A Decision Support Tool for Video Retinal Angiography

2018

Sumit Laha

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

 Part of the Computer Sciences Commons

STARS Citation

Laha, Sumit, "A Decision Support Tool for Video Retinal Angiography" (2018). *Electronic Theses and Dissertations*. 6201.
https://stars.library.ucf.edu/etd/6201

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.

A DECISION SUPPORT TOOL FOR VIDEO RETINAL ANGIOGRAPHY

by

SUMIT LAHA
B.Tech. Future Institute of Engineering and Management, 2015

A thesis submitted in partial fulfilment of the requirements
for the degree of Master of Science
in the Department of Computer Science
in the College of Engineering & Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2018

Major Professor: Ulas Bagci

ABSTRACT

Fluorescein angiogram (FA) is a medical procedure that helps the ophthalmologists to monitor the status of the retinal blood vessels and to diagnose proper treatment. This research is motivated by the necessity of blood vessel segmentation of the retina. Retinal vessel segmentation has been a major challenge and has long drawn the attention of researchers for decades due to the presence of complex blood vessels with varying size, shape, angles and branching pattern of vessels, and non-uniform illumination and huge anatomical variability between subjects. In this thesis, we introduce a new computational tool that combines deep learning based machine learning algorithm and a signal processing based video magnification method to support physicians in analyzing and diagnosing retinal angiogram videos for the first time in the literature.

The proposed approach has a pipeline-based architecture containing three phases - image registration for large motion removal from video angiogram, retinal vessel segmentation and video magnification based on the segmented vessels. In image registration phase, we align distorted frames in the FA video using rigid registration approaches. In the next phase, we use baseline capsule based neural networks for retinal vessel segmentation in comparison with the state-of-the-art methods. We move away from traditional convolutional network approaches to capsule networks in this work. This is because, despite being widely used in different computer vision applications, convolutional neural networks suffer from learning ability to understand the object-part relationships, have high computational times due to additive nature of neurons and, loose information in the pooling layer. Although having these drawbacks, we use deep learning methods like *U-Net* and *Tiramisu* to measure the performance and accuracy of *SegCaps*. Lastly, we apply Eulerian video magnification to magnify the subtle changes in the retinal video. In this phase, magnification is applied to segmented videos to visualize the flow of blood in the retinal vessels.

Dedicated to my parents

ACKNOWLEDGMENTS

I would like to convey my sincere gratitude to my advisor Dr. Ulas Bagci for his continuous guidance and support throughout my master thesis. I would also like to express my gratitude to our collaborators, Dr. Saad Sheikh from the University of Central Florida College of Medicine, Orlando, FL and staffs from Central Florida Retina, Orlando, FL.

I would like to thank Mert Tuna Ozerdem for providing valuable insight into image registration and Rodney LaLonde for image segmentation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AMD Age-related macular degeneration.

ANN Artificial neural network.

CNN Convolutional neural network.

DSC Dice similarity coefficient.

EVM Eulerian video magnification.

FA Fluorescein angiogram.

FCN Fully convolutional network.

FN False negative.

FP False positive.

IRB Institutional review board.

ITK Insight toolkit.

LoG Laplacian of Gaussian.

MSTV Multi scale tensor voting.

OCTA Optical coherence tomography angiography.

ReLU Rectified linear unit.

ROC Receiver operating characteristics.

TN True negative.

TP True positive.

CHAPTER 1: INTRODUCTION

Fluorescein angiogram (FA) is a medical procedure that helps the ophthalmologists to diagnose proper treatment or monitor the status of the retinal blood vessels [1]. In this procedure, a fluorescein dye is injected into the patient's bloodstream to highlight the blood vessels in the retina. A dedicated fundus camera is used to capture the images of the blood vessels to examine the retinal vasculature structure in detecting any abnormalities in the retinal pathology like glaucoma, age-related macular degeneration (AMD) or diabetic retinopathy.

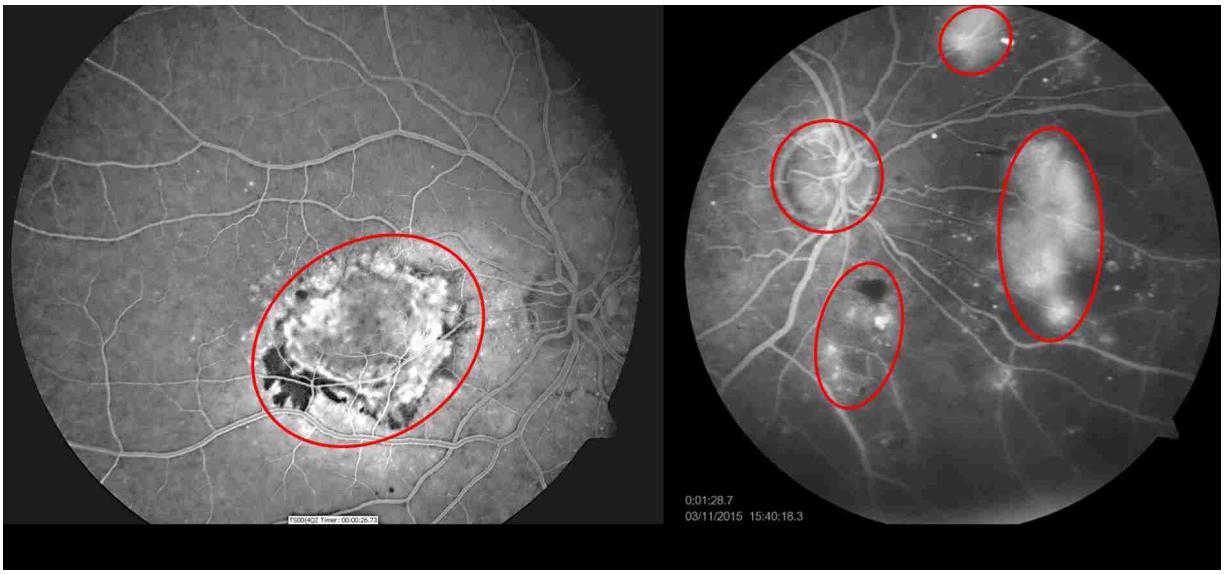


Figure 1.1: FA image of a subject having (a) Age-related macular degeneration [2] and (b) Proliferative diabetic retinopathy [4].

AMD is a disorder that affects the macula in the center of the retina for the people over the age of 55 [2]. The macula is an essential part of the eye, which is responsible for reading and recognizing faces. Figure 1.1 (a) shows the FA where the black area marked by the red circle corresponds to the hemorrhage.

Diabetic patients can have an eye disease called diabetic retinopathy [3]. This occurs due to high blood sugar levels causing damage like swelling, leaking or, closing of the retinal blood vessels. These changes affect the vision. Figure 1.1 (b) shows affected regions marked by red circles of a typical case of proliferative diabetic retinopathy.

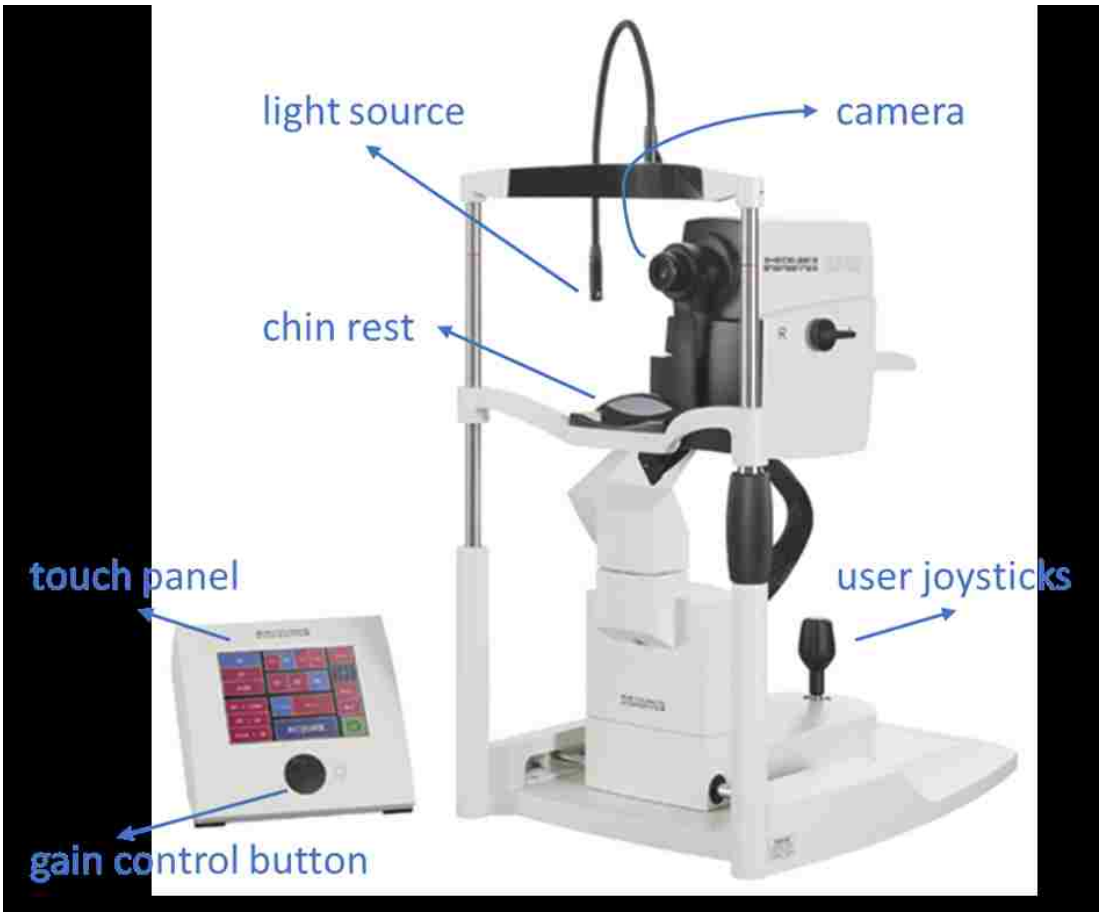


Figure 1.2: Fundus camera (SPECTRALIS OCT) *(Courtesy of Heidelberg Engineering Inc.)*.

A fundus camera as shown in Figure 1.2, can be considered as a low power microscope for the purpose of capturing retina fundus imaging, where the retina is illuminated for imaging by means of the attached camera. Specifically, the fundus camera is used to take an image for the interior parts of the human eye including macula, optic disk, retina and posterior pole [1].

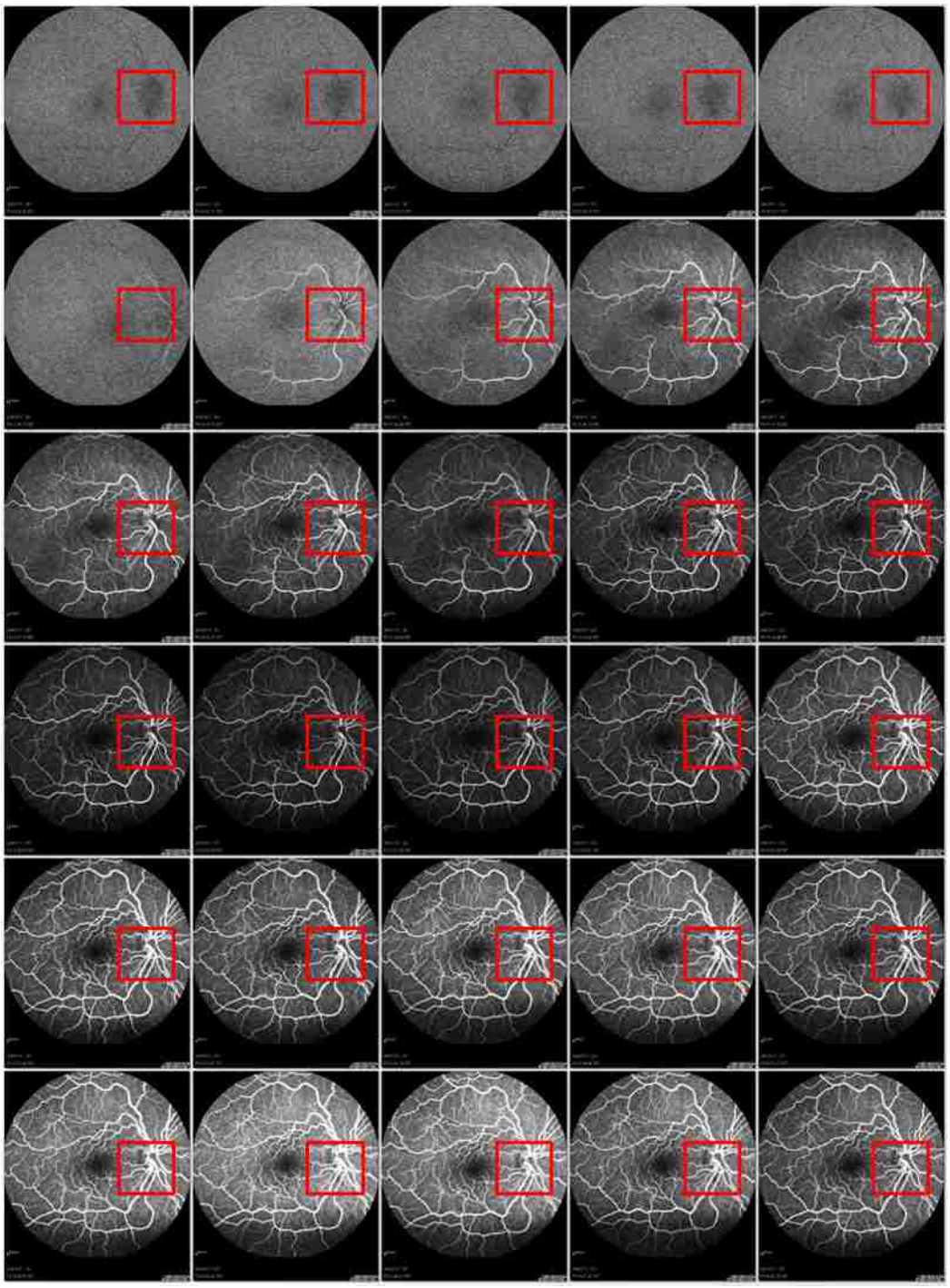


Figure 1.3: The region within the red square box highlighting the presence of motion and noise in FA video frames.

Our research is motivated by the necessity of retinal blood vessel segmentation, which has a remarkable place in the medical image segmentation literature. The blood vessel segmentation has long drawn the attention of researchers and doctors due to the presence of complex features in the retinal vessel structure with respect to the size, shape, angles and branching patterns of vessels, non-uniform illumination and anatomical variability between the subjects [13].

Some of the critical challenges in FA video analysis include motion, noise, complex shape nature of vessels, lack of validated ground truths to train machine learning algorithm and difficulty in blood flow analysis. To our best of knowledge, there is no quantitative tool for analyzing these images. One of the biggest challenges in FA analysis is the inevitable motion between frames of FA video. Because the operator is taking the videos by hand, and there is a presence of huge motion in the videos when doctors analyze the same. Figure 1.3 displays the frames of a FA video of a normal subject. The red box in the figure indicates that the blood vessels are not aligned in the frames due to the motion. This causes the frames to be either translated, rotated or both.

Another challenge is the noise. Most of the areas in the FA video frames are very noisy as shown in Figure 1.3. To develop an image analysis tool, noise should be carefully considered, especially when identifying the vessel structures for quantification purpose. Noise is a widely known problem in image processing field for many decades.

The tree-like appearances of blood vessels is a major hurdle too. Segmenting such structures is relatively more difficult than closed objects (i.e., Jordan Surface) with the same resolution settings. Figure 1.4 exhibits that similar tree-like appearances can be found in bronchial airways [5] as well.

Physicians can analyze the blood vessels quantitatively if the length, shape, and volume information of the vessels are available through image segmentation. However, extracting the image segmentation mask, which we call ground truth for training machine learning algorithms, is tedious and involves a lot of effort. Its reproducibility is very low because when an operator repeats the

experiment to make some annotations, there will be inter-operator and intra-operator variations. In other words, many segmentation problems require precisely labeled output. A small segmentation error can lead to major problems in diagnosing the diseases and other clinical evaluations. Therefore, the automated segmentation method of retinal vessels still remains a challenging task and has been a major research issue.

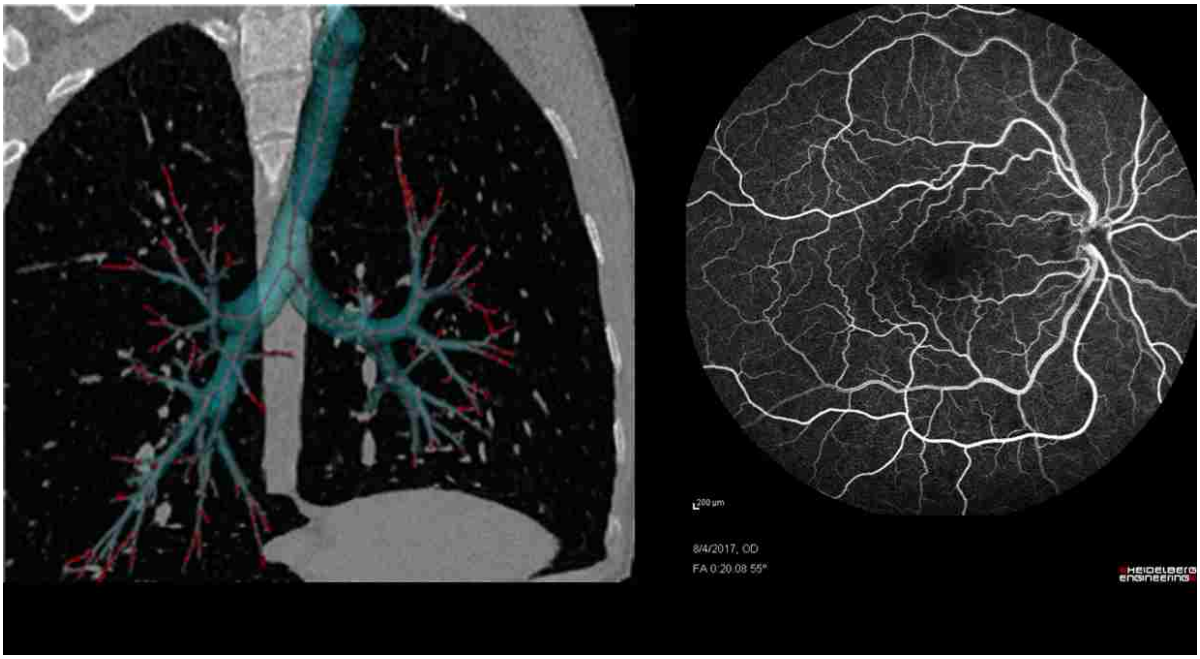


Figure 1.4: Tree-like appearances present in both (a) Bronchial airways [5] and (b) Retinal blood vessels.

Since motion and noise are two important challenges in images, blood flow analysis is not largely adopted in clinics. One of the hypotheses of incorporating computational tools is to design a computational framework for accurate blood flow analysis. This may follow some disease pattern identification too. At this stage, we do not test the clinical utility of such a tool but propose to develop a proof of concept for the first time in the literature. In other words, we make an attempt to prepare something in case doctors are interested in finding the dynamics of the blood flow. Currently, optical coherence tomography angiography (OCTA) images provide dynamic

blood flow information. However, OCTA is not widely used for vessel segmentation analysis.

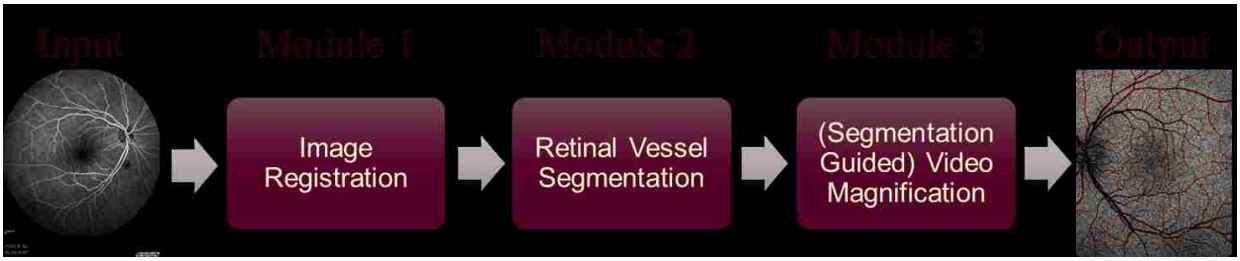


Figure 1.5: Schematic diagram of our proposed approach.

In this research work, we introduce a new tool which has three modules, namely, image registration, retinal blood vessel segmentation, and segmentation guided video magnification as shown in Figure 1.5. The main contribution of the first module is serial rigid registration with binary images instead of grayscale images. The binary images are created through rough thresholding. During registration, a middle frame is chosen as a reference in order to register all the remaining frames to that one. The objective of this module is to get rid of the motion between the frames and some of the noise.

The next module involves the segmentation process. To the best of our knowledge, there is no deep learning based segmentation in FA video study and our work is the first ever deep learning based segmentation in FA videos. We use a capsule-based deep learning approach [6] which is a very recent work in the deep learning area. We also perform an extensive comparison with the baseline networks like *U-Net* [10] and *Tiramisu* [12].

In the final module, we explore Eulerian video magnification [8] on retinal blood vessels. We use a constrained based magnification approach. Instead of using the entire frame, we use segmentation as a prior i.e., we only magnify the blood vessels.

The organization of this thesis is as follows. Section 2 briefly reviews the existing literature on

the applications of different segmentation methods including deep learning to various medical images and the fundus images. Section 3 describes the details of deep learning based segmentation methods in biomedical images. Then, Section 4 describes the proposed approach. Section 5 reports the results of the computational experimentation. Finally, Section 6 summarizes our findings and some fruitful future research directions.

CHAPTER 2: LITERATURE REVIEW

Over the last two decades, a considerable group of researchers has developed and implemented different vessel segmentation methodologies. These techniques can be broadly classified into two categories: supervised [15, 16, 17, 18, 19] and unsupervised [20, 21, 22, 23]. More recent supervised segmentation methods are based on mostly deep neural networks [15, 16, 17, 18, 19]. On the other hand, matched filtering [20], multi-scale approach [21], vessel tracing [22], and thresholding-based approaches [23] are some of the important approaches in the unsupervised segmentation category, pertaining to pre-deep learning era. The supervised segmentation methods rely on a labeled set of training images, whereas, the unsupervised segmentation methods are based on without having any prior knowledge and labeled responses. Usually, the unsupervised methods are faster and lesser time complexity than the supervised methods. However, the supervised methods perform better than the unsupervised methods. Recent research show detailed surveys of the blood vessel segmentation methodologies in retinal fundus images [1, 13, 24].

Currently, convolutional neural networks (CNNs) in deep learning have emerged as the most powerful learning algorithms compared to other machine learning algorithms and perform very well in a variety of fields. However, these deep learning networks suffer from some limitations such as limited learning ability to understand the object-part relationships and requiring higher computational times due to the scalar and additive nature of neurons in CNNs [6]. Additionally, the CNN is losing valuable information as well despite it performs very well by using its max-pooling component. Recently, as an alternative method to the CNNs, a new version of deep learning networks, called capsule networks (*CapsNet*) [7] have attracted growing interest of researchers and have emerged as competitive deep learning tools. These *CapsNet* have shown significant remarkable performance compared to CNNs in many computer science applications including digit recognition and image classification [7]. Recently, the *CapsNet* has been successfully implemented in various medical

image applications [25, 26, 27], and other applications [28].

Table 2.1: Summary of recently used retinal vessel segmentation techniques [1].

Year	Method	Performance Measure	Dataset
2015	Modified Gaussian matched filter + Entropy thresholding [29]	Acc, Sp, Se	DRIVE [40]
2015	Snakes contours [30]	Acc, Sp, Se	DRIVE
2015	Level set without using local region area [31]	Acc, Sp, Se	DRIVE
2015	Fuzzy Logic [32]	Acc	DRIVE
2016	Filter Kernel: Laplacian of Gaussian [33]	Acc, Sp, Se	DRIVE, STARE [41]
2016	Local adaptive thresholding based on multi-scale tensor [34]	Acc, Sp, Se	Erlangen Dataset [42]
2016	Ensemble of 12 convolutional ANNs [35]	Acc	DRIVE
2016	Deep Convolutional ANNs [18]	Area under Recall-Precision Curve	DRIVE, STARE
2016	Deep ANNs [16]	ROC, Acc	DRIVE, STARE
2017	Global thresholding based on morphological operations [36]	Acc, Execution time	DRIVE, STARE
ROC: Receiver Operating Characteristics; Acc: Accuracy; Sp: Specificity; Se: Sensitivity; ANN: Artificial Neural Network.			

There exists many segmentation tools available in the medical image processing and retinal vessel structure segmentation literature [1, 16, 35, 29, 30]. Each retinal vessel segmentation tool consists of three common phases such as pre-processing, processing and post-processing. Apart from the above taxonomy of supervised and unsupervised methods, the vessel structure segmentation techniques used in the processing phase can be classified into two major categories: rule-based and machine learning. Rule-based techniques include kernel-based [29], parametric deformable modeling [30, 31], mathematical morphology-based [36], and multi-scale adaptive local thresholding [34], whereas, machine learning tools include ANNs and CNNs [16, 18, 35]. A summary of comparison of recently used different retinal vessel segmentation techniques are given in Table

2.1.

The following four performance measures, namely, sensitivity, specificity, accuracy, and precision are popular in the medical image processing literature including retinal vessel segmentation [1, 29, 34]. These metrics are defined as follows.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2.1}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{TP}} \tag{2.2}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}} \tag{2.3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}\,, \tag{2.4}$$

where, True Positive (TP) is the number of pixels correctly segmented as blood vessels, True Negative (TN) is the number of pixels correctly detected as background, False Positive (FP) is the number of pixels falsely segmented as blood vessels and False Negative (FN) is the number of pixels falsely detected as background.

Moreover, many researchers use the performance metric as receiver operating characteristics (ROC) curve [16], especially, for the techniques that rely on particular parameters during the segmentation process.

Singh et al. [29] present a segmentation technique based on Gaussian matched filter with a Gaussian kernel for the retinal vascular structure and implement it by means of DRIVE dataset. Through the generated segmentation results, they show that their technique produces the performance met-

rics of 0.9387, and 0.9647 and 0.6721 for average accuracy, average specificity, and average sensitivity, respectively.

Jin et al. [30] propose a snake-based segmentation technique in the category of parametric deformable modeling with the objective of fast convergence of the snakes to the correct vessel edges in some situations like the presence of high noise in the video, empty blood vessels, and low contrast levels.

Gongt et al. [31] propose a novel level set technique in the category of geometry deformable models for retinal image vessel segmentation. The proposed technique is based on local cluster value through bias correction. This method provides more local intensity information at the image pixel level and does not require initialization of level set function. They test their method on DRIVE dataset and report an accuracy of 0.9360, sensitivity of 0.7078 and specificity of 0.9699.

Sharma and Wasson [32] develop a fuzzy-based retinal vessel segmentation method. They use the fuzzy-logic processing considering the difference of low-pass and high-pass filters in the retinal image. The fuzzy logic comprises different sets of fuzzy rules where each fuzzy rule is constructed based on different thresholding values to select and discard pixel values, leading to vessel extraction. Their methodology yields an average accuracy of 95% on DRIVE dataset.

Kumar et al. [33] propose an algorithm based on the inherent zero-crossing property of Laplacian of Gaussian (LoG) filter for the detection of retinal vascular structure for the fundus images. In their algorithm, they apply two-dimensional matched filters with LoG kernel functions. Their proposed methodology achieves the performance measures of 0.9626, 0.7006 and 0.9871 for average accuracy, sensitivity, and specificity respectively.

Christodoulidis et al. [34] propose a scheme based on Multi Scale Tensor Voting (MSTV) for the segmentation of small thin vessels. The proposed technique consists of four major phases, namely,

pre-processing, multiscale vessel enhancement, adaptive thresholding and MSTV processing, and post-processing. The proposed methodology is tested Erlangen dataset [42] and produces average measures of 94.79%, and 85.06% and 95.82% for the accuracy, sensitivity, and specificity respectively.

Maji et al. [35] propose a deep neural network technique by means of an ensemble of twelve CNNSs to discriminate between vessel pixels from non-vessel ones. Each CNN contains three convolutional layers and each one is trained individually based on a set of 60,000 randomly 31 x 31 x 31-sized patches taken from 20 raw color retinal images of DRIVE dataset. Their technique shows superior performance in terms of learning vessel presentation from data than one neural network because it is more powerful and accurate due to the ensemble of many CNNS.

Maninis et al. [18] implement a fast and accurate automated supervised deep CNN algorithm for the segmentation of retinal and optic disc structures. The proposed algorithm is validated on DRIVE and STARE datasets, resulting in an average accuracy of 0.822 for DRIVE dataset and 0.831 for STARE dataset.

Liskowski et al. [16] also present a CNN-based algorithm for segmenting retinal vessels. They show through the results of their algorithm an accuracy of 0.9533 and the area under supremum curve of 0.974.

Jiang et al. [36] apply a mathematical morphology-based technique using global thresholding operations to segment the retinal vascular structure. The technique is tested on DRIVE and STARE datasets and attains an average accuracy of 95.88% for single dataset test and 95.27% for the cross-validation dataset test.

The literature survey reveals that there have been several studies on the applications of various retinal vessels segmentation methods including deep neural networks and CNNs. However, the

segmentation technique based on the *CapsNet* [7] was not reported in the vessels segmentation literature. Therefore, this research is motivated to verify the suitability of the potential applications of the *CapsNet* in retinal blood vessels segmentation and to compare its performance with some state-of-the-art segmentation methods.

CHAPTER 3: DEEP LEARNING FOR SEGMENTATION

There have been numerous studies applying deep learning, demonstrating its superiority over other image segmentation methods in various fields, including in biomedical imaging. *SegNet* [9], *U-Net* [10] and Fully Convolutional Network (*FCN*) [11] are some of the widely used deep learning algorithms for segmentation. Recently, LaLonde and Bagci propose a new method [6] considering a new semantic segmentation based on capsule networks.

3.1 U-Net

The *U-Net* model [10, 38] is the most widely used image segmentation algorithm. It is based on the *FCN*. It is modified to support better segmentation in medical imaging. It differs from the FCN-8 architecture in the following ways. Firstly, the *U-Net* is symmetric. Secondly, instead of a summation operator, the skip connections between the down-sampling and up-sampling path use a concatenation operator. The skip connections provide local information to the global information during up-sampling. Since the network is symmetric, it consists of a large number of feature maps in the up-sampling path. This helps in transferring information. On the other hand, the basic *FCN* architecture contains ”number of classes” as feature maps in the up-sampling path.

The model is named due to its symmetric shape as illustrated in Figure 3.1. It contains three parts - contracting or down-sampling path, bottleneck and expanding or up-sampling path.

The down-sampling path consists of four blocks. Each block contains 3×3 convolution layer and activation function (with batch normalization), another 3 convolution layer and activation function (with batch normalization), and finally, a 2×2 max pooling layer. The number of feature maps increases twice in size at each pooling layer. It starts with 64 feature maps in the first block, then

another 128 for the second one, and so on. This down-sampling path captures the context of the input image. This helps in the segmentation process. The skip connections are used to transfer the coarse contextual information to the up-sampling path.

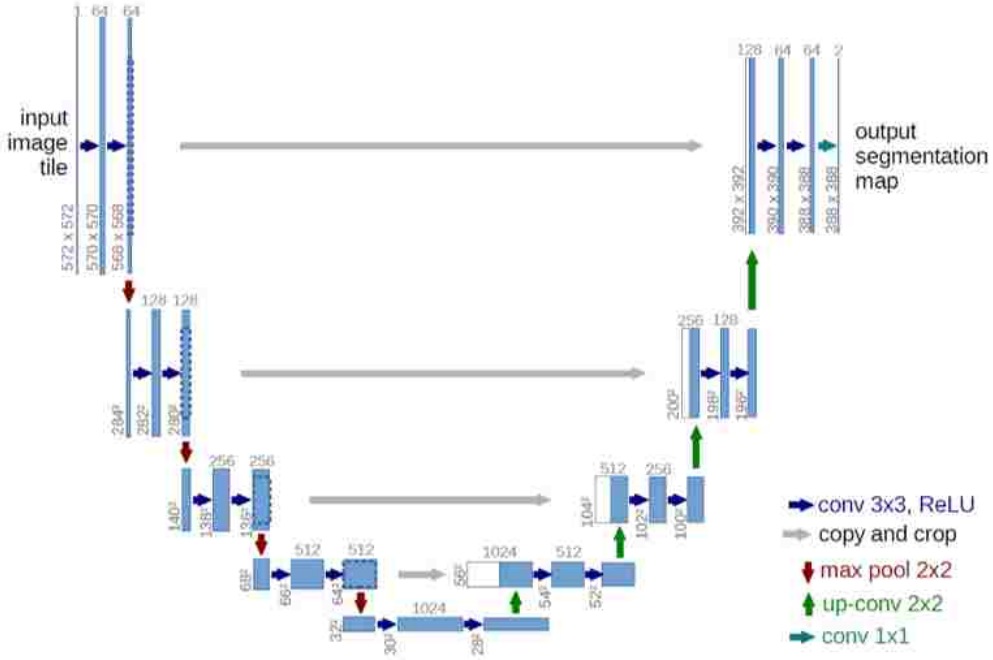


Figure 3.1: Schematic diagram of a typical *U-Net* architecture [10].

The next part of the *U-Net* architecture is the bottleneck. It connects the down-sampling and the up-sampling path. It consists of two convolutional layers with batch normalization and dropout.

The final part of this model is the up-sampling path. Like the down-sampling path, it consists of four blocks. However, these blocks are different from the down-sampling path. Each block contains a deconvolutional layer with stride 2, concatenation of the skip connections with the corresponding cropped feature map from the down-sampling path and, two 3×3 convolutional layers and activation functions with batch normalization. The up-sampling path enables precise localization combined with contextual information from the down-sampling path.

The model obtains the local information from the down-sampling path and combines the contextual information in the up-sampling path. This produces a general information which helps in the prediction of a good segmentation map. Moreover, being devoid of dense layers, the model can take images of different sizes as input.

3.2 Tiramisu

Although the *U-Net* is good in segmentation, it uses a lot of parameters compared to *Tiramisu* model [12]. The name is derived from the tiramisu dessert because of its many layers. In fact, this model contains 103 layers. It is a variant of *U-Net* architecture with down-sampling, bottleneck, and up-sampling paths and skip connections. Instead of using the convolution and max-pooling layers, it uses dense blocks from the DenseNet architecture.

Table 3.1: Building blocks of *Tiramisu* model. From left to right: layer used in the model, Transition Down (TD) and Transition Up (TU) [12].

Layer
Batch Normalization
ReLU
3×3 Convolution
Dropout $p = 0.2$

Transition Down (TD)
Batch Normalization
ReLU
1×1 Convolution
Dropout $p = 0.2$
2×2 Max Pooling

Transition Up (TU)
3×3 Transposed Convolution $stride = 2$

Table 3.1 provides details about the dense block layer, transition down and transition up of the network. The dense block layer consists of batch normalization, followed by a rectified linear unit (ReLU), a 3×3 convolution and a dropout with a probability of 0.2. The growth rate (k) of the layer is fixed at 16. The transition down consists of batch normalization, followed by a rectified linear unit (ReLU), a 1×1 convolution, a dropout with a probability of 0.2 and a max pooling of size 2×2. The transition up consists of a 3×3 transposed convolution with a stride of 2. This

helps in compensation of the pooling operation.

Table 3.2: Architecture details of *Tiramisu* model containing 103 layers [12].

Architecture
Input, $m = 3$
3×3 Convolution, $m = 48$
DB (4 layers) + TD, $m = 112$
DB (5 layers) + TD, $m = 192$
DB (7 layers) + TD, $m = 304$
DB (10 layers) + TD, $m = 464$
DB (12 layers) + TD, $m = 656$
DB (15 layers), $m = 896$
TU + DB (12 layers), $m = 1088$
TU + DB (10 layers), $m = 816$
TU + DB (7 layers), $m = 578$
TU + DB (5 layers), $m = 384$
TU + DB (4 layers), $m = 256$
1×1 Convolution, $m = c$
Softmax

All the layers in the *Tiramisu* model are summarized in Table 3.2. The network contains 103 convolutional layers. The first layer is the input layer. It is followed by 38 layers in the down-sampling path, 15 layers in the bottleneck and another 38 layers in the up-sampling path. The model uses 5 Transition Down (TD), each of which has one extra convolution, and another 5 Transition Up (TU), each of which has a transposed convolution. The final layer in the model contains a 1×1 convolution, followed by a softmax non-linearity activation function. This provides the per class distribution at each pixel. Figure 3.2 illustrates the schematic diagram of the *Tiramisu* architecture.

Although this network is very deep, it has very few parameters compared to other convolutional neural networks. It has been found to be effective in challenging datasets, without requiring any extra post-processing or pre-training.

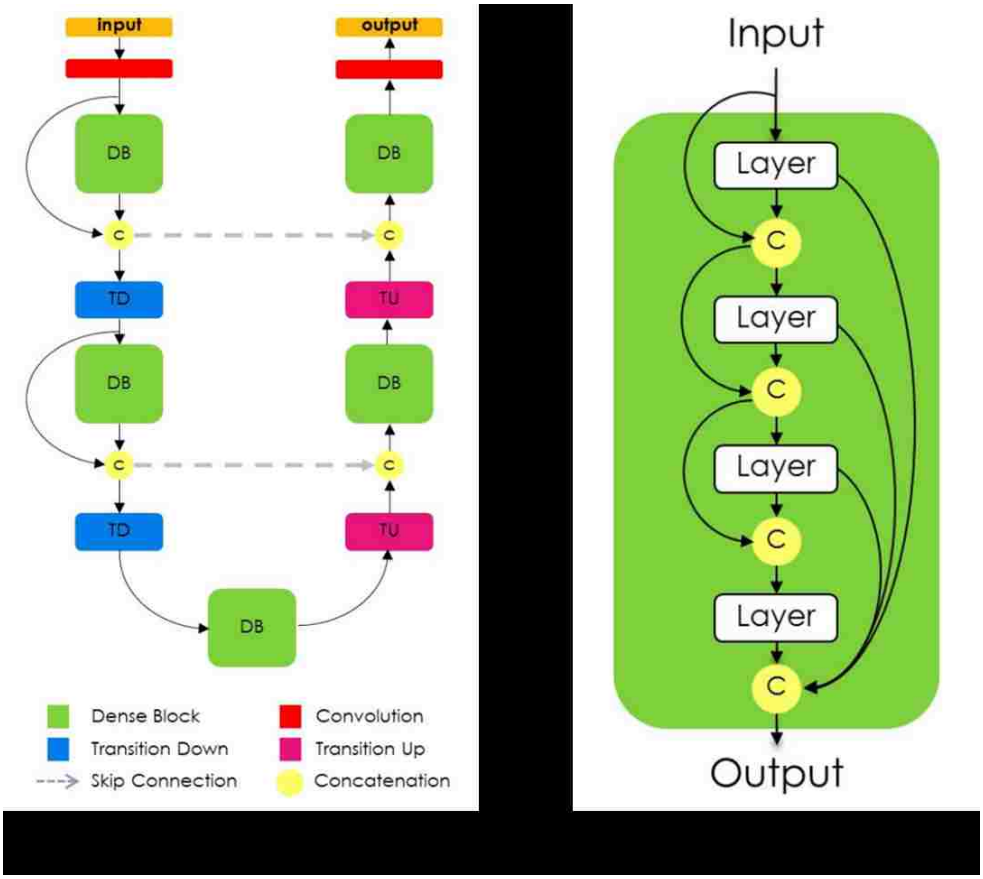


Figure 3.2: Schematic diagram of *Tiramisu* architecture (a) Fully Convolutional DenseNet with 103 layers (b) Content of a dense block [12].

3.3 Capsule networks

The most important part of a CNN is the convolution operation. The purpose of a convolution along with its weights is to identify key features. During training of the CNN, it tunes the weights of the convolution operations to identify certain features in the image. If the model is trained for face detection, the eyes or ears might trigger some convolutions in the network. If all the major parts of the face (like eyes, ears, nose, and mouth) are found together, then the CNN will be able

to predict a face.

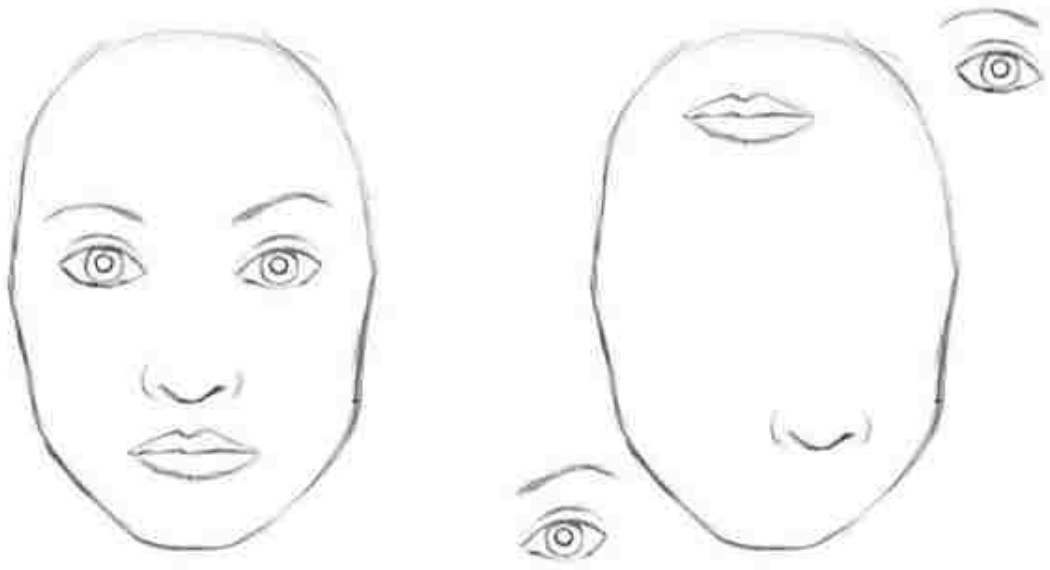


Figure 3.3: A typical scenario where CNNs fail to understand the object-part relationships in face images [39].

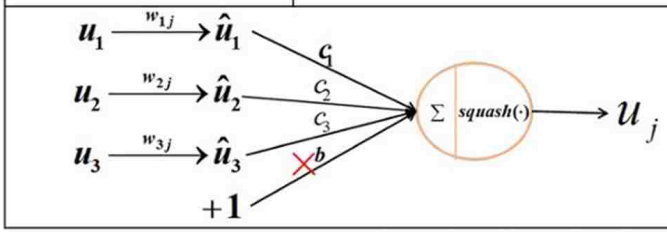
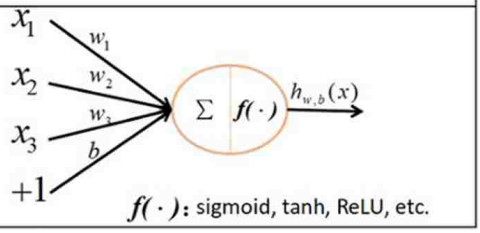
However, the CNNs cannot perform such a task. Figure 3.3 depicts an example where CNNs fail [39]. Although the convolution operation of the CNNs activate all the important features of the face, it concludes both the images as a face.

High-level features are combined with low-level features before being passed through an activation function. The part-whole relation or the relationship between the features are not provided in this information flow. Thus, we can conclude that the main drawback of CNNs is that they fail to recognize the relative relationships among features. This is due to the fact that CNNs rely on scalar values during convolution operation.

For correct image recognition, it is essential to preserve hierarchical pose relationships between features. For a better methodology to model these relationships, capsules [7] introduce a new

building block that can be used in deep learning. This new feature representation can be obtained by using vectors instead of scalars.

Table 3.3: Comparison of capsules with traditional neurons [39].

		capsule	vs.	traditional neuron
Input from low-level neuron/capsule		vector(u_i)		scalar(x_i)
Operation	Affine Transformation	$\hat{u}_{j\|i} = W_{ij} u_i$		–
	Weighting	$s_j = \sum_i c_{ij} \hat{u}_{j\|i}$		$a_j = \sum_{i=1}^{3} W_i x_i + b$
	Sum			
	Non-linearity activation fun	$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|}$		$h_{w,b}(x) = f(a_j)$
output		vector(v_i)		scalar(h)
				

To sum up the basic operations of a convolution in a typical CNN include scalar weighting of input scalars, Sum of weighted input scalars and scalar-to-scalar non-linearity. The capsule networks use vectors instead of scalars. When compared to the convolution of a CNN, the basic operations of capsule networks are matrix multiplication of input vectors, scalar weighting of input vectors, sum of weighted input vectors and vector-to-vector non-linearity. Table 3.3 provides a detailed comparison of the operations and pictorial representation of capsules and traditional neurons. The nomenclature for the parameters used in this table has the usual meanings.

Capsules also have the advantage of being able to produce good accuracy with much less training datasets. Intuitively, a capsule identifies the rotation of face right or left rather than recognizes a rotated variant of a face. Thus, having vector information such as orientation, size, etc, capsules

can use that information to learn a richer representation of the features. Since the model learns the feature variant in a capsule, possible variants are extrapolated more effectively with less training data.

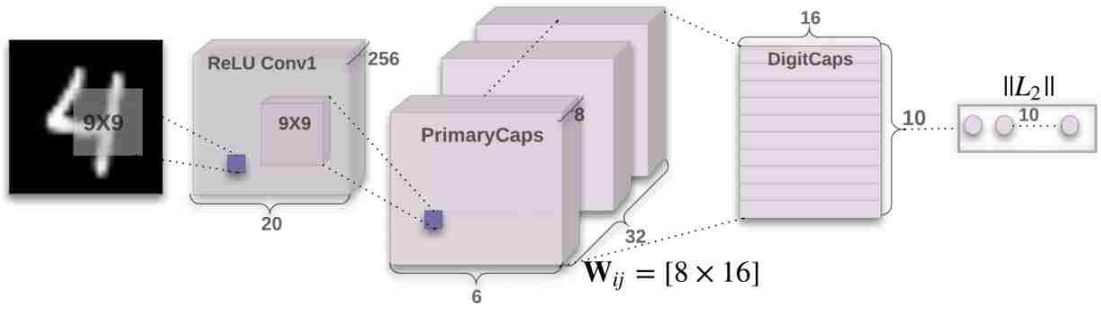


Figure 3.4: Schematic diagram of a simple capsule network with 3 layers [7].

Figure 3.4 shows a simple capsule network with 3 layers. The model is a shallow network that contains only two convolutional layers and one fully connected layer. Conv1 contains 256, 9×9 convolution kernels with a stride of 1 and ReLU activation. The features obtained from this layer are used as inputs to the primary capsules.

The primary capsules are the lowest level of multi-dimensional entities [7]. Since the capsule network is based on inverse graphics rendering like the human brain, the primary capsules invert the rendering process. The second layer (PrimaryCaps), a convolutional capsule layer contains 32 channels of convolutional 8D capsules where each primary capsule consists of 8 convolutional units with a 9×9 kernel and a stride of 2. Each primary capsule output visualizes the outputs of all 256 Conv1 units. The PrimaryCaps contains a total of $32 \times 6 \times 6$ capsule outputs with an 8D vector.

The final Layer (DigitCaps) contains one 16D capsule per digit class and each of these capsules takes input from all the capsules in the layer below. Routing occurs only between two consecutive

capsule layers, PrimaryCaps and DigitCaps.

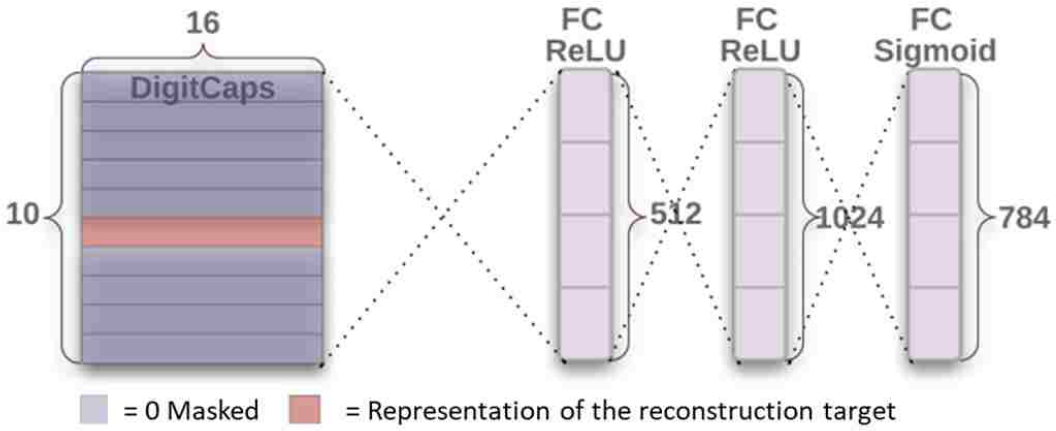


Figure 3.5: Schematic diagram of a decoder structure to reconstruct a digit from the DigitCaps layer representation [7].

Additional reconstruction loss is used to encourage the digit capsules to encode the instantiation parameters of the input digit. During training, everything except the activity vector of the correct DigitCaps is masked out. This activity vector is used to reconstruct the input image. The output of the DigitCaps is fed into a decoder. The decoder comprises 3 fully connected layers. Each connected layer models the pixel intensities as described in Figure 3.5.

3.4 SegCaps

Capsule networks are proposed for small images and for classification. We need to adapt it to do segmentation. The *SegCaps* [6] is the modified version of the capsule network for object segmentation. It differs from the dynamic routing of the original capsule networks in the following ways. First, routing is done within a defined spatially-local window. Second, each member of the grid shares the transformation matrix within a capsule type. These modifications can handle large im-

age sizes, unlike original capsule networks. Additionally, the concept of deconvolutional capsules is introduced. A three-layer convolutional capsule is illustrated in Figure 3.6.

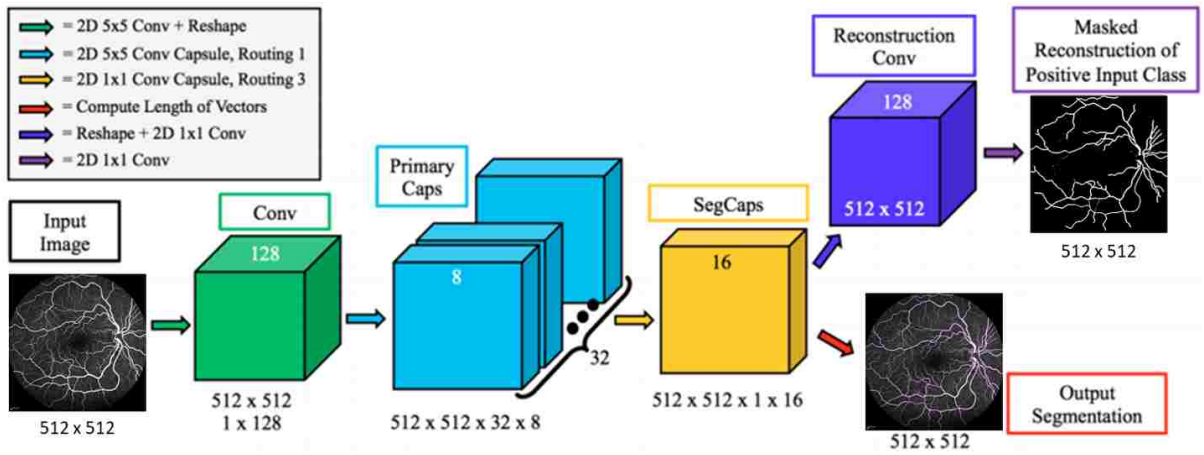


Figure 3.6: Schematic diagram of baseline *SegCaps* [6] used for retinal blood vessel segmentation of FA videos.

Table 3.4: Number of parameters used in training the different neural network models.

Method	Parameters
U-Net	31 M
Tiramisu	2.3 M
Baseline SegCaps	**1.6 M**

In our experiment with FA videos using *U-Net*, *Tiramisu* and baseline *SegCaps*, we find that baseline *SegCaps* uses 94.84% less parameters than *U-Net* and 30.43% less parameters than *Tiramisu*. Table 3.4 gives the total number of parameters required in training the models.

CHAPTER 4: PROPOSED APPROACH

To the best of our knowledge, for the first time in the literature, we introduce a new computational tool that combines deep learning and video magnification allowing quantitative analysis of FA videos to assist doctors in analyzing the diagnostic interpretations of these videos. We propose a pipeline-based approach, which has three modules, namely, image registration, segmentation using baseline *SegCaps* [6] and video magnification using Eulerian video magnification [8].

In module 1, we perform image registration to remove the motion, which is one of the major challenges. Next, in module 2, the segmentation of the vessels is done using some machine learning approaches. Finally, in module 3, we use these segmented regions to magnify only the retinal blood vessel structures for easier diagnostic interpretation of those videos.

4.1 Image Registration

While capturing FA videos, the images are not properly aligned due to motion in the frames caused by the movement of the patient and blinking of the eyes and the presence of noise in the videos. Therefore, proper image registration is required. Moreover, it is required to diminish the artifact error of Eulerian video magnification, which is implemented in the final module of the proposed solution.

In the image registration, we use serial rigid registration with binary images instead of grayscale images. We also consider a middle frame from the video and register all the remaining frames to it. During pre-processing, we crop FA videos to 512×512 size. For the removal of noise from the images, we normalize, blur using the median filter, and threshold them using Otsu's algorithm [37].

We also see that affine registration with one scale may be used, but the rigid registration is found to be more effective. Here, for any pixel location, we find a new location, where the degrees of freedom is three. Since we already know the old and new locations, we try to find the relevant parameters for the image registration. We do not want vessel structures to be changed. Considering two point sets, rigid registration generates a rigid transformation which maps one point set to the other. A rigid transformation is a type of transformation that does not vary the distance between any two points. In rigid registration, it includes only rotation and translation. In some cases, the point set can also get mirrored. Gradient images are fed into rigid registration, because of the heavy noise nature of the images. The rigid transformation is represented by the following Equations (4.1) and (4.2).

$$X' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \sin\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \tag{4.1}$$

$$X' = T(X|P) = T(x, y|t_x, t_y, \theta)\,, \tag{4.2}$$

where, x and y are the old pixel locations and x' and y' are the new pixel locations. t_x and t_y are the translation parameters and θ is the rotational parameter. It has three degrees of freedom.

4.2 Image Segmentation

Retinal vessel segmentation helps doctors to visualize and diagnose accurately, and prescribe early treatment and surgery of ocular diseases. This is the first-ever deep segmentation study in FA videos. We use the capsule based deep network for vessel segmentation. We create masks for the blood vessels in retinal images using AMIRA software (Thermo Fisher Scientific, Massachusetts,

USA). This is the ground truth. Data augmentation is performed to increase the performance of the baseline *SegCaps* [6] algorithm. We also compare the results with the state-of-the-art methods like *U-Net* [10] and *Tiramisu* [12].

4.2.1 Data Augmentation

Since our dataset contains only 10 subjects, we use data augmentation to increase the performance of the Baseline Caps algorithm. We create another ten pseudo-subjects by augmenting the original ten subjects. As a result, the total number of subjects becomes twenty consisting of 2802 frames. In order to obtain more variations in the dataset, we employ several operations including rotation (5 to -5), flipping left and right, flipping top and bottom, random zoom with 80% zoom level and elastic distortion.

4.2.2 Segmentation using baseline SegCaps

In this work, we conduct a comprehensive computational experimentation to compare the baseline *SegCaps* [6] with some state-of-the-art baseline networks such as *U-Net* and *Tiramisu*, which were not used before in this field.

4.3 Video magnification

The final module of the pipeline-based architecture is the video magnification. This is done in order to visualize the change of intensity in the blood vessels. In the literature, it has been found that Lagrangian perspective (with reference to fluid mechanics) based approach is used to amplify subtle motions and visualize deformation of objects that would otherwise be undetectable. However,

this perspective involves excessive computational times and it is difficult to make artifact-free, especially at obstruction regions and complex motions due to its accurate motion estimation. On the other hand, Eulerian based approach does not rely on exact motion estimation, but it amplifies the motion by changing temporal color at fixed positions [8]. In this study, we apply Eulerian video magnification (EVM) to FA videos. The segmentation is used as a prior (constrained magnification) to obtain the magnification of the segmented blood vessels.

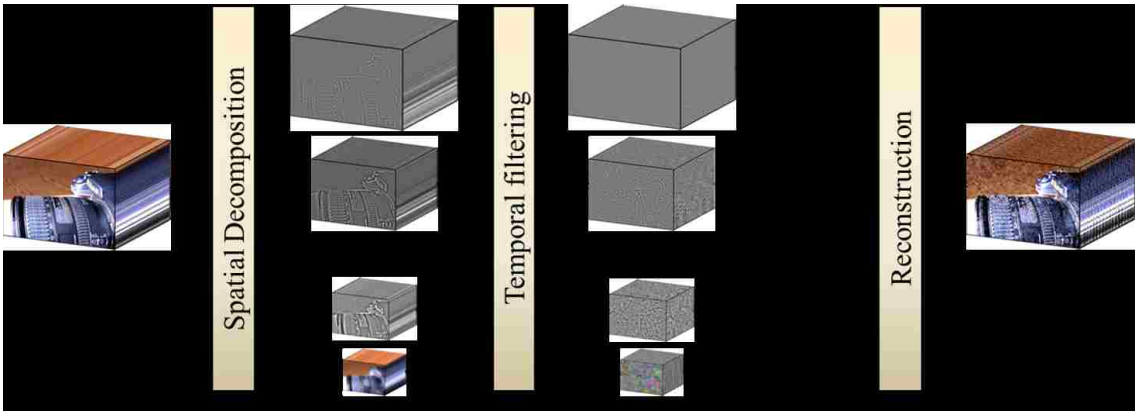


Figure 4.1: Overview of the Eulerian video magnification framework [8].

Eulerian video magnification algorithm combines spatial and temporal processing to give an emphasis to minor changes in a video along the temporal axis. The process involves in the decomposition of the frames into different spatial frequency bands. Then, the frames are converted from the spatial to the frequency domain using Fourier transformation. Considering the time series of a pixel intensity in a frequency domain, a bandpass filter is applied to obtain the frequency bands of interest. The extracted bandpassed signal is amplified by a magnification factor α. This factor is user-defined. Finally, the magnified signal is appended to the original and the spatial pyramid is collapsed to generate the output. Figure 4.1 illustrates the overall architecture of the Eulerian video magnification.

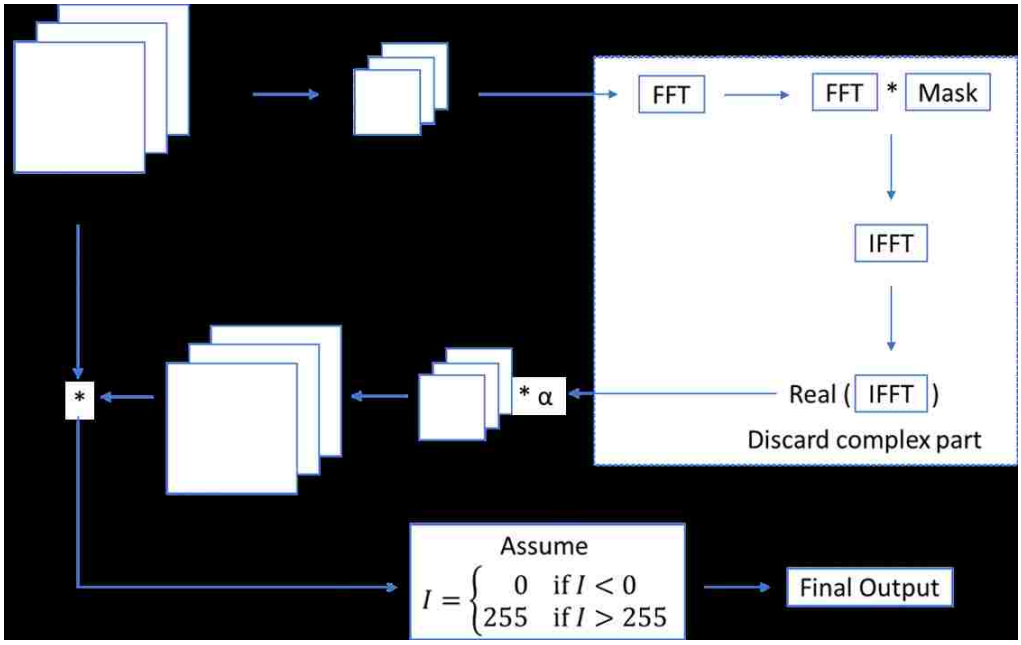


Figure 4.2: Flowchart of modified Eulerian video magnification for FA video.

We use a modified version of Eulerian video magnification as illustrated in Figure 4.2. We modify the algorithm to amplify only the region of interest. Here, in FA video, we consider the region as the blood vessels instead of the whole frame. Accordingly, the segmented frames obtained from the segmentation algorithm are taken as the inputs of the magnification algorithm. We use an amplification factor of 5, a frequency range of 0.5-10 Hz and spatial decomposition levels of 4. The output is a grayscale video, where only the blood vessels are magnified in the red color. It shows the flow of blood in the vessels as the video progresses.

CHAPTER 5: EXPERIMENTS AND RESULTS

We now provide the details of the computational experiments carried out in this work.

5.1 Dataset

A prospective study is conducted with existing IRB at Central Florida Retina, Orlando, FL, in collaboration with University of Central Florida College of Medicine, Orlando, FL. The FA video dataset consists of 10 normal subjects. The frames of each of the 10 subjects are given as: Subject 1 – 143 frames, Subject 2 – 108 frames, Subject 3 – 218 frames, Subject 4 – 119 frames, Subject 5 – 120 frames, Subject 6 – 116 frames, Subject 7 – 190 frames, Subject 8 – 112 frames, Subject 9 – 164 frames, and Subject 10 – 111 frames. Therefore, we have a total of 1402 frames.

5.2 Image registration

The image registration algorithm is coded in C++ language. The Insight Toolkit (ITK) library is used for the rigid registration. The OpenCV library is adopted for the pre-processing the frames of FA videos. In order to optimize the registration parameters, we apply a gradient descent algorithm with a learning rate of 0.125 and step size of 0.001. We run the optimizer for 200 iterations. The parameters of the gradient descent algorithm are obtained through the trial and error method. We visually evaluate the results obtained after image registration. We consider cross-correlation as similarity matrix instead of mutual information because it has been shown in the literature that cross-correlation gives better interpretation than mutual information on images with the same modality or source. It is found that image registration fails when a grayscale image is used as reference image instead of a binary image. This is shown in Figure 5.1.

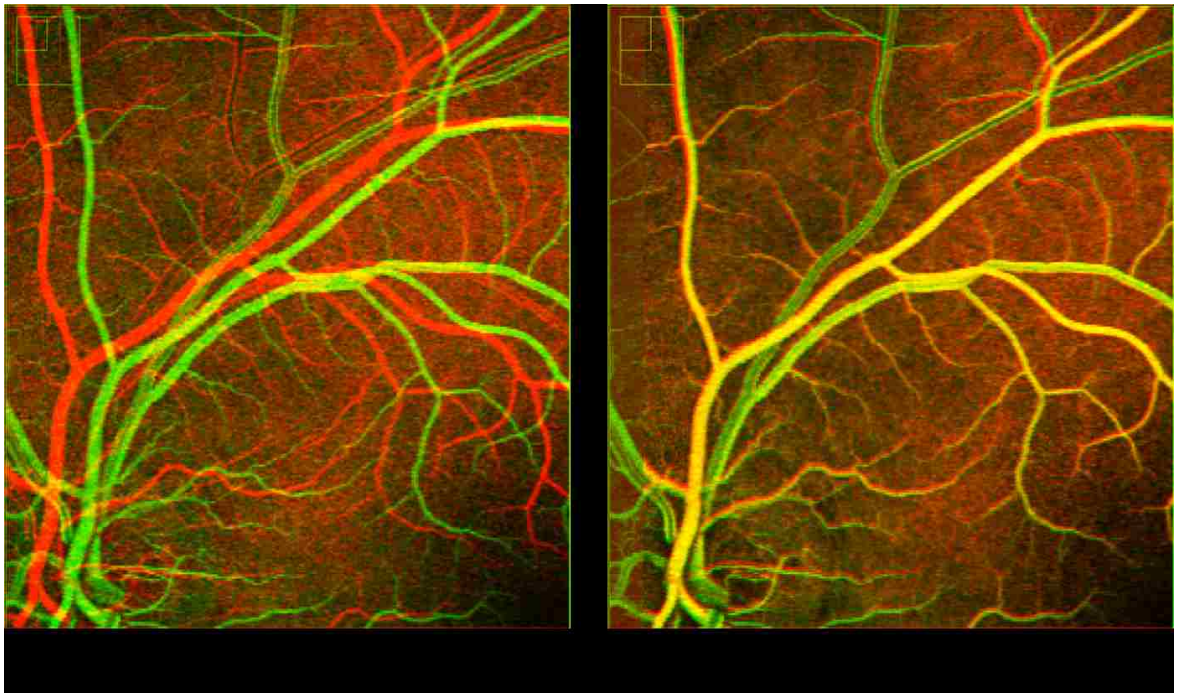


Figure 5.1: Overlapping two successive frames shows registration (a) Fails with grayscale images and (b) Successful with binary images.

5.3 Image segmentation using deep learning algorithms

Manual annotations of the FA video frames are done using AMIRA software. Since physicians primarily look for the changes in the blood flow in the large vessels, only the large vessels are annotated. The networks: *U-Net*, *Tiramisu* and baseline *SegCaps* are coded in python using libraries like Keras and Tensorflow. The experiments are run on a computer equipped with Ubuntu Linux operating system with NVIDIA Tesla GP graphics card.

5.3.1 *Details of training procedure*

All possible experimental parameters play an important role in the performance of different networks. The networks are trained from scratch on the FA images. Inside the network, the input images go through another round of data augmentation methods comprising scale, flip, shift, rotate, elastic deformations and random noise. Adam optimizer is used with an initial learning rate of 0.0001. The learning rate decays by a factor of 0.000001 when the validation loss reaches a saturation level. The batch size is taken as 1 for all the experiments. Weighted binary cross-entropy loss is taken as the loss function. The training continues for 200 epochs (1000 steps per epoch) and early stopping is applied with a patience value of 25 epochs.

To validate the datasets, we use a 10-fold cross-validation, where it divides the entire dataset into ten equally sized subsets and uses 90% of the dataset for training and 10% for testing to evaluate the accuracy of the cross-validation results. This process is repeated by exchanging the training and testing subsets ten times or folds such that all data has a chance of being trained and tested. Since the dataset used for validation in each fold is unknown to the learning algorithm, it is a good benchmark for the trained model for evaluation on a testing dataset and thus avoids over-fitting.

5.3.2 *Metric for algorithm performance*

In order to verify the accuracy of the performance of the baseline *SegCaps*, we use the metric Dice coefficient (DSC). The formula of Dice score is given in Equation (5.1)

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{5.1}$$

Higher the DSC, better the accuracy is.

5.3.3 *Quantitative analysis*

While training the FA video frames with different networks, we use weighted binary cross-entropy as the loss function. Figures 5.2, 5.3 and 5.4 illustrate the loss curve during training and validation for three different models.

Figure 5.2: Loss curves for the *U-Net* model during training and validation.

Figure 5.3: Loss curves for the *Tiramisu* model during training and validation.

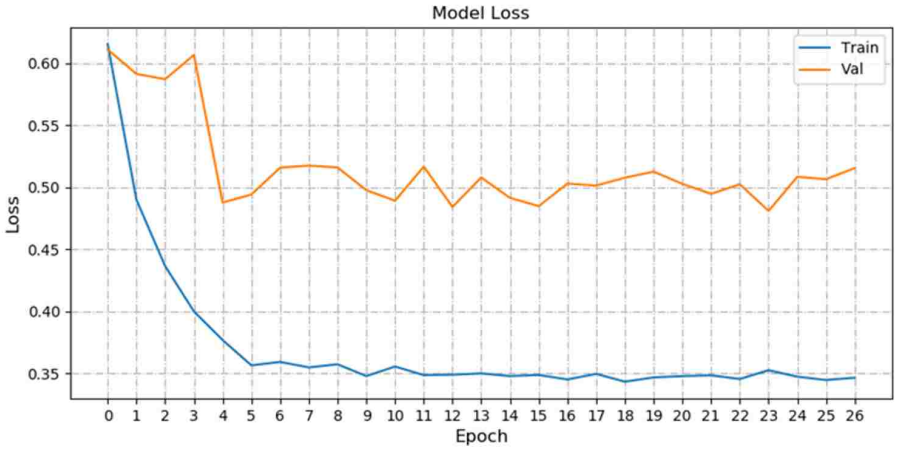


Figure 5.4: Loss curves for the baseline *SegCaps* model during training and validation.

The comparative Dice score results of these experiments are shown in Table 5.1. It is evident from the results that the baseline *SegCaps* network considerably outperforms the other two models.

Table 5.1: Comparison of DSC for different neural network models.

Method	DSC
U-Net	66.83%
Tiramisu	64.99%
Baseline SegCaps	**73.28%**

5.3.4 Qualitative analysis

For qualitative evaluations, three different frames taken from the same subject are used to compare three different models. These are depicted in Figures 5.5 — 5.7. The ground truth is colored in red and the output segmentation in blue. From the figures, it is clear that the segmentation done by the *U-Net* and *Tiramisu* are leaky and also over-segmented. The baseline *SegCaps*, on the other

hand, performs better than the other two models. Although manual annotations are done for large vessels only, it is found that baseline *SegCaps* also achieves segmentation of smaller blood vessels as well.

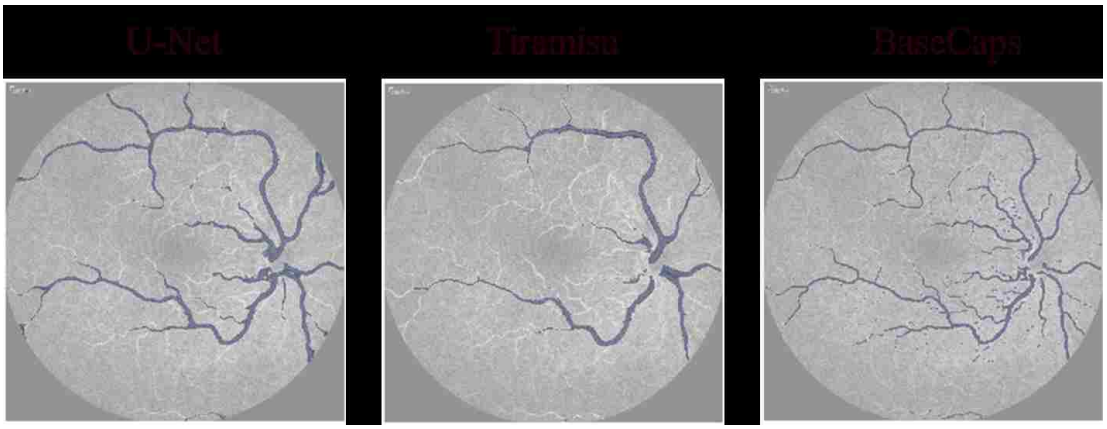


Figure 5.5: Comparison of different test results on a normal FA video (here, Frame 47 from Subject 2).

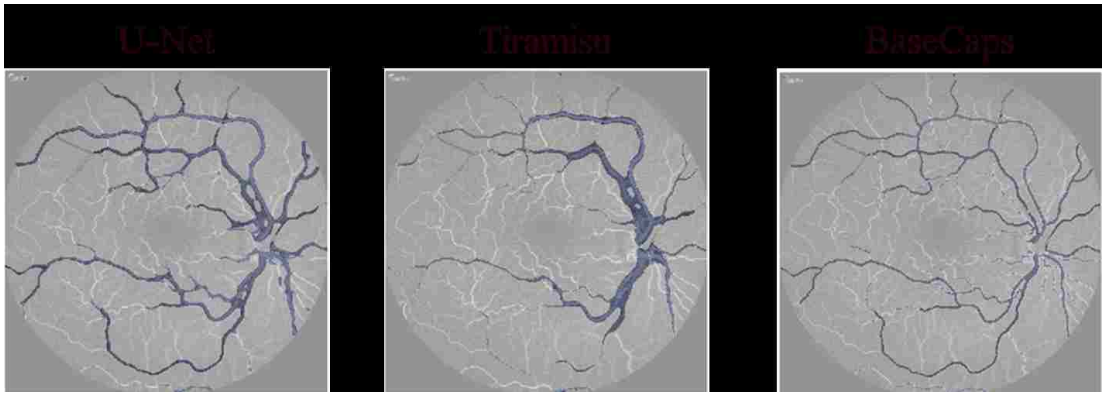


Figure 5.6: Comparison of different test results on a normal FA video (here, Frame 71 from Subject 2).

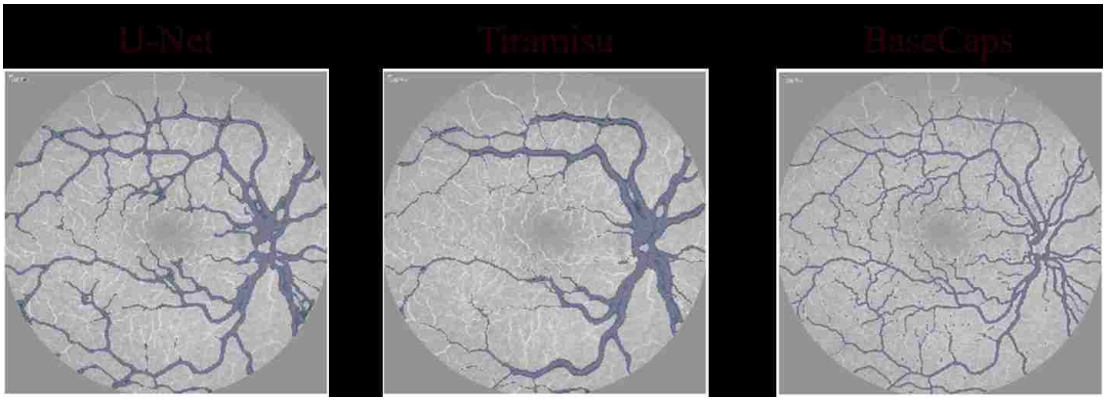


Figure 5.7: Comparison of different test results on a normal FA video (here, Frame 106 from Subject 2).

5.4 Eulerian video magnification

We conduct some preliminary computational experiments to ascertain the best parameters of the proposed EVM method and consider the parameters for the EVM approach: amplification factor (α) of 5, a frequency range between 0.5–10 Hz, and spatial decomposition levels of 4.

Figure 5.8 shows the raw and magnified fluorescein angiogram frame. It is seen from the figure that blood vessels are significantly amplified. In the output of the magnified frame, we see slow blood flow in the vessels.

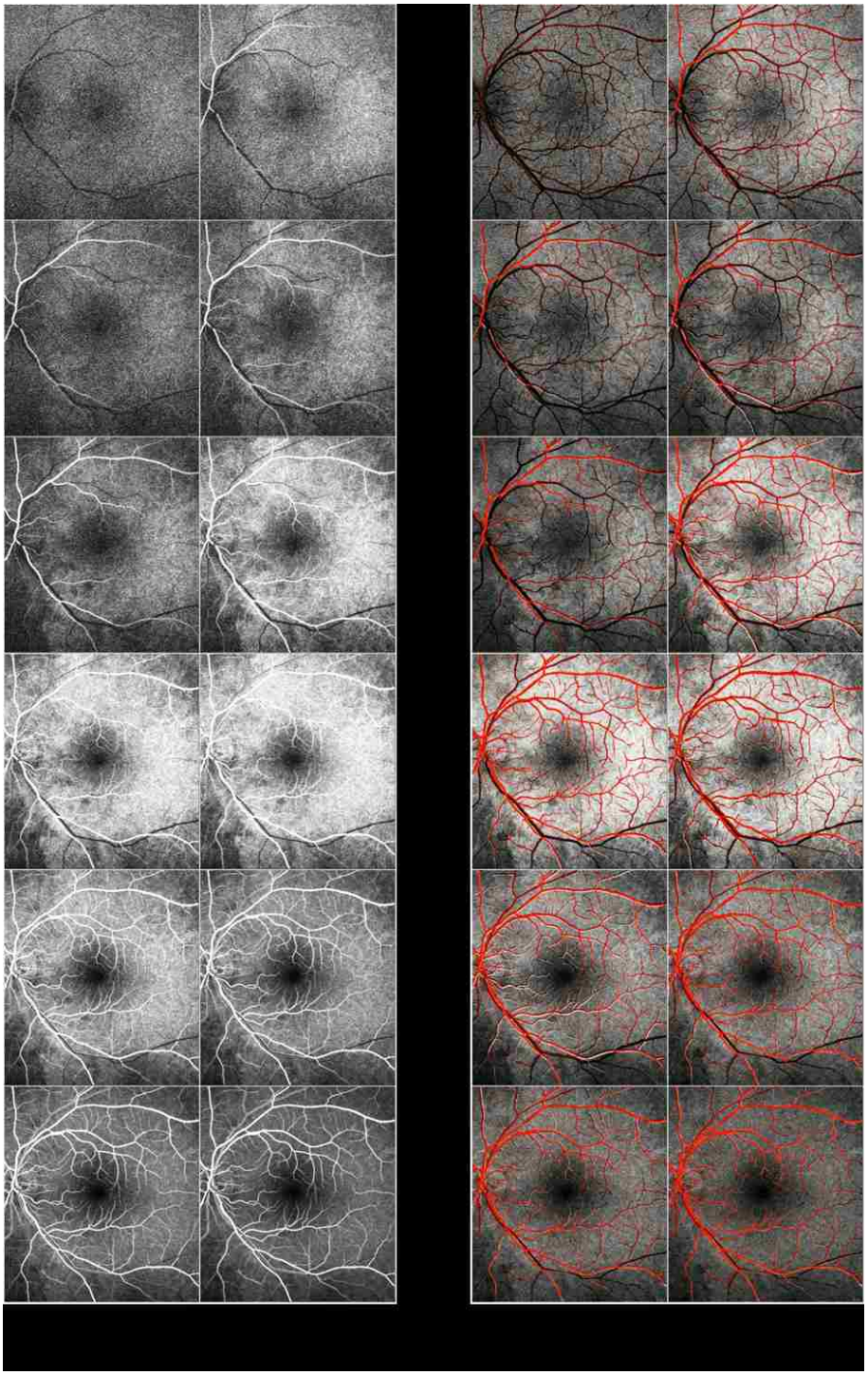


Figure 5.8: (a) Raw FA frames (b) Magnified FA frames using EVM algorithm.

CHAPTER 6: DISCUSSIONS AND CONCLUSIONS

In this research work, we have introduced a new computational tool that helps doctors to analyze and diagnose fluorescein retinal angiogram video. Our proposed method consists of three phases such as image registration, retinal blood vessel segmentation and segmentation guided video magnification. We perform serial rigid registration with binary images instead of grayscale images. For registration, all the frames are registered to the middle frame with the objective of eliminating the motion between the frames and some of the noise.

To the best of our knowledge based on the literature, deep learning based segmentation is the first ever application in FA video study. We apply a capsule-based deep learning approach for retinal video segmentation. The experimental results show the accuracy of more than 70% for the retinal vessel segmentation, which is competitive with the state-of-the-art procedures such as *U-Net* and *Tiramisu* networks.

Finally, we use Eulerian video magnification to amplify the intensities of the blood vessels. The output video provides better analysis of blood flow dynamics from fluorescein retinal angiogram videos.

There are some limitations of the current research worthy to note. The image registration is essential for removal of noise and motion from the video. For better pre-processing, other algorithms can be tried. Although we used rigid registration, better registration algorithm can be applied. For finding the parameters of rigid registration, we used a gradient descent algorithm. To get more optimized parameters, other optimization algorithms can be applied.

We have investigated the performance of the baseline *SegCaps* based on a small dataset although the data used here proves useful to validate the superior performance of the algorithm. However,

larger dataset, once these are available, could be used for more accurate FA video image segmentation.

The parameters used in the magnification algorithm can be further optimized to get more precise amplification. Since we empirically find the parameters, parameter search can be a promising area for future research.

LIST OF REFERENCES

[1] Almotiri, J., Elleithy, K., & Elleithy, A. (2018). *Retinal Vessels Segmentation Techniques and Algorithms: A Survey*. Applied Sciences, 8(2), 155.

[2] Yorston, D. (2006). *What's new in age-related macular degeneration?*. Community eye health, 19(57), 4.

[3] *What Is Diabetic Retinopathy?* https://www.aao.org/eye-health/diseases/what-is-diabetic-retinopathy/

[4] *Proliferative Diabetic Retinopathy in Fluorescein Angiography*. https://medtube.net/ophthalmology/medical-pictures/20397-proliferative-diabetic-retinopathy-in-fluorescein-angiography/

[5] Xu, Z., Bagci, U., Foster, B., Mansoor, A., Udupa, J. K., & Mollura, D. J. (2015). *WA hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT*. Medical image analysis, 24(1), 1-17.

[6] LaLonde, R., & Bagci, U. (2018). *Capsules for Object Segmentation*. arXiv preprint arXiv:1804.04241.

[7] Sabour, S., Frosst, N., & Hinton, G. E. (2017). *Dynamic routing between capsules*. In Advances in Neural Information Processing Systems (pp. 3856-3866).

[8] Wu, H. Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). *Eulerian video magnification for revealing subtle changes in the world*.

[9] Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*. arXiv preprint arXiv:1511.00561.

[10] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). *U-net: Convolutional networks for biomedical image segmentation*. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[11] Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[12] Jgou, S., Drozdzal, M., Vazquez, D., Romero, A., & Bengio, Y. (2017, July). *The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation*. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on (pp. 1175-1183). IEEE.

[13] Srinidhi, C. L., Aparna, P., & Rajan, J. (2017). *Recent advancements in retinal vessel segmentation*. Journal of medical systems, 41(4), 70.

[14] Orlando, J. I., Prokofyeva, E. & Blaschko, M. B. (2017). *A Discriminatively Trained Fully Connected Conditional Random Field Model for Blood Vessel Segmentation in Fundus Images*. IEEE Transactions on Biomedical Engineering, 64.1: 16-27.

[15] Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., & Wang, T. (2016). *A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images*. IEEE Trans. Med. Imaging, 35(1), 109-118.

[16] Liskowski, P., & Krawiec, K. (2016). *Segmenting retinal blood vessels with deep neural networks.*. IEEE transactions on medical imaging, 35(11), 2369-2380.

[17] Wang, S., Yin, Y., Cao, G., Wei, B., Zheng, Y., & Yang, G. (2015). *Hierarchical retinal blood vessel segmentation based on feature and ensemble learning*. Neurocomputing, 149, 708-717.

[18] Maninis, K. K., Pont-Tuset, J., Arbelez, P., & Van Gool, L. (2016, October). *Deep retinal image understanding*. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 140-148). Springer, Cham.

[19] Yang, Y., Li, T., Li, W., Wu, H., Fan, W., & Zhang, W. (2017, September). *Lesion detection and grading of diabetic retinopathy via two-stages deep convolutional neural networks*. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 533-540). Springer, Cham.

[20] Kar, S. S., & Maity, S. P. (2016). *Blood vessel extraction and optic disc removal using curvelet transform and kernel fuzzy c-means*. Computers in biology and medicine, 70, 174-189.

[21] Yin, B., Li, H., Sheng, B., Hou, X., Chen, Y., Wu, W., ... & Jia, W. (2015). *Vessel extraction from non-fluorescein fundus images using orientation-aware detector*. Medical image analysis, 26(1), 232-242.

[22] Bekkers, E., Duits, R., Berendschot, T., & ter Haar Romeny, B. (2014). *A multi-orientation analysis approach to retinal vessel tracking*. Journal of Mathematical Imaging and Vision, 49(3), 583-610.

[23] Mapayi, T., Viriri, S., & Tapamo, J. R. (2015). *Adaptive thresholding technique for retinal vessel segmentation based on GLCM-energy information*. Computational and mathematical methods in medicine, 2015.

[24] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). *Blood vessel segmentation methodologies in retinal imagesa survey*. Computer methods and programs in biomedicine, 108(1), 407-433.

[25] Jimnez-Snchez, A., Albarqouni, S., & Mateus, D. (2018). *Capsule Networks against Medical Imaging Data Challenges*. arXiv preprint arXiv:1807.07559.

[26] Afshar, P., Mohammadi, A., & Plataniotis, K. N. (2018). *Brain tumor type classification via capsule networks*. arXiv preprint arXiv:1802.10200.

[27] Iesmantas, T., & Alzbutas, R. (2018, June). *Convolutional capsule network for classification of breast cancer histology images*. In International Conference Image Analysis and Recognition (pp. 853-860). Springer, Cham.

[28] Deng, F., Pu, S., Chen, X., Shi, Y., Yuan, T., & Pu, S. (2018). *Hyperspectral Image Classification with Capsule Network Using Limited Training Samples*. Sensors, 18(9), 3153.

[29] Singh, N. P., Kumar, R., & Srivastava, R. (2015, May). *Local entropy thresholding based fast retinal vessels segmentation by modifying matched filter*. In Computing, Communication & Automation (ICCCA), 2015 International Conference on (pp. 1166-1170). IEEE.

[30] Jin, Z., Zhaohui, T., Weihua, G., & Jinping, L. (2015, November). *Retinal vessel image segmentation based on correlational open active contours model*. In Proceedings of the 2015 Chinese Automation Congress (CAC), Wuhan, China (pp. 27-29).

[31] Gongt, H., Li, Y., Liu, G., Wu, W., & Chen, G. (2015, October). *A level set method for retina image vessel segmentation based on the local cluster value via bias correction*. In Image and Signal Processing (CISP), 2015 8th International Congress on (pp. 413-417). IEEE.

[32] Sharma, S., & Wasson, E. V. (2015). *Retinal blood vessel segmentation using fuzzy logic*. Journal of Network Communications and Emerging Technologies, 4(3).

[33] Kumar, D., Pramanik, A., Kar, S. S., & Maity, S. P. (2016, June). *Retinal blood vessel segmentation using matched filter and laplacian of gaussian*. In Signal Processing and Communications (SPCOM), 2016 International Conference on (pp. 1-5). IEEE.

[34] Christodoulidis, A., Hurtut, T., Tahar, H. B., & Cheriet, F. (2016). *A multi-scale tensor voting approach for small retinal vessel segmentation in high resolution fundus images*. Computerized Medical Imaging and Graphics, 52, 28-43.

[35] Maji, D., Santara, A., Mitra, P., & Sheet, D. (2016). *Ensemble of deep convolutional neural networks for learning to detect retinal vessels in fundus images*. arXiv preprint arXiv:1603.04833.

[36] Jiang, Z., Yepez, J., An, S., & Ko, S. (2017). *Fast, accurate and robust retinal vessel segmentation system*. Biocybernetics and Biomedical Engineering, 37(3), 412-421.

[37] Otsu, N. (1979). *A threshold selection method from gray-level histograms*. IEEE transactions on systems, man, and cybernetics, 9(1), 62-66.

[38] *U-Net DeepLearning 0.1 documentation*. `http://deeplearning.net/tutorial/unet.html`

[39] *A simple and intuitive explanation of Hintons Capsule Networks* https://towardsdatascience.com/a-simple-and-intuitive-explanation-of-hintons-capsule-networks-b59792ad46b1/

[40] *DRIVE: Digital Retinal Images for Vessel Extraction* https://www.isi.uu.nl/Research/Databases/DRIVE/

[41] *STructured Analysis of the Retina* http://cecas.clemson.edu/ ahoover/stare/

[42] *High-Resolution Fundus (HRF) Image Database* https://www5.cs.fau.de/research/data/fundus-images/