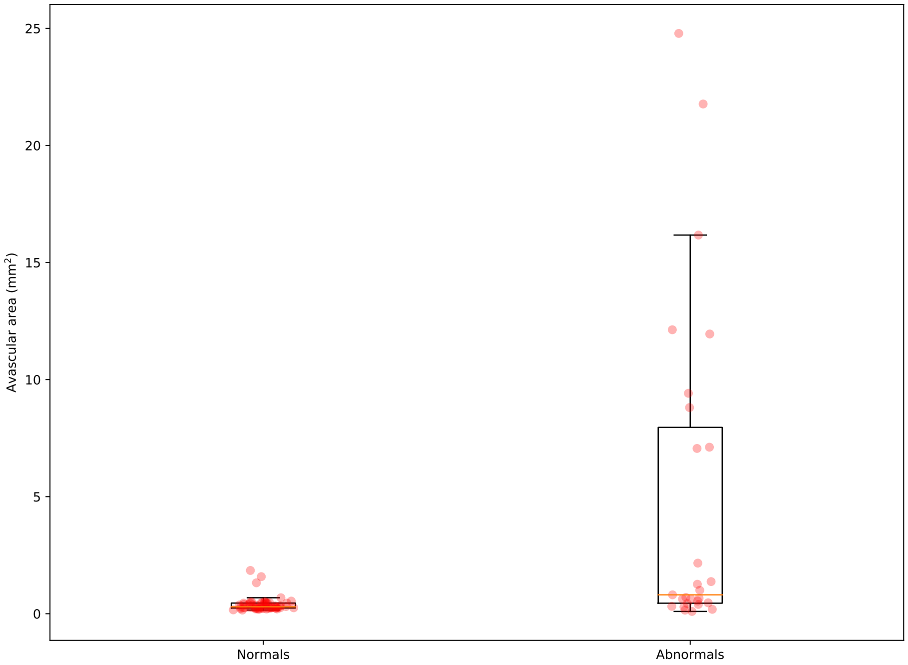


Figure 7.1: A comparison of avascular area between expert's diagnosis of normals and abnormals.

Table 7.3: Sizes of avascular area of expert's diagnosis of normal and abnormal (Mean ± SD).

	Normal	Abnormal
Avasc. Area	$0.39 \pm 0.3\,\mathrm{mm}^2$	$4.86 \pm 6.88\,\mathrm{mm}^2$

7.3 Analysis of the differences between affected and unaffected eyes

7.3.1 Methods

For this study, the same dataset is divided based on healthy control subjects and uveal melanoma patients. The patient OCTA images are further divided into two classes. If both eyes are diagnosed by expert 1 as normal, then the patient is classified as a patient with normal OCTA. These patient are either uveal melanoma patients before radiation treatments or post-treatment patients with no apparent capillary loss. If either eye or both eyes are diagnosed as abnormal, the patient is a patient with abnormal OCTA.

The goal of this section is to explore the differences between the irradiated eye versus the unaffected eye in terms of the avascular area differences. To achieve this goal, the absolute difference between the area (in mm^2) of avascular regions are calculated for each of the 86 subjects. These differences for healthy subjects, uveal melanoma patients, patients with normal OCTA *en-face* images, and patients with abnormal OCTAs are plotted and compared. We expect significant differences between the patients with abnormal OCTAs and the control subjects. We also expect to not obtain a statistically significant difference between controls and patients normal OCTAs.

7.3.2 Results

Fig. 7.2 shows differences of the avascular area in healthy subjects, all uveal melanoma patients, patients with abnormal OCTA *en-face* images, and patients with normal OCTAs. Visually, it shows a very similar distribution between healthy subjects and patients with normal OCTAs. And for healthy subjects and patients with abnormal OCTAs, the range of data and distributions are very different. The differences for each subject is also plotted in the chart, shown as red dots. We can see most patients with abnormal OCTAs show higher differences between left and right eyes in terms of the area of avascular regions. However, it also shows that for a small fraction

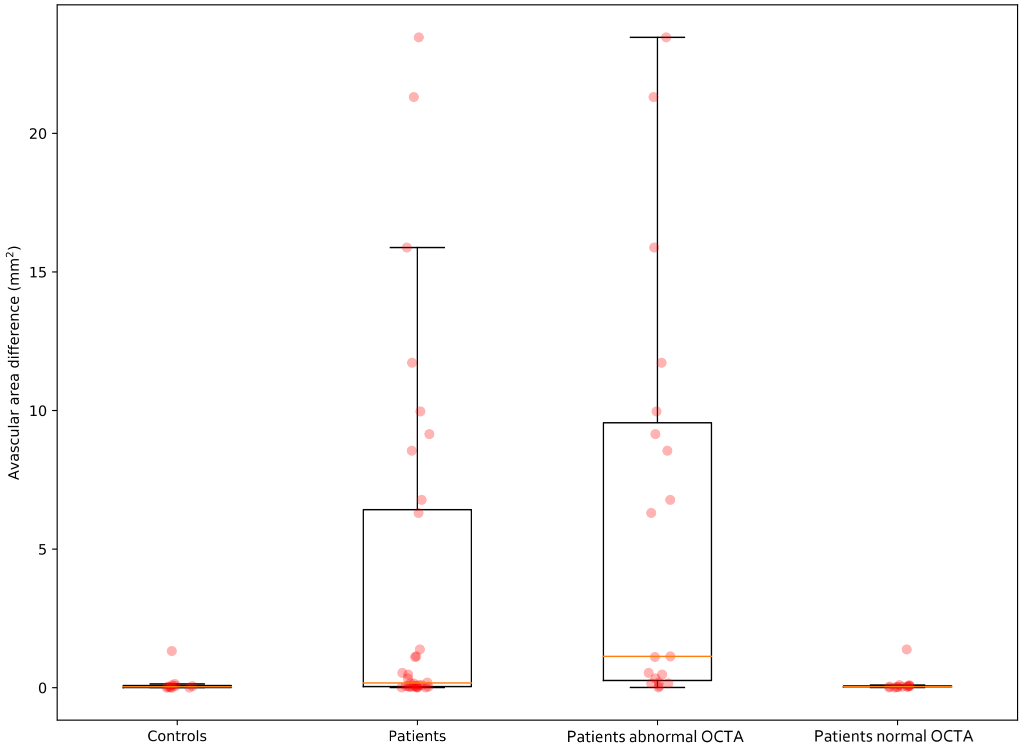


Figure 7.2: Area differences of avascular size in controls, patients, and patients with normal/abnormal OCTA *en-face* according to expert 1 analysis.

of the patients there are only very small differences for the avascular areas similar to normal cases. For these subjects, the main contributor to the diagnosis is the capillary dropped area and further studies are needed to make a better classification.

Table 7.4 confirms the observation with mean and standard deviations. Patients with abnormal OCTAs have a mean avascular area difference of $6.17 \pm 7.31\,\text{mm}^2$ compared to $0.15 \pm 0.35\,\text{mm}^2$ for the patients with normal OCTAs. It also shows that the distribution for patients with normal OCTAs is very similar to healthy subjects with $0.16 \pm 0.4\,\text{mm}^2$ differences. With ANOVA tests, the null hypothesis cannot be rejected with p-value $= 0.08$ between healthy subjects and all patients. After excluding the patients with normal OCTAs, the p-value reduced to 0.01 and smaller than 0.05. The tests also show that the p-value between healthy subjects and RR

Table 7.4: Avascular area differences between two eyes of controls, patients, and patients with normal/abnormal OCTA *en-face* images (Mean ± SD).

	Number of Subjects	Avasc. Area Diff.
Controls	11	$0.16 \pm 0.4\,\mathrm{mm}^2$
Patients	32	$3.72 \pm 6.37\,\mathrm{mm}^2$
w. abnormal OCTA	19	$6.17 \pm 7.31\,\mathrm{mm}^2$
w. normal OCTA	13	$0.15 \pm 0.35\,\mathrm{mm}^2$

patients with normal OCTAs is 0.93 and was consistent with our expectations.

Overall, by analyzing the differences of the avascular area in healthy subjects and uveal melanoma patients, the results show statistical significance between healthy subjects and patients with abnormal OCTAs. This is consistent with our hypothesis.

7.4 Discussion, limitations & conclusions

In this chapter, in order to investigate the method developed in Chapter 6 and its ability to predict the severity of RR disease, we compared both the manual and automated regional vascular measurements to the severity ranking from clinicians. The automated segmented avascular areas show high correlations and sometimes better than manual to expert rankings. We also compare the OS and OD of healthy subjects and uveal melanoma patients in terms of avascular region differences. The results show that when we exclude patients with normal OCTAs diagnosed by an expert, the difference of avascular area shows statistical significance. It confirms the effectiveness of using the avascular area as an indicator of RR disease.

This study also shows some limitations compared to the automated analysis. Comparing to experts' analyses, our approach shows less correlation with expert's diagnosis than the correlation between two experts. We think the main reason for this is that area of avascular is not the only factor for clinicians' diagnosis and capillary dropped regions also at least play a part in human decisions. By improving manual tracing and incorporating both avascular regions and capillary dropped regions, the automated analysis can be improved closer to the human level in future work.

In summary, this pilot study verifies the potential effectiveness of using retinal avascular areas to predict radiation retinopathy. It shows encouraging results for using an automated approach to potentially be used to help analysis and diagnosis of this disease. Although the result of the automated analysis is worse than human experts for diagnosis of radiation retinopathy, it still shows high performance for predicting sickness and can be used as a baseline for future improvements.

CHAPTER 8
CONCLUSIONS

To summarize, this PhD work explored approaches for analyzing retinal microvasculature in mice and humans. By using deep learning, our work shows that many previously hard-to-segment tasks can be directly achieved.

As a review, the aims achieved in this thesis are mainly three-fold, as follows:

- **Aim 1:** Here we develop approaches measuring different retinal structures in mouse FA images. We first develop a model-based approach to roughly segment main vessels and the microvascular network. Then the results are interactively corrected to train a deep learning based approach to segment the optic nerve head (ONH), main vessels, and the microvascular network simultaneously.

- **Aim 2:** We also develop deep-learning based approaches to measure the microvasculature in human optical coherence tomography angiography (OCTA) data. The retinal surfaces in corresponding structural OCT volumes are first segmented. Then, three automated approaches are developed and evaluated for directly segmenting microvasculature in OCTA *en-face* images. Finally, we also develop and evaluate an alternative approach to simultaneously find avascular regions, regions with capillary dropout, and capillary-dense regions.

- **Aim 3:** At last, we analyze the regions segmented in Aim 2 and compare them to experts' analysis of the severity of radiation retinopathy. We also compare the differences of affected and unaffected eyes in radiation retinopathy patients and normal subjects as a clinical related application.

In Chapter 4, we developed an automated approach with semi-automated and deep-learning components to accurately segment retinal structures including ONH region, main vessels, and the microvascular network in mouse FA images (Aim 1). To achieve this goal, a semi-automated approach was first developed to get the retinal structures. We then used the results from the semi-automated part to develop a deep-learning based approach to accurately segment these structures. The results show very high AUCs (> 0.95) for segmenting each of the structures. Also, by generating training data interactively, our approach minimizes the efforts to manually trace the data.

Chapter 5 further explored the microvascular segmentation in the recently developed human OCTA modality. We developed and compared three approaches including two baseline approaches and a deep-learning approach for predicting microvascular network. The results show that with an AUC of 0.95, the deep-learning based approach performs significantly better than the baseline and Hessian-based approaches. It also shows that the baseline approach performs poorly on distinguishing avascular regions from regions with dense microvasculature ($\text{AUC} = 0.77$). To our knowledge, this is also the first study to compare multiple automated approaches and their effectiveness to predict microvasculature in RR patients.

On the other hand, because of the unstable qualities in the images in the data we use, direct segmentation approaches shows limitations finding the precise microvascular network in some of the OCTA *en-face* images. An alternative region-based segmentation method was proposed in Chapter 6. Instead of finding the vessels, regions with different microvascular status were directly segmented. More specifically, avascular regions, regions with capillary dropout, and capillary-dense regions were simultaneously segmented with a deep network. Thus, we avoided segmenting indistinguishable microvasculature in some very low SNR images. Moreover, less postprocessing is needed to obtain abnormal regions. For instance, enlargement of the foveal avascular zone is an early indicator of retinal disease, and the fovea avascular region can directly be assessed with the region-based segmentation results. Our results also show very accurate predictions for avascular regions (AUC of 0.97) in RR study and slightly worse in predicting capillary dropout regions (AUC of 0.93).

Finally, in Chapter 7 (Aim 3), we investigated and analyzed the clinical significance of our region-based OCTA *en-face* segmentations. In order to achieve this goal, we first analyzed the effectiveness of regional vascular measurements as an indication of RR severity. By comparing automated measurements to severity rankings from clinical experts, predicting the area of avasculature shows a high correlation of 0.76 compares to an expert's ranking. Although the result of the automated analysis is moderately worse than the correlation between two experts, this analysis still showed the potential effectiveness of using retinal avascular areas to predict the severity of radiation retinopathy disease. Additionally, the microvascular differences between affected and unaffected eyes of the radiation retinopathy patients were also studied and compared to healthy subjects. The result showed statistical significance (p-value < 0.05) comparing the differences of RR patients and healthy controls.

To summarize, this thesis mainly focused on segmentation and analysis of microvascular network in mouse and human retinal images. Our studies showed that the usage of deep-learning networks provides accurate results for the analysis of retinal vasculature. We think our developed approaches will further help study and diagnose retinal diseases in mice and humans.

REFERENCES

[1] A. Gupta, F. Dhawahir-Scala, A. Smith, L. Young, and S. Charles, “Radiation retinopathy: case report and review,” *BMC Ophthalmology*, vol. 7, no. 1, p. 6, 2007.

[2] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[3] J. Flammer, S. Orgül, V. P. Costa, N. Orzalesi, G. K. Krieglstein, L. M. Serra, J.-P. Renard, and E. Stefánsson, “The impact of ocular blood flow in glaucoma,” *Progress in Retinal and Eye Research*, vol. 21, no. 4, pp. 359–393, 2002.

[4] Y. Jia, E. Wei, X. Wang, X. Zhang, J. C. Morrison, M. Parikh, L. H. Lombardi, D. M. Gattey, R. L. Armour, B. Edmunds *et al.*, “Optical coherence tomography angiography of optic disc perfusion in glaucoma,” *Ophthalmology*, vol. 121, no. 7, pp. 1322–1332, 2014.

[5] Y. Jia, S. T. Bailey, D. J. Wilson, O. Tan, M. L. Klein, C. J. Flaxel, B. Potsaid, J. J. Liu, C. D. Lu, M. F. Kraus *et al.*, “Quantitative optical coherence tomography angiography of choroidal neovascularization in age-related macular degeneration,” *Ophthalmology*, vol. 121, no. 7, pp. 1435–1444, 2014.

[6] M. D. Abràmoff, M. K. Garvin, and M. Sonka, “Retinal imaging and image analysis,” *IEEE Rev. Biomed. Eng.*, vol. 3, pp. 169–208, 2010.

[7] S. Zahid, R. Dolz-Marco, K. B. Freund, C. Balaratnasingam, K. Dansingani, F. Gilani, N. Mehta, E. Young, M. R. Klifto, B. Chae *et al.*, “Fractal dimensional analysis of optical coherence tomography angiography in eyes with diabetic retinopathy,” *Investigative Ophthalmology & Visual Science*, vol. 57, no. 11, pp. 4940–4947, 2016.

[8] Q. Chen, Q. Ma, C. Wu, F. Tan, F. Chen, Q. Wu, R. Zhou, X. Zhuang, F. Lu, J. Qu *et al.*, “Macular vascular fractal dimension in the deep capillary layer as an early indicator of microvascular loss for retinopathy in type 2 diabetic patients,” *Investigative Ophthalmology & Visual Science*, vol. 58, no. 9, pp. 3785–3794, 2017.

[9] C. L. Shields, E. A. T. Say, W. A. Samara, C. T. Khoo, A. Mashayekhi, and J. A. Shields, “Optical coherence tomography angiography of the macula after plaque radiotherapy of choroidal melanoma: comparison of irradiated versus nonirradiated eyes in 65 patients,” *Retina*, vol. 36, no. 8, pp. 1493–1505, 2016.

[10] A. Matet, A. Daruich, and L. Zografos, "Radiation maculopathy after proton beam therapy for uveal melanoma: optical coherence tomography angiography alterations influencing visual acuity," *Investigative Ophthalmology & Visual Science*, vol. 58, no. 10, pp. 3851–3861, 2017.

[11] H. R. Novotny and D. L. Alvis, "A method of photographing fluorescence in circulating blood in the human retina," *Circulation*, vol. 24, no. 1, pp. 82–86, 1961.

[12] Y. Jia, O. Tan, J. Tokayer, B. Potsaid, Y. Wang, J. J. Liu, M. F. Kraus, H. Subhash, J. G. Fujimoto, J. Hornegger *et al.*, "Split-spectrum amplitude-decorrelation angiography with optical coherence tomography," *Optics Express*, vol. 20, no. 4, pp. 4710–4725, 2012.

[13] T. Spencer, R. P. Phillips, P. F. Sharp, and J. V. Forrester, "Automated detection and quantification of microaneurysms in fluorescein angiograms," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 230, no. 1, pp. 36–41, 1992.

[14] M. B. Vickerman, P. A. Keith, T. L. McKay, D. J. Gedeon, M. Watanabe, M. Montano, G. Karunamuni, P. K. Kaiser, J. E. Sears, Q. Ebrahem *et al.*, "VESGEN 2D: Automated, user-interactive software for quantification and mapping of angiogenic and lymphangiogenic trees and networks," *The Anatomical Record*, vol. 292, no. 3, pp. 320–332, 2009.

[15] S. G. Gadde, N. Anegondi, D. Bhanushali, L. Chidambara, N. K. Yadav, A. Khurana, and A. S. Roy, "Quantification of vessel density in retinal optical coherence tomography angiography images using local fractal dimension-vessel density in OCTA images," *Investigative Ophthalmology & Visual Science*, vol. 57, no. 1, pp. 246–252, 2016.

[16] P. Prentašić, M. Heisler, Z. Mammo, S. Lee, A. Merkur, E. Navajas, M. F. Beg, M. Šarunić, and S. Lončarić, "Segmentation of the foveal microvasculature using deep learning networks," *Journal of Biomedical Optics*, vol. 21, no. 7, pp. 075 008–075 008, 2016.

[17] R. Klein, B. E. Klein, and S. E. Moss, "How many steps of progression of diabetic retinopathy are meaningful?: The wisconsin epidemiologic study of diabetic retinopathy," *Archives of Ophthalmology*, vol. 119, no. 4, pp. 547–553, 2001.

[18] L. D. Hubbard, R. J. Brothers, W. N. King, L. X. Clegg, R. Klein, L. S. Cooper, A. R. Sharrett, M. D. Davis, and J. Cai, “Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the atherosclerosis risk in communities study1,” *Ophthalmology*, vol. 106, no. 12, pp. 2269–2280, 1999.

[19] T. Y. Wong, A. Shankar, R. Klein, B. E. Klein, and L. D. Hubbard, “Prospective cohort study of retinal vessel diameters and risk of hypertension,” *BMJ*, vol. 329, no. 7457, p. 79, 2004.

[20] J. Parr and G. Spears, “Mathematic relationships between the width of a retinal artery and the widths of its branches,” *American Journal of Ophthalmology*, vol. 77, no. 4, pp. 478–483, 1974.

[21] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M. D. Abramoff, “Comparative study of retinal vessel segmentation methods on a new publicly available database,” in *Medical Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 648–656.

[22] M. Melinščak, P. Prentašić, and S. Lončarić, “Retinal vessel segmentation using deep neural networks,” in *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications)*, 2015.

[23] M. Sonka and M. D. Abràmoff, “Quantitative analysis of retinal OCT,” 2016.

[24] K. K. Veverka, J. E. AbouChehade, R. Iezzi Jr, and J. S. Pulido, “Noninvasive grading of radiation retinopathy: the use of optical coherence tomography angiography,” *Retina*, vol. 35, no. 11, pp. 2400–2410, 2015.

[25] L. Liu, S. S. Gao, S. T. Bailey, D. Huang, D. Li, and Y. Jia, “Automated choroidal neovascularization detection algorithm for optical coherence tomography angiography,” *Biomedical Optics Express*, vol. 6, no. 9, pp. 3564–3576, 2015.

[26] H. Giannakaki-Zimmermann, D. Kokona, S. Wolf, A. Ebneter, and M. S. Zinkernagel, “Optical coherence tomography angiography in mice: Comparison with confocal scanning laser microscopy and fluorescein angiography,” *Translational Vision Science & Technology*, vol. 5, no. 4, pp. 11–11, 2016.

[27] R. F. Spaide, J. M. Klancnik, and M. J. Cooney, “Retinal vascular layers imaged by fluorescein angiography and optical coherence tomography angiography,” *JAMA Ophthalmology*, vol. 133, no. 1, pp. 45–50, 2015.

[28] J. J. Chen, J. E. AbouChehade, R. Iezzi Jr, J. A. Leavitt, and R. H. Kardon, "Optical coherence angiographic demonstration of retinal changes from chronic optic neuropathies," *Neuro-Ophthalmology*, pp. 1–8, 2017.

[29] S. Wang, M. P. Villegas-Pérez, M. Vidal-Sanz, and R. D. Lund, "Progressive optic axon dystrophy and vascular changes in rd mice," *Investigative Ophthalmology & Visual Science*, vol. 41, no. 2, pp. 537–545, 2000.

[30] M. Paques, R. Tadayoni, R. Sercombe, P. Laurent, O. Genevois, A. Gaudric, and E. Vicaut, "Structural and hemodynamic analysis of the mouse retinal microcirculation," *Investigative Ophthalmology & Visual Science*, vol. 44, no. 11, pp. 4960–4967, 2003.

[31] C. for Disease Control, Prevention *et al.*, "National diabetes statistics report: estimates of diabetes and its burden in the United States, 2014," *Atlanta, GA: US Department of Health and Human Services*, vol. 2014, 2014.

[32] R. Klein, B. E. Klein, S. E. Moss, M. D. Davis, and D. L. DeMets, "The Wisconsin epidemiologic study of diabetic retinopathy: II. prevalence and risk of diabetic retinopathy when age at diagnosis is less than 30 years," *Archives of Ophthalmology*, vol. 102, no. 4, pp. 520–526, 1984.

[33] R. L. Engerman, "Pathogenesis of diabetic retinopathy," *Diabetes*, vol. 38, no. 10, pp. 1203–1206, 1989.

[34] D. C. Klonoff and D. M. Schwartz, "An economic analysis of interventions for diabetes." *Diabetes Care*, vol. 23, no. 3, pp. 390–404, 2000.

[35] C. Biallosterski, M. E. Van Velthoven, R. P. Michels, R. O. Schlingemann, J. H. DeVries, and F. D. Verbraak, "Decreased optical coherence tomography-measured pericentral retinal thickness in patients with diabetes mellitus type 1 with minimal diabetic retinopathy," *British Journal of Ophthalmology*, vol. 91, no. 9, pp. 1135–1138, 2007.

[36] T. Oshitari, K. Hanawa, and E. Adachi-Usami, "Changes of macular and RNFL thicknesses measured by Stratus OCT in patients with early stage diabetes," *Eye*, vol. 23, no. 4, p. 884, 2009.

[37] H. Stallard, "Radiant energy as (a) a pathogenetic and (b) a therapeutic agent in ophthalmic disorders," *British Journal of Ophthalmology*, vol. 6, pp. 67–79, 1933.

[38] D. Archer, W. Amoaku, and T. Gardiner, “Radiation retinopathyclinical, histopathological, ultrastructural and experimental correlations,” *Eye*, vol. 5, no. 2, pp. 239–251, 1991.

[39] D. Archerz and T. Gardiner, “Ionizing radiation and the retina,” *Current Opinion in Ophthalmology*, vol. 5, no. 3, pp. 59–65, 1994.

[40] K. Gündüz, C. L. Shields, J. A. Shields, J. Cater, J. E. Freire, and L. W. Brady, “Radiation retinopathy following plaque radiotherapy for posterior uveal melanoma,” *Archives of Ophthalmology*, vol. 117, no. 5, pp. 609–614, 1999.

[41] C. L. Shields, J. A. Shields, J. Cater, K. Gündüz, C. Miyamoto, B. Micaily, and L. W. Brady, “Plaque radiotherapy for uveal melanoma: long-term visual outcome in 1106 consecutive patients,” *Archives of Ophthalmology*, vol. 118, no. 9, pp. 1219–1228, 2000.

[42] N. Horgan, C. L. Shields, A. Mashayekhi, L. F. Teixeira, M. A. Materin, and J. A. Shields, “Early macular morphological changes following plaque radiotherapy for uveal melanoma,” *Retina*, vol. 28, no. 2, pp. 263–273, 2008.

[43] P. Finger and M. Kurli, “Laser photocoagulation for radiation retinopathy after ophthalmic plaque radiation therapy,” *British Journal of Ophthalmology*, vol. 89, no. 6, pp. 730–738, 2005.

[44] M. E. Martínez-Pérez, A. D. Hughes, A. V. Stanton, S. A. Thom, A. A. Bharath, and K. H. Parker, “Retinal blood vessel segmentation by means of scale-space analysis and region growing,” in *Medical Image Computing and Computer-Assisted Intervention–MICCAI1999*. Springer, 1999, pp. 90–97.

[45] X. Jiang and D. Mojon, “Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 1, pp. 131–137, 2003.

[46] F. Zana and J.-C. Klein, “Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation,” *Image Processing, IEEE Transactions on*, vol. 10, no. 7, pp. 1010–1019, 2001.

[47] J. V. Soares, J. J. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, “Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification,” *Medical Imaging, IEEE Transactions on*, vol. 25, no. 9, pp. 1214–1222, 2006.

[48] C. A. Lupascu, D. Tegolo, and E. Trucco, “FABC: retinal vessel segmentation using AdaBoost,” *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 5, pp. 1267–1274, 2010.

[49] C. Sinthanayothin, J. F. Boyce, H. L. Cook, and T. H. Williamson, “Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images,” *British Journal of Ophthalmology*, vol. 83, no. 8, pp. 902–910, 1999.

[50] C. Sinthanayothin, J. Boyce, T. Williamson, H. Cook, E. Mensah, S. Lal, and D. Usher, “Automated detection of diabetic retinopathy on digital fundus images,” *Diabetic Medicine*, vol. 19, no. 2, pp. 105–112, 2002.

[51] E. Ricci and R. Perfetti, “Retinal blood vessel segmentation using line operators and support vector classification,” *Medical Imaging, IEEE Transactions on*, vol. 26, no. 10, pp. 1357–1365, 2007.

[52] P. Liskowski and K. Krawiec, “Segmenting retinal blood vessels with deep neural networks,” *Medical Imaging, IEEE Transactions on*, vol. 35, no. 11, pp. 2369–2380, 2016.

[53] K. Li, X. Wu, D. Z. Chen, and M. Sonka, “Optimal surface segmentation in volumetric images-a graph-theoretic approach,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 119–134, 2006.

[54] M. K. Garvin, M. D. Abràmoff, X. Wu, S. R. Russell, T. L. Burns, and M. Sonka, “Automated 3-d intraretinal layer segmentation of macular spectral-domain optical coherence tomography images,” *Medical Imaging, IEEE Transactions on*, vol. 28, no. 9, pp. 1436–1447, 2009.

[55] X. Xu, M. Niemeijer, Q. Song, M. Sonka, M. K. Garvin, J. M. Reinhardt, and M. D. Abràmoff, “Vessel boundary delineation on fundus images using graph-based approach,” *Medical Imaging, IEEE Transactions on*, vol. 30, no. 6, pp. 1184–1191, 2011.

[56] Q. Song, J. Bai, M. K. Garvin, M. Sonka, J. M. Buatti, and X. Wu, “Optimal multiple surface segmentation with shape and context priors,” *Medical Imaging, IEEE Transactions on*, vol. 32, no. 2, pp. 376–386, 2013.

[57] K. Lee, M. Niemeijer, M. K. Garvin, Y. H. Kwon, M. Sonka, and M. D. Abràmoff, “3-D segmentation of the rim and cup in spectral-domain optical coherence tomography volumes of the optic nerve head,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2009, pp. 72 622D–72 622D.

[58] B. J. Antony, W. Jeong, M. D. Abramoff, J. Vance, E. H. Sohn, and M. K. Garvin, “Automated 3D segmentation of intraretinal surfaces in SD-OCT volumes in normal and diabetic mice,” *Translational Vision Science & Technology*, vol. 3, no. 5, pp. 8–8, 2014.

[59] Y. Yin, X. Zhang, R. Williams, X. Wu, D. Anderson, and M. Sonka, “LOGISMOS: Layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint,” *IEEE Trans. Med. Imag.*, vol. 29, no. 12, pp. 2023–2037, 2010.

[60] Q. Song, J. Bai, M. K. Garvin, M. Sonka, J. M. Buatti, and X. Wu, “Optimal multiple surface segmentation with shape and context priors,” *Medical Imaging, IEEE Transactions on*, vol. 32, no. 2, pp. 376–386, Feb. 2013.

[61] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[62] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[63] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA*, vol. 316, no. 22, pp. 2402–2410, 2016.

[64] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[65] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, “Deep learning with non-medical training used for chest pathology identification,” in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2015, pp. 94 140V–94 140V.

[66] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “CNN features off-the-shelf: an astounding baseline for recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806–813.

[67] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *Medical Imaging, IEEE Transactions on*, vol. 35, no. 5, pp. 1285–1298, 2016.

[68] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *Advances in Neural Information Processing Systems*, 2012, pp. 2843–2851.

[69] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1915–1929, 2013.

[70] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, “Overfeat: Integrated recognition, localization and detection using convolutional networks,” *arXiv preprint arXiv:1312.6229*, 2013.

[71] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[72] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[73] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. DAnastasi *et al.*, “Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 415–423.

[74] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, “Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation,” *Medical Imaging, IEEE Transactions on*, vol. 35, no. 5, pp. 1229–1239, 2016.

[75] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Medical image analysis*, vol. 36, pp. 61–78, 2017.

[76] C. S. Lee, A. J. Tyring, N. P. Deruyter, Y. Wu, A. Rokem, and A. Y. Lee, “Deep-learning based, automated segmentation of macular edema in optical coherence tomography,” *bioRxiv*, p. 135640, 2017.

[77] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, “Multiscale vessel enhancement filtering,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 130–137.

[78] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. ter Haar Romeny, J. B. Zimmerman, and K. Zuiderveld, “Adaptive histogram equalization and its variations,” *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[79] T. Zhang and C. Y. Suen, “A fast parallel algorithm for thinning digital patterns,” *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.

[80] R. O. Duda and P. E. Hart, “Use of the Hough transformation to detect lines and curves in pictures,” *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[81] Q. Song, X. Wu, and Y. Liu, “Simultaneous searching of globally optimal interacting surfaces with shape priors,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2010.

[82] F. L. Bookstein, “Principal warps: Thin-plate splines and the decomposition of deformations,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 11, no. 6, pp. 567–585, 1989.

[83] U. Ramer, “An iterative procedure for the polygonal approximation of plane curves,” *Computer Graphics and Image Processing*, vol. 1, no. 3, pp. 244–256, 1972.

[84] D. H. Douglas and T. K. Peucker, “Algorithms for the reduction of the number of points required to represent a digitized line or its caricature,” *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 10, no. 2, pp. 112–122, 1973.

[85] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.

[86] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *3D Vision (3DV), 2016 Fourth International Conference on.* IEEE, 2016, pp. 565–571.

[87] T. M. Quan, D. G. Hilderbrand, and W.-K. Jeong, “Fusionnet: A deep fully residual convolutional neural network for image segmentation in connectomics,” *arXiv preprint arXiv:1612.05360*, 2016.

[88] Y. Han and J. C. Ye, “Framing u-net via deep convolutional framelets: Application to sparse-view CT,” *arXiv preprint arXiv:1708.08333*, 2017.

[89] Z. Zhang, Q. Liu, and Y. Wang, “Road extraction by deep residual u-net,” *IEEE Geoscience and Remote Sensing Letters*, 2018.

[90] V. Iglovikov and A. Shvets, “Ternausnet: U-net with VGG11 encoder pre-trained on ImageNet for image segmentation,” *arXiv preprint arXiv:1801.05746*, 2018.

[91] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in *International Conference on Machine Learning*, 2015, pp. 448–456.

[92] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026–1034.

[93] Y. Wu and K. He, “Group normalization,” *arXiv preprint arXiv:1803.08494*, 2018.

[94] D. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

[95] X. Guo, D. Dean, S. Denman, C. Fookes, and S. Sridharan, “Evaluating automatic road detection across a large aerial imagery collection,” in *Digital Image Computing Techniques and Applications (DICTA), 2011 International Conference on*. IEEE, 2011, pp. 140–145.

[96] R. Youssef, A. Ricordeau, S. Sevestre-Ghalila, and A. Benazza-Benyahya, “Evaluation protocol of skeletonization applied to grayscale curvilinear structures,” in *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*. IEEE, 2015, pp. 1–6.

[97] A. Camino, M. Zhang, C. Dongye, A. D. Pechauer, T. S. Hwang, S. T. Bailey, B. Lujan, D. J. Wilson, D. Huang, and Y. Jia, “Automated registration and enhanced processing of clinical optical coherence tomography angiography,” *Quantitative Imaging in Medicine and Surgery*, vol. 6, no. 4, p. 391, 2016.

[98] C. Dongye, M. Zhang, T. S. Hwang, J. Wang, S. S. Gao, L. Liu, D. Huang, D. J. Wilson, and Y. Jia, “Automated detection of dilated capillaries on optical coherence tomography angiography,” *Biomedical Optics Express*, vol. 8, no. 2, pp. 1101–1109, 2017.

[99] A. Valverde-Megías, E. A. Say, S. R. Ferenczy, and C. L. Shields, "Differential macular features on optical coherence tomography angiography in eyes with choroidal nevus and melanoma," *Retina*, vol. 37, no. 4, pp. 731–740, 2017.

[100] J. Carpenter and J. Bithell, "Bootstrap confidence intervals: when, which, what? a practical guide for medical statisticians," *Statistics in Medicine*, vol. 19, no. 9, pp. 1141–1164, 2000.

[101] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Müller, "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, p. 77, 2011.

[102] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.

[103] P. Y. Simard, D. Steinkraus, J. C. Platt *et al.*, "Best practices for convolutional neural networks applied to visual document analysis." in *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*, Aug 2003, pp. 958–963.

[104] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2016, pp. 4898–4906.

[105] F. Yu, V. Koltun, and T. Funkhouser, "Dilated residual networks," *arXiv preprint arXiv:1705.09914*, 2017.

[106] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2881–2890.

[107] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 40, no. 4, pp. 834–848, 2018.

[108] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Computer Vision (ICCV), 2017 IEEE International Conference on.* IEEE, 2017, pp. 2980–2988.

[109] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.