UNIVERSITY OF CALIFORNIA

Los Angeles

An Artificial Intelligence Framework for the Automated Segmentation and

Quantitative Analysis of Retinal Vasculature

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Computer Science

by

Ali Hatamizadeh

2020

ABSTRACT OF THE THESIS

An Artificial Intelligence Framework for the Automated Segmentation and
Quantitative Analysis of Retinal Vasculature

by

Ali Hatamizadeh
Master of Science in Computer Science
University of California, Los Angeles, 2020
Professor Demetri Terzopoulos, Chair

The reliable segmentation and quantification of retinal vasculature can provide the means
to diagnose and monitor the progression of a variety of diseases affecting the blood
vessel network, including diabetes and hypertension. We address this problem in depth,
leveraging the power of artificial intelligence to devise automated approaches for the
segmentation and width estimation of vessels in two ophthalmological image modalities.

First, we investigate the automated segmentation of retinal vessels in color fundus
images. We propose a novel, fully convolutional deep neural network with an encoder-
decoder architecture that employs dilated spatial pyramid pooling with multiple dilation
rates to recover the lost content in the encoder and add multiscale contextual information
to the decoder. We also propose a simple yet effective way of quantifying and tracking
the widths of retinal vessels through direct use of the segmentation predictions. Our
methodology takes a whole-image approach and is tested on two publicly available datasets.

Second, we introduce the first deep-learning based method for the semantic segmenta-
tion of retinal arteries and veins in infrared imaging along with a novel dataset dubbed
AVIR, and propose an innovative encoder-decoder that is regularized by variational au-
toencoders. Additionally, our method automatically quantifies the morphological changes
of the segmented arteries and veins, which is important for establishing automated vessel
tracking systems.

The thesis of Ali Hatamizadeh is approved.

Song-Chun Zhu

Hamid Hosseini

Quanquan Gu

Demetri Terzopoulos, Committee Chair

University of California, Los Angeles

2020

This thesis is dedicated to my mother, Elham Pirhadi, whose wisdom, love, and unconditional support made it possible for me to become the person I am today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

CHAPTER 1

Introduction

The retina and its vasculature are directly visible due to the optically clear media of the human eye. As the only part of the central nervous system that can be rapidly and non-invasively imaged with a variety of modalities in the out-patient setting, the retina provides a window into the human body, thus offering the opportunity to assess changes associated with systemic diseases such as hypertension, diabetes, and neurodegenerative disorders. The sequelae of these conditions, specifically stroke, heart disease, and dementia represent major causes of morbidity and mortality in the developed world. To date, all classification schemes for retinal vascular changes in these conditions, particularly in the early stages of disease, have been based on qualitative changes based on human assessment.

We and others hypothesize that biomarkers of seriously adverse health events exist in the quantitative assessment of retinal vasculature changes associated with early, even asymptomatic, diabetes, hypertension, or neurodegenerations. Specifically, high blood pressure, for example, causes structural changes in the macro- and micro-vasculature of vital organs throughout the body, including the brain, heart, and kidney. The retinal vasculature is similarly impacted, but has the advantage of accessibility to multimodal imaging, providing the opportunity to quantitatively assess prognosis, risk, and response to treatment.

For example, the narrowing of retinal vessels has been described as an early, classic sign of hypertension. However, this early sign is difficult to use in everyday clinical practice, which usually includes only non-quantitative, subjective visual assessment of the retina by examination, photograph, or even angiography. An automated, quantitative, reliable,

reproducible tool that measures changes in the retinal vasculature in response to disease and intervention might augment and disrupt current evaluation and treatment paradigms by allowing physicians to detect disease, predict outcomes, and assess interventions much earlier in the course of disease, thereby opening the potential for improved outcomes in major unmet public health needs.

A critical step in tracking important structural changes of the retinal vasculature is segmentation of the retinal vessels, as it enables locating the veins and arteries and extracting relevant information such as a profile of the width changes of the vessels. Since the manual segmentation of vessels by clinicians is a notoriously laborious and error-prone process, it is important to establish fully automated and reliable segmentation methods that can be leveraged to extract the aforementioned information with minimal supervision.

Since the advent of deep learning, Convolutional Neural Networks (CNNs) have been widely used in various computer vision related applications, including object classification, detection, and segmentation, due to their powerful nonlinear feature extraction capabilities as well as their data-driven nature, which enables learning the underlying distribution of a dataset and fitting a model in an automated manner. In medical imaging, CNNs have become ubiquitous for the task of image segmentation and have been successfully leveraged in many 2D and 3D applications such brain, liver, kidney, and general lesion segmentation.

In ophthalmology and for the task of retinal vessel segmentation, researchers have pursued various models that utilize CNNs as a standalone approach or combined with other techniques. In general, there exists two common approaches to the application of CNNs to this task. The first uses a patch-wise sliding strategy in which the image is split into smaller sub-sections and, consequently, the network is trained and tested on the patches via multiple inferences. By contrast, the second is a whole-image approach for both training and inference. Both approaches have their own advantages and disadvantages. The first has the advantage of providing more data for training, but it leads to a global context loss during training and has longer testing times. The second preserves the global content, but suffers from insufficient training data, since the currently available datasets

2

Figure 1.1: Example segmentations of images in the DRIVE (top) and AVIR (bottom) datasets. The DRIVE dataset comprises of binary masks of retinal vessels, whereas the AVIR dataset presents pixel-wise labels for retinal arteries (red) and veins (blue). (a) Input images. (b) Ground truth. (c) Segmentation predictions output by our networks.

in this area consist of at most 40 images, which is very small relative to large-scale deep learning datasets.

1.1   Contributions

In this thesis, we exploit the power of CNNs using both the patch-wise and whole-image approaches to create a reliable, fully automated, method that can accurately detect and segment retinal vessels.[1] Additionally, we explore the problem of automated quantification of retinal vessels by exploring robust methods for measuring vessel width

---

[1]The author has a number of prior publications on the topic of CNN-based deep learning models (Hatamizadeh et al., 2019a,d; Myronenko and Hatamizadeh, 2019, 2020; Hatamizadeh et al., 2020).

profiles from the segmentation masks. Such an effort can be employed toward the creation the aforementioned biomarkers.

In particular, as part of a collaborative effort between the UCLA Samueli School of Engineering Computer Science Department and the UCLA David Geffen School of Medicine Stein Eye Institute, we study different image segmentation models for two important image modalities in ophthalmology.

First, we study the problem of binary retinal vessel segmentation in color fundus images and propose a novel CNN architecture using a whole-image approach. We demonstrate that our network outperforms the state-of-the-art benchmarks on two popular datasets of DRIVE and CHASE-DB1 with very moderate data augmentation and without utilizing any pre-trained backbones. This work was published as (Hatamizadeh et al., 2019b,c).

Second, we study the problem of semantic segmentation of retinal arteries and veins in infrared imaging and propose a novel CNN encoder-decoder model capable of identifying and delineating these vessels. In doing so, we introduce a novel dataset, dubbed AVIR, which contains pixel-wise semantic masks of retinal arteries and veins. We also explore the use of variational auto-encoders as a means of regularizing the main encoder-decoder CNN by learning how to reconstruct the input images. In addition, we study the width profile changes of the segmented veins and arteries. Note that this thesis introduces the first semantic segmentation model capable of separately segmenting retinal arteries and veins.

Figure 1.1 shows examples of the images, ground truth masks, and predictions of the networks devised in this thesis. The top row shows a fundus image from the DRIVE dataset, its corresponding binary ground truth, and our network's prediction. The bottom row shows an image from the AVIR dataset, its corresponding semantic ground truth consisting of arteries and veins, and our network's prediction.

1.2 Overview

The remainder of the thesis is organized as follows:

In Chapter 2, we review the relevant literature in the area of retinal vessel segmentation with an emphasis on deep learning-based methods.

In Chapter 3, we devise a novel CNN architecture for retinal vessel segmentation in color fundus images.

In Chapter 4, we propose SegAVIR, a methodology for the semantic segmentation of retinal arteries and veins, and devise a model that comprises a CNN architecture and variational auto-encoder for regularization.

In Chapter 5, we report our experiments with the models developed in the previous two chapters and benchmark our results.

Chapter 6 presents the conclusions of the thesis as well as an outlook to promising future research directions.

CHAPTER 2

Related Work

In this chapter, we review relevant efforts in the area of retinal vessel segmentation with an emphasis on deep learning-based methods and CNN architectures.

2.1 General Retinal Vessel Segmentation

2.1.1 Traditional Approaches

Prior to deep learning, methods for segmenting retinal vessels typically included image processing operations for enhancing the image and increasing the contrast between the vessels and background, followed by a segmentation technique, thresholding, and post-processing to remove false positives (Singh and Kaur, 2015).

Zana and Klein (2001) used mathematical morphology and cross-curvature evaluation to segment vessel-like patterns. Soares et al. (2005) constructed feature vectors consisting of pixel's intensity and Morlet wavelet transform responses and a Bayesian classifier to classify whether each pixel belongs to vessel or non-vessel classes. Nguyen et al. (2013) used line detectors and different scales, linearly combining the detection responses to create the final segmentation map. Xu and Luo (2010) proposed an adaptive local thresholding to construct a binary mask, which is further refined by classifying thin vessels using a Support Vector Machine (SVM).

Lupascu et al. (2010) created a 41-dimensional feature vector corresponding to each pixel of the image and used an AdaBoost classifier to classify vessel and non-vessel classes. Delibasis et al. (2010) proposed a parametric model of retinal vessels and used an
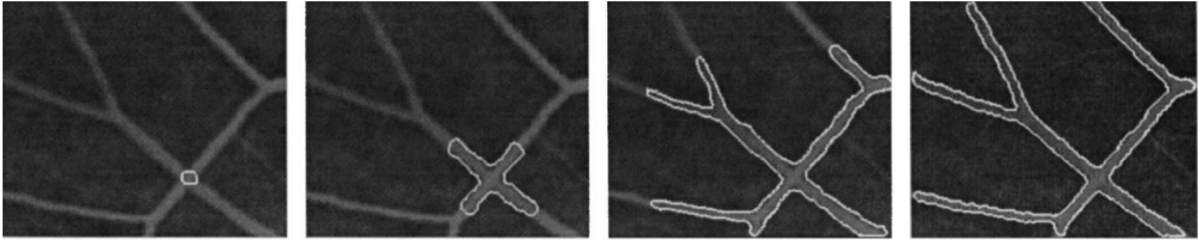
Figure 2.1: T-snakes perfectly conforms to the boundaries of vessels. From (McInerney and Terzopoulos, 2000).

automatic algorithm to track and eventually segment the vessels.

Some researchers have employed deformable models, especially active contours (a.k.a. "snakes") (Kass et al., 1988), for the task of retinal vessel segmentation. In active contour models, the curve of interest is iteratively evolved to conform to the boundaries of objects of interest in the image according to an energy functional that depends on the image content. The contour can be evolved explicitly by directly evolving the points on the contour, or implicitly by evolving the zero level set of an implicit function. If initialized properly, active contours are more powerful than classical methods for feature extraction and classification in retinal vessel segmentation.

McInerney and Terzopoulos (2000) introduced T-snakes for segmentation of complex biological structures such as retinal vessels, as it provided an improved topological flexibility in comparison to the original snakes formulation. Figure 2.1 shows several iterations of T-snakes evolution and how the model perfectly delineates the boundaries of retinal vessels. Espona et al. (2007) presented an explicit active contour model that was initialized automatically by approximation of the vessel skeletons. Al-Diri et al. (2009) proposed an implicit active contour model for segmenting and measuring retinal vessels by using two pairs of contours that capture vessel edges. Sum and Cheung (2007) introduced an implicit active contour without edges by proposing a novel loss term to take into account the local image contrast. Karn et al. (2018) presented a hybrid active contour model that combines gradient vector flow-based snakes and a balloon method to address the difficulties associated with conforming to closed-curvature vessels.

2.1.2 Deep Learning-Based Approaches

One of the earliest attempts in leveraging CNNs for the task of retinal vessel segmentation was by Melinščak et al. (2015), who proposed a 10-layer CNN architecture based on a patch-wise technique, but their results suffer from low sensitivity compared to other techniques. Fu et al. (2016) treated the problem of segmentation as a boundary detection problem and combined a CNN with a conditional random field to address the segmentation of retinal vessels, but their method is slower than whole-image CNNs and is outperformed by a number of other proposed methods in different metrics.

Oliveira et al. (2018) leveraged stationary wavelet transforms as a means of augmenting new input channels into a fully convolutional neural network and used a joint strategy for data augmentation and prediction by learning the the arrangement and orientation of the vessels during training. Using a VGG16 architecture with deep supervision, Lin et al. (2018) proposed to integrate a holistically nested edge detector with conditional random field as a global smoothness regularization mechanism.

After the success of the seminal U-Net architecture for medical and biomedical image segmentation, many efforts have been proposed to leverage the encoder-decoder paradigm for retinal vessel segmentation. Zhuang (2018) proposed an architecture based on U-Net (Ronneberger et al., 2015), which utilizes multiple-path networks that leverages a path-wise formulation in segmenting retinal vessels. Conversely, Wang et al. (2019b) proposed a Dense U-Net architecture with a patched-based training strategy and extensive data augmentation, which has outperformed the baseline U-Net by a small margin. Jiang et al. (2019) developed a multi-path encoder-decoder architecture that fused the outputs of the network at different scales to obtain the final predictions.

Alom et al. (2019) proposed a recurrent residual U-Net architecture (Figure 2.2) in which residual convolutional layers are used for optimal feature representation learning. Jin et al. (2019) proposed a deformable U-Net architecture that captures various scales and shapes of the vessels by adjusting the adaptive receptive field of the deformable convolutional layers. Wu et al. (2019c) used a U-like encoder-decoder architecture that
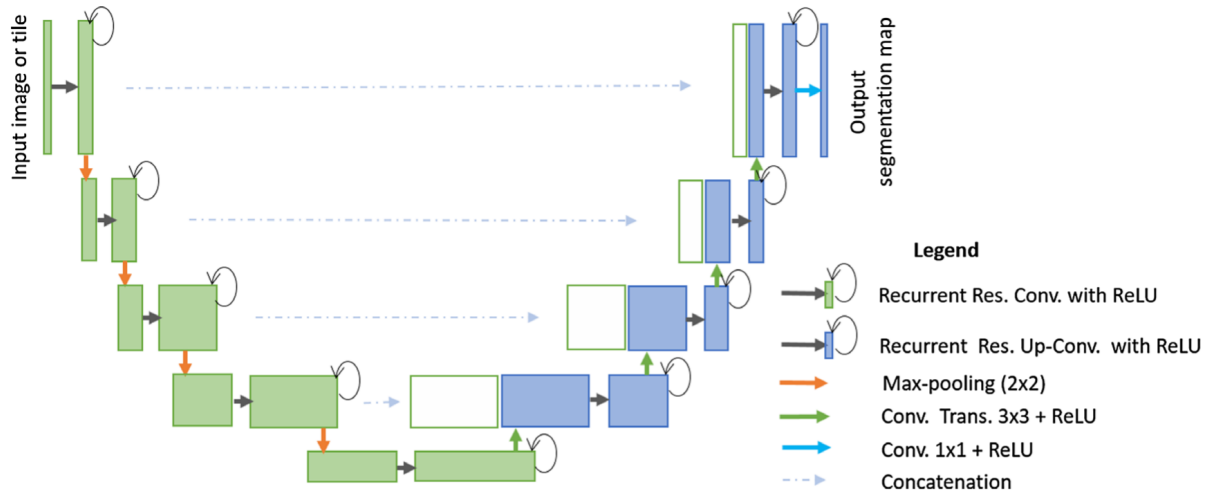
Figure 2.2: Recurrent residual convolutions are used to ensure optimal feature representation learning. From (Jin et al., 2019).

leverages Inception residual blocks for improved feature representation and is supervised in a multi-scale manner to preserve features of different sizes. Wang et al. (2019a) proposed to use two encoders, in a U-Net like architecture, to preserve the spatial information and semantic representations separately.

Recently, the application of cascaded architectures where several encoder-decoders are chained together to refine the segmentation outputs has gained traction. Although these approaches typically result in more refined boundaries, they are memory-intensive and the inference time is substantially increased due to the complexity of the model. Yan et al. (2018) proposed a methodology in which a three-step CNN was employed to separately segment thick and thin vessels and learn how to effectively fuse the results and refine the predictions. Francia et al. (2020) proposed to use an architecture consisting of two residual U-Nets where the second refines the feature maps of the first and resolves the ambiguities. Li et al. (2020) proposed to use a cascade of U-Net architectures, as shown in Figure 2.3, which share their weights and are connected by mid-level skip connections to allow for a deeper CNN that can effectively disambiguate the uncertain regions in the output segmented vessels. Guo and Peng (2020) introduced a bidirectional symmetric cascade with adaptive supervision of each layer in accordance to the diameter scale of each vessel.
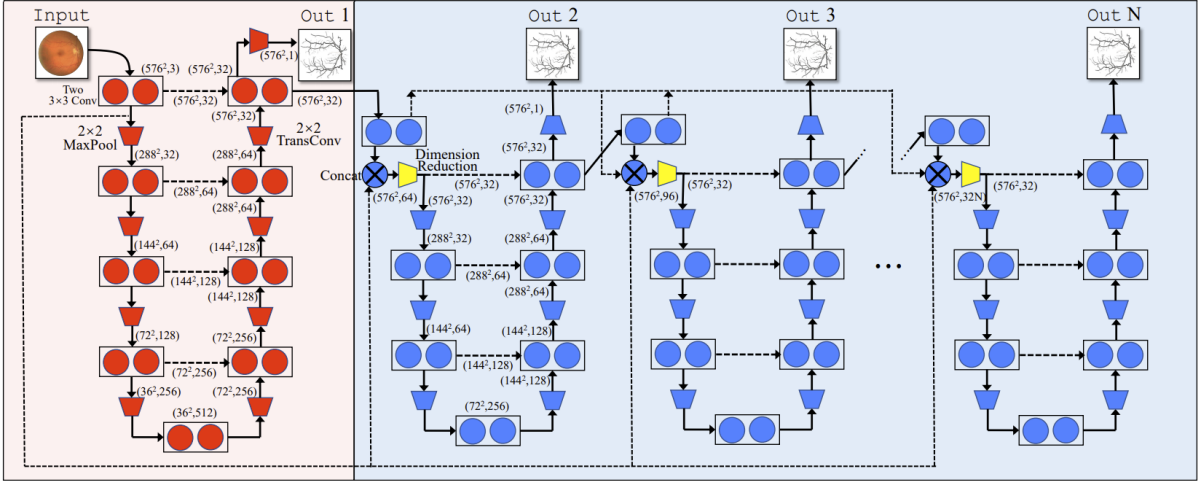
Figure 2.3: The IterNet leverages a cascade of U-Net architectures to refine the output predictions at each stage and disambiguate the uncertain regions in the output segmented vessels. From (Li et al., 2020).

Several efforts have proposed to leverage attention layers to effectively target the regions that contain vessels. Zhang et al. (2019a) proposed guided filters for preserving the learned structural representations and attention blocks to reduce background noise and focus on important features. Wu et al. (2019b) proposed a dense U-Net that is equipped with attention blocks in every resolution of the decoder sub-network. Zhang et al. (2019b) developed an edge-guided network with tailored attention blocks to effectively learn the edge representation of the vessels.

Generative Adversarial Networks (GANs) have also been exploited for the task of retinal vessel segmentation. Wu et al. (2019a) proposed a GAN with dense attention U-Net as the generator that predicts the output probability maps. Andreini et al. (2019) proposed a multi-stage architecture in which a GAN was used to synthesize high quality images during the training process. Furthermore, Son et al. (2019) proposed an image translation scheme in which the output probability map is generated from the input images by constraining the outputs to the annotation's image space.

To date, few efforts have utilized deep learning-based methods for the semantic segmentation of retinal arteries and vessels. Girard et al. (2019) proposed a whole-image approach with U-Net like encoder-decoder where output predictions of the CNN

is propagated through a graph representation of the retinal vasculature by introducing a likelihood score for detecting arteries. Girard et al. (2019) proposed a patched-based U-Net with categorical cross-entropy as the loss function to segment arteries and vessels in images from the DRIVE and HRF datasets.

2.2 Retinal Artery and Vein Segmentation and Classification

Some efforts have focused on classifying arteries and veins in accordance to intensity and color information. Vázquez et al. (2010) leveraged an explicit active contour model to locate the vessels and a K-means clustering algorithm to distinguish detected arteries and veins. Relan et al. (2013) classified arteries and veins by using a Gaussian mixture model, an Expectation-Maximization by taking into account colour features. Dashtbozorg et al. (2013) developed a classification scheme by considering a set of intensity features and graph-based labeling.

Another common approach is to exploit the structural properties for classifying the vessels. Estrada et al. (2015) leveraged a graph-theoretic framework for classifying small and averaged-sized vessels by establishing a global likelihood model via the incorporation of domain-specific features. Huang et al. (2018) proposed to use a genetic-search based feature selection technique based on features such as centerline pixels to optimally classify arteries and vessels. Joshi et al. (2014) proposed an approach in which the graph of the vessels is found by the width, orientation, and intensity of each vessel and further classified by color attributes of the vessels.

CHAPTER 3

Retinal Vessel Segmentation in Color Fundus Images

In this chapter, we propose a fully convolutional encoder-decoder architecture for the automated segmentation of retinal vessels in color fundus images.

3.1   Encoder-Decoder Architecture

We introduce a novel encoder-decoder architecture, depicted in Figure 3.1, that leverages dilated residual blocks along with deep supervision at multiple scales for effectively learning the multiresolution details of retinal vessels. Our encoder-decoder architecture spans four different resolutions. In the encoder, each path consist of 2 consecutive 3×3 convolutional layers, followed by a dilated residual unit with a dilation rate of 2. Before being fed into the dilated residual unit, the output of these convolutional layers are added with the output feature maps of another 2 consecutive 3×3 convolutional layers that learn additional multiscale information from the re-sized input image in that resolution. At the third stage of our architecture, we utilize a series of 3 consecutive dilated residual blocks with dilation rates of 1, 2, and 4, respectively. Finally, we incorporate a dilated spatial pyramid pooling layer with 4 different dilation rates of 1, 6, 12 and 18 in order to recover the content lost in the learned feature maps during the encoding process.

Subsequently, the decoder in our architecture receives the learned multiscale contextual information of the dilated spatial pyramid pooling layer and is connected to the dilated residual units at each resolution via skip connections. In each path of the decoder, the image is up-sampled and 2 consecutive 3×3 convolutional layers are used before proceeding to the next resolution. Moreover, each scale branches to an additional convolutional
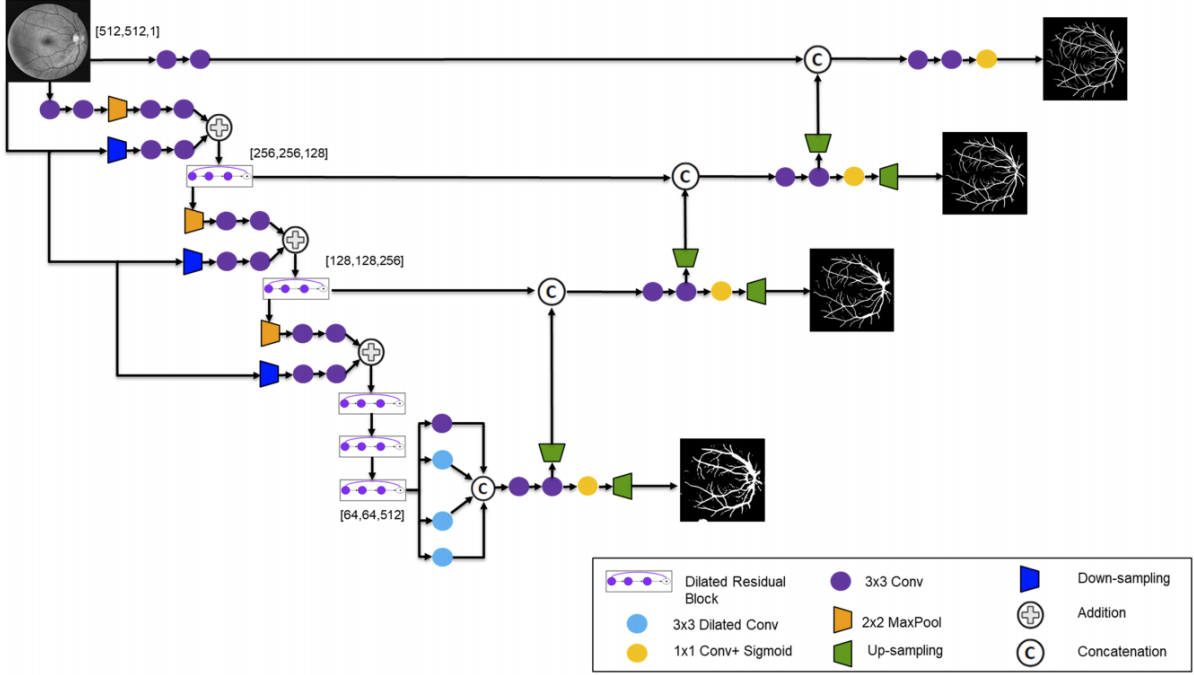
Figure 3.1: Our fully convolutional architecture. Dilated spatial pyramid pooling aggregates the outputs of the multiple stages.

layer whose output is resized to the original input image size and is followed by another convolutional layer with sigmoid activation function.

3.1.1 Dilated Convolutional Layers

We next detail the convolutional layers that are used in our architecture. Each convolutional layer, with kernel W, is followed by a Rectified Linear Unit (ReLU) defined as

$$Re(X) = \max(0, X). \tag{3.1}$$

The activation outputs are then fed into a batch normalization layer $BN_{\gamma,\beta}(X)$ with parameters γ, β that are learned during training. Consequently, every location i in the output of a convolutional layer followed by ReLU and batch normalization can be represented as

$$Y(i) = BN_{\gamma,\beta}\left(Re\left(\sum_{j=1} X[i + jr]W[j]\right)\right), \tag{3.2}$$

13
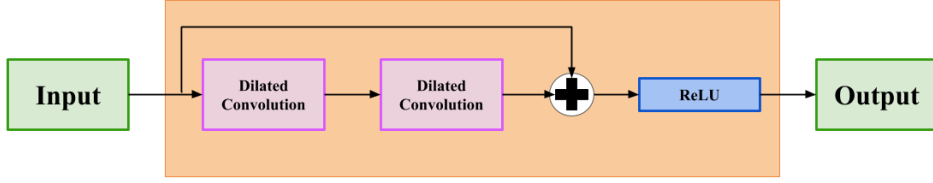
Figure 3.2: Dilated Residual Block Diagram.

where r is the dilation rate. We employ both standard and dilated convolutional layers for which the value of r in the former is 1 and in the latter depends on where it is used.

3.1.2 Residual Blocks

The output of each resolution path in the encoder of the backbone CNN is fed into a dilated residual block. The input to the dilated residual block, as shown in Figure 3.2, is first passed to 2 consecutive dilated convolutions with kernel size of 3×3 and dilation rates of 6 and 12, respectively, and is then fused with itself before being passed to a ReLU activation function.

3.1.3 Loss Layer

As was discussed earlier, the multiscale prediction maps are supervised and as a result contribute to the final loss layer. Without loss of generality, we utilize a soft Sørensen-Dice loss function as our basis and aggregate throughout each of the four resolutions:

$$Loss = \sum_{m=1}^{5} \left(1 - \sum_{n=1}^{N} \frac{2G_n P_{n,m}}{G_n + P_{n,m} + \epsilon} \right) + \lambda \|w\|_2^2, \tag{3.3}$$

where N, $P_{n,m}$, and G_n denote the total number of pixels, the label prediction of pixel n in scale m, and the ground truth label of pixel n, respectively, ϵ is a smoothing constant, and λ is the weight decay regularization hyper-parameter.
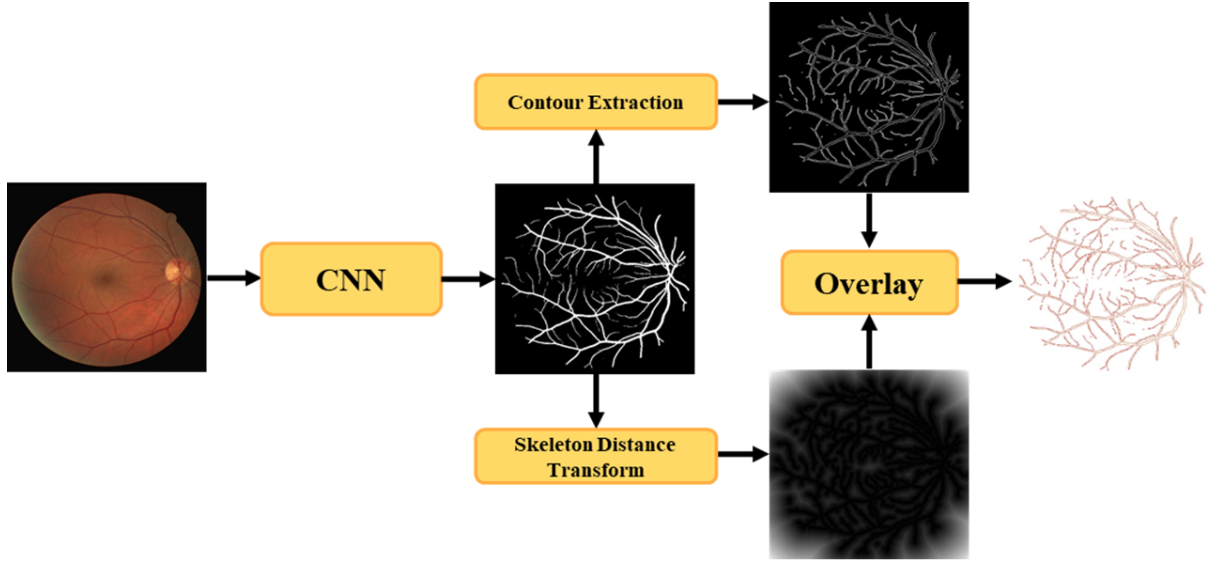
Figure 3.3: Estimating the width profile of retinal vessels from segmentation.

3.2   Vessel Width Estimation

We propose a simple, yet effective method for the automatic estimation of vessel width profiles by leveraging the segmentation masks obtained by our CNN. Following Zhang and Suen (1984), we first obtain the topological skeleton of the image by successively identifying the borderline pixels and removing the corresponding pixels that maintain the connectivity of the vessels. This operation approximates the skeleton of the vessel and represents its topology. We then calculate the distance of each pixel to the derived skeleton by applying an Euclidean distance transform to the generated feature map. Finally, we extract the contour of the original segmentation mask and overlay the generated distance transform onto this map to create the final width map of the retinal vessels. Needless to say, our formulation is only valid in areas where these vessels exist, otherwise the width is set to zero. Figure 3.3 illustrates our width estimation algorithm in more detail. Unlike competing methods, our method does not rely on hand-crafted geometric equations or on user interaction.

CHAPTER 4

Semantic Segmentation of Retinal Arteries and Veins in IR Images

Infrared (IR) reflectance imaging enables the highlighting of specific sub-retinal features that facilitates the detection of morphological changes in early stages of disease more accurately than traditional imaging modalities. In this chapter, we introduce a CNN architecture along with tailored loss functions for the semantic segmentation of retinal arteries and veins in the IR retinal imaging modality.

4.1 SegAVIR Architecture

We devise an encoder-decoder architecture, named SegAVIR, which leverages residual blocks and skip connections that connect the encoder and decoder. We employ a whole-image approach in order to take full advantage of global content in the image. Each residual block consists of two convolutions with batch normalization, ReLU, and an identity skip connection that adds the input of the block to the output of convolutions. As illustrated in Figure 4.1, each stage of the encoder includes residual blocks followed by a convolutional layer with stride of two. Similarly, each stage of the decoder has a residual block followed by a a transposed convolution with the stride of two. The encoder is connected to the decoder via two residual blocks followed by an Dilated Spatial Pyramid Pooling (DSPP) at the lowest resolution as well as skip connections at every stage. The output of the decoder is connected to a convolution with three output channels that correspond to background, vein, and artery classes. We use channel-wise sigmoid functions as the activation of the last convolution layer. Tables 4.1 and 4.2 present detailed
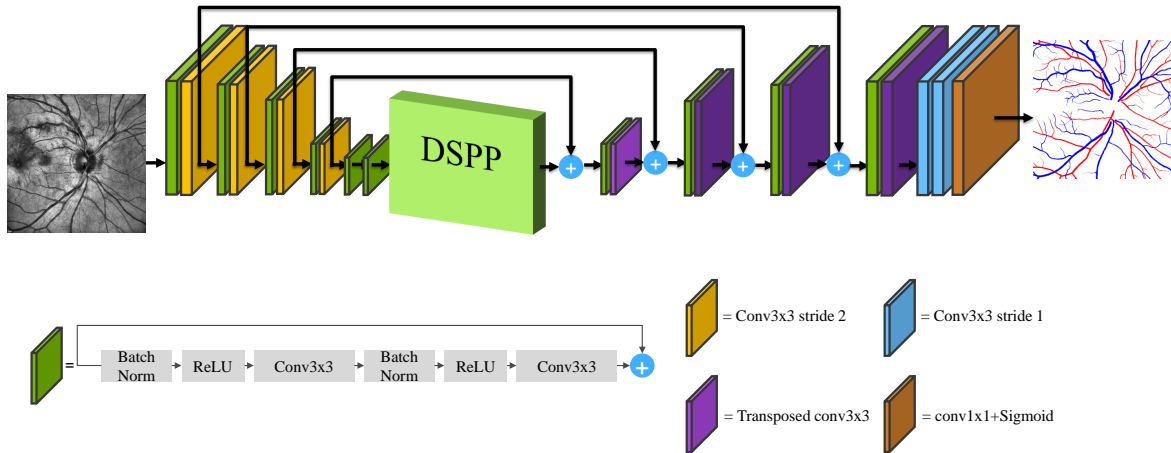
Figure 4.1: Our CNN architecture leverages a symmetric encoder-decoder with residual units as well as an Dilated Spatial Pyramid Pooling (DSPP).

Table 4.1: Detailed information regarding the encoder of SegAVIR. BN and Add denote batch normalization and additive identity skip connections. Conv and DSPP denote a 3×3 convolutional layer and Dilated Spatial Pyramid Pooling.

Operations	Output size
Input	$768 \times 768 \times 1$
Conv	$768 \times 768 \times 64$
Conv stride 2	$384 \times 384 \times 128$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$384 \times 384 \times 128$
Conv stride 2	$192 \times 192 \times 256$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$192 \times 192 \times 256$
Conv stride 2	$96 \times 96 \times 512$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$96 \times 96 \times 512$
Conv stride 2	$48 \times 48 \times 1024$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$48 \times 48 \times 1024$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$48 \times 48 \times 1024$
DSPP	$48 \times 48 \times 1024$

Table 4.2: Detailed information regarding the decoder of SegAVIR. BN and Add denote batch normalization and additive identity skip connections. Conv and Conv1 denote 3×3 and 1×1 convolutional layers, respectively. TransConv and Dense represent 3×3 transposed convolutional layers with a kernel size of 2.

Operations	Output size
Input	$48 \times 48 \times 1024$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$48 \times 48 \times 1024$
TransConv stride 2	$96 \times 96 \times 512$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$96 \times 96 \times 512$
TransConv stride 2	$192 \times 192 \times 256$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$192 \times 192 \times 256$
TransConv stride 2	$384 \times 384 \times 128$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$384 \times 384 \times 128$
TransConv stride 2	$768 \times 768 \times 64$
Conv, Conv	$768 \times 768 \times 64$
Conv1, Sigmoid	$768 \times 768 \times 3$

information about the SegAVIR encoder and decoder architectures, respectively.

4.2 SegAVIR Loss Function

To supervise the output of SegAVIR, we use a hybrid loss function consisting of the following terms:

$$\mathcal{L} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}. \tag{4.1}$$

Assuming that the total number of pixels is N,

$$L_{\text{Dice}} = 1 - \frac{2 \sum_{j=1}^{N} G_j P_j}{\sum_{j=1}^{N} G_j + \sum_{j=1}^{N} P_j + \epsilon}, \tag{4.2}$$

where P_j and G_j denote output prediction and corresponding ground truth at pixel j, and

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{j=1}^{N} \left[G_j \log P_j + (1 - G_j) \log(1 - P_j) \right]. \tag{4.3}$$
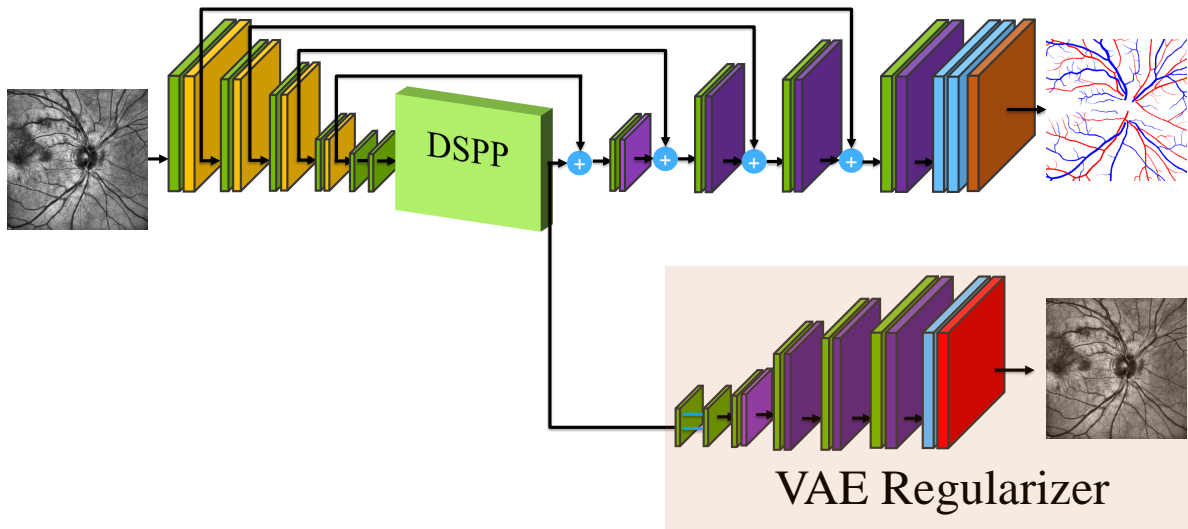
Figure 4.2: We leverage a VAE that learns how to reconstruct the input image as a means of regularizing our SegAVIR.

4.3 Variational Autoencoder Regularization

The dataset used in this effort is relatively small compared to the large datasets commonly used in deep learning. As such, our model would simply overfit the training set and perform sub-optimally on the testing set if certain precautions are not taken. In an effort to mitigate the overfitting problem, we first tried to use dropout after each convolutional layer, but extensive experiments demonstrated the inefficacy of dropout as a means of regularizing our CNN. Inspired by (Myronenko, 2018), we instead employ kernel regularization as well as a Variational AutoEncoder (VAE) that learns to reconstruct the input images in a self-supervised manner. Instead of using a separate module to implement the VAE, we share the encoder of SegAvIR with the VAE, but leverage a dedicated decoder for the VAE. The SegAVIR architecture with the added VAE branch is shown in Figure 4.2.

Table 4.3 presents detailed information about the SegAVIR VAE decoder architecture. The output of the DSPP in SegAVIR's encoder is first converted to a low-dimensional space of size 256 with mean and standard deviations of 128. Consequently, the input image is reconstructed by taking a sample from a Gaussian distribution with the aforementioned

Table 4.3: Detailed information regarding the VAE decoder of SegAVIR. BN and Add stand for batch normalization and additive identity skip connections. Conv and Conv1 denote 3×3 and 1×1 convolutional layers, respectively. TransConv and Dense represent 3×3 transposed convolutional layers with kernel size of 2 and fully connected layers, respectively.

Operations	Output size
BN, ReLU, Conv, Dense (256)	1×256
Sample $\sim \mathcal{N} \, (\mu = 128, \, \sigma^2 = 128)$	1×128
BN, ReLU, Conv, BN, ReLU, Conv, Add	$48 \times 48 \times 256$
TransConv stride 2	$96 \times 96 \times 128$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$96 \times 96 \times 128$
TransConv stride 2	$192 \times 192 \times 64$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$192 \times 192 \times 64$
TransConv stride 2	$384 \times 384 \times 32$
BN, ReLU, Conv, BN, ReLU, Conv, Add	$384 \times 384 \times 32$
TransConv stride 2	$768 \times 768 \times 16$
Conv, Conv, ReLU	$768 \times 768 \times 1$

mean and standard deviations.

4.3.1 VAE Loss

In addition to the loss terms that are introduced in (4.1), we further add a VAE loss

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{L2}} + \mathcal{L}_{\text{KL}}, \tag{4.4}$$

where the L2 reconstruction loss of the input image is defined as

$$\mathcal{L}_{\text{L2}} = \|X_i - X_r\|_2^2, \tag{4.5}$$

in which X_i and X_r denote the input and reconstructed images, respectively, and the Kullback–Leibler (KL) loss is (Myronenko, 2018)

$$\mathcal{L}_{\text{KL}} = \frac{1}{N} \sum \left[\sigma^2 + \mu^2 - \log(\sigma^2) - 1 \right]. \tag{4.6}$$

Algorithm 1: Width Measurement Algorithm

Data: X: Input image; f: CNN for semantic segmentation; r: binarizing threshold; Y: Segmentation mask; r: Binarizing threshold; s: Skeletonizing Morphological operations; d: Distance transform operations; C: Medial curve; D: Distance transform of medial curve; k: Pixel distance scaling factor.

Result: Z: Pixel-wise width distance map

1 for $t = 1$ **to** N **do**
2 $Y_t = h(f(X_t), r)$
3 $C_t = s(Y_t)$
4 $D_t = d(C_t)$
5 $Z_t = D_t * Y_t$
6 $Z_t = kZ_t$
7 end

4.4   Vessel Width Estimation

Using the output segmentation of our SegAVIR network, we further measure the width of the arteries and veins according to Algorithm 1. We first threshold the output of the SegAVIR network and obtain a pixel-wise mask that identifies the location of the medial curve of the segmented vessel. We then obtain the distance transform of the medial curve mask and multiply it with the thresholded segmentation mask in a pixel-wise manner. As a result, the output provides the width distance map with respect to the medial curve of the segmented vessels.

CHAPTER 5

Experiments and Results

This chapter presents our experiments and results with the models that we developed in the previous two chapters.

5.1 Color Fundus Image Segmentation

In this section, we demonstrate the architecture and loss function presented in Chapter 3 and test our model on two publicly available image datasets, DRIVE and CHASDE-DB1.

5.1.1 The DRIVE and CHASE-DB1 Datasets

We have tested our model on two publicly available retinal vessel segmentation datasets— DRIVE and CHASE-DB1. The DRIVE dataset consists of 40 two-dimensional RGB images, with each image having a resolution of 565×584 pixels, divided into a training set and test set, each comprising 20 images. The CHASE-DB1 dataset includes 28 images, collected from both eyes of each of 14 children. Each image has a resolution of 999×960 pixels. We divided the CHASE-DB1 dataset into a training set of 20 images and a testing set of 8 images.

5.1.2 Implementation Details

We have implemented our CNN in TensorFlow (Abadi et al., 2016). All the input images are converted to gray-scale, transformed by contrast-limited adaptive histogram equalization, resized to a predefined size of 512×512, and intensity normalized between

0 and 1. Our model is trained, with a batch size of 2, on an Nvidia Titan XP GPU and an Intel® Core™ i7-7700K CPU @ 4.20 GHz. We use the Adam optimization algorithm with an initial learning rate of 0.001 and exponentially decay its rate. The smoothing constant in the loss function and the weight decay hyper-parameter are set to 10^{-5} and 8^{-4}, respectively. Since the number of images is limited, we perform common data augmentation techniques such as rotating, flipping horizontally and vertically, and transposing the image.

5.1.2.1 Evaluation Metrics

In order to assess the accuracy of our predictions, we used various evaluation metrics, including *Sensitivity (SE)*, *Specificity (SP)*, *Accuracy (Acc)*, *Precision*, *Recall*, and *F1* score which is the same as the Dice coefficient in the context of image segmentation. Letting *TP*, *TN*, *FP*, *FN* denote true positive, true negative, false positive, and false negative, respectively, the Sensitivity is defined as

$$SE = \frac{TP}{TP + FN}. \tag{5.1}$$

Similarly the Specificity is given as

$$SP = \frac{TN}{TN + FP}. \tag{5.2}$$

In addition, the Accuracy is defined as

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}. \tag{5.3}$$

Accordingly, the Precision can be defined as

$$Precision = \frac{TP}{TP + FP}, \tag{5.4}$$

Table 5.1: Segmentation Evaluations on the DRIVE and CHASE-DB1 datasets.

Method	DRIVE				CHASE-DB1			
	SE	SP	Acc	F1	SE	SP	Acc	F1
(Melinščak et al., 2015)	0.7276	0.9785	0.9466	-	-	-	-	-
(Li et al., 2016)	0.7569	0.9816	0.9527	-	0.7507	0.9793	0.9581	-
(Liskowski and Krawiec, 2016)	0.7520	0.9806	0.9515	-	-	-	-	-
(Fu et al., 2016)	0.7603	-	0.9523	-	0.7130	-	0.9489	-
(Oliveira et al., 2018)	0.8039	0.9804	0.9576	-	0.7779	**0.9864**	0.9653	-
Vessel-Net (Wu et al., 2019c)	0.8038	0.9802	0.9578	-	0.8132	0.9814	0.9661	-
U-Net w/Inception (Wu et al., 2019c)	0.7873	0.9794	0.9550	-	0.7982	0.9787	0.9623	-
M2U-Net (Laibacher et al., 2018)	-	-	0.9630	0.8091	-	-	0.9703	0.8006
U-Net (Alom et al., 2019)	0.7537	0.9820	0.9531	0.8142	0.8288	0.9701	0.9578	0.7783
Recurrent U-Net (Alom et al., 2019)	0.7751	0.9816	0.9556	0.8155	0.7459	0.9836	0.9622	0.7810
R2U-Net (Alom et al., 2019)	0.7792	0.9816	0.9556	0.8171	0.7756	0.9820	0.9634	0.7928
LadderNet (Zhuang, 2018)	0.7856	0.9810	0.9561	0.8202	0.7978	0.9818	0.9656	0.8031
DU-Net (Jin et al., 2019)	0.7894	**0.9870**	**0.9697**	0.8203	0.8229	0.9821	0.9724	0.7853
Ours	**0.8197**	0.9819	0.9686	**0.8223**	**0.8300**	0.9848	**0.9750**	**0.8073**

and the Recall as

$$Recall = \frac{TP}{TP + FN}. \tag{5.5}$$

Finally, the *F1* score can be defined in terms of precision and recall according to

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \tag{5.6}$$

We will present our benchmarks in terms of these metrics and compare to other performance benchmarks in Section 5.1.3.

5.1.3 Results and Analysis

We present the benchmarks of our model's predictions in Table 5.1. As is evident, for both DRIVE and CHASE-DB1 datasets, the model compares favorably to several competing methods.

5.1.3.1 DRIVE Dataset Results

On the DRIVE dataset, our model exceeds the state-of-the-art performance on the F1 score and Sensitivity while being very competitive to (Jin et al., 2019) in Specificity and
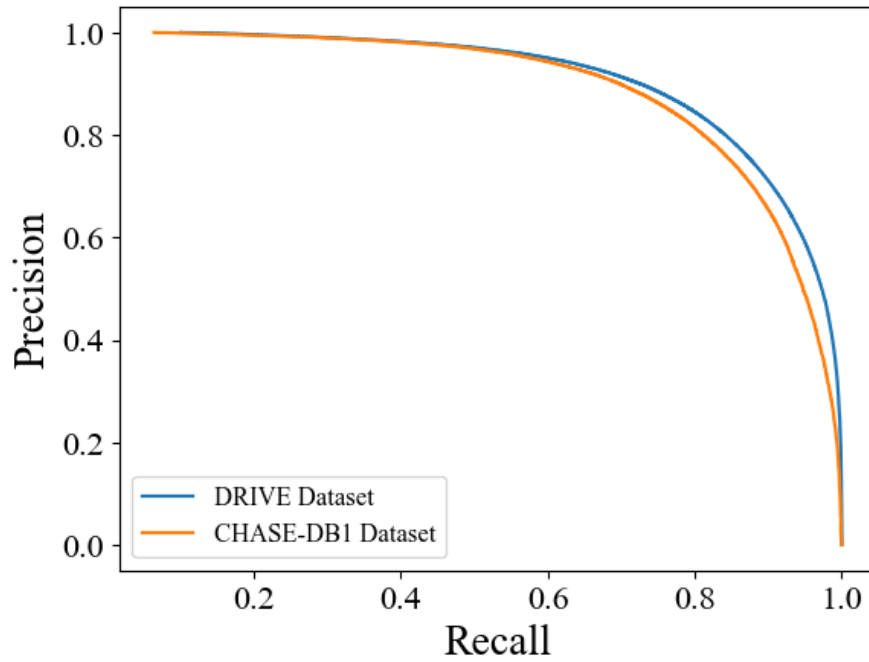
Figure 5.1: Precision-Recall curves for our model's segmentation performance on the DRIVE and CHASE-DB1 datasets.
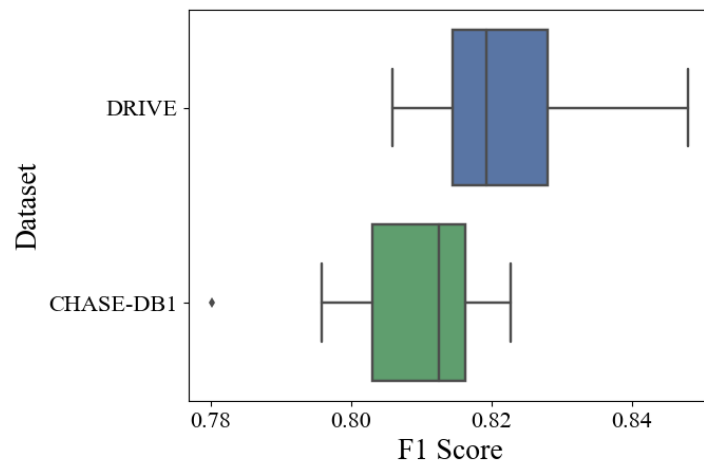

Figure 5.2: Box and whisker plot for evaluation of the model's segmentation performance on the DRIVE and CHASE-DB1 datasets.

Accuracy. Figure 5.1 presents the precision-recall curve for the performance of the model on this dataset. In addition, the box and whisker plot of our model's performance, as shown in Figure 5.2 demonstrates a robust performance with *F1* scores ranging between 0.80 to 0.84 with no outliers falling outside the interquartile range.

Moreover, Figure 5.3 demonstrates the final segmentation of our model compared to the masks that were manually created by clinicians as well as the generated width profiles. Clearly, our model successfully captures the intricate arteries and veins without the presence of any additional false positives as are present in the outputs of the competing methods.

In addition, we have presented the width estimation maps of the detected vessels. Note that areas with greater intensity of red denote lower diameter and the background is colored white to denote the non-existence of vessels. It is evident that the success of width estimation module is highly dependent on the model's segmentation performance.

5.1.3.2 CHASE-DB1 Dataset Results

On the CHASE-DB1 dataset, our model exceeds the state-of-the-art performance for all metrics except Specificity, for which (Oliveira et al., 2018) performs slightly better. Figure 5.2 shows that the *F1* scores are between 0.78 to 0.82, demonstrating a robust performance. The precision-recall curve shows a similar performance to the results obtained for the DRIVE dataset, as in Figure 5.1. We demonstrate the visual segmentation of our model's output as well as width profile measurement on two cases from the CHASE-DB1 test set in Figure 5.4.

5.1.4 Discussion

Our substantially improved state-of-the-art results on two publicly available datasets, DRIVE and CHASE-DB1, confirm the effectiveness of our model. Unlike the competing patch-wise approaches, our method operates on the entire image. This has proven to be beneficial as our model is significantly faster than the patch-wise approaches that must
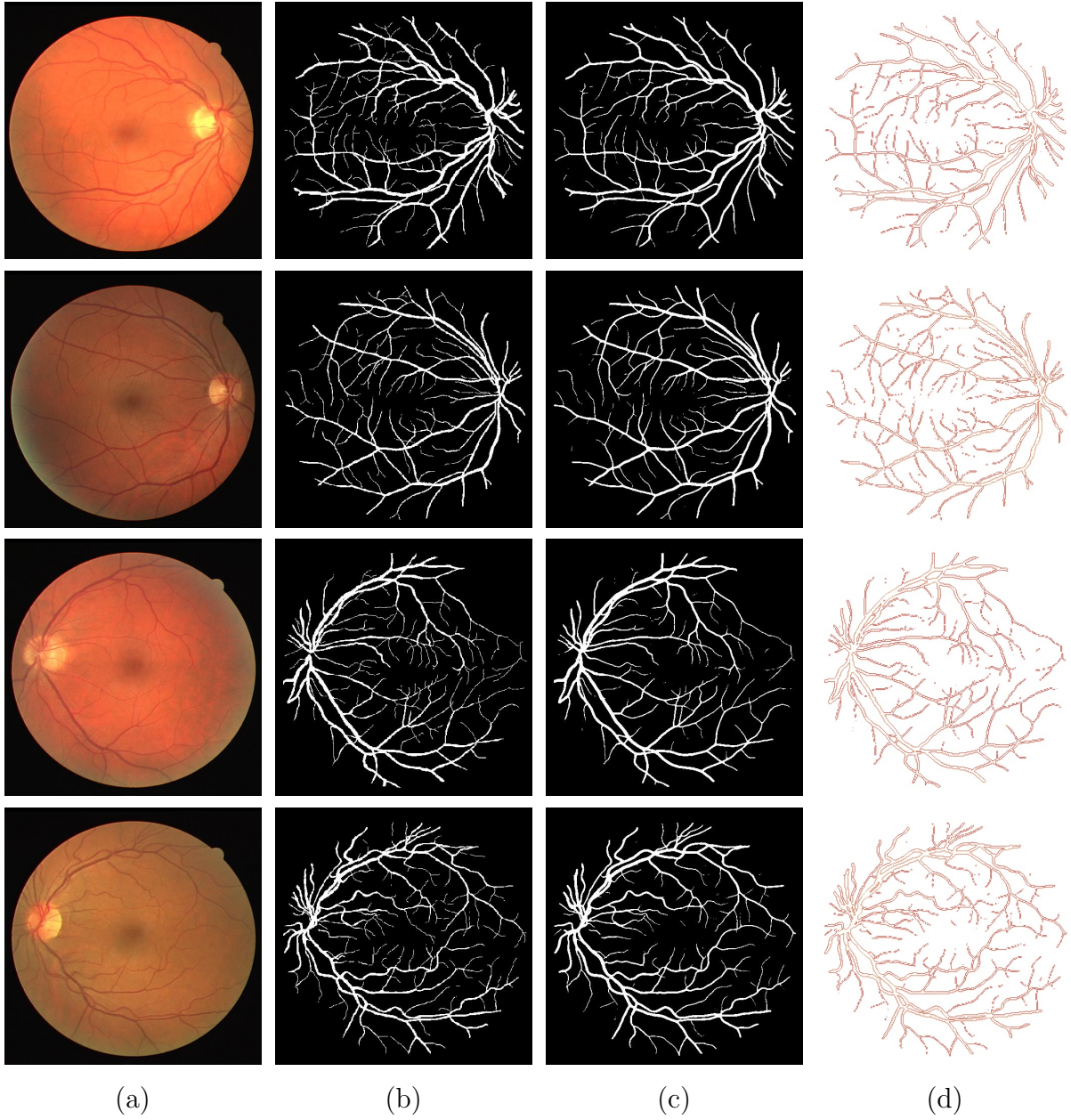
Figure 5.3: (a) Input images, (b) labels, (c) segmentation outputs, and (d) width estimation profiles.
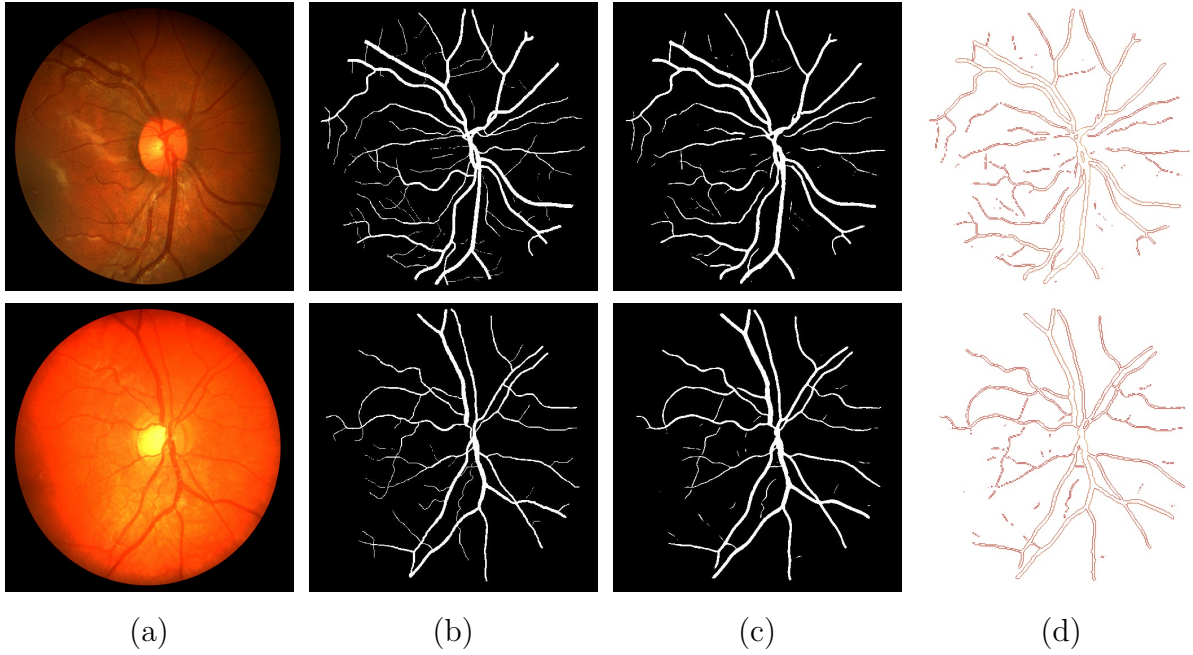
| (a) | (b) | (c) | (d) |

Figure 5.4: (a) Input images, (b) labels, (c) segmentation outputs, and (d) width estimation profiles.

slide moving windows of multiple sizes over the image. Additionally, our model produces more natural and continuous segmentation masks and is able to capture finer details because it benefits from a dilated spatial pyramid pooling layer that recovers the content lost during encoding by leveraging different dilation rates and aggregating the multiscale feature maps into the decoder. Furthermore, introducing the input image at multiple scales throughout our architecture and introducing supervision at each of these scales helps our model to effectively aggregate the outputs of different stages.

Our technique for estimating the width profiles of retinal vessels by leveraging the generated segmentation masks has also proven to be effective. Its accuracy promises to help in quantifying new relevant biomarkers that correlate with the narrowing and structural changes of vessels.
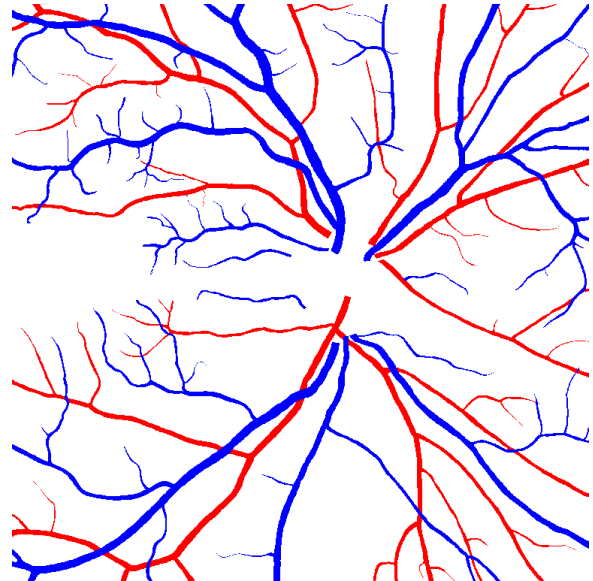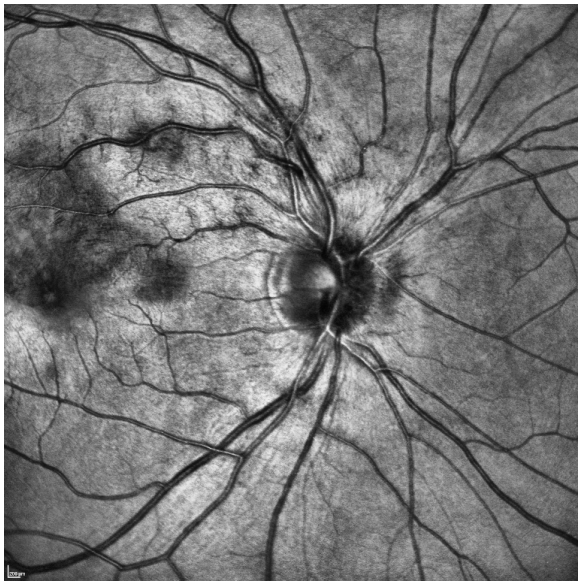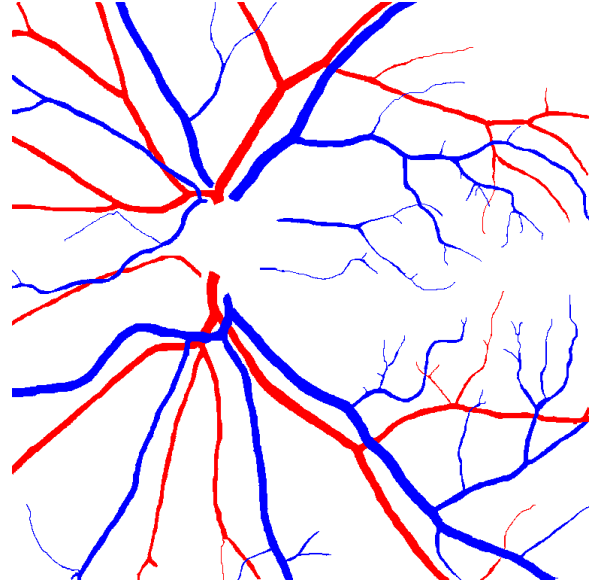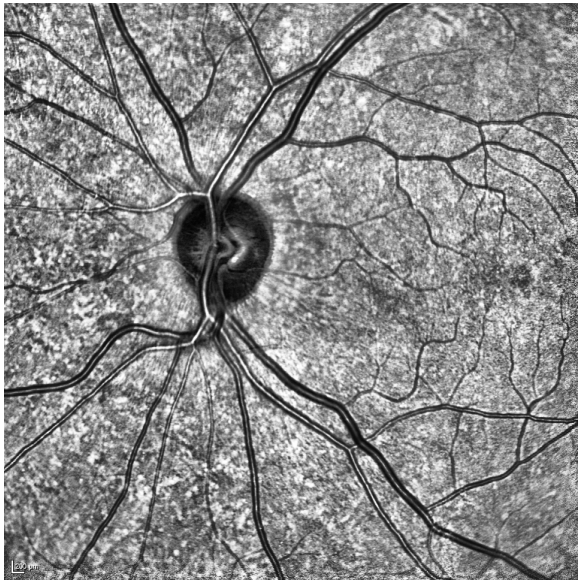
5.2 IR Image Segmentation

In this section, we test our IR image segmentation model, SegAVIR that was developed in Chapter 4, on a novel dataset, which was created as part of a collaborative effort between the UCLA Samueli School of Engineering Computer Science Department and the UCLA David Geffen School of Medicine Stein Eye Institute. We use our methodology to directly measure the structural properties of detected arteries and veins.

5.2.1 The AVIR Dataset

Arteries and Veins in InfraRed (AVIR) is a novel dataset that contains IR images of 59 patients, from both left and right eyes, with pixel-wise semantic labels identifying retinal arteries and veins. The resolution of the images is 768×768. The dataset includes per-pixel semantic masks of arteries and veins, which enables the creation of a model that distinguishes between the type of extracted vessels. Figure 5.5 shows examples of the IR images analyzed along with their annotated segmentation masks. In addition, 20 cases include unhealthy patients who suffer from a variety of retinal anomalies. The top and bottom rows of Figure 5.5 show examples of healthy and unhealthy cases in the AVIR dataset, respectively.

5.2.2 Implementation Details

SegAVIR is implemented in TensorFlow and is compatible with both TensorFlow 1.x and 2.0 versions. The size of the images input into SegAVIR is 768×768 with an initial filter size of 64 that is doubled at each stage. All images are transformed by contrast-limited adaptive histogram equalization. We have performed training of all experiments using an Nvidia Titan RTX GPU, and an Intel® Core™ i7-7700K CPU @ 4.20 GHz. We used minibatches of size 3 and employed the Adam optimization algorithm (Kingma and Ba, 2014) with an initial learning rate of $\alpha_0 = 0.001$ decreasing according to (Myronenko and

(a) Input Images
(b) Semantic Labels

Figure 5.5: Example IR images analyzed. (a) Input images. (b) Ground truth segmentation masks distinguishing artery (red) and vein (blue) classes. (top) Healthy case. (bottom) unhealthy case.

Table 5.2: Segmentation Evaluations of Veins on the AVIR dataset.

Method	F1 Score	Accuracy	Recall	Precision	Sensitivity	Specificity
U-Net-1	0.7557	0.9688	0.7467	0.7649	0.7467	0.9841
U-Net-2	0.7880	0.9719	0.8101	0.7672	0.8101	0.9830
SegAVIR	0.8272	0.9783	0.8054	0.8503	0.8054	0.9902
SegAVIR+VAE	**0.8504**	**0.9810**	**0.8345**	**0.8669**	**0.8345**	**0.9911**

Table 5.3: Segmentation Evaluations of Arteries on the AVIR dataset.

Method	F1 Score	Accuracy	Recall	Precision	Sensitivity	Specificity
U-Net-1	0.7314	0.9711	0.7257	0.7372	0.7257	0.9851
U-Net-2	0.7623	0.9758	0.7158	0.8153	0.7158	0.9907
SegAVIR	0.8159	0.9794	0.8394	0.7937	0.8394	0.9874
SegAVIR+VAE	**0.8389**	**0.9830**	**0.8141**	**0.8653**	**0.8141**	**0.9927**

Hatamizadeh, 2019)

$$\alpha = \alpha_0 \left(1 - e/N_e\right)^{0.9} \tag{5.7}$$

with epoch counter e and the total number of epochs N_e. In addition, we used 70%, 20%, and 10% of the AVIR dataset for training, testing, and validation, respectively. Due to the very small size of the dataset, we dynamically augmented the training set by flipping and rotating the images with a probability of 50% for each augmentation.

5.2.3   Results and Analysis

In this section, we quantitatively study the performance of our SegAVIR in the AVIR dataset and compare it against different variations of the gold-standard U-Net (Ronneberger et al., 2015), which has become a standard CNN-based image segmentation technique, as well as SegAVIR without VAE regularization.

The segmentation results of these experiments are summarized in Table 5.2 and Table 5.3 for the vein and artery classes, respectively. Generally, our SegAVIR model performs equally well for both arteries and veins, although there seems to be a slight advantage for veins since they usually manifest as larger vessels and are easier to segment.

| (a) | (b) | (c) | (d) |
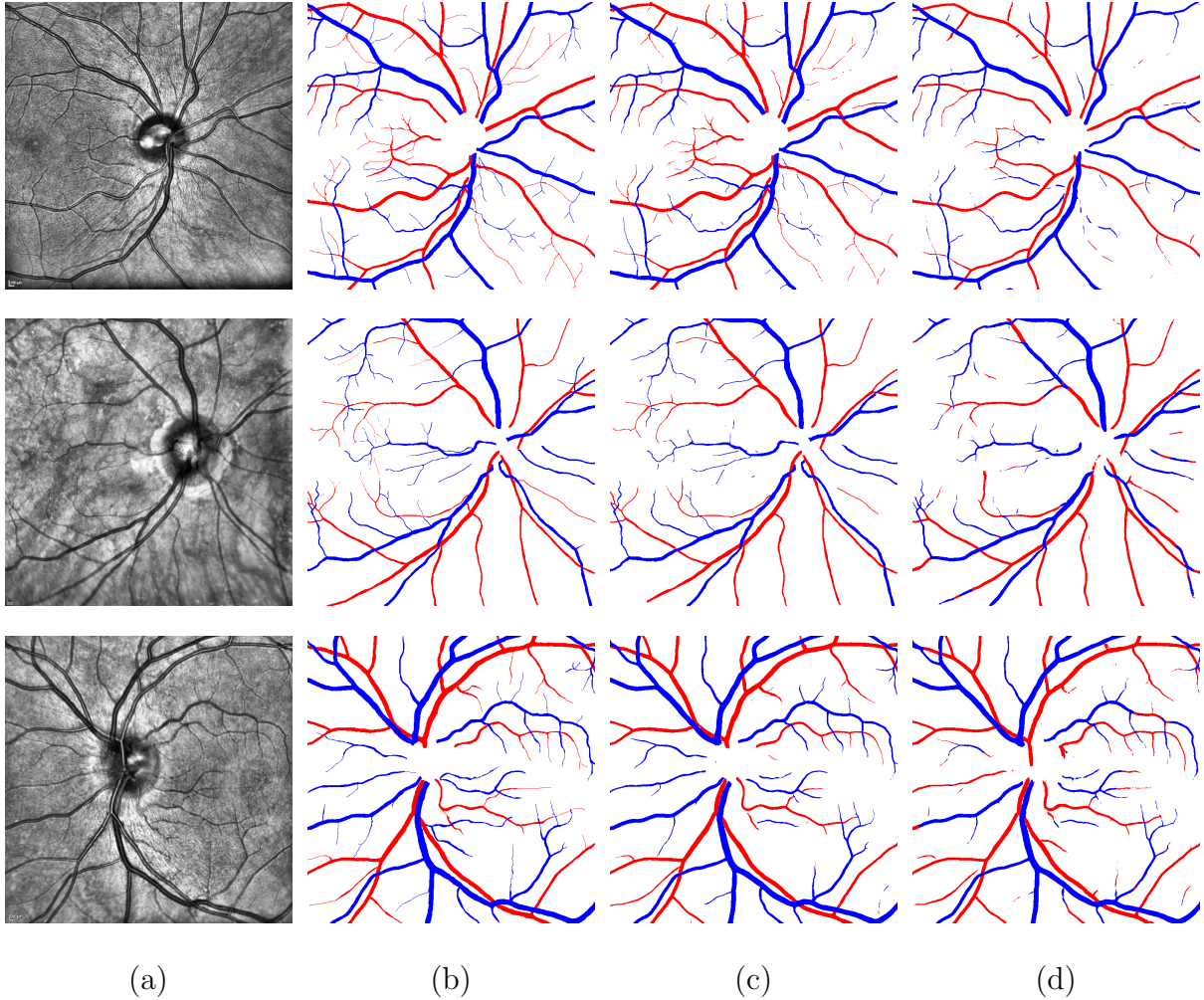
Figure 5.6: Visual comparison of the segmentation performance of SegAVIR+VAE and U-Net. (a) Input image. (b) Semantic Labels. (c) SegAVIR+VAE. (d) U-Net-2. SegAVIR evidently detects more arteries and veins with better structural consistency.

5.2.3.1 Comparison to U-Net

We have trained and tested two U-Nets with different filter sizes. As reported Table 5.2 and Table 5.3, U-Net-2 closely follows the original U-Net (Ronneberger et al., 2015) with minor adjustments to accommodate for the size of the input images in the AVIR dataset. U-Net-1 has the same architecture as U-Net-2 except for the fact that the size of input filter is reduced four times to create a more lightweight network that is suitable for small datasets.

Benchmarks demonstrate that our SegAVIR framework with and without VAE regular-

32

ization comfortably outperforms both U-Nets by a considerable margin. In particular, the *F1* score (Dice) clearly demonstrates the dominance of our method. Figure 5.6 presents a visual comparison between the output segmentation of VAE, regularized SegAVIR, and U-Net-2. As is evident by the benchmarks and these visualizations, SegAVIR has a higher rate of artery and vein detection and generally leads to better vessel delineation.

Although the U-Net has successfully segmented major arteries and veins, it struggles with smaller vessels, whose fine-grained structural details are not captured properly. Moreover, the output segmentation of the U-Net fails to maintain class consistency across the vessel.

5.2.3.2 Effect of VAE Regularization

Our benchmarks attest that VAE regularization yields a slight improvement in the performance of the model. In essence, learning to reconstruct the image, which impacts the encoder of SegAVIR, has been shown to avoid over-fitting and learn the structural details of the vessels.

5.2.3.3 Vessel Width Estimation

Using the method detailed in Algorithm 1, we have obtained the pixel-wise width profiles of segmented arteries and veins as well as their ground truth masks. The width distance map needs to be scaled with an appropriate factor to reflect the sizes in their true scale. Since every pixel in the AVIR dataset images has a length of 12.5 microns, we scale our measurements in accordance to this scaling factor. Figures 5.7 and 5.8 show visualizations of some profiles obtained for the ground truth masks and predictions, respectively.

We have also quantitatively compared the average values of the widths obtained for every branch of arteries and veins. In doing so, the width of the vessel is represented as twice the maximum distance to the medial axis. To asses the performance of our model in predicted width estimations, we have employed two common metrics, Mean Absolute

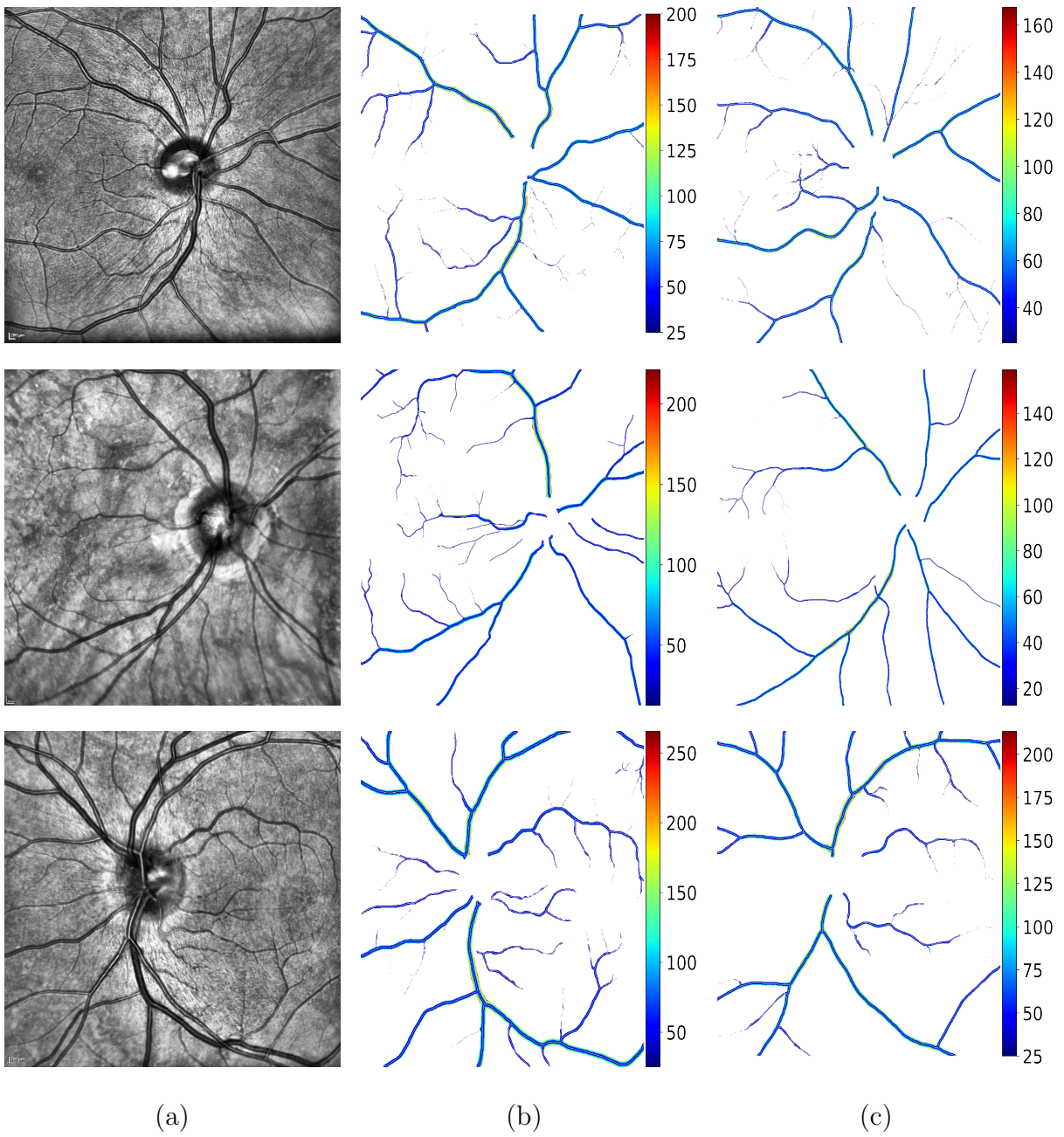Figure 5.7: Example width profile measurements of ground truth masks in the AVIR test set. All measurements are in microns. (a) Input image. (b) Width profile of veins. (c) Width profile of arteries.
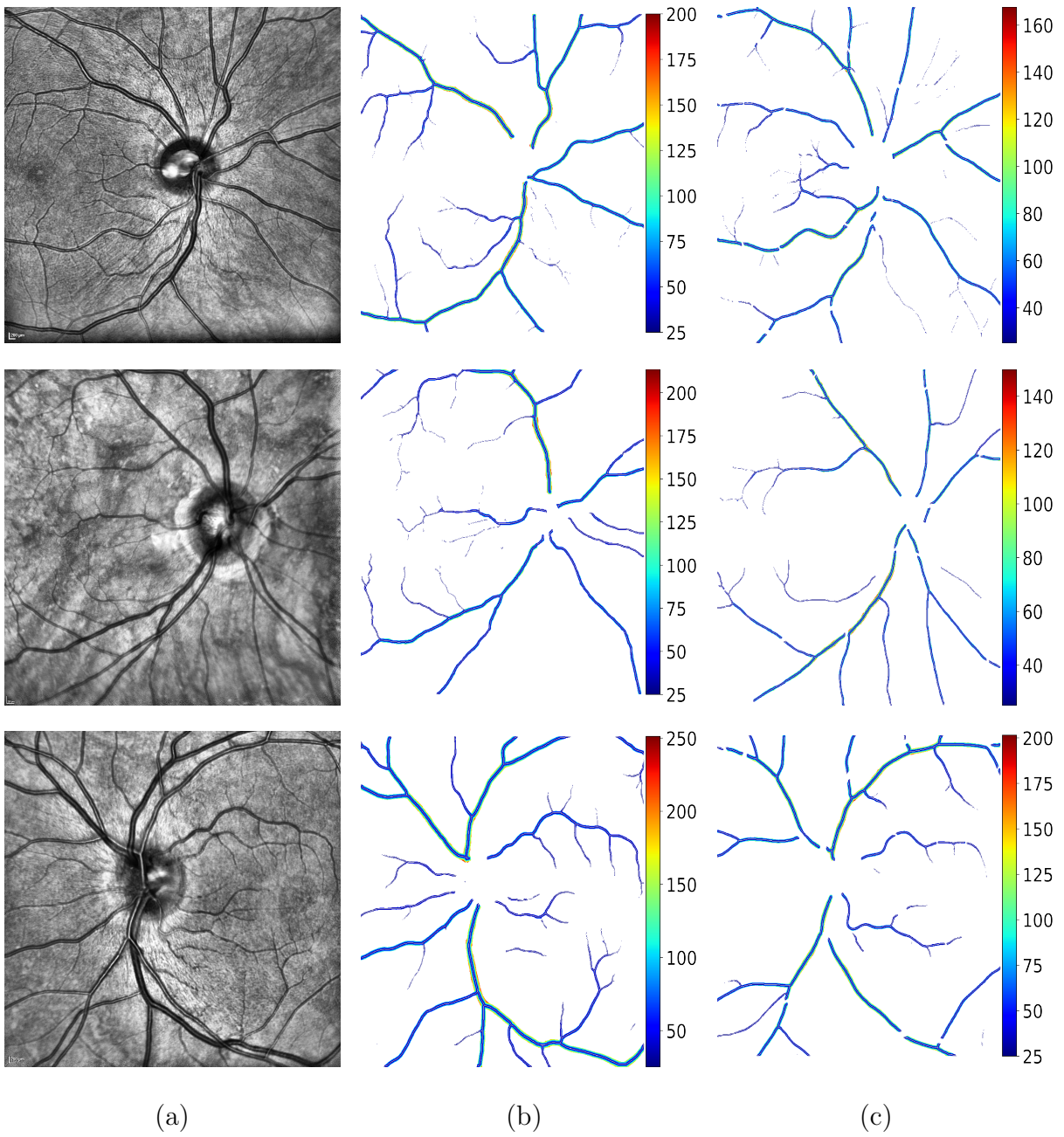
Figure 5.8: Example width profile measurements of SegAVIR predictions in the AVIR test set. All measurements are in microns. (a) Input image. (b) Width profile of veins. (c) Width profile of arteries.

Table 5.4: Comparison of average artery and vein width estimations between the ground truth masks and SegAVIR outputs. The average is estimated over patients in the test set of the AVIR dataset. All measurements are in microns.

Vessel Type	Artery	Vein
Ground Truth	92.4208	104.7287
SegAVIR	96.1492	108.7485

Table 5.5: Error analysis of average artery and vein width estimations using MAPE and RMSE metrics.

Vessel Type	Artery	Vein
MAPE (%)	9.6910	5.5515
RMSE	10.5476	7.9405

Percentage Error (MAPE) and Root Mean Square Error (RMSE), defined as

$$MAPE = \frac{1}{N} \sum_{m=1}^{N} \left| \frac{G_m - P_m}{G_m} \right| \tag{5.8}$$

and

$$RMSE = \sqrt{\frac{1}{N} \sum_{m=1}^{N} (G_m - P_m)^2}, \tag{5.9}$$

where N, P_m, and G_m denote the total number of cases, predictions, and ground truth values. The mean estimated width values of arteries and veins for ground truth and SegAVIR outputs predictions are reported in Table 5.4. Table 5.5 reports error analysis of the obtained results using the MAPE and RMSE metrics.

5.2.4 Discussion

The benchmarks presented in Tables 5.4 and 5.5 demonstrate that our network's predictions yield precise quantitative width measurements for both artery and vein classes. The average width estimation of segmented veins and arteries are 108.7485 and 96.1492 microns respectively. Using the MAPE and RMSE metrics, the estimated error for these estimations are 5.5515 and 7.9405 for veins and 9.6919 and 10.5476 for arteries. Thus,

the width estimations of veins are more accurate than arteries, following the trend of segmentation performance accuracy as presented in Tables 5.2 and 5.3. In essence, as also validated here, the accuracy of initial segmentation plays a crucial role in the correctness of quantitative width estimation.

The width profiles that have been obtained in this chapter of our significant importance for the automated or interactive tracking of the morphological changes across different branches of arteries and veins. Various anomalies such as Arteriovenous nicking can be identified in a precise manner. This provides a powerful quantitative measure for various studies that involve the early diagnosis and progress prediction of diseases such as hypertensive retinopathy.

CHAPTER 6

Conclusions and Future Work

6.1 Conclusions

This thesis has made two contributions to the automated delineation and quantification of retinal vasculature, based on color fundus imaging and infrared reflectance retinal imaging.

First, we introduced a novel, fully automated method for retinal vessel segmentation and width estimation. Our novel deep CNN employs spatial dilated pyramid pooling to recover lost content of the encoding process and leverages a multi-scale supervision of prediction maps in order to capture the smallest structural details of vessels. We tested our segmentation method on two publicly available datasets.

On the DRIVE dataset, it achieved better than state-of-the-art results in sensitivity and accuracy while being comparable in F1 score. It also achieves competitive results on the CHASE-DB1 dataset. In addition, we developed a method that employs the vessel segmentation maps to estimate the width profiles of retinal vessels.

Second, we studied the problem of semantic segmentation of retinal vessels in IR imaging. In this context, we introduced a novel dataset dubbed AVIR that entails pixel-wise ground truth labels of arteries and veins. Furthermore, we introduced the first end-to-end CNN, dubbed SegAVIR, that can directly segment retinal arteries and veins without additional post-processing.

Our extensive benchmarks demonstrated that SegAVIR can reliably segment arteries and veins and outperforms the seminal U-Net architecture and its variants by a wide margin. We also showed that regularizing SegAVIR with variational autoencoders improves

38

its performance, leading to more precise vessel delineation.

Additionally, we quantitatively studied the width profile of arteries and veins and provided pixel-wise maps that facilitate automated or interactive analysis of morphological changes.

The quantitative information automatically gleaned by our algorithms promises to be helpful to clinicians as they explore novel biomarkers and in the quantitative assessment of retinal vasculature changes associated with diseases such as diabetes and hypertension.

6.2 Future Work

The work presented in this thesis can be further developed in at least four different research directions, as follows:

1. It is a well-known fact that the predictions of CNNs are typically sub-optimal around edges. Consequently, the quality of the segmentation in these regions can be much improved. A promising approach would be to leverage the powerful nonlinear feature extraction capabilities of our SegAVIR and the versatility of energy-based techniques such as active contours in order to improve the segmentation performance particularly around vessel boundaries. Another potential remedy would be a loss function that enforce simultaneous learning of vessel textures and edges.

2. Once a reliable output segmentation map is obtained by a CNN, there are many opportunities for the automated analysis of desired patterns that have clinical significance. For instance, learning the connectivity patterns of vessels and branching is important and can be extensively studied. In this regard, an interesting solution would be to create a cascaded framework that consists of an initial encoder-decoder that outputs the probabilities of vessel locations, followed by a graph convolutional neural network that locates the nodes of desired patterns that constitute a graph. Each stage can be trained separately or in an end-to-end manner that would potentially lead to better segmentation accuracy due to the leveraging of synergies

39

between the two modules.

3. An important step to integrate the results of an AI model into a clinical pipeline is to estimate the uncertainty of predictions with the goal of providing enough information on how to reliably utilize the results. In this regard, an interesting research direction is to examine data and model imperfection by estimating aleatoric and epistemic uncertainties, respectively. For instance, the data uncertainty can be estimated by averaging the entropy of each sampled probability distribution, whereas model uncertainty is captured by Monte Carlo dropout—dropout layers function as approximate Bayesian inference in deep Gaussian processes. As such, it is important to investigate both data and model uncertainty to identify the best measure for highlighting the mis-segmentation of vessels.

4. Our framework for the automated delineation and quantification of vessels can be readily integrated with clinical studies aimed towards the discovery of novel biomarkers for the early detection of hypertensive retinopathy. With the availability of more temporal data, our framework could be augmented with features that make it useful in the early prediction of this disease.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). TensorFlow: A system for large-scale machine learning. In *OSDI*, volume 16, pages 265–283. 22

Al-Diri, B., Hunter, A., and Steel, D. (2009). An active contour model for segmenting and measuring retinal vessels. *IEEE Transactions on Medical imaging*, 28(9):1488–1497. 7

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., and Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *Journal of Medical Imaging*, 6(1):014006. 8, 24

Andreini, P., Bonechi, S., Bianchini, M., Mecocci, A., Scarselli, F., and Sodi, A. (2019). A two stage gan for high resolution retinal image generation and segmentation. *arXiv preprint arXiv:1907.12296.* 10

Dashtbozorg, B., Mendonça, A. M., and Campilho, A. (2013). An automatic graph-based approach for artery/vein classification in retinal images. *IEEE Transactions on Image Processing*, 23(3):1073–1083. 11

Delibasis, K. K., Kechriniotis, A. I., Tsonos, C., and Assimakis, N. (2010). Automatic model-based tracing algorithm for vessel segmentation and diameter estimation. *Computer Methods and Programs in Biomedicine*, 100(2):108–122. 6

Espona, L., Carreira, M. J., Ortega, M., and Penedo, M. G. (2007). A snake for retinal vessel segmentation. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 178–185. Springer. 7

Estrada, R., Allingham, M. J., Mettu, P. S., Cousins, S. W., Tomasi, C., and Farsiu, S. (2015). Retinal artery-vein classification via topology estimation. *IEEE Transactions on Medical Imaging*, 34(12):2518–2534. 11

Francia, G. A., Pedraza, C., Aceves, M., and Tovar-Arriaga, S. (2020). Chaining a U-Net with a residual U-Net for retinal blood vessels segmentation. *IEEE Access*, 8:38493–38500. 9

Fu, H., Xu, Y., Lin, S., Wong, D. W. K., and Liu, J. (2016). Deepvessel: Retinal vessel segmentation via deep learning and conditional random field. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 132–139. Springer. 8, 24

Girard, F., Kavalec, C., and Cheriet, F. (2019). Joint segmentation and classification of retinal arteries/veins from fundus images. *Artificial Intelligence in Medicine*, 94:96–109. 10, 11

Guo, Y. and Peng, Y. (2020). BSCN: Bidirectional symmetric cascade network for retinal vessel segmentation. *BMC Medical Imaging*, 20(1):1–22. 9

Hatamizadeh, A., Hoogi, A., Sengupta, D., Lu, W., Wilcox, B., Rubin, D., and Terzopoulos, D. (2019a). Deep active lesion segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 98–105. Springer. 3

Hatamizadeh, A., Hosseini, H., Liu, Z., Schwartz, S. D., and Terzopoulos, D. (2019b). Deep dilated convolutional nets for the automatic segmentation of retinal vessels. *arXiv preprint arXiv:1905.12120.* 4

Hatamizadeh, A., Hosseini, H., Liu, Z., Schwartz, S. D., and Terzopoulos, D. (2019c). Deep dilated convolutional nets for the automatic segmentation of retinal vessels. In *Proc. 15th International Conference on Machine Learning and Data Mining (MLDM'19)*, pages 39–48. 4

Hatamizadeh, A., Terzopoulos, D., and Myronenko, A. (2019d). End-to-end boundary aware networks for medical image segmentation. In *International Workshop on Machine Learning in Medical Imaging*, pages 187–194. Springer. 3

Hatamizadeh, A., Terzopoulos, D., and Myronenko, A. (2020). Edge-gated CNNs for volumetric semantic segmentation of medical images. *arXiv preprint arXiv:2002.04207.* 3

Huang, F., Dashtbozorg, B., Tan, T., and ter Haar Romeny, B. M. (2018). Retinal artery/vein classification using genetic-search feature selection. *Computer Methods and Programs in Biomedicine*, 161:197–207. 11

Jiang, Y., Zhang, H., Tan, N., and Chen, L. (2019). Automatic retinal blood vessel segmentation based on fully convolutional neural networks. *Symmetry*, 11(9):1112. 8

Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., and Su, R. (2019). DUNet: A deformable network for retinal vessel segmentation. *Knowledge-Based Systems*, 178:149–162. 8, 9, 24

Joshi, V. S., Reinhardt, J. M., Garvin, M. K., and Abramoff, M. D. (2014). Automated method for identification and artery-venous classification of vessel trees in retinal vessel networks. *PloS One*, 9(2). 11

Karn, P. K., Biswal, B., and Samantaray, S. R. (2018). Robust retinal blood vessel segmentation using hybrid active contour model. *IET Image Processing*, 13(3):440–450. 7

Kass, M., Witkin, A., and Terzopoulos, D. (1988). Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331. 7

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* 29

Laibacher, T., Weyde, T., and Jalali, S. (2018). M2U-Net: Effective and efficient retinal vessel segmentation for resource-constrained environments. *arXiv preprint arXiv:1811.07738.* 24

Li, L., Verma, M., Nakashima, Y., Nagahara, H., and Kawasaki, R. (2020). IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks. In *The IEEE Winter Conference on Applications of Computer Vision (WACV).* 9, 10

Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., and Wang, T. (2016). A cross-modality learning approach for vessel segmentation in retinal images. *IEEE Transactions on Medical Imaging*, 35(1):109–118. 24

Lin, Y., Zhang, H., and Hu, G. (2018). Automatic retinal vessel segmentation via deeply supervised and smoothly regularized network. *IEEE Access*, 7:57717–57724. 8

Liskowski, P. and Krawiec, K. (2016). Segmenting retinal blood vessels with deep neural networks. *IEEE Transactions on Medical Imaging*, 35(11):2369–2380. 24

Lupascu, C. A., Tegolo, D., and Trucco, E. (2010). FABC: Retinal vessel segmentation using adaboost. *IEEE Transactions on Information Technology in Biomedicine*, 14(5):1267–1274. 6

McInerney, T. and Terzopoulos, D. (2000). T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4(2):73–91. 7

Melinščak, M., Prentašić, P., and Lončarić, S. (2015). Retinal vessel segmentation using deep neural networks. In *VISAPP 2015 (10th International Conference on Computer Vision Theory and Applications).* 8, 24

Myronenko, A. (2018). 3D MRI brain tumor segmentation using autoencoder regularization. In *BrainLes, Medical Image Computing and Computer Assisted Intervention (MICCAI)*, LNCS, pages 311–320. Springer. 19, 20

Myronenko, A. and Hatamizadeh, A. (2019). 3D kidneys and kidney tumor semantic segmentation using boundary-aware networks. *arXiv preprint arXiv:1909.06684.* 3, 29

Myronenko, A. and Hatamizadeh, A. (2020). Robust semantic segmentation of brain tumor regions from 3D MRIs. *arXiv preprint arXiv:2001.02040.* 3

Nguyen, U. T., Bhuiyan, A., Park, L. A., and Ramamohanarao, K. (2013). An effective retinal blood vessel segmentation method using multi-scale line detection. *Pattern Recognition*, 46(3):703–715. 6

Oliveira, A., Pereira, S., and Silva, C. A. (2018). Retinal vessel segmentation based on fully convolutional neural networks. *Expert Systems with Applications*, 112:229–242. 8, 24, 26

Relan, D., MacGillivray, T., Ballerini, L., and Trucco, E. (2013). Retinal vessel classification: Sorting arteries and veins. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 7396–7399. IEEE. 11

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer. 8, 31, 32

Singh, N. and Kaur, L. (2015). A survey on blood vessel segmentation methods in retinal images. In *2015 International Conference on Electronic Design, Computer Networks & Automated Verification (EDCAV)*, pages 23–28. IEEE. 6

Soares, J., Cree, M., Leandro, J., Cesar, R., and Jelinek, H. (2005). Retinal vessel segmentation using the 2-D morlet wavelet and supervised classification. Technical report, cs/0510001. 6

Son, J., Park, S. J., and Jung, K.-H. (2019). Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. *Journal of digital imaging*, 32(3):499–512. 10

Sum, K. and Cheung, P. Y. (2007). Vessel extraction under non-uniform illumination: A level set approach. *IEEE Transactions on Biomedical Engineering*, 55(1):358–360. 7

Vázquez, S., Barreira, N., Penedo, M. G., Ortega, M., and Pose-Reino, A. (2010). Improvements in retinal vessel clustering techniques: Towards the automatic computation of the arterio venous ratio. *Computing*, 90(3-4):197–217. 11

Wang, B., Qiu, S., and He, H. (2019a). Dual encoding U-Net for retinal vessel segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 84–92. Springer. 9

Wang, C., Zhao, Z., Ren, Q., Xu, Y., and Yu, Y. (2019b). Dense U-net based on patch-based learning for retinal vessel segmentation. *Entropy*, 21(2):168. 8

Wu, C., Zou, Y., and Yang, Z. (2019a). U-GAN: Generative adversarial networks with U-Net for retinal vessel segmentation. In *2019 14th International Conference on Computer Science & Education (ICCSE)*, pages 642–646. IEEE. 10

Wu, C., Zou, Y., and Zhan, J. (2019b). DA-U-Net: Densely connected convolutional networks and decoder with attention gate for retinal vessel segmentation. In *IOP Conference Series: Materials Science and Engineering*, volume 533, page 012053. IOP Publishing. 10

Wu, Y., Xia, Y., Song, Y., Zhang, D., Liu, D., Zhang, C., and Cai, W. (2019c). Vessel-Net: Retinal vessel segmentation under multi-path supervision. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 264–272. Springer. 8, 24

Xu, L. and Luo, S. (2010). A novel method for blood vessel detection from retinal images. *Biomedical Engineering Online*, 9(1):14. 6

Yan, Z., Yang, X., and Cheng, K.-T. (2018). A three-stage deep learning model for accurate retinal vessel segmentation. *IEEE journal of Biomedical and Health Informatics*, 23(4):1427–1436. 9

Zana, F. and Klein, J.-C. (2001). Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on Image Processing*, 10(7):1010–1019. 6

Zhang, S., Fu, H., Yan, Y., Zhang, Y., Wu, Q., Yang, M., Tan, M., and Xu, Y. (2019a). Attention guided network for retinal image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 797–805. Springer. 10

Zhang, T. Y. and Suen, C. Y. (1984). A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239. 15

Zhang, Z., Fu, H., Dai, H., Shen, J., Pang, Y., and Shao, L. (2019b). ET-Net: A generic edge-attention guidance network for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–450. Springer. 10

Zhuang, J. (2018). LadderNet: Multi-path networks based on U-Net for medical image segmentation. *arXiv preprint arXiv:1810.07810*. 8, 24